Electronic Journal of Statistics
Vol. 17 (2023) 1316–1343
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2128

Sieve estimation of semiparametric accelerated mean models with panel count data*

Xiangbin Hu

Department of Applied Mathematics,
The Hong Kong Polytechnic University
e-mail: xiang-bin.hu@connect.polyu.hk

Wen Su

Department of Statistics and Actuarial Science,
University of Hong Kong
e-mail: jenna.wen.su@connect.hku.hk

Xingqiu Zhao

Department of Applied Mathematics,
The Hong Kong Polytechnic University
e-mail: xingqiu.zhao@polyu.edu.hk

Abstract: A widely adopted semiparametric model for analyzing panel count data is a proportional mean model, which may be deemed inappropriate when the proportionality assumption is violated. Motivated by the popular accelerated failure time model that relaxes such assumption, we investigate accelerated mean models for semiparametric regression analysis of panel count data. For estimation of bundled parameters, we develop a sieve least squares estimation procedure, which is robust in the sense that no distributional assumption is required for the underlying recurrent event process. Overcoming the theoretical challenges from bundled parameters, we establish the consistency and convergence rate of the proposed estimators, and derive the asymptotic normality of both the finite-dimensional estimator and the functionals of the infinite-dimensional estimator. Simulation studies demonstrate promising performances of the proposed approach, and an application to a skin cancer chemoprevention trial yields some new findings.

MSC2020 subject classifications: Primary 62N01, 62N02; secondary 62F12.
Keywords and phrases: Accelerated mean model, counting process, empirical process, sieve least squares estimation, panel count data.

Received January 2022.

Contents

*This research was supported in part by the Research Grant Council of Hong Kong (15306521) and the National Natural Science Foundation of China (12271459).

1. Introduction

Statistical inference on recurrent event processes has been gaining an increasing amount of research interests during the past few decades for its far-reaching applicability to various fields, such as econometrics, medicine, and epidemiology. In longitudinal follow-up studies with recurrent events, panel count data often occur when participants are observed at discrete time points where the numbers of recurrent events incurred between two observation points are available but the exact times these events happened remain unknown. A typical example is the skin cancer chemoprevention trial conducted by the University of Wisconsin Comprehensive Cancer Center, which included 291 patients that were randomly assigned to either placebo or 0.5 $g/m^2/day$ PO difluoromethylornithine (DFMO) treatment [1, 6, 16]. In the study, patients returned to the hospital roughly every six months for follow-up assessments, where the number of basal cell carcinoma and squamous cell carcinoma accumulated between two observation time points were recorded for analysis. The exact times when each tumor cell developed are unavailable since it is unrealistic to place patients under continuous computed tomography (CT) scans. The analysis of recurrent event processes with panel count data plays an important role in modern medicine, however, such unique characteristics bring additional challenges to statistical modeling. Luckily, numerous advanced methodologies emerged in recent years that significantly enhanced performance and predictability, including parametric modeling [3, 15, 27], nonparametric approaches [2, 10, 13, 19, 25, 30, 35, 36], and semiparametric modeling [5, 11, 20, 21, 26, 31].

In particular, many researches focused on the proportional mean/rate model in developing various classes of inference procedures, including estimating equation approaches [5, 11, 12, 14, 17, 26, 29], semiparametric likelihood approaches [20, 31], likelihood and estimating equation approaches [13, 21]. However, approaches based on the proportional mean/rate model become futile under some commonly encountered situations where the proportionality assumption is violated. Thus, to address more general situations, we consider an accelerated mean model, which is a natural extension of the well-known accelerated failure time

(AFT) model to recurrent event processes [9, 18]. The accelerated mean model possesses a unique feature which allows covariates to accelerate or decelerate the length of time elapsed to each recurrence [18, 32]. Recently, [33] studied a general class of semiparametric scale-change models for recurrent event data. To the best of our knowledge, the accelerated mean model with panel count data remained unexplored until a recent study by [6], who proposed a novel conditional likelihood and estimating equation method for model parameters and established the consistency of estimators. Meanwhile, they pointed out that the asymptotic distribution of the estimators for accelerated mean model parameters remains an open problem. The main challenge in deriving the asymptotics of the semiparametric accelerated mean model concerns the mean function with regression coefficients embedded as its argument resulting in "bundled" parameters, which requires more tactical development in empirical process theory.

The major contributions of this paper are fourfold. First, we develop a sieve least squares estimation approach for estimating unknown functional and scale parameters bundled in a semiparametric accelerated mean model. Second, we overcome the theoretical challenges from bundled parameters, derive the proposed estimators having the overall convergence rate slower than $n^{-1/2}$ but the regression estimator achieving the standard $n^{-1/2}$-convergence rate, and also establish the asymptotic normality of both the finite-dimensional estimator and functionals of the infinite-dimensional estimator. Compared with [8], our new contribution is that we first established the asymptotic normality for functionals of the proposed estimator with complex panel count data and then derived the asymptotic normality for the finite-dimensional estimator and functionals of the infinite-dimensional estimator. However, [8] only provided the asymptotic normality for the finite-dimensional estimator in general semiparametric sieve M-estimation with bundled parameters. Third, in contrast to most existing methods, the proposed estimation approach and the asymptotic results of the resulting estimator are robust in the sense that no distributional assumption is required for the underlying recurrent event process. Fourth, we apply the proposed method to a skin cancer chemoprevention trial and have some new findings.

The remainder of this paper is organized as follows. In Section 2, we introduce the accelerated mean model for panel count data and describe the enhanced least squares estimation procedure. The consistency, the convergence rate, and the asymptotic normality of the estimator are established in Section 3. We include some simulation studies in Section 4, and in Section 5, we apply the proposed approach to the skin cancer chemoprevention trial data. Finally, some discussions and potential future research directions are included in Section 6. All technical proofs are given in Appendix.

2. Methodology

Consider a longitudinal follow-up study containing n subjects. Let $N(t)$ denote the number of recurrent events occurred up to time t for $0 \leq t \leq \tau$, where τ

is the length of the study. Suppose each subject is observed at K discrete time points denoted by $\underline{\mathbf{T}} = (T_1, \dots, T_K)$, and let $\underline{\mathbf{N}} = (N(T_1), \dots, N(T_K))$. Let $\mathbf{X}$ be a p-dimensional covariate vector associated with recurrent events. Then for the i^{th} subject, the observation is represented by $\mathbf{O}_i = \{K_i, \underline{\mathbf{T}_i}, \underline{\mathbf{N}_i}, \mathbf{X}_i\}$ with $\underline{\mathbf{T}_i} = (T_{i1}, T_{i2}, \dots, T_{iK_i})$ and $\underline{\mathbf{N}_i} = (N_i(T_{i1}), N_i(T_{i2}), \cdots, N_i(T_{iK_i}))$, for $i = 1, \dots, n$. For recurrent event process $N(\cdot)$, we consider a semiparametric accelerated mean model:

$$E(N_i(t)|\mathbf{X}_i) = \Lambda(te^{\mathbf{X}_i^T\boldsymbol{\beta}}), \tag{2.1}$$

where $\Lambda(\cdot)$ is an unspecified non-negative and non-decreasing baseline mean function with $\Lambda(0) = 0$, and $\boldsymbol{\beta}$ is a p-dimensional parameter representing covariate effects. To estimate $\Lambda(\cdot)$ and $\boldsymbol{\beta}$, we propose to use the following loss function:

$$\ell_n(\boldsymbol{\beta}, \Lambda) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left\{N_i(T_{ij}) - \Lambda(T_{ij}e^{\mathbf{X}_i^T\boldsymbol{\beta}})\right\}^2.$$

In particular, we adopt the monotone I-spline function to approximate $\Lambda(\cdot)$. Set $b = \sup_{t,\mathbf{x},\boldsymbol{\beta}} t\exp(\mathbf{x}^T\boldsymbol{\beta})$. Take $\{t_i : i = 1, \cdots, m_n + 2d\}$ as a sequence of knots that partitions $[0, b]$ into $m_n + 1$ subintervals with

$$0 = t_1 = \cdots = t_d < t_{d+1} < \cdots < t_{m_n+d} < t_{m_n+d+1} = \cdots = t_{m_n+2d} = b.$$

Let $\{I_l(s), l = 1, \cdots, q_n\}$ be the I-spline basis functions [22] with $q_n = m_n + d$. According to [7], a reasonable approach to selecting d and m_n is treating them as turning parameters and using the BIC or cross-validation methods. For the simplicity of calculations, we choose the cubic I-spline function with $d = 4$ and $m_n = O(n^{1/7})$ by [19] and our simulation experiences. Set the functional space linearly spanned by the I-spline basis functions as follows

$$\mathcal{G}_n = \left\{\Lambda(s) = \sum_{l=1}^{q_n}\xi_l I_l(s) : \xi_l \geq 0 \text{ for } l = 1, \cdots, q_n\right\}.$$

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \Lambda)$ and $\boldsymbol{\theta}_0$ denote the true value of $\boldsymbol{\theta}$. Let $\mathcal{R}$ be a compact set of $\mathbb{R}^p$. Then we define the estimator of $\boldsymbol{\theta}$ to be the value that minimizes $\ell_n(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} \in \mathcal{R} \times \mathcal{G}_n$, denoted by $\widehat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$ with $\hat{\Lambda}_n(s) = \mathbf{I}(s)^T\hat{\boldsymbol{\xi}}_n$. This proposed estimator can be computed by L-BFGS-B algorithm [4].

3. Asymptotic results

We utilize the empirical process theory [28] to derive the asymptotic properties of the proposed estimator. First, we introduce more notation. Define the loss function for each individual based on observed data $\mathbf{O}$ as

$$L(\boldsymbol{\beta}, \Lambda; \mathbf{O}) = \sum_{j=1}^{K}\left\{N(T_j) - \Lambda\big(T_j \exp\left(\mathbf{X}^T\boldsymbol{\beta}\right)\big)\right\}^2,$$

and rewrite $\ell_n(\boldsymbol{\beta}, \Lambda) = \mathbb{P}_n L(\boldsymbol{\beta}, \Lambda; \mathbf{O})$. Let

$$\mathcal{H}_r = \{g : |g^{(r-1)}(s) - g^{(r-1)}(t)| \leq c_0|s-t| \ for\ all\ 0 \leq s, t \leq b,\ g(0) = 0, \\ g(\cdot) \text{ is nonnegative and nondecreasing function on } [0, b]\}$$

denote the function space of the baseline mean function Λ. Set the accelerated time of T_j to be $T_j^*(\boldsymbol{\beta}) = T_j \exp(\mathbf{X}^T\boldsymbol{\beta})$, and set $T_{0j}^* = T_j^*(\boldsymbol{\beta}_0) = T_j \exp(\mathbf{X}^T\boldsymbol{\beta}_0)$. Let $\mathcal{B}_p$ and $\mathcal{B}$ denote the collection of Borel sets in $\mathbb{R}^p$ and $\mathbb{R}$, respectively. We define the measure

$$\mu_1(B \times C) = \int_C \sum_{k=1}^{\infty} P(K = k|\mathbf{X} = \mathbf{x}) \sum_{j=1}^{k} P\big(T_{0j}^* \in B|K = k, \mathbf{X} = \mathbf{x}\big) dF_{\mathbf{X}}(\mathbf{x})$$

for any $B \in \mathcal{B}$ and $C \in \mathcal{B}_p$, where $F_{\mathbf{X}}$ is the distribution function of $\mathbf{X}$. Furthermore, set $\mu_2(B) = \mu_1(B \times \mathcal{X})$ for any $B \in \mathcal{B}$, where $\mathcal{X} \subseteq \mathbb{R}^p$ is a bounded set representing the domain of $\mathbf{X}$. For any $\Lambda_1, \Lambda_2 \in \mathcal{H}_r$, we define the metric:

$$d_1(\Lambda_1, \Lambda_2) = ||\Lambda_1 - \Lambda_2||_{L_2(\mu_2)} = \mathcal{P}\left[\sum_{j=1}^{K} \left\{\Lambda_1(T_{0j}^*) - \Lambda_2(T_{0j}^*)\right\}^2\right]^{1/2}.$$

For any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, we define the metric $d_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ as

$$d_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \Big[||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||_2^2 + ||\Lambda_1 - \Lambda_2||_{L_2(\mu_2)}^2\Big]^{1/2}.$$

To establish the asymptotic properties of the proposed estimator, we assume:

(C1) Let the true parameter $\boldsymbol{\beta}_0 \in \mathcal{R}^\circ$, where $\mathcal{R}^\circ$ is the interior of $\mathcal{R}$.

(C2) The true baseline function $\Lambda_0 \in \mathcal{H}_r$ satisfies that $\Lambda_0(b) \leq M_1$ for a constant $M_1 > 0$ with $r \geq 3$.

(C3) If there exists a constant vector $\boldsymbol{\eta}$ and a deterministic function g such that $\boldsymbol{\eta}^T\mathbf{X} = g(T_{0j}^*), j = 1, \ldots, K$ with probability 1, then $\boldsymbol{\eta} = \mathbf{0}$ and $g = 0$.

(C4) The number of subintervals in $[0, b]$ satisfies $m_n = O(n^\nu)$ for $0 < \nu < 1/2$. Furthermore, $\max_{d+1 \leq i \leq m_n+d+1} |t_i - t_{i-1}| = O(n^{-\nu})$, and there is a constant $M_2 > 0$ such that

$$\frac{\max_{d+1 \leq i \leq m_n+d+1} |t_i - t_{i-1}|}{\min_{d+1 \leq i \leq m_n+d+1} |t_i - t_{i-1}|} \leq M_2.$$

(C5) There is a constant $M_3 > 0$ such that $P(K \leq M_3) = 1$.

(C6) $E(e^{M_4 N(t)})$ is uniformly bounded for $t \in [0, \tau]$ and a constant $M_4 > 0$.

(C7) μ_2 is absolutely continuous with respect to Lebesgue measure with a derivative $\dot{\mu}_2$, and $\dot{\mu}_2$ has a uniform positive lower bound.

(C8) There is a constant $M_5 > 0$ such that $1/M_5 < \Lambda_0'(s) < M_5$ for all $s \in [b', b]$ with the positive constant b' satisfying $\Lambda_0(b') > 0$, where Λ_0' is the first order derivative of Λ_0.

(C9) $P\big(\bigcap_{j=1}^{K}\{T_{0j}^* \in [b', b]\}\big) = 1$.

(C10) There is a constant $0 < M_6 < 1$ such that

$$\mathbf{a}^T Var[\mathbf{X}|U]\mathbf{a} \geq M_6 \mathbf{a}^T E[\mathbf{X}\mathbf{X}^T|U]\mathbf{a}$$

a.s. for all $\mathbf{a} \in \mathbb{R}^p$, where $(U, \mathbf{X})$ has distribution $\mu_1/\mu_1(\mathbb{R}^+ \times \mathcal{X})$. Furthermore, $E[\mathbf{X}\mathbf{X}^T]$ is nonsingular.

Remark 1. Conditions (C1) and (C2) are common in the analysis of the semiparametric model. Similar to Condition (C.2) of [34], Condition (C3) ensures the identifiability of the proposed model. By [19, 20], Condition (C4) is regular for monotone spline estimation. Condition (C5) implies that the number of observation times is bounded, which is regular in medical follow-up studies. According to [31], Condition (C6) holds when the counting process is uniformly bounded or from a Poisson-type process, which is common in practice. Condition (C7) holds when the metric μ_2 has a strictly positive intensity. Condition (C8) is regular according to [31], meaning that the true baseline mean function is absolutely continuous with a bounded intensity function. Condition (C9) indicates that all the accelerated observation times at $\boldsymbol{\beta}_0$ fall into a fixed interval in which the baseline mean function is bounded away from zero. Conditions (C10) is a technical assumption, which is satisfied in many applications (see Condition (C13) of [31]).

Theorem 3.1 (Consistency)**.** *Suppose that Conditions* (C1)–(C6) *hold. Then* $d_2(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = o_p(1)$.

Theorem 3.2 (Convergence Rate)**.** *Suppose that Conditions* (C1)–(C10) *hold. Taking* $\nu = 1/(1+2r)$*, we have* $d_2(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p(n^{-r/(1+2r)})$.

Although the overall convergence rate of $\hat{\boldsymbol{\theta}}_n$ is slower than $n^{1/2}$, our proposed estimator for $\boldsymbol{\beta}_0$ is still asymptotically normal at the rate of $n^{1/2}$. We use Theorem 3.1 in [36] to establish the asymptotical distributions of $\hat{\boldsymbol{\beta}}_n$ and functionals of $\hat{\Lambda}_n$, respectively. Let $\tilde{\mathcal{H}} = \{(\mathbf{h}_1, h_2) : ||\mathbf{h}_1||_2 \leq 1, ||h_2||_\infty \leq 1, \mathbf{h}_1 \in \mathcal{R}, h_2 \in \mathcal{H}_r\}$. For all $(\mathbf{h}_1, h_2) \in \tilde{\mathcal{H}}$, define

$$m(\boldsymbol{\beta}, \Lambda; \mathbf{O})[\mathbf{h}_1, h_2] = \left. \frac{dL\big(\boldsymbol{\beta} + \eta\mathbf{h}_1, \Lambda + \eta h_2; \mathbf{O}\big)}{d\eta} \right|_{\eta=0}.$$

After some algebraic calculations, we have

$$\begin{aligned} & m(\boldsymbol{\beta}, \Lambda; \mathbf{O})[\mathbf{h}_1, h_2] \\ =& \sum_{j=1}^{K} \Big[\big\{\Lambda\big(T_j^*(\boldsymbol{\beta})\big) - N(T_j)\big\} \big\{\Lambda'\big(T_j^*(\boldsymbol{\beta})\big) T_j^*(\boldsymbol{\beta}) X_j^T \mathbf{h}_1 + h_2\big(T_j^*(\boldsymbol{\beta})\big)\big\} \Big]. \end{aligned}$$

Let

$$G(\boldsymbol{\beta}, \Lambda)[\mathbf{h}_1, h_2] = \mathcal{P} m(\boldsymbol{\beta}, \Lambda; \mathbf{O})[\mathbf{h}_1, h_2]$$

and

$$G_n(\boldsymbol{\beta}, \Lambda)[\mathbf{h}_1, h_2] = \mathbb{P}_n m(\boldsymbol{\beta}, \Lambda; \mathbf{O})[\mathbf{h}_1, h_2].$$

Then we have the following theorem.

Theorem 3.3 (Asymptotic Normality)**.** *Suppose that Conditions* (C1)–(C10) *hold.*

(i) For all $(\mathbf{h}_1, h_2) \in \tilde{\mathcal{H}}$*, we have*

$$\sqrt{n}(\sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1, h_2] + \sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2]) \xrightarrow{d} N(0, \sigma_0[\mathbf{h}_1, h_2]^2),$$

where $\sigma_0[\mathbf{h}_1, h_2]^2 = E[m(\boldsymbol{\beta}, \Lambda; \mathbf{O})[\mathbf{h}_1, h_2]^2]$*, and* $\sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1, h_2]$ *and* $\sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2]$ *are defined in* (B.4)*.*

(ii) Furthermore, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^T), \qquad \text{and}$$
$$\sqrt{n}\sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{R}(h_2), h_2] \xrightarrow{d} N(0, \sigma_0[\mathbf{R}(h_2), h_2]^2),$$

for all $h_2 \in \mathcal{H}_r$*, where* $\mathbf{A}$ *and* $\mathbf{B}$ *are defined in* (B.6)*, and* $\mathbf{R}(h_2)$ *is defined in* (B.7)*.*

In the above theorem, part (i) presents the joint asymptotic normality of the regression coefficient estimate and the functionals of the baseline mean function estimate, while part (ii) provides the asymptotic normality of the regression coefficient estimate and the asymptotic normality of the functionals of the baseline mean function estimate, respectively. These results are useful for conducting hypothesis testing and constructing confidence intervals of regression coefficients.

The proofs of Theorems 3.1-3.3 are sketched in Appendix.

4. Simulation study

We conducted extensive simulation studies to evaluate the finite-sample performance of the proposed estimators. Consider covariates $\mathbf{X}_i = (X_{1i}, X_{2i})^T$, with X_{1i} from Bernoulli(0.5) and X_{2i} from Uniform$(0, 1)$. Let the true $\boldsymbol{\beta}_0 = (1, -0.5)^T$. For each of the four cases of the counting process $N(\cdot)$ below, we considered $\Lambda_0(t) = t$ and $3\sqrt{t}$.

Case 1: $N(\cdot)$ followed a standard a Poisson process satisfying model (1). That is, $N_i(T_{i1})$ was generated from the Poisson distribution with mean

$\Lambda_0(T_{i1}e^{\mathbf{X}_i^T\boldsymbol{\beta}_0})$, and $\triangle N_i(T_{ij})$ was generated from the Poisson distribution with mean $\Lambda_0(T_{ij}e^{\mathbf{X}_i^T\boldsymbol{\beta}_0}) - \Lambda_0(T_{i(j-1)}e^{\mathbf{X}_i^T\boldsymbol{\beta}_0})$, for $j = 2, \ldots, K_i$.

Case 2: $N(\cdot)$ followed a mixed Poisson process satisfying model (1). That is, $N_i(T_{i1})$ was generated from the Poisson distribution with mean $\gamma_i\Lambda_0(T_{i1}e^{\mathbf{X}_i^T\boldsymbol{\beta}_0})$ and $\triangle N_i(T_{ij})$ was generated from the Poisson distribution with mean

$\gamma_i\Lambda_0(T_{ij}e^{\mathbf{X}_i^T\boldsymbol{\beta}_0}) - \gamma_i\Lambda_0(T_{i(j-1)}e^{\mathbf{X}_i^T\boldsymbol{\beta}_0})$, for $j = 2, \ldots, K_i$,
where $\gamma_i \sim$ Gamma(4,1/4).

Case 3: $N(\cdot)$ followed a negative binomial process satisfying model (1). That is, $N_i(T_{i1})$ was generated from the negative binomial distribution with parameters $\Lambda_0(T_{i1}e^{\mathbf{X}_i^T\boldsymbol{\beta}_0})$ and 1/2, and $\triangle N_i(T_{ij})$ was generated from the negative binomial distribution with parameters $\Lambda_0(T_{ij}e^{\mathbf{X}_i^T\boldsymbol{\beta}_0}) - \Lambda_0(T_{i(j-1)}e^{\mathbf{X}_i^T\boldsymbol{\beta}_0})$ and $1/2$, for $j = 2, \ldots, K_i$.

Case 4: $N(\cdot)$ followed a mixed negative binomial process satisfying model (1). That is, $N_i(T_{i1})$ was generated from the negative binomial distribution with parameters $\gamma_i \Lambda_0(T_{i1} e^{\mathbf{X}_i^T \boldsymbol{\beta}_0})$ and 1/2, and $\triangle N_i(T_{ij})$ was generated from the negative binomial distribution with parameters $\gamma_i(\Lambda_0(T_{ij} e^{\mathbf{X}_i^T \boldsymbol{\beta}_0}) - \Lambda_0(T_{i(j-1)} e^{\mathbf{X}_i^T \boldsymbol{\beta}_0}))$ and 1/2, for $j = 2, \ldots, K_i$, where $\gamma_i \sim$ Gamma(4,1/4).

In Cases 1 and 3, the number of observation times K_i was from a uniform distribution with equal probability 1/6 on $\{1, \cdots, 6\}$. In Cases 2 and 4, the number of observation times K_i was from a uniform distribution with equal probability 1/8 on $\{1, \cdots, 8\}$ for $\gamma_i > 1$, and was from a uniform distribution with equal probability 1/6 on $\{1, \cdots, 6\}$ for $\gamma_i \leq 1$. For given K_i, we set observation times $(T_{i,1}, \ldots, T_{i,K_i})$ as the order statistics from Uniform(0, τ) with $\tau = 6$. For the upper bound of the domain of Λ, we suggest selecting a large b within a rational range. We chose $b = 200$ such that $b \gg \tau_0^* = \sup_{\mathbf{X}} \tau \exp(\mathbf{X}^T \boldsymbol{\beta}_0) \approx 16.31$.

For the computation details, we implemented the L-BFGS-B algorithm with initial values $\widetilde{\boldsymbol{\beta}} = \mathbf{0}$ and $\widetilde{\boldsymbol{\xi}} = (1, 1, 1, 1, 1, 1, 1, 1)$. For the selection of location of knots, we considered the two most common methods by dividing the time interval equally and using the percentiles of the observation times, respectively. The details for the two methods are given as follows.

Method 1: We first calculated $T^*_{\max}(\widetilde{\boldsymbol{\beta}}) = \max_i T^*_{iK_i}(\widetilde{\boldsymbol{\beta}})$ and set the knots $t_{d+1}, \cdots, t_{d+m_n}$ to be $T^*_{\max}(\widetilde{\boldsymbol{\beta}})/(m_n + 1), \cdots, (m_n T^*_{\max}(\widetilde{\boldsymbol{\beta}}))/(m_n + 1)$, which divided $[0, T^*_{\max}(\widetilde{\boldsymbol{\beta}})]$ equally. Using the L-BFGS-B algorithm with initial value $(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\xi}})$, we obtained an initial estimate $(\check{\boldsymbol{\beta}}_n, \check{\boldsymbol{\xi}}_n)$. Then we selected the knots based on $\check{\boldsymbol{\beta}}_n$. That is, we calculated $T^*_{\max}(\check{\boldsymbol{\beta}}_n) = \max_i T^*_{iK_i}(\check{\boldsymbol{\beta}}_n)$ and set the knots $t_{d+1}, \cdots, t_{d+m_n}$ to be $T^*_{\max}(\check{\boldsymbol{\beta}}_n)/(m_n + 1), \cdots, (m_n T^*_{\max}(\check{\boldsymbol{\beta}}_n))/(m_n + 1)$.

Method 2: We first set the knots $t_{d+1}, \cdots, t_{d+m_n}$ to be the $1/(m_n + 1), \cdots, m_n/(m_n + 1)$ percentiles of $\{T^*_{ij}(\widetilde{\boldsymbol{\beta}}) : i = 1, \cdots n, j = 1, \cdots, K_i\}$. Using the L-BFGS-B algorithm with initial value $(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\xi}})$, we obtained an initial estimate $(\check{\boldsymbol{\beta}}_n, \check{\boldsymbol{\xi}}_n)$. Then we selected the knots based on $\check{\boldsymbol{\beta}}_n$. That is, we set the knots $t_{d+1}, \cdots, t_{d+m_n}$ to be the $1/(m_n+1), \cdots, m_n/(m_n+1)$ percentiles of $\{T^*_{ij}(\check{\boldsymbol{\beta}}_n) : i = 1, \cdots n, j = 1, \cdots, K_i\}$.

The estimator $(\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\xi}}_n)$ was finally obtained by the L-BFGS-B algorithm with initial value $(\check{\boldsymbol{\beta}}_n, \check{\boldsymbol{\xi}}_n)$.

For each of the four cases, we applied the proposed sieve I-spline estimation with knots selected by Methods 1 and 2. The Monte Carlo simulation was repeated 500 times for sample sizes $n = 100, 200, 300$. For the simulation results on Λ, we reported the mean, the 2.5% and 97.5% pointwise percentiles of the functional estimate $\hat{\Lambda}_n$ based on 500 replications. For the simulation results on β, we reported the estimated bias (BIAS) given by the average of the estimates minus the true value, the sample standard deviation of the estimates (SSE), the estimated standard errors of the estimates (ESE), and the estimated 95% coverage probabilities (CP) obtained from 500 replications, where the standard error of the estimate was estimated by 100 bootstrap samples.

Figure 1 shows the simulation results for the estimated function $\hat{\Lambda}(\cdot)$ on $[0, \tau_0^*]$ under Case 1 for each sample size. One can see that the results with knots se-

Table 1
Simulation results of $\boldsymbol{\beta}$ for Case 1.

	$\Lambda_0(t) = t$		$\Lambda_0(t) = 3\sqrt{t}$	
	Method 1	Method 2	Method 1	Method 2
	Sample Size: n =100			
Bias	(0.007,−0.021)	(0.019,−0.005)	(0.017,−0.022)	(0.020,−0.011)
SSE	(0.177,0.193)	(0.179,0.188)	(0.231,0.332)	(0.221,0.325)
ESE	(0.144,0.187)	(0.142,0.179)	(0.223,0.334)	(0.242,0.344)
CP	(0.948,0.950)	(0.942,0.946)	(0.922,0.926)	(0.948,0.948)
	Sample Size: $n = 200$			
Bias	(0.007,−0.010)	(0.014,−0.005)	(0.001,0.006)	(0.014,0.011)
SSE	(0.101,0.129)	(0.100,0.127)	(0.156,0.225)	(0.153,0.221)
ESE	(0.099,0.127)	(0.097,0.121)	(0.152,0.229)	(0.156,0.226)
CP	(0.936,0.940)	(0.936,0.930)	(0.942,0.956)	(0.954,0.944)
	Sample Size: $n = 300$			
Bias	(0.002,−0.006)	(0.005,−0.004)	(−0.002,0.008)	(0.002,0.013)
SSE	(0.084,0.102)	(0.085,0.095)	(0.115,0.181)	(0.114,0.176)
ESE	(0.083,0.104)	(0.080,0.099)	(0.123,0.185)	(0.125,0.184)
CP	(0.934,0.944)	(0.932,0.952)	(0.948,0.956)	(0.954,0.952)

Table 2
Simulation results of $\boldsymbol{\beta}$ for Case 2.

	$\Lambda_0(t) = t$		$\Lambda_0(t) = 3\sqrt{t}$	
	Method 1	Method 2	Method 1	Method 2
	Sample Size: $n = 100$			
Bias	(0.031,−0.026)	(0.057,−0.007)	(0.030,0.021)	(0.080,−0.076)
SSE	(0.343,0.374)	(0.328,0.363)	(0.367,0.599)	(0.465,0.628)
ESE	(0.315,0.347)	(0.330,0.349)	(0.348,0.568)	(0.408,0.636)
CP	(0.902,0.932)	(0.930,0.938)	(0.924,0.940)	(0.946,0.942)
	Sample Size: $n = 200$			
Bias	(0.003,−0.006)	(0.017,0.006)	(0.022,0.019)	(0.023,0.005)
SSE	(0.160,0.235)	(0.155,0.226)	(0.266,0.404)	(0.256,0.418)
ESE	(0.152,0.241)	(0.152,0.234)	(0.242,0.408)	(0.268,0.432)
CP	(0.936,0.950)	(0.942,0.948)	(0.916,0.938)	(0.952,0.960)
	Sample Size: $n = 300$			
Bias	(−0.019,−0.006)	(−0.005,0.004)	(−0.021,−0.002)	(−0.012,0.011)
SSE	(0.119,0.197)	(0.119,0.189)	(0.193,0.329)	(0.191,0.323)
ESE	(0.122,0.194)	(0.124,0.190)	(0.194,0.338)	(0.204,0.349)
CP	(0.950,0.938)	(0.962,0.954)	(0.940,0.938)	(0.954,0.956)

lected by Methods 1 and 2 are similar, all estimates of mean functions are very close to the true function, and the deviations in the tail regions reduce with sample size, suggesting that the proposed approach is appropriate. Moreover, the variability of the functional estimator decreases as the sample size increases, which means that our estimates are consistent. The figures in Cases 2–4 are similar and thus not presented here. Tables 1–4 summarize the simulation results for the estimates of regression parameters under each case. Across all simulation scenarios, we observed a small average bias and CP close to the theoretical level of 95%. Moreover, ESE and SSE are within proximity of each other, and both metrics decrease with increasing sample size. In summary, the proposed estimation procedures perform well for all situations considered, which demonstrate the robustness.

Table 3
Simulation results of β for Case 3.

	$\Lambda_0(t)=t$		$\Lambda_0(t)=3\sqrt{t}$	
	Method 1	Method 2	Method 1	Method 2
	Sample Size: $n=100$			
Bias	(0.030,−0.016)	(0.036,−0.003)	(0.024,−0.026)	(0.025,−0.056)
SSE	(0.252,0.270)	(0.232,0.256)	(0.367,0.472)	(0.359,0.506)
ESE	(0.206,0.276)	(0.214,0.269)	(0.321,0.485)	(0.361,0.525)
CP	(0.922,0.944)	(0.946,0.944)	(0.906,0.948)	(0.940,0.946)
	Sample Size: $n=200$			
Bias	(0.004,−0.009)	(0.014,0.002)	(0.006,−0.001)	(0.018,−0.001)
SSE	(0.148,0.176)	(0.144,0.168)	(0.219,0.309)	(0.216,0.320)
ESE	(0.142,0.183)	(0.142,0.175)	(0.218,0.330)	(0.233,0.340)
CP	(0.938,0.948)	(0.932,0.952)	(0.944,0.946)	(0.956,0.956)
	Sample Size: $n=300$			
Bias	(−0.001,−0.014)	(0.005,−0.009)	(0.006,−0.022)	(0.011,−0.017)
SSE	(0.114,0.152)	(0.118,0.145)	(0.175,0.263)	(0.175,0.262)
ESE	(0.118,0.152)	(0.116,0.145)	(0.177,0.274)	(0.184,0.278)
CP	(0.934,0.948)	(0.934,0.950)	(0.928,0.954)	(0.950,0.956)

Table 4
Simulation results of β for Case 4.

	$\Lambda_0(t)=t$		$\Lambda_0(t)=3\sqrt{t}$	
	Method 1	Method 2	Method 1	Method 2
	Sample Size: $n=100$			
Bias	(0.055,−0.023)	(0.077,−0.009)	(0.074,0.024)	(0.085,−0.066)
SSE	(0.359,0.433)	(0.368,0.412)	(0.406,0.642)	(0.399,0.770)
ESE	(0.345,0.399)	(0.366,0.406)	(0.374,0.619)	(0.442,0.701)
CP	(0.928,0.934)	(0.936,0.954)	(0.932,0.940)	(0.958,0.938)
	Sample Size: $n=200$			
Bias	(0.016,−0.010)	(0.022,0.001)	(0.004,−0.005)	(0.011,−0.013)
SSE	(0.205,0.289)	(0.178,0.278)	(0.259,0.419)	(0.258,0.430)
ESE	(0.172,0.269)	(0.174,0.266)	(0.257,0.444)	(0.285,0.475)
CP	(0.948,0.936)	(0.962,0.942)	(0.930,0.952)	(0.958,0.970)
	Sample Size: $n=300$			
Bias	(−0.010,−0.014)	(−0.002,−0.005)	(−0.009,−0.001)	(−0.001,0.003)
SSE	(0.131,0.222)	(0.134,0.221)	(0.207,0.345)	(0.200,0.345)
ESE	(0.137,0.215)	(0.139,0.212)	(0.210,0.361)	(0.223,0.376)
CP	(0.938,0.940)	(0.940,0.946)	(0.946,0.936)	(0.966,0.964)

For comparison, we implemented the method in Chiou et al. (2018) using "panelReg" with method "AMM" in the R package "spef". Since their method in in [6] was time-consuming, we only run 200 replications on a server with Intel Xeon Gold 6148 CPU @ 2.40GHz. Figures 2 and 3 show the pointwise mean squared errors (MSEs) of the estimated nonparametric function for Cases 1–4 with $\Lambda_0(t)=t$. In these figures, the MSEs decrease as the sample size increases, and the MSEs of the I-spline estimates are smaller than those in [6] when the sample size is large, say $n=200$ or 300. That means all the estimates converge to the true vales as the sample size increases, and the proposed I-spline estimates converge faster than those in [6]. Table 5 reports the averages of biases, the MSEs of $\hat{\boldsymbol{\beta}}_n$, and the averages of the consuming time (in minutes) based on the 200 replications. We can see that all the biases are small, and we spend much

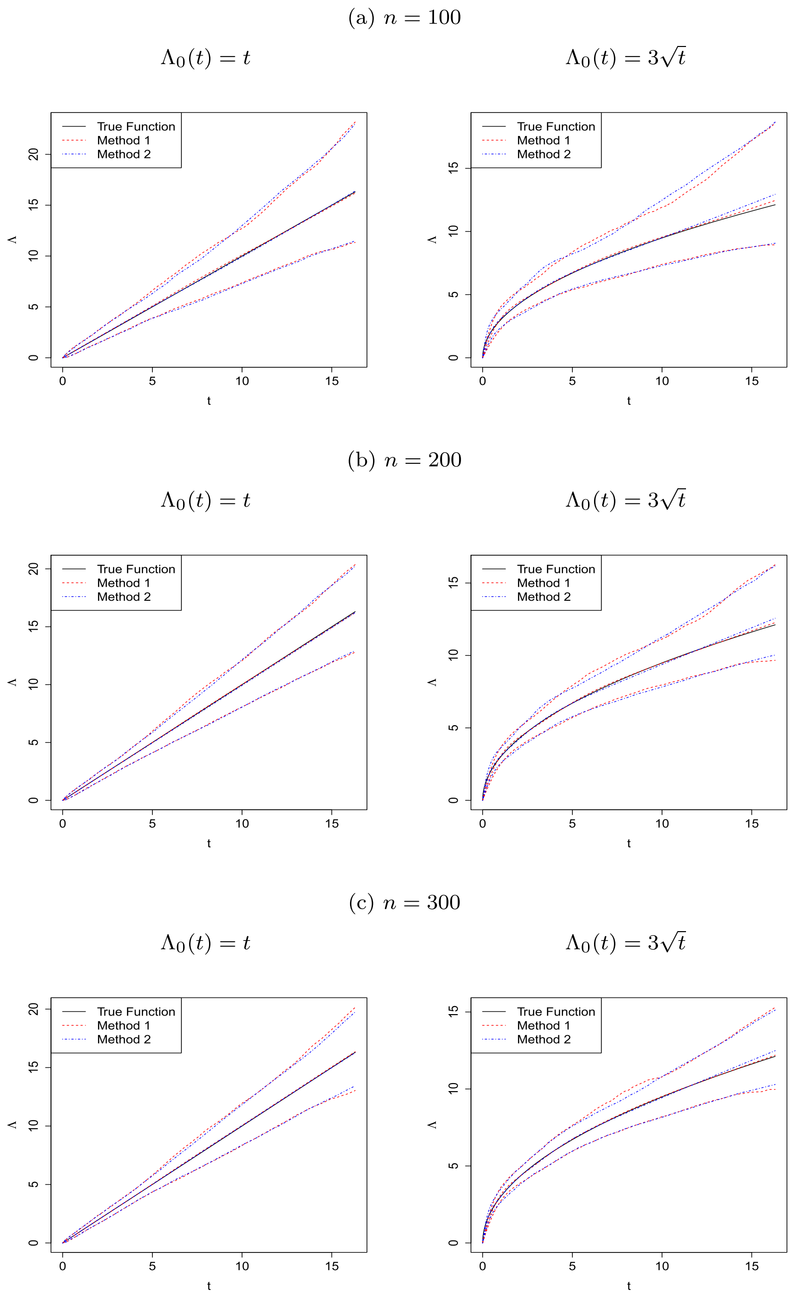

FIG 1. *Estimates of the baseline mean function* Λ *for Case 1.*

TABLE 5
Simulation results for the comparison of the proposed method with Chiou et al. (2018)'s method (CXYH).

	$\Lambda_0(t) = t$			$\Lambda_0(t) = 3\sqrt{t}$		
	CXYH	Method 1	Method 2	CXYH	Method 1	Method 2
	Case 1 and Sample Size: $n = 100$					
Bias	(−0.036,0.027)	(0.014,−0.012)	(0.031,0.007)	(−0.033,-0.003)	(0.018,0.015)	(0.010,0.014)
MSE	(0.034,0.048)	(0.054,0.039)	(0.056,0.038)	(0.052,0.082)	(0.049,0.105)	(0.049,0.105)
Time	1.543	0.168	0.189	0.873	0.130	0.084
	Case 1 and Sample Size: $n = 200$					
Bias	(−0.029,0.030)	(0.006,0.004)	(0.012,0.004)	(−0.032,0.017)	(0.002,0.021)	(0.014,0.012)
MSE	(0.034,0.050)	(0.011,0.015)	(0.011,0.014)	(0.056,0.063)	(0.021,0.046)	(0.024,0.049)
Time	5.441	0.355	0.435	2.529	0.259	0.200
	Case 1 and Sample Size: $n = 300$					
Bias	(−0.019,−0.007)	(0.002,−0.013)	(0.006,−0.010)	(−0.013,0.010)	(−0.019,0.002)	(−0.017,0.005)
MSE	(0.025,0.021)	(0.007,0.011)	(0.007,0.010)	(0.014,0.033)	(0.013,0.032)	(0.013,0.032)
Time	6.777	0.583	0.581	6.015	0.518	0.473
	Case 2 and Sample Size: $n = 100$					
Bias	(−0.010,0.053)	(0.033,0.001)	(0.041,0.019)	(−0.004,0.024)	(0.054,0.026)	(0.071,−0.018)
MSE	(0.041,0.109)	(0.079,0.121)	(0.042,0.113)	(0.070,0.220)	(0.144,0.380)	(0.228,0.435)
Time	1.815	0.194	0.142	0.915	0.079	0.100
	Case 2 and Sample Size: $n = 200$					
Bias	(−0.001,0.020)	(−0.002,−0.009)	(0.019,0.006)	(−0.014,0.049)	(−0.003,0.022)	(0.004,0.013)
MSE	(0.020,0.064)	(0.023,0.055)	(0.023,0.054)	(0.038,0.105)	(0.062,0.159)	(0.058,0.173)
Time	5.406	0.385	0.275	2.832	0.258	0.210
	Case 2 and Sample Size: $n = 300$					
Bias	(−0.008,0.022)	(−0.001,−0.014)	(0.014,−0.003)	(−0.028,0.022)	(−0.024,−0.007)	(−0.012,0.010)
MSE	(0.028,0.055)	(0.012,0.034)	(0.011,0.031)	(0.030,0.075)	(0.038,0.116)	(0.038,0.108)
Time	11.225	0.875	0.937	8.985	0.390	0.815
	Case 3 and Sample Size: $n = 100$					
Bias	(−0.030,0.015)	(0.046,−0.016)	(0.047,0.002)	(−0.043,-0.009)	(0.042,0.010)	(0.038,0.001)
MSE	(0.050,0.074)	(0.096,0.081)	(0.080,0.072)	(0.103,0.164)	(0.169,0.214)	(0.161,0.204)
Time	1.591	0.198	0.183	1.335	0.116	0.138
	Case 3 and Sample Size: $n = 200$					
Bias	(−0.016,0.036)	(0.017,−0.015)	(0.017,−0.006)	(−0.023,0.001)	(0.011,0.002)	(0.025,0.003)
MSE	(0.019,0.042)	(0.025,0.033)	(0.021,0.029)	(0.071,0.084)	(0.044,0.093)	(0.044,0.093)
Time	8.183	0.361	0.299	2.946	0.265	0.169
	Case 3 and Sample Size: $n = 300$					
Bias	(−0.037,0.035)	(0.006,−0.007)	(0.010,−0.006)	(−0.028,−0.009)	(0.009,−0.024)	(0.014,−0.029)
MSE	(0.048,0.033)	(0.015,0.022)	(0.016,0.020)	(0.029,0.049)	(0.029,0.060)	(0.029,0.064)
Time	11.810	0.552	0.637	8.442	0.341	0.564
	Case 4 and Sample Size: $n = 100$					
Bias	(−0.023,0.038)	(0.040,−0.008)	(0.055,−0.001)	(0.005,0.003)	(0.080,−0.048)	(0.085,−0.066)
MSE	(0.056,0.106)	(0.062,0.186)	(0.068,0.178)	(0.103,0.256)	(0.330,0.510)	(0.449,0.501)
Time	1.977	0.245	0.123	1.385	0.080	0.098
	Case 4 and Sample Size: $n = 200$					
Bias	(−0.009,0.042)	(−0.006,−0.013)	(0.012,−0.003)	(0.037,0.017)	(0.036,−0.012)	(0.044,−0.026)
MSE	(0.039,0.067)	(0.035,0.074)	(0.034,0.067)	(0.056,0.151)	(0.090,0.214)	(0.084,0.210)
Time	7.202	0.414	0.253	4.172	0.184	0.216
	Case 4 and Sample Size: $n = 300$					
Bias	(−0.025,0.006)	(−0.012,0.006)	(0.004,0.014)	(−0.033,0.018)	(−0.010,−0.001)	(−0.001,−0.006)
MSE	(0.028,0.061)	(0.018,0.062)	(0.018,0.055)	(0.047,0.100)	(0.050,0.135)	(0.051,0.133)
Time	16.794	0.775	0.553	7.978	0.375	0.441

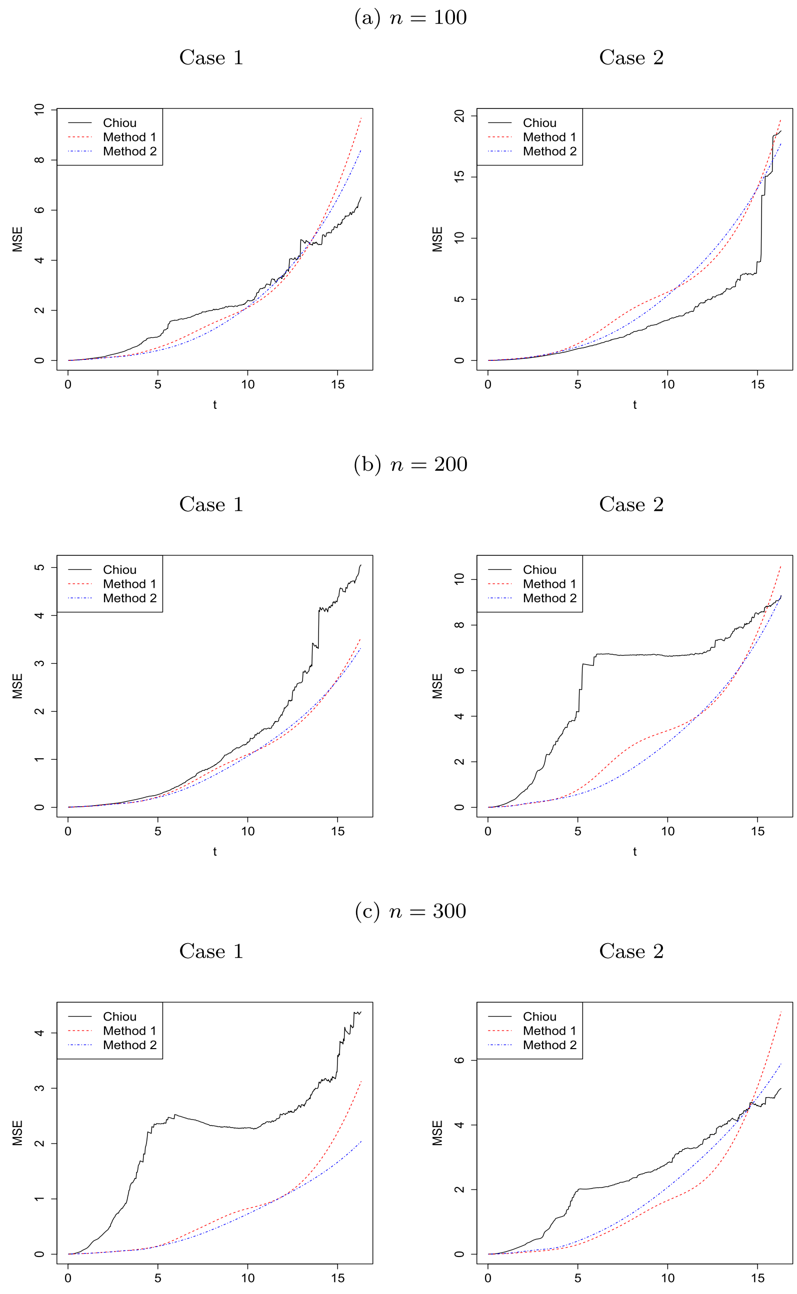

Fig 2. *Pointwise MSEs of $\hat{\Lambda}_n$ for Cases 1 and 2 with $\Lambda_0(t) = t$.*

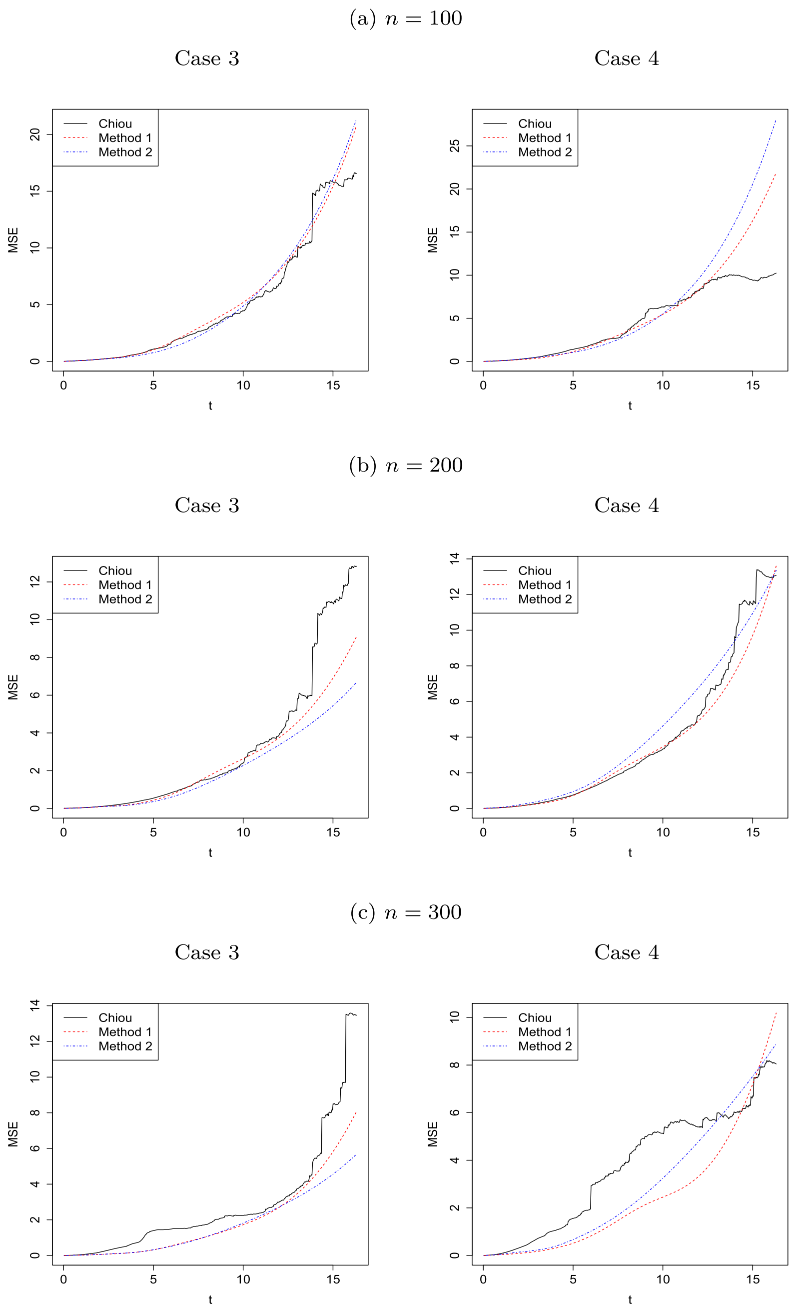

FIG 3. *Pointwise MSEs of $\hat{\Lambda}_n$ for Cases 3 and 4 with $\Lambda_0(t) = t$.*

more time for the method in [6] than our methods. Furthermore, the MSEs of $\hat{\boldsymbol{\beta}}_n$ in our methods are smaller than those in [6] when $\Lambda_0(t) = t$ and the sample size is large.

5. Application

We applied the proposed approach to analyzing the data from a skin cancer trial conducted by the University of Wisconsin Comprehensive Cancer Center [1, 16]. This double-blind and placebo-controlled clinical trial aimed to investigate the treatment effect of 0.5 g/m2/day PO difluoromethylornithine (DFMO) on the occurrence rates of combined two types of non-melanoma skin cancers (NMSC): basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). Trial participants were selected from a population of patients with a history of non-melanoma skin cancers. During the initial screening, existing tumors were completely eradicated. Subsequently, patients were scheduled follow-up assessments every six months, where the number of new tumors grown between each observation would be recorded. In reality, the true observation times of each patient could differ drastically due to a variety of reasons.

The dataset is available as *skinTumor* from the R package "spef", which includes 290 patients, with 147 assigned to receive placebo and 143 treated with DFMO. We applied the proposed approach to analyzing the treatment effect on the recurrence rate of non-melanoma tumors, with tumor relapse between observations as the recurrent event process of interest and observation times the number of days elapsed from the previous clinical visit to the next. We set $\tau = 1879$, which is the longest follow-up time. Four covariates were included in the analysis: the treatment indicator (1 for DFMO, 0 for placebo), the number of prior tumor, gender (1 for male, 0 for female), and age.

To investigate the proportional mean assumption, we first used the nonparametric least squares estimation procedure to analyze the occurrence rates of skin cancers for patients with DFMO and placebo, respectively. Let

$$E(N(t)) = \Lambda(t).$$

Then the estimator of the true mean function was obtained by minimizing the following loss function

$$\ell_n(\Lambda) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left\{ N_i(T_{ij}) - \Lambda(T_{ij}) \right\}^2 .$$

We also approximated $\Lambda(t)$ by the monotone cubic I-spline function with $d = m_n = 4$, and the knots $t_{d+1}, \cdots, t_{d+m_n}$ were the $1/(m_n+1), \cdots, m_n/(m_n+1)$ percentiles of $\{T_{ij} : i = 1, \cdots n, j = 1, \cdots, K_i\}$. Figure 4 shows the estimates for the occurrence means of BCC, SCC, and NMSC for patients with DFMO and placebo, respectively, and indicates that the proportional mean assumption is not satisfied. Thus, we considered the accelerated mean model to analyze the data.

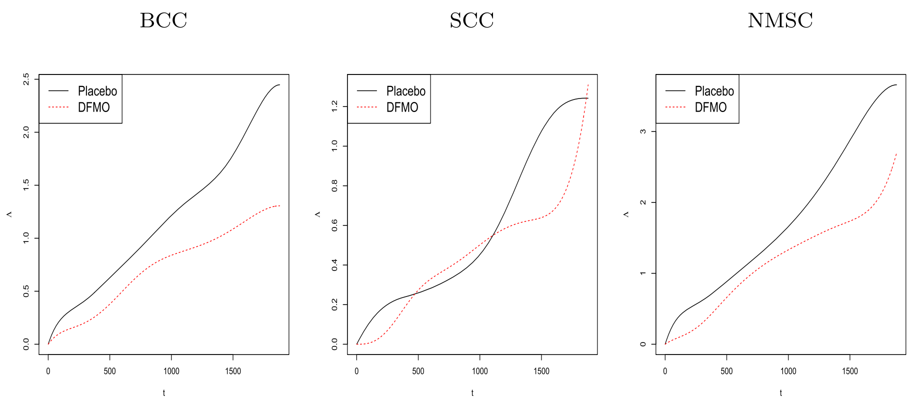

FIG 4. *Estimates for the baseline mean functions of skin cancers in different treatment groups under the nonparametric model.*

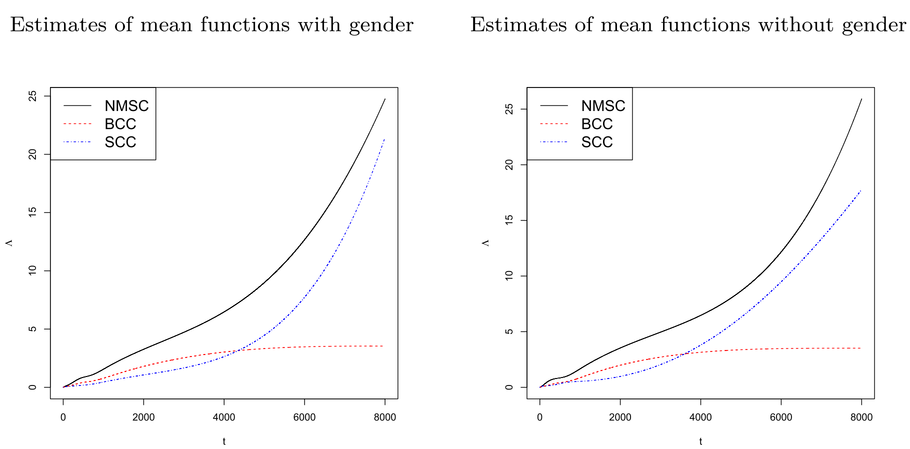

FIG 5. *Estimates of the baseline mean functions for the occurrence rates of NMSC, BCC, and SCC under the accelerated mean model.*

TABLE 6
Estimation results of regression coefficients in the skin cancer data.

	BCC			SCC			NMSC		
Risk Factor	Estimate	ESE	p-value	Estimate	ESE	p-value	Estimate	ESE	p-value
DFMO	−0.167	0.430	0.699	−0.445	0.344	0.196	−0.391	0.205	0.056
Prior Tumor	0.945	0.280	0.001***	0.268	0.062	$<$ 0.001***	0.293	0.069	$<$ 0.001***
Gender	0.146	0.242	0.548	−0.007	0.376	0.986	0.094	0.192	0.627
Age	−0.355	0.158	0.025*	0.398	0.153	0.009**	0.230	0.138	0.096
DFMO	−0.135	0.305	0.659	−0.565	0.395	0.153	−0.398	0.169	0.018*
Prior Tumor	0.930	0.254	0.003**	0.270	0.061	$<$ 0.001***	0.306	0.026	$<$ 0.001***
Age	−0.336	0.150	0.025*	0.380	0.129	0.003***	0.260	0.116	0.025*

Level of significance: * 0.05   ** 0.01   *** 0.005

Suppose three recurrent event processes from NMSC, BCC and SCC follow the accelerated mean model (2.1). We implemented the L-BFGS-B algorithm with initial values $\widetilde{\boldsymbol{\beta}} = \mathbf{0}$ and $\widetilde{\boldsymbol{\xi}} = (1, 1, 1, 1, 1, 1, 1, 1)$ as well as constant $b = 10000$ for the interval $[0, b]$. We considered the cubic spline functions with the knots t_{d+1}, t_{d+2}, t_{d+3}, and t_{d+4} being the 20%, 40%, 60% and 80% quartiles of $\{T^*_{ij}(\widetilde{\boldsymbol{\beta}}) : i = 1, \cdots n, j = 1, \cdots, K_i\}$. Figure 5 displays the estimated baseline mean functions based on the data with and without gender. The black solid line, red dashed line, and blue dotted line represent estimates for NMSC, BCC and SCC, respectively. Table 6 summarizes the coefficient estimates and the corresponding p-values. Under the fitted accelerated mean model, the covariate vector $\mathbf{X}$ affects the frequency of recurrences over time by expanding or contracting the time scale on which the events occur by a multiplicative factor of $\exp(\hat{\boldsymbol{\beta}}_n^T \mathbf{X})$ relative to that of a zero-valued covariate vector [18]. For example, for fixed values of Prior Tumor and Age, DFMO reduced the frequency of tumor recurrences over time by contracting the time scale on which the events occur by a multiplicative factor of $\exp(-0.398)$ relative to that of covariate vector (0, Prior Tumor, Age). In particular, we have the following findings:

(i) The recurrent event processes arising from BCC, SCC and NMSC were significantly associated with the number of prior skin tumors, while gender did not have any significant relationship on three recurrent event processes. These results are consistent with those obtained by [6, 16].
(ii) The DFMO significantly reduced the recurrence of NMSC.
(iii) Three recurrent event processes were significantly associated with age. The recurrent rates of SCC and NMSC increased with age, while the recurrent rate of BCC decreased with age.

6. Concluding remarks

An accelerated mean model provides an important alternative when the proportionality assumption in a proportional mean model is violated. For analyzing recurrent event processes with panel count data, we proposed to use semiparametric accelerated mean models and developed a novel sieve least squares estimation procedure for model parameters. We overcome the theoretical challenges from bundled parameters and established the asymptotic properties of the proposed estimators using empirical process theory. Results from simulation studies and applications to the skin cancer dataset confirmed that the proposed method performs reasonably well. In particular, it seems that we are the first to derive the limiting distributions of both finite-dimensional estimator and functionals of infinite-dimensional estimator for an accelerated mean model based on panel count data.

Compared with [6], our method has advantages in the following three aspects. First, we establish the asymptotic properties for the accelerated mean model with panel count data in Theorem 3.1–3.3. Our idea can be extend to dealing with the model considered by [6]. Under their model, the conditional log-likelihood function is given by

$$\ell_n(\boldsymbol{\beta}, \Lambda) = -\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Big\{\big(N_i(T_{ij}) - N_i(T_{i(j-1)})\big) \times \log\frac{\Lambda(T_{ij}\exp(\mathbf{X}^T\boldsymbol{\beta})) - \Lambda(T_{i(j-1)}\exp(\mathbf{X}^T\boldsymbol{\beta}))}{\Lambda(T_{iK_i}\exp(\mathbf{X}^T\boldsymbol{\beta}))}\Big\},$$

where $T_{i0} \equiv 0$, and Λ can be approximated by a spline function. Then we can treat it as a semiparametric sieve M-estimation with bundled parameters. Define $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$ be the estimator obtained by maximizing $\ell_n(\boldsymbol{\beta}, \Lambda)$ with respect to $\mathcal{R} \times \mathcal{G}_n$. Similar to the proofs of theorems in our paper, we can show that the estimator $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$ converges to the true value $(\boldsymbol{\beta}_0, \Lambda_0)$ in probability at an overall convergence rate $n^{-r/(1+2r)}$ and further establish the asymptotic normality for $\hat{\boldsymbol{\beta}}_n$ and functionals of $\hat{\Lambda}_n$. Second, the proposed least squares estimation procedure relaxed the Poisson assumption for the recurrent event process required in [6]. Third, our sieve I-spline estimator has a faster convergence rate than the step function estimator in [6] and our estimation procedure is easy to implement and takes a shorter time. The simulation results in Table 5 and Figures 2 and 3 coincide with the theoretical conclusions.

A further research is to study recurrent event processes with informative observation times and terminal events. For these general situations, we can consider two modeling approaches. One is to directly characterize their relationships by treating observation times and terminal events as covariates; another is to use shared frailty variables for modeling correlated recurrent event processes, observation times, and terminal events. We can explore sieve least squares or estimating equation methods for model parameters. The new challenges arising from bundled parameters will definitely require more innovative research efforts.

Appendix A: Lemmas

Lemma 1. *Suppose that Conditions* (C7)–(C9) *hold. Then for any* $\Lambda \in \mathcal{H}_r$ *with* $d_1(\Lambda, \Lambda_0) \le \eta$*, we have* $||\Lambda - \Lambda_0||_\infty \lesssim \eta^{2/3}$*, where* $||\cdot||_\infty$ *denotes the infinity norm of a function.*

Proof. The proof of this lemma is similar to the proof of Lemma 7.1 of [31] and is omitted here. □

Lemma 2. *Suppose that Condition* (C4) *holds. For*

$$\mathcal{G}_{n,\eta} := \{\Lambda : \Lambda \in \mathcal{G}_n, d_1(\Lambda, \Lambda_0) \le \eta\},$$

set $N_{[]}(\varepsilon, \mathcal{G}_{n,\eta}, ||\cdot||_\infty)$ *to be the* ε*-bracketing number of* $\mathcal{G}_{n,\eta}$ *under the norm* $||\cdot||_\infty$*. Then we have* $\log N_{[]}(\varepsilon, \mathcal{G}_{n,\eta}, ||\cdot||_\infty) \lesssim q_n \log(\eta/\varepsilon)$*.*

Proof. By the calculation of [23], there exists an $\varepsilon/2$-net

$$\{\Lambda_i : i = 1, \cdots, \lceil (2\eta/\varepsilon)^{cq_n} \rceil\}$$

with a constant c such that for all $\Lambda \in \mathcal{G}_{n,\eta}$, we can find a Λ_i satisfying $||\Lambda_i - \Lambda||_\infty \leq \varepsilon/2$. Set $\Lambda_i^L = \Lambda_i - \varepsilon/2$ and $\Lambda_i^U = \Lambda_i + \varepsilon/2$ for $i = 1, \cdots, \lceil (2\eta/\varepsilon)^{cq_n} \rceil$. Then for all $\Lambda \in \mathcal{G}_{n,\eta}$, we can find a $(\Lambda_i^L, \Lambda_i^U)$ from the set

$$\left\{ (\Lambda_i^L, \Lambda_i^U) : i = 1, \cdots, \lceil (2\eta/\varepsilon)^{cq_n} \rceil \right\}$$

such that $\Lambda_i^L \leq \Lambda \leq \Lambda_i^U$. Noting that $||\Lambda_i^U - \Lambda_i^L||_\infty \equiv \varepsilon$, we have $N_{[]}(\varepsilon, \mathcal{G}_{n,\eta}, ||\cdot||_\infty) \leq (2\eta/\varepsilon)^{cq_n}$. It follows that $\log N_{[]}(\varepsilon, \mathcal{G}_{n,\eta}, ||\cdot||_\infty) \lesssim q_n \log(\eta/\varepsilon)$. □

Lemma 3. *Suppose that Conditions* (C1), (C2) *and* (C4)–(C9) *hold. Define the class*

$$\mathcal{L}_\eta = \{ L(\boldsymbol{\beta}, \Lambda; \mathbf{O}) - L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) : d_2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq \eta, \boldsymbol{\beta} \in \mathcal{R}, \Lambda \in \mathcal{G}_n \}.$$

Then for sufficiently small ε, the ε-bracketing number with $||\cdot||_{P,B}$ for $\mathcal{L}_\eta$ satisfies

$$\log N_{[]}(\varepsilon, \mathcal{L}_\eta, ||\cdot||_{P,B}) \lesssim q_n \log(\eta/\varepsilon),$$

where $||\cdot||_{P,B}$ is the Bernstein norm defined as $||f||_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$.

Proof. This lemma can be showed by Lemmas 1 and 2 with the conditions. □

Appendix B: Proofs

B.1. Proof of Theorem 3.1

Proof. By direct calculations, we obtain

$$\mathcal{P}L(\boldsymbol{\beta}, \Lambda; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) = \mathcal{P}\left[\sum_{j=1}^{K} \left\{ \Lambda(T_j^*(\boldsymbol{\beta})) - \Lambda_0(T_{0j}^*) \right\}^2 \right]. \quad \text{(B.1)}$$

Then we have $\mathcal{P}L(\boldsymbol{\beta}, \Lambda; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) \geq 0$, and it equals 0 if and only if $\Lambda(T_j^*(\boldsymbol{\beta})) = \Lambda_0(T_{0j}^*)$ for $j = 1, \cdots, K$ with probability 1. It follows that $T_j e^{\mathbf{X}^T \boldsymbol{\beta}} = g(T_j e^{\mathbf{X}^T \boldsymbol{\beta}_0})$, where $g = \Lambda^{-1} \circ \Lambda_0$ is an increasing and differentiable function. Taking the derivative with respect to T_j at both sides, we obtain $\mathbf{X}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = \log g'(T_{0j}^*)$. Condition (C3) implies that $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\Lambda = \Lambda_0$. That is, $\mathcal{P}L(\boldsymbol{\beta}, \Lambda; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) = 0$ if and only if $(\boldsymbol{\beta}, \Lambda) = (\boldsymbol{\beta}_0, \Lambda_0)$. Since $\mathcal{P}L(\boldsymbol{\beta}, \Lambda; \mathbf{O})$ is continuous with respect to $(\boldsymbol{\beta}, \Lambda)$, for any small constant $\delta > 0$, there exists an $\varepsilon > 0$ such that

$$\inf_{d_2((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0)) \geq \delta} \mathcal{P}L(\boldsymbol{\beta}, \Lambda; \mathbf{O}) \geq \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) + \varepsilon,$$

which yields that

$$\{\hat{\boldsymbol{\theta}}_n : d_2(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) \geq \delta\} \subset \{\hat{\boldsymbol{\theta}}_n : \mathcal{P}L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) \geq \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) + \varepsilon\}.$$

Then we turn to consider the term $\mathcal{P}L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})$. Under Conditions (C1), (C2) and (C4), Lemma A1 of [19] implies that there is a function $\Lambda_n^* \in \mathcal{G}_n$ such that $||\Lambda_0 - \Lambda_n^*||_\infty = O(n^{-\nu r}) = o(1)$. By the definition of $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$, we have $\mathbb{P}_n L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - \mathbb{P}_n L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) \leq 0$. Moreover, by (B.1), we have $\mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) \lesssim ||\Lambda_0 - \Lambda_n^*||_\infty^2 = o(1)$. This implies that

$$
\begin{aligned}
0 \leq & \mathcal{P}L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) \\
\lesssim & \mathcal{P}L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - \mathbb{P}_n L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) + \mathbb{P}_n L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) + o(1).
\end{aligned}
$$

Since the I-spline basis functions are the summation of some B-spline basis functions [22], and the B-spline function space is compact [23], we have $\mathcal{G}_n$ is compact. Note that $L(\boldsymbol{\beta}, \Lambda; \mathbf{O})$ is continuous with respect to $(\boldsymbol{\beta}, \Lambda)$, and $\mathcal{P}L(\boldsymbol{\beta}, \Lambda; \mathbf{O}) \geq \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})$ under Condition (C5). Following the one-sided Gilvenko-Cantelli theorem [30], we obtain $\lim_{n\to\infty}(\mathbb{P}_n - \mathcal{P})L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) \geq 0$. Furthermore, $\mathcal{L}_\eta$ is Donsker by Conditions (C1), (C2), (C6) and (C7). By the Taylor expansion, under Conditions (C1), (C2) and (C5), for any $\boldsymbol{\theta} = (\boldsymbol{\beta}, \Lambda)$ with $d_2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq \eta$, we can obtain $\mathcal{P}\left[\sum_{j=1}^K \left\{\Lambda(T_j^*(\boldsymbol{\beta})) - \Lambda_0(T_{0j}^*)\right\}^2\right] \lesssim d_2^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + \eta^2 \lesssim \eta^2$.Thus, $\mathcal{P}f^2 \lesssim \eta^2$ for all $f \in \mathcal{L}_\eta$. This implies that $\sup_{f\in\mathcal{L}_\eta} \rho_{\mathcal{P}}(f) \leq \sup_{f\in\mathcal{L}_\eta}\{\mathcal{P}f^2\}^{1/2} \to 0$ as $\eta \to 0$ for the seminorm $\rho_{\mathcal{P}}(f) = \{\mathcal{P}(f - \mathcal{P}f)^2\}^{1/2}$. By Corollary 2.3.12 of [28], $(\mathbb{P}_n - \mathcal{P})\big\{L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) - L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})\big\} = o_p(1)$. Since $(\mathbb{P}_n - \mathcal{P})L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) = o_p(1)$, it follows that $(\mathbb{P}_n - \mathcal{P})L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) = o_p(1)$. Hence, we obtain

$$
0 \leq \mathcal{P}L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) \leq o_p(1).
$$

Therefore, for all $\varepsilon > 0$, we have $P\{\hat{\boldsymbol{\theta}}_n : \mathcal{P}L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) \geq \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) + \varepsilon\} \to 0$ as $n \to \infty$.

Noting that $\{\hat{\boldsymbol{\theta}}_n : d_2(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) \geq \delta\} \subset \{\hat{\boldsymbol{\theta}}_n : \mathcal{P}L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) \geq \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) + \varepsilon\}$, we have $P\{\hat{\boldsymbol{\theta}}_n : d_2(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) \geq \delta\} \to 0$ for all $\delta > 0$. Thus, $d_2(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = o_p(1)$. □

B.2. Proof of Theorem 3.2

Proof. We use Theorem 3.2.5 of [28] to prove this theorem.

For $\boldsymbol{\theta} = (\boldsymbol{\beta}, \Lambda)$ in the neighborhood of $\boldsymbol{\theta}_0$, by the Taylor expansion,

$$
\begin{aligned}
& \mathcal{P}\left[\sum_{j=1}^K \left\{\Lambda(T_j^*(\boldsymbol{\beta})) - \Lambda_0(T_{0j}^*)\right\}^2\right] \\
= & \mathcal{P}\left[\sum_{j=1}^K \left\{\Lambda'(T_{0j}^*)T_{0j}^*\mathbf{X}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\}^2\right] + d_1^2(\Lambda, \Lambda_0) \\
& + 2\mathcal{P}\left[\sum_{j=1}^K \left\{\Lambda'(T_{0j}^*)T_{0j}^*\mathbf{X}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)(\Lambda(T_{0j}^*) - \Lambda_0(T_{0j}^*))\right\}\right] + o\big(d_2^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\big).
\end{aligned} \tag{B.2}
$$

By the Cauchy-Schwarz inequality and Condition (C10), we obtain

$$\begin{aligned}
&\left|\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Lambda'(T_{0j}^*)T_{0j}^*\mathbf{X}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)(\Lambda(T_{0j}^*)-\Lambda_0(T_{0j}^*))\right\}\right]\right| \\
=&\left|E_1\left[\Lambda'(U)U\mathbf{X}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)(\Lambda(U)-\Lambda_0(U))\right]\right| \\
=&\left|E_1\left[E\left[\Lambda'(U)U\mathbf{X}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|U\right](\Lambda(U)-\Lambda_0(U))\right]\right| \\
\le& E_1^{\frac{1}{2}}\left[E^2\left[\Lambda'(U)U\mathbf{X}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|U\right]\right]E_1^{\frac{1}{2}}\left[(\Lambda(U)-\Lambda_0(U))^2\right] \\
=& d_1(\Lambda,\Lambda_0)E_1^{\frac{1}{2}}\left[\Lambda'(U)^2U^2(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^TE\left[\mathbf{X}|U\right]E\left[\mathbf{X}|U\right]^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right] \\
\le&(1-M_6)^{\frac{1}{2}}d_1(\Lambda,\Lambda_0)E_1^{\frac{1}{2}}\left[\Lambda'(U)^2U^2(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^TE\left[\mathbf{X}\mathbf{X}^T|U\right](\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right] \\
\le&\frac{(1-M_6)^{\frac{1}{2}}}{2}\left(d_1^2(\Lambda,\Lambda_0)+\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Lambda'(T_{0j}^*)T_{0j}^*\mathbf{X}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\}^2\right]\right),
\end{aligned} \tag{B.3}$$

where E_1 is the expectation under μ_1. By (B.2) and (B.3), and Conditions (C8)–(C10), we have

$$\begin{aligned}
&\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Lambda(T_j^*(\boldsymbol{\beta}))-\Lambda_0(T_{0j}^*)\right\}^2\right] \\
\gtrsim& d_1^2(\Lambda,\Lambda_0)+E\left[\left\{\mathbf{X}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\}^2\right]+o\big(d_2^2(\boldsymbol{\theta},\boldsymbol{\theta}_0)\big) \\
\gtrsim& d_1^2(\Lambda,\Lambda_0)+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_2^2+o\big(d_2^2(\boldsymbol{\theta},\boldsymbol{\theta}_0)\big)=d_2^2(\boldsymbol{\theta},\boldsymbol{\theta}_0)+o\big(d_2^2(\boldsymbol{\theta},\boldsymbol{\theta}_0)\big).
\end{aligned}$$

It follows that $\mathcal{P}L(\boldsymbol{\beta},\Lambda;\mathbf{O})-\mathcal{P}L(\boldsymbol{\beta}_0,\Lambda_0;\mathbf{O})\gtrsim d_2^2(\boldsymbol{\theta},\boldsymbol{\theta}_0)$.

Note that for any $\eta>0$,

$$\mathcal{L}_\eta=\{L(\boldsymbol{\beta},\Lambda;\mathbf{O})-L(\boldsymbol{\beta}_0,\Lambda_0;\mathbf{O}):d_2((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))\le\eta,\boldsymbol{\beta}\in\mathcal{R},\Lambda\in\mathcal{H}_r\}.$$

Then we turn to find a function $\phi_n(\eta)$ such that $E||n^{1/2}(\mathbb{P}_n-\mathcal{P})||_{\mathcal{L}_\eta}\lesssim\phi_n(\eta)$, where $\|n^{1/2}(\mathbb{P}_n-\mathcal{P})\|_{\mathcal{L}_\eta}=\sup_{f\in\mathcal{L}_\eta}|n^{1/2}(\mathbb{P}_n-\mathcal{P})f|$. Lemma 3 implies that $\log N_{[]}(\varepsilon,\mathcal{L}_\eta,||\cdot||_{P,B})\lesssim q_n\log(\eta/\varepsilon)$. Note that for all $\boldsymbol{\beta}\in\mathcal{R}$ and $\Lambda\in\mathcal{H}_r$,

$$|L(\boldsymbol{\beta},\Lambda;\mathbf{O})-L(\boldsymbol{\beta}_0,\Lambda_0;\mathbf{O})|\lesssim\sum_{j=1}^{K}(N(T_j)+1)|\Lambda(T_j^*(\boldsymbol{\beta}))-\Lambda_0(T_{0j}^*)|$$

holds with probability 1. Thus, $||f||_{P,B}^2\le\eta^2$ for all $f\in\mathcal{L}_\eta$. By Lemma 3.4.3 of [28], we have

$$E||n^{1/2}(\mathbb{P}_n-\mathcal{P})||_{\mathcal{L}_\eta}\lesssim J_{[]}(\eta,\mathcal{L}_\eta,||\cdot||_{P,B})\left\{1+\frac{J_{[]}(\eta,\mathcal{L}_\eta,||\cdot||_{P,B})}{\eta^2n^{1/2}}\right\},$$

where $J_{[]}(\eta,\mathcal{L}_\eta,||\cdot||_{P,B}):=\int_0^\eta\{1+\log N_{[]}(\varepsilon,\mathcal{L}_\eta,||\cdot||_{P,B})\}^{1/2}d\varepsilon\lesssim q_n^{1/2}\eta$. It follows that $E||n^{1/2}(\mathbb{P}_n-\mathcal{P})||_{\mathcal{L}_\eta}\lesssim\phi_n(\eta)$, where $\phi_n(\eta)=q_n^{1/2}\eta+q_nn^{-1/2}$ and

$\phi_n(\eta)/\eta$ decreases with η. For a sequence $r_n = O(n^a)$, since $q_n = O(n^\nu)$ with $0 < \nu < 1/2$, we have $r_n^2\phi(1/r_n) = O(n^{a+\frac{\nu}{2}} + n^{2a+\nu-\frac{1}{2}})$. Then $a \leq (1-\nu)/2$ ensures $r_n^2\phi(1/r_n) \lesssim n^{1/2}$, and so we choose $r_n = O(n^{(1-\nu)/2})$. Finally, we find ν such that $\mathbb{P}_n(L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})) \leq O_p(r_n^{-2})$. Since $\mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) \leq ||\Lambda_0 - \Lambda_n^*||_\infty^2$ and by the definition of $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$, we have

$$
\begin{aligned}
&\mathbb{P}_n\big(L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})\big) \\
=&\mathbb{P}_n L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - \mathbb{P}_n L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) + \mathbb{P}_n L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) \\
&+\mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) - \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) + \mathcal{P}L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) - \mathbb{P}_n L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O}) \\
\leq& n^{-\nu r+\varepsilon}(\mathbb{P}_n - \mathcal{P}) \left(\frac{L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) - L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})}{n^{-\nu r+\varepsilon}}\right) + O(n^{-2\nu r}),
\end{aligned}
$$

where $0 < \varepsilon < 1/2 - r\nu$. Furthermore, note that under Conditions (C1), (C2), (C5) and (C6),

$$
\mathcal{L} = \left\{\frac{L(\boldsymbol{\beta}_0, \Lambda; \mathbf{O}) - L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})}{n^{-\nu r+\varepsilon}} : \Lambda \in \mathcal{G}_n, ||\Lambda - \Lambda_0||_\infty = O(n^{-\nu r})\right\}
$$

is Donsker. Moreover, for the seminorm $\rho_{\mathcal{P}}(f) = \{\mathcal{P}(f - \mathcal{P}f)^2\}^{1/2}$, we have $\sup_{f\in\mathcal{L}} \rho_{\mathcal{P}}(f) \leq \sup_{f\in\mathcal{L}}\{\mathcal{P}f^2\}^{1/2} \to 0$, as $n \to \infty$ for all $f \in \mathcal{L}$. By Corollary 2.3.12 of [28],

$$
n^{-\nu r+\varepsilon}(\mathbb{P}_n - \mathcal{P}) \left(\frac{L(\boldsymbol{\beta}_0, \Lambda_n^*; \mathbf{O}) - L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})}{n^{-\nu r+\varepsilon}}\right) = o_p(n^{-\nu r+\varepsilon-1/2}) = o_p(n^{-2\nu r}).
$$

It follows that $\mathbb{P}_n\big(L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})\big) \leq O_p(n^{-2\nu r})$. Then $\nu \geq 1/(1+2r)$ ensures $\mathbb{P}_n\big(L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O}) - L(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})\big) \leq O_p(r_n^{-2})$. Thus, taking $\nu = 1/(1+2r)$, we obtain $d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p(n^{-r/(1+2r)})$. □

B.3. Proof of Theorem 3.3

Proof. (i) Following Theorem 1 of [36], it suffices to verify the following conditions (A1)–(A5) to prove this theorem.

(A1) $\sqrt{n}(G_n - G)(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{h}_1, h_2] - \sqrt{n}(G_n - G)(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1, h_2] = o_p(1)$.

(A2) $\sqrt{n}G_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{h}_1, h_2] = o_p(1)$ and $G(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1, h_2] = 0$.

(A3) The function $G(\boldsymbol{\beta}, \Lambda)[\mathbf{h}_1, h_2]$ is Fréchet-differentiable at $(\boldsymbol{\beta}_0, \Lambda_0)$ with a continuous derivative $\dot{G}_{\boldsymbol{\beta}_0,\Lambda_0}[\mathbf{h}_1, h_2]$.

(A4) $G(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{h}_1, h_2] - G(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1, h_2] - \dot{G}_{(\boldsymbol{\beta}_0,\Lambda_0)}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2] = o_p(n^{-1/2})$.

(A5) $\sqrt{n}(G_n - G)(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1, h_2]$ converges in distribution to a tight Gaussian process.

For (A1), we define

$$
\begin{aligned}
\mathcal{M}(\eta)[\mathbf{h}_1, h_2] = \Big\{&m(\boldsymbol{\beta}, \Lambda; \mathbf{O})[\mathbf{h}_1, h_2] - m(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})[\mathbf{h}_1, h_2] : d_2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq \eta, \\
&d_1(\Lambda', \Lambda_0') \leq \eta, \boldsymbol{\beta} \in \mathcal{R}, \Lambda \in \mathcal{G}_n\Big\}
\end{aligned}
$$

for fixed $(\mathbf{h}_1, h_2) \in \tilde{\mathcal{H}}$. Similar to Lemma 3, we get the upper bound of the bracketing entropy of $\mathcal{M}(\eta)[\mathbf{h}_1, h_2]$, and conclude that $\mathcal{M}(\eta)[\mathbf{h}_1, h_2]$ is Donsker. Note that $\mathcal{P}(f^2) \lesssim \eta^2$ for $f \in \mathcal{M}(\eta)[\mathbf{h}_1, h_2]$, By Corollary 2.3.12 of [28], for sufficiently small η, we have

$$\sup_{d_2((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))\leq\eta, d_1(\Lambda',\Lambda_0')\leq\eta} (\mathbb{P}_n-\mathcal{P})\big[m(\boldsymbol{\beta},\Lambda;\mathbf{O})[\mathbf{h}_1,h_2]-m(\boldsymbol{\beta}_0,\Lambda_0;\mathbf{O})[\mathbf{h}_1,h_2]\big]$$
$$=o_p(n^{-1/2}).$$

Since $d_1(\hat{\Lambda}_n, \Lambda_0) = O_p(n^{-r/(1+2r)})$, by Lemma 8 of [24], we obtain $d_1(\hat{\Lambda}_n', \Lambda_0') = O_p(n^{-(r-1)/(1+2r)}) = o_p(1)$. This yields that

$$m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O})[\mathbf{h}_1, h_2] - m(\boldsymbol{\beta}_0, \Lambda_0; \mathbf{O})[\mathbf{h}_1, h_2] \in \mathcal{M}(\eta)[\mathbf{h}_1, h_2],$$

and (A1) holds.

For (A2), note that $G(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1, h_2] = 0$ under our model. For any $\mathbf{h}_1 \in \mathcal{R}$ and $h_{2n} \in \mathcal{G}_n$, by the definition of $\hat{\boldsymbol{\beta}}_n$ and $\hat{\Lambda}_n$, we have

$$\lim_{\varepsilon\to 0} \frac{\mathbb{P}_n L\big(\hat{\boldsymbol{\beta}}_n + \varepsilon\mathbf{h}_1, \hat{\Lambda}_n + \varepsilon h_{2n}; \mathbf{O}\big) - \mathbb{P}_n L(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n; \mathbf{O})}{\varepsilon} = 0,$$

which implies that $\sqrt{n} G_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{h}_1, h_{2n}] = 0$. For any $h_2 \in \mathcal{H}_r$, there exists an $h_{2n} \in \mathcal{G}_n$ such that $||h_2 - h_{2n}||_\infty = O(n^{-\nu r})$. Thus, to verify $\sqrt{n} G_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{h}_1, h_2] = o_p(1)$, we need to prove $\sqrt{n} G_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{0}, h_2 - h_{2n}] = o_p(1)$. Note that

$$\begin{aligned} &G_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{0}, h_2 - h_{2n}] \\ =&\mathbb{P}_n\left[\sum_{j=1}^K \big\{\hat{\Lambda}_n(T_j^*(\hat{\boldsymbol{\beta}}_n)) - \Lambda_0(T_{0j}^*)\big\}\big\{(h_2 - h_{2n})(T_j^*(\hat{\boldsymbol{\beta}}_n))\big\}\right] \\ &+\mathbb{P}_n\left[\sum_{j=1}^K \big\{\Lambda_0(T_{0j}^*) - N(T_j)\big\}\big\{(h_2 - h_{2n})(T_j^*(\hat{\boldsymbol{\beta}}_n))\big\}\right] \\ =&: I_{1n} + I_{2n}. \end{aligned}$$

For I_{1n}, we have

$$\begin{aligned} \mathcal{P}|I_{1n}| \leq& \mathcal{P}\left|\sum_{j=1}^K \big\{\hat{\Lambda}_n(T_j^*(\hat{\boldsymbol{\beta}}_n)) - \Lambda_0(T_{0j}^*)\big\}\big\{(h_2 - h_{2n})(T_j^*(\hat{\boldsymbol{\beta}}_n))\big\}\right| \\ \lesssim& d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0)||h_2 - h_{2n}||_\infty. \end{aligned}$$

For I_{2n}, since $\mathcal{P}\left[\sum_{j=1}^K \big\{\Lambda_0(T_{0j}^*) - N(T_j)\big\}\big\{(h_2 - h_{2n})(T_j^*(\hat{\boldsymbol{\beta}}_n))\big\}\right] = 0$ and $\mathbf{O}_i$ are independent and identically distributed, we obtain

$$\mathcal{P} I_{2n}^2 \lesssim \frac{1}{n}\mathcal{P}\left[\sum_{j=1}^K \big\{\Lambda_0(T_{0j}^*) - N(T_j)\big\}^2\right] ||h_2 - h_{2n}||_\infty^2 = o(n^{-1}).$$

Thus, $\sqrt{n}G_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{h}_1, h_2] = o_p(1)$, and (A2) holds.

For (A3), since $G(\boldsymbol{\beta}, \Lambda)[\mathbf{h}_1, h_2]$ is a smooth function with respect to $(\boldsymbol{\beta}, \Lambda)$, we have the Fréchet derivative

$$\begin{aligned}
&\dot{G}_{(\boldsymbol{\beta}_0,\Lambda_0)}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2] \\
=& \left.\frac{dG(\boldsymbol{\beta}_0 + \eta(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0), \Lambda_0 + \eta(\hat{\Lambda}_n - \Lambda_0); \mathbf{O})[\mathbf{h}_1, h_2]}{d\eta}\right|_{\eta=0} \\
=& \sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1, h_2] + \sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2],
\end{aligned}$$

where

$$\begin{aligned}
&\sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1, h_2] \\
=& \mathcal{P}\left[\sum_{j=1}^{K}\left\{\left(\Lambda_0'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\mathbf{h}_1 + h_2(T_{0j}^*)\right)\Lambda_0'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\right\}\right](\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0), \\
&\sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2] \\
=& \mathcal{P}\left[\sum_{j=1}^{K}\left\{\left(\Lambda_0'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\mathbf{h}_1 + h_2(T_{0j}^*)\right)\left(\hat{\Lambda}_n(T_{0j}^*) - \Lambda_0(T_{0j}^*)\right)\right\}\right].
\end{aligned} \tag{B.4}$$

Then (A3) holds.

For (A4), since $\hat{\Lambda}_n \in \mathcal{G}_n \subseteq \mathcal{H}_r$ with $r \geq 3$, using the Taylor expansion, we have

$$\begin{aligned}
&G(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{h}_1, h_2] \\
=& \mathcal{P}\left[\sum_{j=1}^{K}\left\{\left(\hat{\Lambda}_n(T_{0j}^*) - \Lambda_0(T_{0j}^*)\right)\left(\hat{\Lambda}_n'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\mathbf{h}_1 + h_2(T_{0j}^*)\right)\right\}\right] \\
&+ \mathcal{P}\left[\sum_{j=1}^{K}\left\{\hat{\Lambda}_n'(T_{0j}^*)T_{0j}^*\mathbf{X}^T(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\left(\hat{\Lambda}_n'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\mathbf{h}_1 + h_2(T_{0j}^*)\right)\right\}\right] + o_p(n^{-1/2}).
\end{aligned}$$

Setting

$$\begin{aligned}
G_{1n} =& \mathcal{P}\left[\sum_{j=1}^{K}\left\{\left(\hat{\Lambda}_n'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\mathbf{h}_1 + h_2(T_{0j}^*)\right)\hat{\Lambda}_n'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\right\}\right](\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0), \\
G_{2n} =& \mathcal{P}\left[\sum_{j=1}^{K}\left\{\left(\hat{\Lambda}_n'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\mathbf{h}_1 + h_2(T_{0j}^*)\right)\left(\hat{\Lambda}_n(T_{0j}^*) - \Lambda_0(T_{0j}^*)\right)\right\}\right],
\end{aligned}$$

by (B.4), we obtain

$$G_{1n} - \sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1, h_2] \lesssim d_1(\hat{\Lambda}_n', \Lambda_0')||\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0||_2 = o_p(n^{-1/2}),$$

and

$$G_{2n} - \sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2] \lesssim d_1(\hat{\Lambda}_n', \Lambda_0')d_1(\hat{\Lambda}_n, \Lambda_0) = o_p(n^{-1/2}).$$

Since $G(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1, h_2] = 0$, it follows that

$$\begin{aligned} &G(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)[\mathbf{h}_1, h_2] - G(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1, h_2] - \dot{G}_{(\boldsymbol{\beta}_0,\Lambda_0)}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2] \\ =&G_{1n} - \sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1, h_2] + G_{2n} - \sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2] = o_p(n^{-1/2}). \end{aligned}$$

Then we have (A4).

Note that $\mathcal{R}_\beta \subseteq \mathbb{R}^p$ is bounded and $\mathcal{H}_r$ is a Donsker class. Furthermore, $m(\beta_0, \Lambda_0; \mathbf{O})[\mathbf{h}_1, h_2]$ is bounded Lipschitz function with respect to $\mathbf{h}_1$ and h_2. Thus,

$$\{m(\beta_0, \Lambda_0; \mathbf{O})[\mathbf{h}_1, h_2] : \mathbf{h}_1 \in \mathcal{H}_r \text{ and } h_2 \in \mathcal{R}_\beta\}$$

is a Donsker class, and (A5) holds.

By Theorem 1 of [36], (A1)–(A5) yield that

$$\begin{aligned} &-\sqrt{n}\dot{G}_{(\boldsymbol{\beta}_0,\Lambda_0)}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2] \\ =&\sqrt{n}(G_n - G)(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1, h_2] + o_p(1) = \sqrt{n}G_n(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1, h_2] + o_p(1) \end{aligned} \tag{B.5}$$

converges in distribution to $N(0, \sigma_0[\mathbf{h}_1, h_2]^2)$. It follows that

$$\begin{aligned} &\sqrt{n}\left(\sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1, h_2] + \sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2]\right) \\ &\quad = \sqrt{n}\dot{G}_{(\boldsymbol{\beta}_0,\Lambda_0)}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1, h_2] \end{aligned}$$

converges in distribution to $N(0, \sigma_0[\mathbf{h}_1, h_2]^2)$.

(ii) First, we consider $(\mathbf{h}_1^*, h_2^*) \in \tilde{\mathcal{H}}$ such that $\sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{h}_1^*, h_2^*] = 0$. That is

$$\begin{aligned} &\mathcal{P}\left[\sum_{j=1}^K \left\{\big(\Lambda_0'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\mathbf{h}_1^* + h_2^*(T_{0j}^*)\big)(\hat{\Lambda}_n(T_{0j}^*) - \Lambda_0(T_{0j}^*))\right\}\right] \\ =&\mathcal{P}\left[\sum_{j=1}^K \left\{\big(\Lambda_0'(T_{0j}^*)T_{0j}^*E(\mathbf{X}^T\mathbf{h}_1^*|K, T_{0j}^*) + h_2^*(T_{0j}^*)\big)(\hat{\Lambda}_n(T_{0j}^*) - \Lambda_0(T_{0j}^*))\right\}\right] = 0. \end{aligned}$$

It follows that $h_2^*(T_{0j}^*) = -\Lambda_0'(T_{0j}^*)T_{0j}^*E(\mathbf{X}|K, T_{0j}^*)^T\mathbf{h}_1^*$. Then we have

$$\sqrt{n}\sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1^*, h_2^*] = -\sqrt{n}G_n(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1^*, h_2^*] + o_p(1),$$

where

$$\begin{aligned} &\sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1^*, h_2^*] \\ &\quad = \mathbf{h}_1^{*T}\mathcal{P}\left[\sum_{j=1}^K \left\{\Lambda_0'(T_{0j}^*)^2(T_{0j}^*)^2\big(\mathbf{X} - E(\mathbf{X}|K, T_{0j}^*)\big)^{\otimes 2}\right\}\right](\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0), \\ &\quad\; - G_n(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{h}_1^*, h_2^*] \\ &\quad = \mathbf{h}_1^{*T}\mathbb{P}_n\left[\sum_{j=1}^K \left\{(N(T_j) - \Lambda_0(T_{0j}^*))\Lambda_0'(T_{0j}^*)T_{0j}^*\big(\mathbf{X} - E(\mathbf{X}|K, T_{0j}^*)\big)\right\}\right]. \end{aligned}$$

Since the above equation holds for all $\mathbf{h}_1^* \in \mathcal{R}$ with $\|\mathbf{h}_1^*\|_2 \le 1$, setting

$$\begin{aligned} A &= E\left[\sum_{j=1}^{K}\left\{\Lambda_0'(T_{0j}^*)^2(T_{0j}^*)^2\big(\mathbf{X}-E(\mathbf{X}|K,T_{0j}^*)\big)^{\otimes 2}\right\}\right], \\ B &= E\left[\left\{\sum_{j=1}^{K}\Big((N(T_j)-\Lambda(T_{0j}^*))\Lambda'(T_{0j}^*)T_{0j}^*\big(\mathbf{X}-E(\mathbf{X}|K,T_{0j}^*)\big)\Big)\right\}^{\otimes 2}\right], \end{aligned} \tag{B.6}$$

we obtain $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} N(0, A^{-1}B(A^{-1})^T)$.

Then we turn to consider $(\mathbf{h}_1^{**}, h_2^{**})$ such that $\sigma_1(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)[\mathbf{h}_1^{**}, h_2^{**}] = 0$. That is

$$\begin{aligned} &\mathcal{P}\left[\sum_{j=1}^{K}\left\{\big(\Lambda_0'(T_{0j}^*)T_{0j}^*\mathbf{X}^T\mathbf{h}_1^{**} + h_2^{**}(T_{0j}^*)\big)\Lambda_0'(T_{0j}^*)T_{0j}^*\mathbf{X}\right\}\right] \\ =&\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Lambda_0'(T_{0j}^*)^2(T_{0j}^*)^2\mathbf{X}^{\otimes 2}\mathbf{h}_1^{**} + h_2^{**}(T_{0j}^*)\Lambda_0'(T_{0j}^*)T_{0j}^*\mathbf{X}\right\}\right] = 0. \end{aligned}$$

Thus, we have $\mathbf{h}_1^{**} = \mathbf{R}(h_2^{**})$, where

$$\mathbf{R}(h_2) = -\left\{E\left[\sum_{j=1}^{K}\{\Lambda_0'(T_{0j}^*)^2(T_{0j}^*)^2\mathbf{X}^{\otimes 2}\}\right]\right\}^{-1} E\left[\sum_{j=1}^{K}\{\Lambda_0'(T_{0j}^*)T_{0j}^*\mathbf{X}h_2(T_{0j}^*)\}\right] \tag{B.7}$$

for all $h_2 \in \mathcal{H}_r$. It follows that

$$\sqrt{n}\sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{R}(h_2), h_2] = -\sqrt{n}G_n(\boldsymbol{\beta}_0, \Lambda_0)[\mathbf{R}(h_2), h_2] + o_p(1),$$

and we have $\sqrt{n}\sigma_2(\hat{\Lambda}_n - \Lambda_0)[\mathbf{R}(h_2), h_2] \xrightarrow{d} N(0, \sigma_0[\mathbf{R}(h_2), h_2]^2)$. □

Acknowledgments

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.

References

[1] Bailey, H. H., Kim, K., Verma, A. K., Sielaff, K., Larson, P. O., Snow, S., et al. (2010). A randomized, double-blind, placebo-controlled phase 3 skin cancer prevention study of α-difluoromethylornithine in subjects with previous history of skin cancer. *Cancer Prev. Res.* **3** 35–47.

[2] Balakrishnan, N. and Zhao, X. (2009). New multi-sample nonparametric tests for panel count data. *Ann. Stat.* **37** 1112–1149. MR2509069

[3] Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *J. R. Stat. Soc. Ser. C* **33,** 38–44.

[4] Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16,** 1190–1208. MR1346301

[5] Cheng, S. C., and Wei, L. J. (2000). Inferences for a semiparametric model with panel data. *Biometrika* **87,** 89–97. MR1766830

[6] Chiou, S. H., Xu, G., Yan, J., and Huang, C. Y. (2018). Semiparametric estimation of the accelerated mean model with panel count data under informative examination times. *Biometrics* **74,** 994–953. MR3860715

[7] Deng, S., Liu, L., and Zhao, X. (2015). Monotone spline-based least squares estimation for panel count data with informative observation times. *Biom. J.* **57,** 743–765. MR3394808

[8] Ding, Y., and Nan, B. (2011). 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,** 3032–3061. MR3012400

[9] Ghosh, D. (2004). Accelerated rates regression models for recurrent failure time data. *Lifetime Data Anal.* **10,** 247–261. MR2086959

[10] Hu, X. J., Lagakos, S. W., and Lockhart, R. A. (2009). Marginal analysis of panel counts through estimating functions. *Biometrika* **96,** 445–456. MR2507154

[11] Hu, X. J., Sun, J., and Wei, L. J. (2003). Regression parameter estimation from panel counts. *Scand. J. Stat.* **30,** 25–43. MR1963891

[12] Hua, L., and Zhang, Y. (2012). Spline-based semiparametric projected generalized estimating equation method for panel count data. *Biostatistics* **13,** 440–454. MR2941442

[13] Huang, C. Y., Wang, M. C., and Zhang, Y. (2006). Analysing panel count data with informative observation times. *Biometrika* **93,** 763–775. MR2285070

[14] Jiang, H., Su, W., and Zhao, X. (2020). Robust estimation for panel count data with informative observation times and censoring times. *Lifetime Data Anal.* **26,** 65–84. MR4051838

[15] Kalbfleisch, J. D., and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *J. Am. Stat. Assoc.* **80,** 863–871. MR0819585

[16] Li, N., Park, D-H., Sun, J., and Kim, K. (2011). Semiparametric transformation models for multivariate panel count data with dependent observation process. *Can. J. Stat.* **39,** 458–474. MR2842424

[17] Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Stat. Soc. Ser. B* **62,** 711–730. MR1796287

[18] Lin, D. Y., Wei, L. J., and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85,** 605–618. MR1665810

[19] Lu, M., Zhang, Y., and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94,** 705–718. MR2410018

[20] Lu, M., Zhang, Y., and Huang, J. (2009). Semiparametric estimation methods for panel count data using monotone B-splines. *J. Am. Stat. Assoc.* **104,** 1060–1070. MR2750237

[21] Ma, L., and Sundaram, R. (2018). Analysis of gap times based on panel count data with informative observation times and unknown start time. *J. Am. Stat. Assoc.* **113,** 294–305. MR3803465

[22] Ramsay, J. O. (1988). Monotone regression splines in action. *Stat. Sci.* **3,** 425–441.

[23] Shen, X., and Wong, W. H. (1994). Convergence rate of sieve estimates. *Ann. Stat.* **22,** 580–615. MR1292531

[24] Stone, C. J. (1985) Additive regression and other nonparametric models. *Ann. Stat.* **13,** 689–705. MR0790566

[25] Sun, J., and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Stat. Sin.* **5,** 279–290. MR1329298

[26] Sun, J., and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. Am. Stat. Assoc.* **62,** 293–302. MR1749540

[27] Thall, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44,** 197–209. MR0931635

[28] van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, Springer–Verlag, New York. MR1385671

[29] Wang, M.-C., Qin, J. and Chiang, C.-T. (2001). Analyzing recurrent event data with informative censoring. *J. Am. Stat. Assoc.* **96,** 1057–1065. MR1947253

[30] Wellner, J. A., and Zhang, Y. (2000). Two estimator of the mean of a counting process with panel count data. *Ann. Stat.* **28,** 779–814. MR1792787

[31] Wellner, J. A., and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Stat.* **35,** 2106–2142. MR2363965

[32] Xu, G., Chiou, S. H., Huang, C.-Y., Wang, M. C., and Yan, J. (2017). Joint scale-change models for recurrent events and failure time. *J. Am. Stat. Assoc.* **112,** 794–805. MR3671771

[33] Xu, G., Chiou, S. H., Yan, J., Marr, K. and Huang, C.-Y. (2020). Generalized scale-change models for recurrent event processes under informative censoring. *Stat. Sin.* **30,** 1773–1795. MR4260744

[34] Zeng, D., and Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *J. Am. Stat. Assoc.* **102,** 1387–1396. MR2412556

[35] Zhao, X., and Sun, J. (2011). Nonparametric comparison for panel count data with unequal observation processes. *Biometrics* **67,** 770–779. MR2829131

[36] Zhao, X., and Zhang, Y. (2017). Asymptotic normality of nonparametric M-estimators with applications to hypothesis testing for panel count data. *Stat. Sin.* **27,** 931–950. MR3675037