Monotone spline-based least squares estimation for panel count data with informative observation times

Shirong Deng[1], **Li Liu**[1], and **Xingqiu Zhao**[*,2]

[1] School of Mathematics and Statistics, Wuhan University, Wuhan, China
[2] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China

Received 2 October 2014; revised 14 March 2015; accepted 21 April 2015

This article discusses the statistical analysis of panel count data when the underlying recurrent event process and observation process may be correlated. For the recurrent event process, we propose a new class of semiparametric mean models that allows for the interaction between the observation history and covariates. For inference on the model parameters, a monotone spline-based least squares estimation approach is developed, and the resulting estimators are consistent and asymptotically normal. In particular, our new approach does not rely on the model specification of the observation process. The proposed inference procedure performs well through simulation studies, and it is illustrated by the analysis of bladder tumor data.

Keywords: Informative observation process; Least squares estimation; Monotone B-splines; Panel count data; Semiparametric mean models.

1 Introduction

In many longitudinal follow-up studies, each subject may be observed at several distinct times and only the numbers of events between two adjacent times are available. It may be impossible to record the exact event times because of too expensive examination cost or too frequent occurrence of the events. Moreover, the observation times may vary from subject to subject. Such complex data are called panel count data, which often occur in many fields such as demographic studies, industrial reliability, and clinical trials; see Kalbfleisch and Lawless (1985), Gaver and O'Muicheartaigh (1987), Thall and Lachin (1988), and Sun and Kalbfleisch (1995). For such data, important information includes the observation times, counts of recurrent events, censoring or follow-up times, and covariates related to the study for each study subject. A typical example is bladder tumor data that will be discussed below.

For the analysis of panel count data, Sun and Zhao (2013) provide a relatively complete review for the analysis of panel count data wherein more references can be found. The existing research mainly focuses on the assumption that the underlying observation process and recurrent event process are independent completely or conditional on covariates (e.g., Kalbfleisch and Lawless, 1985; Thall and Lachin, 1988; Sun and Kalbfleisch, 1995; Staniswalis et al., 1997; Sun and Wei, 2000; Wellner and Zhang, 2000, 2007; Zhang, 2002; Hu et al., 2003, 2009; Lu et al., 2007, 2009; He et al., 2008). However, this assumption may be violated in practice. Such an example is a set of panel count data arising from a bladder cancer follow-up study conducted by the Veterans Administration Cooperative Urological Research Group (Byar, 1980). Many patients had multiple recurrences of new tumors during the study. The patients in the thiotepa group had significantly more clinical visits than those in the placebo group since thiotepa needed be instilled in the bladder (Lu et al., 2009; Li et al., 2010; Zhao and

*Corresponding author: e-mail: xingqiu.zhao@polyu.edu.hk, Phone: +852-2766-6921, Fax: +852-2362-9045

Tong, 2011). The patients in the thiotepa group received more medical treatment so that these patients had the lower tumor recurrence rate. This indicates that the number of clinical visits may contain some information about the tumor recurrence rate—that is, the observation process is informative.

In contrast, there is limited research on the analysis of panel count data with informative observation times. Li et al. (2010) incorporated the observation history to the mean models of the underlying recurrent event process that follows a class of semiparametric transformation models. Zhao and Tong (2011) proposed a joint modeling approach through an unobserved frailty variable and a completely unspecified link function to characterize the correlation between the underlying recurrent event process and the observation times. A commonly used assumption in their analysis mentioned above is that the observation process follows a Poisson-type with a proportional intensity function. However, the fit of the Poisson model may be inadequate when the observation process displays underdispersion or overdispersion. For example, Zhao et al. (2013) tested the Poisson process assumption about the observation process for the bladder tumor data and concluded that the Poisson process model cannot be acceptable for the observation process in the data. In addition, the relation between the observation and recurrent event processes may vary with some covariates. For example, in the bladder cancer study, patients who received the thiotepa treatment may make clinical visits more often than those in the placebo group and thus may have less superficial bladder tumors, which means that the correlation between the observation times and the tumor recurrent process as mentioned above may be different for different treatment groups. Motivated by the characteristics of the bladder cancer data, we propose a new class of semiparametric regression models by incorporating the interaction between the observation history and some covariates to the mean model of the recurrent event process, where the observation process model is completely unspecified.

The remainder of this article is organized as follows. We begin in Section 2 by introducing notation and describing models for panel count data. In Section 3, a spline-based least squares method is proposed for estimation of regression parameters and the baseline mean function, and the asymptotic properties of the proposed estimators are presented in Appendix. In Section 4, we present some simulation results to assess the finite-sample performance of the proposed inference procedure. In Section 5, the proposed approach is illustrated through the analysis of a dataset from a bladder tumor study. Some concluding remarks are made in Section 6.

2 Panel count mean models

Consider a study involving n independent subjects who may experience some recurrent events and let $N_i(t)$ denote the total number of occurrences of the event of interest up to time t for subject i $(i = 1, \cdots, n)$ for $0 \le t \le \tau$, where τ is a known constant time point. Also suppose that for subject i, $N_i(t)$ is observed only at discrete potential time points $0 < T_{K_i,1} < T_{K_i,2} < \cdots < T_{K_i,K_i}$, where the total number of observations K_i is an integer-valued random variable. In general, not every subject can be followed until τ and there exists a follow-up time C_i for subject i. That is, $N_i(T_{K_i,j})$ is observed only if $T_{K_i,j} \le C_i \le \tau$. Define $H_i(t) = \sum_{j=1}^{K_i} I(T_{K_i,j} \le t)$ to be the counting process that records the number of observations for subject i up to time t, where $I(\cdot)$ is the indicator function. Then define $\tilde{H}_i(t) = H_i(t \wedge C_i)$ as the actual observation process for subject i, where $a \wedge b = \min(a, b)$. Let $\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})'$ denote a p-dimensional vector of covariates that may not depend on t, $i = 1, \cdots, n$. Define $\mathcal{F}_{it} = \{H_i(s) : 0 \le s < t\}$ as the observation history just before t.

By relaxing the model assumption on the observation process, we assume that given $\mathbf{X}_i$, $\mathcal{F}_{it}$, and the covariate W_i, which is allowed to be a component of the vector $\mathbf{X}_i$, the mean function of $N_i(t)$ has the form

$$\mu_i(t) = \mu_0(t) \exp\{\boldsymbol{\beta}'\mathbf{X}_i + \boldsymbol{\alpha}'\phi(\mathcal{F}_{it}, W_i)\}, \tag{1}$$

where $\mu_0(t)$ is an unspecified smooth and nondecreasing nonnegative function of t; $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are p-dimensional and q-dimensional vectors of unknown regression coefficients, respectively; and $\phi(\cdot)$ is a q-dimensional vector of known functions of the counting process $H_i(t)$ up to time $t-$ and the covariates W_i, representing the interaction between the observation history and some covariates. In particular, for panel count data arising from the longitudinal follow-up clinical studies with different treatments, W_is can be considered as the treatment indicators, and thus $\boldsymbol{\alpha}$ represents the effect of interaction between the frequency of observation times and the treatment group on the underlying recurrent event process. If $\boldsymbol{\alpha} = \mathbf{0}$, then model (1) reduces to the model considered by Sun and Wei (2000), Zhang (2002), and Wellner and Zhang (2007) for regression analysis of panel count data. In fact, our modeling approach is different from the existing approaches. Here, the possible effect of the observation process is directly incorporated into the conditional model about the underlying recurrent event process, and no additional model assumption is needed for the observation process.

For making valid inference, we assume that

$$E\{N_i(t)|\mathbf{X}_i, H_i(s), 0 \le s \le t, C_i\} = E\{N_i(t)|\mathbf{X}_i, \mathcal{F}_{it}, C_i\},$$

which means that the mean of the recurrent event process valued at time point t only depends on the observation history before t, conditional on the covariates and the censoring time. Let $\mathbf{O} = (K, \bar{\boldsymbol{T}}_K, \bar{\boldsymbol{N}}_K, \bar{\boldsymbol{H}}_K, \mathbf{X}, \boldsymbol{C})$, with $\bar{\boldsymbol{T}}_K = (T_{K,1}, \cdots, T_{K,K})$, $\bar{\boldsymbol{N}}_K = (N(T_{K,1}), \cdots, N(T_{K,K}))$, $\bar{\boldsymbol{H}}_K = (H(T_{K,1}), \cdots, H(T_{K,K}))$. Throughout this article, we will assume that we observe n i.i.d. copies, $\mathbf{O}_1, \cdots, \mathbf{O}_n$ of $\mathbf{O}$. The goal is to estimate the unknown nondecreasing function $\mu_0(t)$ and the regression parameters β and α.

3 Estimation procedure

For inference about model (1), we denote $\tilde{\mu}(t) = \log(\mu(t))$ and, using the generalized least square estimation approach given in Hu et al. (2009), define the following objective function:

$$\begin{aligned} L_n(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tilde{\mu}) &= \\ &= \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left[N_i(T_{K_i,j}) - \exp\{\tilde{\mu}(T_{K_i,j}) + \boldsymbol{\beta}'\mathbf{X}_i + \boldsymbol{\alpha}'\phi(\mathcal{F}_{iT_{K_i,j}}, W_i)\}\right]^2 \xi_i(T_{K_i,j}) = \\ &= \sum_{i=1}^{n}\int_0^{\tau}\left[N_i(t) - \exp\{\tilde{\mu}(t) + \boldsymbol{\beta}'\mathbf{X}_i + \boldsymbol{\alpha}'\phi(\mathcal{F}_{it}, W_i)\}\right]^2 d\tilde{H}_i(t), \end{aligned} \tag{2}$$

where $\xi_i(t) = I(C_i \ge t)$.

To make inference about $\mu_0(t)$, we propose to use B-splines to approximate its logarithm $\tilde{\mu}_0(t) = \log(\mu_0(t))$. For a finite closed interval $[0, \tau]$, let $\mathcal{I} = \{t_i\}_1^{m_n+2l}$, with

$$0 = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m_n+l} < t_{m_n+l+1} = \cdots = t_{m_n+2l} = \tau,$$

be a sequence of knots that partition $[0, \tau]$ into $m_n + 1$ subintervals and $m_n = O(n^{\nu})$, for $0 < \nu < 1/2$. Let $\Psi_{l,\mathcal{I}}$ be the class linearly spanned by the B-splines basis functions $\{B_{il}, 1 \le i \le q_n(q_n = m_n + l)\}$ with order l and knots $\mathcal{I}$, that is

$$\Psi_{l,\mathcal{I}} = \left\{\sum_{i=1}^{q_n}\gamma_i B_{il} : \gamma_i \in \mathbb{R}, i = 1, \cdots, q_n\right\}.$$

We now define a subclass of $\Psi_{l,\mathcal{I}}$, as $\Phi_{l,\mathcal{I}} = \{\sum_{i=1}^{q_n} \gamma_i B_{il} : \gamma_1 \leq \cdots \leq \gamma_{q_n}\}$. According to the variation-diminishing properties of B-splines (Schumaker, 1981), $\Phi_{l,\mathcal{I}}$ is a class of nondecreasing splines on $[0, \tau]$. Then, we can approximate the smooth monotone function $\tilde{\mu}_0(t)$ by $\sum_{i=1}^{q_n} \gamma_i B_{il}(t)$ and estimate the coefficients $\gamma_1 \leq \cdots \leq \gamma_{q_n}$ and regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ jointly through minimizing the approximated expression $L_n(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tilde{\mu})$ subject to nondecreasing constraints.

Using the spline-based sieve approximation $\tilde{\mu}_n(t) = \sum_{i=1}^{q_n} \gamma_i B_{il}(t)$ of $\tilde{\mu}_0(t)$, $L_n(\beta, \alpha, \tilde{\mu})$ in Eq. (2) can be approximated by

$$L_n(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \int_0^{\tau} \left[N_i(t) - \exp\{\boldsymbol{\gamma}'\boldsymbol{B}_l(t) + \boldsymbol{\beta}'\mathbf{X}_i + \boldsymbol{\alpha}'\phi(\mathcal{F}_{it}, W_i)\}\right]^2 d\tilde{H}_i(t), \tag{3}$$

where $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_{q_n})'$ and $\boldsymbol{B}_l(t) = (B_{1l}(t), \cdots, B_{q_n l}(t))'$.

Let $\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\alpha}}_n, \hat{\boldsymbol{\gamma}}_n$ be the values that minimize $L_n(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ in Eq. (3) with the constraint $\gamma_1 \leq \cdots \leq \gamma_{q_n}$. Then the monotone spline estimator for $\tilde{\mu}_0(t)$ is $\hat{\tilde{\mu}}_n(t) = \hat{\boldsymbol{\gamma}}_n \boldsymbol{B}_l(t)$. Thus $\mu_0(t)$ can be estimated by $\hat{\mu}_n(t) = \exp\{\hat{\tilde{\mu}}_n(t)\}$. The spline estimation problem can be formulated as the linear inequality constrained minimization problem $\min_{\eta \in \mathbb{R}^{p+q} \times \Theta_\gamma} L_n(\eta)$, where $\boldsymbol{\eta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \boldsymbol{\gamma}')'$ with $\boldsymbol{\gamma} \in \Theta_\gamma = \{\boldsymbol{\gamma} : \gamma_1 \leq \cdots \leq \gamma_{q_n}\}$. Jamshidian (2004) proposed a generalized gradient projection (GP) algorithm for optimizing a nonlinear objective function with linear inequality constraints, based on the generalized Euclidean metric $\|\boldsymbol{x}\| = \boldsymbol{x}^T V \boldsymbol{x}$ with V being a positive definite matrix and possibly varying from iteration to iteration. Zhang and Jamshidian (2004) applied the GP algorithm to large-scale nonparametric maximum-likelihood estimation problems by choosing $V = D_U$, the matrix containing only the diagonal elements of the negative Hessian matrix U, in order to avoid the storage problem in updating U. However, this will increase the number of iterations and thereby the computing time. Lu et al. (2007, 2009) used the generalized GP algorithm in Zhang and Jamshidian (2004) with $V = U$ directly because the dimension of unknown parameter space is usually small in their applications due to the use of polynomial splines, which would also substantially reduce the number of iterations. Here we consider using the generalized GP algorithm for the monotone polynomial spline estimation with V unequal to the negative Hessian matrix U.

Let $\nabla L_n(\boldsymbol{\eta})$ be the negative gradient of $L_n(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ and

$$V = \sum_{i=1}^{n} \int_0^{\tau} \exp\{\boldsymbol{\eta}' Z_{li}(t)\}^2 Z_{li}^{\otimes 2}(t) d\tilde{H}_i(t),$$

which is a positive definite matrix with $Z_{li}(t) = (\mathbf{X}_i', \phi(\mathcal{F}_{it}, W_i)', \boldsymbol{B}_l'(t))'$. Furthermore, $n^{-1}(2V)$ has the same limit as the negative Hessian matrix of $-L_n(\boldsymbol{\eta})$.

Let $\mathcal{A} = \{i_1, i_2, \cdots, i_m\}$ denote the index set of active constraints, that is $\gamma_{i_j} = \gamma_{i_j+1}$, for $j = 1, \cdots, m$, during the numerical computation. $\mathcal{A}$ is allowed to be empty when $m = 0$. We define an m by $p + q + q_n$ working matrix corresponding to $\mathcal{A}$ as

$$A = \begin{pmatrix} 0 & \cdots & 0 & 0 & \cdots & -1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & -1 & 1 \end{pmatrix},$$

that is, the j-th row ($j = 1, \cdots, m$) consists of $(p + q + q_n)$ components with its $(p + q + i_j)$-th and $(p + q + i_j + 1)$-th elements equal to -1 and 1, respectively, and the remaining components zero. The *generalized GP algorithm* is summarized as follows.

Start with a feasible initial value $\boldsymbol{\eta} \in \mathbb{R}^{p+q} \times \Theta_\gamma$, and cycle through the following steps until convergence.

S0. (**Computing the feasible search direction**) Calculate

$$\boldsymbol{d} = \left(I - V^{-1}A^T(AV^{-1}A^T)^{-1}A\right) V^{-1}\nabla L_n(\boldsymbol{\eta}).$$

When there is no active constraint, take $\boldsymbol{d} = V^{-1}\nabla L_n(\boldsymbol{\eta})$.

S1. (**Forcing the updated $\boldsymbol{\eta}$ to fulfill the constraints**) If the resulted direction $\boldsymbol{d}$ is not nondecreasing in its components, compute

$$\varphi = \min_{i \notin \mathcal{A}, d_i > d_{i+1}} \left(-\frac{\gamma_{i+1} - \gamma_i}{d_{i+1} - d_i}\right).$$

Doing so guarantees that $\gamma_{i+1} + \varphi d_{i+1} \geq \gamma_i + \varphi d_i$, for $i = 1, \cdots, q_n$.

S2. (**Step-halving line search**) Find the smallest integer k starting from 0 such that

$$\|\nabla L_n(\boldsymbol{\eta} + (1/2)^k \boldsymbol{d})\| < \|\nabla L_n(\boldsymbol{\eta})\|.$$

S3. (**Updating the solution**) If $\varphi > (1/2)^k$, replace $\boldsymbol{\eta}$ by $\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta} + (1/2)^k \boldsymbol{d}$ and check the stopping criterion (S5).

S4. (**Updating the active constraint set**) If $\varphi \leq (1/2)^k$, in addition to replace $\boldsymbol{\eta}$ by $\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta} + \varphi \boldsymbol{d}$, modify $\mathcal{A}$ by adding indexes of all the newly active constraints to $\mathcal{A}$ and accordingly modify the working matrix A.

S5. (**Checking the stopping criterion**) If $\|\boldsymbol{d}\| \geq \varepsilon$ for a small $\varepsilon > 0$, go to S0, otherwise, compute the Lagrange multipliers $\lambda = (AV^{-1}A^T)^{-1} AV^{-1}\nabla L_n(\boldsymbol{\eta})$.
  - (i) If $\lambda_i \leq 0$ for all $i \in \mathcal{A}$, set $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}$ and stop.
  - (ii) If at least one $\lambda_i > 0$, for $i \in \mathcal{A}$, remove the index corresponding to the largest λ_i from $\mathcal{A}$, and update $\mathcal{A}$ and go to S0.

Let $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\alpha}}_n, \hat{\mu}_n)$ and $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$. Under some regularity conditions stated in Appendix, $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}_0$ and $\sqrt{n}\binom{\hat{\beta}_n - \beta_0}{\hat{\alpha}_n - \alpha_0}$ converges in distribution to $N(\mathbf{0}, \Sigma)$, where Σ is given in Theorem 3 of Appendix.

4 Simulation study

In this section, we conducted a simulation study to assess the finite-sample properties of the proposed estimators. We considered the situation where there were two covariates and for each subject i, X_{1i}s and X_{2i}s were generated from Bernoulli distribution with success probability 0.5 and the uniform distribution over interval $(-1, 1)$, respectively. The follow-up time C_i was generated from the uniform distribution over interval $(\tau/2, \tau)$ with $\tau = 6$. Given the covariate $\mathbf{X}_i = (X_{1i}, X_{2i})'$ and C_i, two setups for the observation process $H_i(t)$ were considered as follows:

(i) The number of observation times m_i was assumed to follow the Poisson distribution with mean $(2C_i/\tau)\exp(-0.5X_{1i} + 0.5X_{2i})$ and the observation times $(T_{m_i1}, \ldots, T_{m_im_i})$ were taken to be the order statistics of a random sample of size m_i from the uniform distribution over $(0, C_i)$.

(ii) The number of observation times m_i was assumed to follow the uniform distribution over $\{1, 2, 3, 4, 5, 6\}$ and the observation times $(T_{m_i1}, \ldots, T_{m_im_i})$ were generated in the same way as in setup (i).

Then, given $\mathbf{X}_i$, m_i, and the observation times $(T_{m_i,1}, \cdots, T_{m_i,m_i})$, we generated recurrent event counts $\tilde{\boldsymbol{N}}_{m_i} = (N_i(T_{m_i,1}), \cdots, N_i(T_{m_i,m_i}))$ from a Poisson process by taking

$$N_i(T_{m_i,j}) = N_i(T_{m_i,1}) + \{N_i(T_{m_i,2}) - N_i(T_{m_i,1})\} + \cdots + \{N_i(T_{m_i,j}) - N_i(T_{m_i,j-1})\},$$

with $N_i(T_{m_i,1}) \sim \text{Poisson}(\exp\{\tilde{\mu}_0(T_{m_i,1}) + \boldsymbol{\beta}_0'\mathbf{X}_i + \alpha_0 H_i(T_{m_i,1}-)W_i\})$, and

$$N_i(T_{m_i,j}) - N_i(T_{m_i,j-1}) \sim \text{Poisson}(\exp\{\tilde{\mu}_0(T_{m_i,j}) + \boldsymbol{\beta}_0'\mathbf{X}_i + \alpha_0 H_i(T_{m_i,j}-)W_i\} - \exp\{\tilde{\mu}_0(T_{m_i,j-1}) + \boldsymbol{\beta}_0'\mathbf{X}_i + \alpha_0 H_i(T_{m_i,j-1}-)W_i\}),$$

for $j \geq 2$.

We took $\boldsymbol{\beta}_0 = (-0.5, 0.5)$; $\alpha_0 = 0, 0.3$, or 0.5; and $\tilde{\mu}_0(t) = \sqrt{t}$ or $\log(t+1)$, and used the cubic B-splines to compute the spline estimators. To choose the number of interior knots, we partitioned the range of ν, $(0, 0.5)$ to 20 equally subintervals and chose ν to be the partition points. For each value of ν, we took $m_n = n^\nu$ as the number of interior knots. To determine locations of knots, we considered two commonly used data-driven methods. One is the equally spaced knots given by $T_{\min} + k(T_{\max} - T_{\min})/(m_n + 1)$, $k = 0, 1, \cdots, m_n + 1$, where $T_{\min}$ and $T_{\max}$ denote the minimum and maximum values of distinct observation times, respectively. Another is the partitions corresponding to quantiles of the observation times, that is, the $k/(m_n + 1)$ quantiles ($k = 0, 1, \cdots, m_n + 1$) of the distinct observation times as the knots. The value of ν that minimizes the Bayesian information criteria (BIC) was selected. Here $\text{BIC} = 2\log(L) + \log(n)(q_n + 3)$, where $L = L_n(\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\alpha}}_n, \hat{\tilde{\mu}}_n)$ as defined in Eq. (2) and q_n is the number of the B-spline basis functions. We carried out simulations for the different situations of the number and placement of knots with $W = X_1$, $\alpha = 0.3$, $\tilde{\mu}_0(t) = \log(t+1)$, and $n = 100$, and found that the estimation results are very similar and the method is insensitive to the selection of number and placement of knots, where the value of ν was selected by the BIC as 1/8. Thus in the following, we took the number of interior knots as $n^{1/8}$ and used the equally spaced knots.

To initialize the algorithm, we choose $\boldsymbol{\gamma} = (1, 2, \cdots, q_n)'$, while $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ were all generated from the uniform distribution over interval $(-0.5, 0.5)$. Tables 1 and 2 present the simulation results on estimation of $\boldsymbol{\beta}_0$ and α_0 under Poisson and non-Poisson observation processes with sample size $n = 100$ or 200 for $\tilde{\mu}_0(t) = \sqrt{t}$ and $\log(t+1)$, respectively. The tables include the estimated bias (BIAS) given by the average of the estimates minus the true value, the sample standard deviation error of estimates (SSE), the mean of the bootstrap standard errors of the estimates (BSE), and the bootstrap 95% coverage probability (CP) obtained from 1000 independent runs. Here we used 100 replications in bootstrap to estimate the standard errors. It can be seen from the tables that the proposed estimators are unbiased for different situations considered, which means that our estimation approach does not rely on the Poisson distributional assumption about the observation process.

Also, the SSE and BSE are quite close to each other and become smaller as the sample size increases, which indicates that proposed bootstrap variance estimation procedure provides reasonable estimates. In addition, the 95% bootstrap CP is consistent with the nominal level, which suggests that the normal approximation seems to be appropriate.

To estimate the smooth function $\tilde{\mu}_0(t)$, we considered cubic B-splines and took $m_n = n^\nu$ with $\nu = 1/8$. For a given number of interior knots m_n, the equally spaced knots are chosen. Figure 1 shows the estimation results of $\tilde{\mu}_0(t) = \log(t+1)$ for simulated panel count data with Possion and non-Poisson observation processes, $\phi(\mathcal{F}_t, W) = H(t-)X_1$ and $\alpha = 0.3$. In the figure, the solid line represents the real curve of $\tilde{\mu}_0(t)$, and the point line and the dotted line represent the B-spline

Table 1 Estimation results of $(\boldsymbol{\beta}_0', \alpha_0')'$ for simulated panel count data with Poisson and non-Poisson observation processes, $\phi(\mathcal{F}_t, W) = H(t-)X_1$, and $\tilde{\mu}_0(t) = \sqrt{t}$.

n	α		$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
			BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
100	0	P	–0.0019	0.1531	0.1523	0.9470	0.0092	0.1359	0.1279	0.9280	–0.0080	0.1204	0.1305	0.9630
		NP	–0.0011	0.1370	0.1305	0.9300	0.0050	0.0980	0.0934	0.9370	–0.0013	0.0411	0.0393	0.9350
	0.3	P	0.0101	0.1541	0.1501	0.9470	0.0056	0.1283	0.1232	0.9450	–0.0064	0.0990	0.1108	0.9590
		NP	0.0020	0.1262	0.1234	0.9410	0.0035	0.0745	0.0741	0.9510	0.0006	0.0316	0.0295	0.9330
	0.5	P	–0.0028	0.1627	0.1536	0.9380	0.0143	0.1328	0.1207	0.9260	–0.0035	0.0978	0.1056	0.9680
		NP	0.0037	0.1169	0.1188	0.9470	–0.0005	0.0684	0.0757	0.9740	–0.0019	0.0525	0.0585	0.9960
200	0	P	–0.0117	0.1095	0.1057	0.9400	0.0025	0.0909	0.0897	0.9410	–0.0035	0.0807	0.0799	0.9530
		NP	0.0046	0.0931	0.0921	0.9540	0.0043	0.0674	0.0657	0.9470	–0.0008	0.0278	0.0275	0.9500
	0.3	P	–0.0047	0.1069	0.1069	0.9490	0.0077	0.0909	0.0864	0.9380	–0.0024	0.0638	0.0637	0.9480
		NP	–0.0027	0.0913	0.0866	0.9370	0.0024	0.0565	0.0525	0.9260	0.0012	0.0216	0.0206	0.9320
	0.5	P	0.0015	0.1101	0.1077	0.9520	0.0045	0.0899	0.0845	0.9430	–0.0030	0.0606	0.0602	0.9620
		NP	–0.0004	0.0850	0.0847	0.9420	0.0027	0.0494	0.0535	0.9710	0.0007	0.0214	0.0390	0.9990

Notes: P and NP represent observation times generated from Poisson and non-Poisson processes, respectively.

Table 2 Estimation results of $(\boldsymbol{\beta}_0', \alpha_0')'$ for simulated panel count data with Poisson and non-Poisson observation processes, $\phi(\mathcal{F}_t, W) = H(t-)X_1$, and $\tilde{\mu}_0(t) = \log(t+1)$.

n	α		$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
			BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
100	0	P	0.0020	0.1835	0.1803	0.9490	0.0111	0.1620	0.1574	0.9420	–0.0162	0.1516	0.1581	0.9520
		NP	–0.0015	0.1595	0.1564	0.9480	0.0039	0.1168	0.1144	0.9550	–0.0019	0.0501	0.0475	0.9350
	0.3	P	0.0017	0.1867	0.1813	0.9400	0.0029	0.1514	0.1488	0.9500	–0.0069	0.1248	0.1365	0.9640
		NP	–0.0015	0.1476	0.1460	0.9540	–0.0020	0.0960	0.0893	0.9330	0.0006	0.0369	0.0349	0.9340
	0.5	P	–0.0033	0.1928	0.1843	0.9340	0.0112	0.1573	0.1472	0.9370	–0.0040	0.1131	0.1257	0.9590
		NP	–0.0009	0.1518	0.1385	0.9280	0.0038	0.0803	0.0776	0.9470	0.0016	0.0339	0.0354	0.9670
200	0	P	–0.0072	0.1254	0.1275	0.9560	0.0017	0.1105	0.1075	0.9450	–0.0085	0.0983	0.0967	0.9450
		NP	–0.0053	0.1117	0.1104	0.9390	0.0053	0.0833	0.0801	0.9360	0.0004	0.0334	0.0328	0.9440
	0.3	P	0.0005	0.1317	0.1283	0.9480	0.0067	0.1052	0.1049	0.9490	–0.0040	0.0770	0.0782	0.9550
		NP	–0.0016	0.1039	0.1037	0.9440	0.0016	0.0629	0.0634	0.9600	0.0011	0.0251	0.0247	0.9430
	0.5	P	0.0018	0.1288	0.1289	0.9570	0.0042	0.1032	0.1014	0.9450	–0.0012	0.0692	0.0713	0.9550
		NP	0.0004	0.0991	0.0985	0.9490	0.0013	0.0532	0.0541	0.9570	0.0004	0.0216	0.0241	0.9650

Notes: P and NP represent observation times generated from Poisson and non-Poisson processes, respectively.

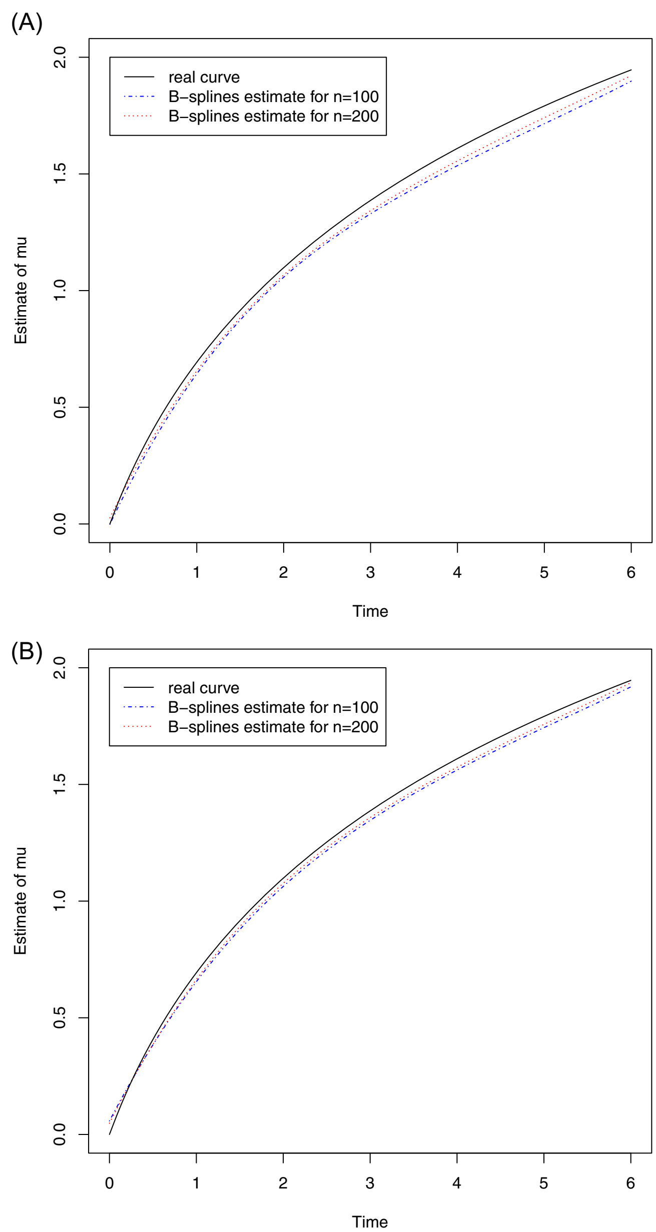

Figure 1 Estimates of $\tilde{\mu}_0(t) = \log(t+1)$ for simulated panel count data with (A) Poisson and (B) non-Poisson observation processes, $\phi(\mathcal{F}_t, W) = H(t-)X_1$, and $\alpha = 0.3$.

estimates of $\tilde{\mu}_0(t)$ for the sample size $n = 100$ and $n = 200$, respectively. From the figure, one can see that the B-spline estimate is close to its real curve with the moderate sample size and especially closer as the sample size increases for all the situations considered here, indicating that the B-spline estimator for $\tilde{\mu}_0(t)$ performs well.

To compare our estimators with the maximum spline pseudo-likelihood estimators (MSPLE) and maximum spline likelihood estimators (MSLE) in Lu et al. (2009), we considered the same setup as

Table 3 Comparison of estimates for $\boldsymbol{\beta}_0$, based on generated data from mixed Poisson recurrent event process with $\phi(\mathcal{F}_t, W) = H(t-)X_1$ and $\tilde{\mu}_0(t) = \log(t+1)$.

n	α	Method	Poisson observation				Non-Poisson observation			
			$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_1$		$\hat{\beta}_2$	
			BIAS	SSE	BIAS	SSE	BIAS	SSE	BIAS	SSE
100	0	Proposed	–0.0031	0.2213	0.0147	0.2037	–0.0054	0.1829	0.0104	0.1461
		MSPLE	–0.0023	0.1968	0.0001	0.1707	0.0031	0.1486	0.0003	0.1290
		MSLE	0.0028	0.1723	–0.0019	0.1511	0.0020	0.1307	0.0021	0.1108
	0.3	Proposed	–0.0078	0.2271	0.0317	0.1989	–0.0105	0.1967	0.0070	0.1376
		MSPLE	0.2372	0.2008	0.0457	0.1665	0.7080	0.1380	–0.0015	0.1218
		MSLE	0.2702	0.1816	0.0499	0.1516	0.9150	0.1377	–0.0001	0.1271
	0.5	Proposed	–0.0174	0.2426	0.0303	0.2083	–0.0023	0.2291	0.0103	0.1608
		MSPLE	0.4343	0.2501	0.0769	0.1980	1.3087	0.1528	–0.0010	0.1478
		MSLE	0.5008	0.2376	0.0861	0.1869	1.6514	0.1736	0.0031	0.1862
200	0	Proposed	0.0002	0.1584	0.0153	0.1488	–0.0026	0.1251	0.0073	0.0978
		MSPLE	0.0038	0.1348	0.0039	0.1207	–0.0012	0.1062	0.0023	0.0895
		MSLE	0.0031	0.1194	0.0060	0.1021	–0.0015	0.0920	–0.0000	0.0794
	0.3	Proposed	–0.0076	0.1704	0.0124	0.1507	0.0034	0.1398	0.0109	0.0972
		MSPLE	0.2563	0.1447	0.0404	0.1268	0.7192	0.0975	0.0050	0.0873
		MSLE	0.2902	0.1287	0.0461	0.1122	0.9248	0.0974	0.0047	0.0911
	0.5	Proposed	–0.0098	0.1848	0.0196	0.1499	0.0034	0.1647	0.0039	0.1162
		MSPLE	0.4699	0.1783	0.0886	0.1572	1.3208	0.1080	0.0044	0.1043
		MSLE	0.5343	0.1696	0.0978	0.1489	1.6620	0.1252	0.0054	0.1323

above except that for given $\mathbf{X}_i$, m_i, and $(T_{m_i,1}, \cdots, T_{m_i,m_i})$, the recurrent event counts were generated from a mixed Poisson process by taking

$$N_i(T_{m_i,j}) = N_i(T_{m_i,1}) + \{N_i(T_{m_i,2}) - N_i(T_{m_i,1})\} + \cdots + \{N_i(T_{m_i,j}) - N_i(T_{m_i,j-1})\},$$

with $N_i(T_{m_i,1})|Q_i \sim \text{Poisson}(Q_i \exp\{\tilde{\mu}_0(T_{m_i,1}) + \boldsymbol{\beta}_0'\mathbf{X}_i + \alpha_0 H_i(T_{m_i,1}-)W_i\})$, and

$$[N_i(T_{m_i,j}) - N_i(T_{m_i,j-1})]|Q_i \sim \text{Poisson}(Q_i\big[\exp\{\tilde{\mu}_0(T_{m_i,j}) + \boldsymbol{\beta}_0'\mathbf{X}_i + \alpha_0 H_i(T_{m_i,j}-)W_i\} + \\ - \exp\{\tilde{\mu}_0(T_{m_i,j-1}) + \boldsymbol{\beta}_0'\mathbf{X}_i + \alpha_0 H_i(T_{m_i,j-1}-)W_i\}\big]),$$

for $j \geq 2$ with Q_i generating from the gamma distribution with mean 1 and variance 0.1. The simulation results are shown in Table 3. From the table, we have the following findings: (i) With both Poisson and non-Poisson observation processes, the proposed estimator, MSPLE , and MSLE in Lu et al. (2009) are approximately unbiased when $\alpha = 0$. (ii) When $\alpha \neq 0$, the proposed estimator is approximately unbiased while the MSPLE and MSLE in Lu et al. (2009) yield biased estimates and the biases could be larger as α diverges from 0. This means that our proposed estimation procedure seems to be more robust.

We also conducted some sensitivity analysis to evaluate the performance of the proposed estimators for $\boldsymbol{\beta}$ when the interaction term of the recurrent event process and observation process was misspecified. Specifically, we generated the recurrent event counts from a mixed Poisson process as above with $\mu_0(t) = \log(t+1)$ and the interaction term $\phi(\mathcal{F}_{it}, W_i) = (H_i(t-)+1)X_{1i}$, where X_{1i}, X_{2i}, C_i, and the observation times were generated from the same setup as given above. Here we take $\boldsymbol{\beta}_0 = (-1, 1)$. We considered the interaction term ϕ as misspecified by three possible forms: (i) $\phi^{(1)}(\mathcal{F}_{it}, W_i) = (H_i(t-)+0.75)X_{1i}$, (ii) $\phi^{(2)}(\mathcal{F}_{it}, W_i) = \log[\exp\{H_i(t-)+0.5\}+1]X_{1i}$, and (iii) $\phi^{(3)}(\mathcal{F}_{it}, W_i) = \sqrt{\{H_i(t-)\}^2 + 1.2X_{1i}}$. We applied the proposed estimation procedure to the true and misspecified models by using the generated data from the true model. The simulation results are summarized in Table 4. It can be seen from the table that the estimates for $\boldsymbol{\beta}$ are still approximately unbiased for the misspecified situations considered here.

5 Application

In this section, we applied our proposed method to analyze the bladder cancer data. There were 116 subjects with superficial bladder tumors, and they were randomized into one of three treatment groups: placebo, thiotepa, and pyridoxine. Following Sun and Wei (2000), we restricted our attention to the placebo and thiotepa groups with respective sizes of 47 and 38. For each patient, all the clinical visits and the numbers of new tumors between clinical visits were recorded. In addition, two baseline covariates were considered and they were the number of initial tumors and the size of the largest initial tumor.

For the analysis, we define the response process $N_i(t)$ and observation process $H_i(t)$ as the cumulated tumor numbers and the accumulated observation numbers of patient i up to time t, respectively. Further, for patient i, define X_{1i} to be equal to 1 if the i-th patient was given the thiotepa treatment and 0 otherwise, X_{2i} the number of initial tumors and X_{3i} the size of the largest initial tumor, $i = 1, \ldots, 85$. Take $H^*(t) = (H(t-) - 8)/8$ and assume that $\{N_i(t)\}$ can be described by model (1) with $\phi(\mathcal{F}_{it}, W_i) = H_i^*(t-)X_{1i} + H_i^*(t-)X_{2i}$, meaning that the relation between the recurrence rate of bladder tumors and the observation times may vary with different treatments and different number of initial tumors, that is,

$$E\{N_i(t)|X_{1i}, X_{2i}, X_{3i}, \mathcal{F}_{it}\} = \exp\{\tilde{\mu}_0(t) + \beta_1' X_{1i} + \beta_2' X_{2i} + \beta_3' X_{3i} + \\ + \alpha_1' H_i^*(t-)X_{1i} + \alpha_2' H_i^*(t-)X_{2i}\}.$$

Here, we took the last visit time of patient i as C_i in the analysis. For estimation of $\tilde{\mu}_0(t)$, we use the cubic B-spline approximation by taking the number of interior knots m_n as n^ν with $\nu = 1/8$ and the equally spaced knots.

The application of the estimation procedure proposed in the previous sections gave $\hat{\beta}_1 = -1.7143$, $\hat{\beta}_2 = 0.1061$, $\hat{\beta}_3 = -0.0473$, $\hat{\alpha}_1 = -0.6100$, and $\hat{\alpha}_2 = 0.3070$ with the bootstrap standard errors (100 replications) being 0.3205, 0.0465, 0.0335, 0.1060, and 0.1079, which correspond to p-values of $<$0.0000, 0.0225, 0.1584, $<$0.0000, and 0.0044, respectively, based on the asymptotic results of the estimators. Here $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ represent the estimated regression coefficients corresponding to the treatment indicator, the number of initial tumors, and the size of the largest initial tumor, respectively, while $\hat{\alpha}_1$ and $\hat{\alpha}_2$ represent the estimated effects of the interaction between the observation process and the treatment indicator and the interaction between the observation process and the number of initial tumors, respectively, on the tumor recurrence rate. These results indicate that the tumor recurrence process and the interaction between the observation process and the thiotepa treatment indicator are significantly negatively correlated. The reason for this may be that the patients in the thiotepa group needed more clinical visits, had tumors removed, and received treatment, which reduced the tumor occurrence rate. In addition, the interaction between the observation process and the number of initial tumors significantly positively influences the tumor recurrence process, which can be explained by the reason that patients with more initial tumors tend to have more clinical visits and more new tumors will arise due to the more serious illness. Furthermore, the thiotepa treatment significantly reduces the occurrence rate of the bladder tumors, and the number of initial tumors has a significantly positive effect on the tumor recurrence rate. However, the recurrence rate of the bladder tumors does not seem

Table 4 Sensitivity analysis for the misspecification of the interaction term ϕ with $\tilde{\mu}_0(t) = \log(t+1)$ and $\boldsymbol{\beta}_0 = (-1, 1)$.

Obs	n	α	$\hat{\alpha}$				$\hat{\beta}_1$				$\hat{\beta}_2$			
			BIAS	BIAS^1	BIAS^2	BIAS^3	BIAS	BIAS^1	BIAS^2	BIAS^3	BIAS	BIAS^1	BIAS^2	BIAS^3
P	100	0	–0.0146	–0.0140	–0.0167	–0.0091	–0.0040	–0.0040	–0.0004	–0.0102	0.0411	0.0345	0.0193	0.0189
		0.2	–0.0152	–0.0073	–0.0411	–0.0741	0.0083	–0.0576	0.0114	0.0410	0.0252	0.0553	0.0312	0.0247
		0.3	–0.0051	–0.0087	–0.0413	–0.0959	–0.0017	–0.0657	–0.0041	0.0684	0.0156	0.0198	0.0264	0.0118
	200	0	–0.0085	–0.0036	–0.0089	–0.0085	0.0022	–0.0049	0.0003	0.0034	0.0102	0.0082	0.0095	0.0032
		0.2	–0.0037	–0.0059	–0.0387	–0.0629	–0.0019	–0.0442	0.0169	0.0488	0.0222	0.0051	0.0218	0.0220
		0.3	–0.0007	–0.0046	–0.0387	–0.0794	–0.0040	–0.0677	–0.0014	0.0468	0.0173	0.0186	0.0204	0.0071
NP	100	0	–0.0027	–0.0070	–0.0035	–0.0013	–0.0098	0.0158	–0.0066	–0.0159	0.0150	0.0152	0.0196	0.0241
		0.2	–0.0010	0.0002	–0.0119	–0.0260	0.0039	–0.0621	–0.0482	–0.0467	0.0188	0.0327	0.0311	0.0239
		0.3	–0.0013	–0.0008	–0.0162	–0.0323	0.0057	–0.0744	–0.0660	–0.0929	0.0175	0.0151	0.0225	0.0178
	200	0	–0.0012	–0.0032	–0.0025	–0.0012	–0.0005	–0.0015	–0.0013	–0.0116	0.0136	0.0189	0.0130	0.0177
		0.2	–0.0010	–0.0009	–0.0130	–0.0262	0.0011	–0.0514	–0.0339	–0.0497	0.0163	0.0105	0.0112	0.0128
		0.3	–0.0002	0.0009	–0.0152	–0.0314	–0.0042	–0.0813	–0.0731	–0.0903	0.0085	0.0106	0.0110	0.0144

Notes: P and NP represent observation times generated from Poisson and non-Poisson processes, respectively. BIAS, the estimated bias of the parameter with the true interaction term H; BIAS^1, BIAS^2, and BIAS^3, the estimated biases of the parameter with misspecified interaction terms $H^{(1)}$, $H^{(2)}$, and $H^{(3)}$, respectively.

to be significantly related to the size of the largest initial tumor. These conclusions are consistent with those made in Lu et al. (2009), Li et al. (2010), and Zhao and Tong (2011), among others. For example, Lu et al. (2009) analyzed the bladder tumor data, and obtained that the number of initial tumors has a positive effect on the tumor recurrence rate with p-values of 0.012 and 0.01 and the thiotepa treatment significantly reduces the recurrence of tumors with p-values of 0.015 and 0.019 by MSPLE and MSLE, respectively. Compared to the models in Lu et al. (2009), Li et al. (2010), and Zhao and Tong (2011), our fitted model may provide more information about the correlation between the tumor recurrence rate and observation times over treatment groups and also could be useful to estimate the future recurrence rate based on the observation history.

6 Concluding remarks

This article studied a conditional model for the underlying recurrent event process of the panel count data that allows for the interaction between the informative observation times and covariates, leaving the distributional form of the observation process to be arbitrary. We proposed to use the easy implemented monotone spline-based least squares estimation method to estimate the regression parameters and the unknown smooth monotone function in the model simultaneously, and also established the asymptotical results including consistency, rate of convergence, and asymptotic normality for the estimators. As demonstrated by simulation studies, our proposed model and inference procedure are more flexible and applicable since they can overcome the underdispersion or overdispersion problem resulting from the model specification for the observation process.

In the proposed estimation approach, weights have not been considered. Following the discussions made in Hu et al. (2009), our proposed objection function in Eq. (2) may be related to the MSPLE and MSLE (Lu et al., 2009) by taking some specific weight functions. Further studies are needed to develop monotone spline-based weighted least squares methods and provide weight selection procedure for finding optimal weight function in practical applications.

Motivated by Li et al. (2010), our proposed models can also be extended to a class of transformation models

$$E\{N_i(t)|\mathbf{X}_i, W_i, \mathcal{F}_{it}\} = G\{\mu_0(t) + \boldsymbol{\beta}'\mathbf{X}_i + \boldsymbol{\alpha}'\phi(\mathcal{F}_{it}, W_i)\},$$

with a given monotone smooth function G. Then for making inference on these models, the same procedure as presented in this article can be used to obtain the estimators for the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ and the nonparametric monotone function $\mu_0(t)$, and the asymptotic properties of the spline-based estimators could be established by using the similar arguments.

Acknowledgment The authors would like to thank the editor, Professor Mauro Gasparini, associate editor, and two reviewers for their constructive and insightful comments and suggestions that greatly improved the article. Deng's research is partly supported by the Natural Science Foundation of China (nos. 11101315, 11401443). Liu's research is partly supported by the Natural Science Foundation of China (nos. 11101315, 11171262), and Zhao's research is partly supported by the Research Grant Council of Hong Kong (nos. 504011, 503513), Natural Science Foundation of China (no. 11371299), and The Hong Kong Polytechnic University.

Conflict of interest
The authors have declared no conflict of interest.

Appendix: Asymptotic results

To establish the asymptotic properties of the estimators, we need the following regularity conditions.

C1. The maximum spacing of the knots satisfies $\triangle = \max_{l+1<i<m_n+l+1} \mid t_i - t_{i-1} \mid = O(n^{-\nu})$.

C2. The parameter spaces of $(\boldsymbol{\beta}', \boldsymbol{\alpha}')'$, $\mathcal{R}$ is bounded and convex on $\mathbb{R}^{p+q}$, and the true parameter $\boldsymbol{\theta_0} = (\boldsymbol{\beta_0}, \boldsymbol{\alpha_0}, \mu_0) \in \mathcal{R}^\circ \times \mathcal{F}_r$, where $\mathcal{R}^\circ$ is the interior of $\mathcal{R}$, and

$$\mathcal{F}_r \equiv \left\{ \mu : [0, \infty) \longrightarrow \mathbb{R} \middle| \mu \text{ is monotone and } |\mu^{(k)}(s) - \mu^{(k)}(t)| \leq M|s-t|^{\varsigma} \right\},$$

where k is a nonnegative integer, $\varsigma \in (0, 1]$ such that $r = k + \varsigma > 0.5$, M is a positive constant and $f^{(k)}$ is the k-th derivative of function f.

C3. $N_i(\tau)(i = 1, \cdots, n)$ are bounded by a constant, $h(\cdot)$ has a bounded total variation, and there exist a positive integer M_1, such that $P(\|\mathbf{X}\| \leq M_1) = 1$, that is, the covariate vector is uniformly bounded.

C4. There exists a positive integer M_2 such that $P(K \leq M_2) = 1$, that is, the number of the observation is finite.

C5. If with probability 1, $\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_t, W) + h_3(t) = 0$ for $\mathbf{h}_1 \in \mathbb{R}^p$, $\mathbf{h}_2 \in \mathbb{R}^q$ and some deterministic function h_3, then $\mathbf{h}_1 = 0$, $\mathbf{h}_2 = 0$, $h_3(t) = 0$.

Next, we introduce more notation. Let

$$\mathcal{F} = \left\{ f : [0, \infty] \longrightarrow \mathbb{R} \middle| \|f\|_2 = \left[E\left\{ \int_0^\tau |f(t)|^2 \xi(t) dH(t) \right\} \right]^{1/2} < \infty \right\}.$$

Let $Z = \{Z(t, W) \equiv \phi(\mathcal{F}_t, W), 0 \leq t \leq \tau\}$ represent a q-dimensional bounded random process indexed by t. Here, without loss of generality, we assume that W is one-dimensional. Define

$$\mathcal{G} \equiv \{ z(t, w) : [0, \tau] \times [-M_1, M_1] \longrightarrow \mathcal{M} \},$$

where $\mathcal{M}$ is a bounded set on $\mathbb{R}^q$, and for function $f(\mathbf{x}, z, t) : [-M_1, M_1]^p \times \mathcal{G} \times [0, \tau] \longrightarrow \mathbb{R}$, define

$$\|f\|_2 \equiv \left[E\left\{ \sum_{j=1}^K |f(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|^2 \xi(T_{K,j}) \right\} \right]^{1/2}.$$

Set $M_n(g) = n^{-1} L_n(\theta) = \mathbb{P}_n m_g(\mathbf{O})$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \mu)$, $g(\mathbf{x}, z, t) = \exp\{\boldsymbol{\beta}'\mathbf{x} + \boldsymbol{\alpha}' z(t, w) + \tilde{\mu}(t)\}$ and

$$m_g(\mathbf{O}) = \sum_{j=1}^K [N(T_{K,j}) - g(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})]^2 \xi(T_{K,j}),$$

and $M(g) = P m_g(\mathbf{O})$, where Pf and $\mathbb{P}_n f$ represent $\int f dP$ and $n^{-1} \sum_{i=1}^n f(\mathbf{O}_i)$, respectively.

Since $\mathcal{F}$ is a Hilbert space, and $\mathcal{F}_r \subset \mathcal{F}$, by the Hilbert Projection Theorem (Stakgold, 1998, p. 288), for $x_j \in \mathcal{F}$, there is a unique $a_j^* \in \mathcal{F}_r$, s.t. $(x_j - a_j^*) \perp \mathcal{F}_r$, for $j = 1, \cdots, p$. Let $z_l(t, w)$ be the l-th component of $h(\mathcal{F}_t, W)$, $l = 1, \cdots, q$. Then for $z_l(t, w) \in \mathcal{F}$, there is a unique $b_l^*(t) \in \mathcal{F}_r$, s.t. $(z_l - b_l^*) \perp \mathcal{F}_r$, for $l = 1, \cdots, q$. Let $\mathbf{a}^* = (a_1^*, \cdots, a_p^*)'$ and $\mathbf{b}^* = (b_1^*, \cdots, b_q^*)'$. Then we need the following condition.

C6. $E[\int_0^\tau \begin{pmatrix} \mathbf{X} - \mathbf{a}^* \\ \phi(\mathcal{F}_t, W) - \mathbf{b}^*(t) \end{pmatrix}^{\otimes 2} d\tilde{H}(t)]$ is nonsingular, where $a^{\otimes 2} = a'a$ for a vector a.

Here, Condition C1 is similar to those required by Stone (1986) and Zhou et al. (1998); Condition C2 is a common assumption in nonparametric smoothing estimation problems; Conditions C3 and C4 are mild and easily justified in many applications; Condition C5 is needed to establish the identifiability of the model; Condition C6 is a technical condition used in the proof for consistency of estimators.

Let $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\alpha}}_n, \hat{\mu}_n)$ be the estimator of $\boldsymbol{\theta}_0$. For $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \boldsymbol{\alpha}_1, \mu_1)$, $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, \boldsymbol{\alpha}_2, \mu_2)$, define

$$\rho\big(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\big) = \big\{\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|^2 + \|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|^2 + \|\mu_1 - \mu_2\|_2^2\big\}^{1/2}.$$

Then the asymptotic properties of the estimators $\hat{\boldsymbol{\theta}}_n$ are summarized as follows.

Theorem 1 (Consistency). *Under conditions C1–C4 and C6,* $\rho\big(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0\big) \to 0$*, almost surely.*

Theorem 2 (Rate of Convergence). *Under conditions C1–C6,*

$$\rho\big(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0\big) = O_p(n^{-\min\{\nu r, \frac{1-\nu}{2}\}}).$$

Remark 1. When $\nu = 1/(1+2r)$, $n^{-\min\{\nu r, \frac{1-\nu}{2}\}} = n^{-\frac{r}{1+2r}}$, we conclude from Stone (1980, 1982) that the rate of convergence of the estimator $\hat{\mu}_n$ is the optimal rate in nonparametric regression.

To state the asymptotic normality, we define

$$A = E\left\{\sum_{j=1}^{K} \xi(T_{K,j}) g_0^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \times \right.$$
$$\left. \times\, a^{-2}(K, T_{K,j}) \mathbf{c}(\mathbf{X}, Z(T_{K,j}, W), K, T_{K,j})^{\otimes 2}\right\}$$

and

$$B = E\left\{\sum_{j=1}^{K}\sum_{j'=1}^{K} \xi(T_{K,j})\xi(T_{K,j'}) b(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) b(\mathbf{X}, Z(T_{K,j'}, W), T_{K,j'}) \times \right.$$
$$\times\, g_0(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) g_0(\mathbf{X}, Z(T_{K,j'}, W), T_{K,j'}) a^{-1}(K, T_{K,j}) a^{-1}(K, T_{K,j'}) \times$$
$$\left. \times\, \mathbf{c}(\mathbf{X}, Z(T_{K,j}, W), K, T_{K,j}) \mathbf{c}(\mathbf{X}, Z(T_{K,j'}, W), K, T_{K,j'})'\right\},$$

where $g_0(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) = \exp\{\boldsymbol{\beta}_0'\mathbf{X} + \boldsymbol{\alpha}_0'\phi(\mathcal{F}_{T_{K,j}}, W) + \tilde{\mu}_0(T_{K,j})\}$,

$$a(K, T_{K,j}) \equiv E\Big\{\xi(T_{K,j}) g_0^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \Big| K, T_{K,j}\Big\},$$
$$b(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \equiv N(T_{K,j}) - g_0(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}),$$
$$\mathbf{c}(\mathbf{X}, Z(T_{K,j}, W), K, T_{K,j}) \equiv \begin{pmatrix} \mathbf{X} \\ \phi(\mathcal{F}_{T_{K,j}}, W) \end{pmatrix} a(K, T_{K,j}) - E\left\{\begin{pmatrix} \mathbf{X} \\ \phi(\mathcal{F}_{T_{K,j}}, W) \end{pmatrix} \times \right.$$
$$\left. \times\, \xi(T_{K,j}) g_0^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \Big| K, T_{K,j}\right\}.$$

Theorem 3 (Asymptotic normality). *Under conditions C1–C6,* $\sqrt{n}\begin{pmatrix}\hat{\beta}_n-\beta_0\\ \hat{\alpha}_n-\alpha_0\end{pmatrix}$ *converges in distribution to* $N(\mathbf{0},\Sigma)$ *with* $\Sigma=A^{-1}B(A^{-1})'$.

Proof of Theorem 1. According to Lemma 5 in Stone (1985), for $\tilde{\mu}_0\in\mathcal{F}_r$, there exist a $\tilde{\mu}_n\in\Phi_{l,\mathcal{I}}$ with order $l\geq k+1$ and knots $\mathcal{I}$ such that $\|\tilde{\mu}_n-\tilde{\mu}_0\|_\infty=O(n^{-\nu r})$. Let $g_n(\mathbf{x},z,t)=\exp\{\boldsymbol{\beta}_0'\mathbf{x}+\boldsymbol{\alpha}_0'z(t,w)+\tilde{\mu}_n(t)\}$, $\hat{g}_n(\mathbf{x},z,t)=\exp\{\hat{\boldsymbol{\beta}}_n'\mathbf{x}+\hat{\boldsymbol{\alpha}}_n'z(t,w)+\hat{\tilde{\mu}}_n(t)\}$, and $g_0(\mathbf{x},z,t)=\exp\{\boldsymbol{\beta}_0'\mathbf{x}+\boldsymbol{\alpha}_0'z(t,w)+\tilde{\mu}_0(t)\}$. Then by using the arguments similar to those used in the proof of Theorem 1 of Lu et al. (2007), we can show that $\|\hat{g}_n-g_0\|_2^2=O(n^{-\nu r}+n^{-\frac{1-\nu}{2}})$ and hence $\|\log\hat{g}_n-\log g_0\|_2^2=O(n^{-\nu r}+n^{-\frac{1-\nu}{2}})$. Also note that

$$\begin{aligned}\|\log\hat{g}_n-\log g_0\|_2&=\|(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)'\mathbf{x}+(\hat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0)'z+(\hat{\tilde{\mu}}_n-\tilde{\mu}_0)\|_2=\\&=\|(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)'(\mathbf{x}-\mathbf{a}^*)+(\hat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0)'(z-\mathbf{b}^*)+\\&\quad+(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)'\mathbf{a}^*+(\hat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0)'\mathbf{b}^*+(\hat{\tilde{\mu}}_n-\tilde{\mu}_0)\|_2=\\&=\|(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)'(\mathbf{x}-\mathbf{a}^*)+(\hat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0)'(z-\mathbf{b}^*)\|_2+\\&\quad+\|(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)'\mathbf{a}^*+(\hat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0)'\mathbf{b}^*+(\hat{\tilde{\mu}}_n-\tilde{\mu}_0)\|_2.\end{aligned}$$

By C6, we can get $\|\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\|\longrightarrow 0$, and $\|\hat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\|\longrightarrow 0$ from the first term of the right-hand side of the above equality, and thus it follows that $\|\hat{\tilde{\mu}}_n-\tilde{\mu}_0\|_2\longrightarrow 0$. Therefore, $\|\hat{\mu}_n-\mu_0\|_2=\|\exp\{\hat{\tilde{\mu}}_n\}-\exp\{\tilde{\mu}_0\}\|_2\longrightarrow 0$ by Taylor expansion. This completes the proof of Theorem 1. □

Proof of Theorem 2. Let $\mathcal{F}_{nr}=\Phi_{l,\mathcal{I}}$ and $\Theta_n^r=\mathcal{R}^0\times\mathcal{F}_{nr}$. It is easy to see that $\Theta_n^r\subseteq\Theta_{n+1}^r\subseteq\cdots\subseteq\Theta^r=\mathcal{R}^0\times\mathcal{F}_r$ for $n\geq 1$. Note that the sieve estimator $\hat{\boldsymbol{\theta}}_n=(\hat{\boldsymbol{\beta}}_n,\hat{\boldsymbol{\alpha}}_n,\hat{\mu}_n)$ is the minimizer of $L_n(\boldsymbol{\theta})$ over the sieve space Θ_n^r.

For each n, $g_n(\mathbf{x},\boldsymbol{z},t)=\exp\{\boldsymbol{\beta}_0'\mathbf{x}+\boldsymbol{\alpha}_0'z(t,w)+\tilde{\mu}_n(t)\}$ with $\boldsymbol{\theta}_n=(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tilde{\mu}_n)\in\Theta_n^r$. Then for any $\eta>0$, let

$$\mathcal{F}_{n,\eta}\equiv\{g=\exp\{\boldsymbol{\beta}'\mathbf{x}+\boldsymbol{\alpha}'\boldsymbol{z}+\tilde{\mu}\}:\ \boldsymbol{\theta}=(\boldsymbol{\beta},\boldsymbol{\alpha},\tilde{\mu})\in\Theta_n^r\text{ and }\frac{\eta}{2}<\|g-g_n\|_2\leq\eta\}.$$

Similar to Lemma A.2 in Huang (1999, p. 1557), for any $\varepsilon\leq\eta$,

$$\log N_{[]}(\varepsilon,\mathcal{F}_{n,\eta},\|\cdot\|_2)\leq c_4q_n\log(\eta/\varepsilon),$$

for a constant c_4. Thus, for $\varepsilon>0$, there exists a set of brackets $\{[g_i^l,g_i^r],i=1,\cdots,(\frac{\eta}{\varepsilon})^{c_4q_n}\}$ such that, for each $g\in\mathcal{F}_\eta$, there is a $[g_s^l,g_s^r]$ with $g_s^l(\mathbf{x},\boldsymbol{z},t)\leq g(\mathbf{x},\boldsymbol{z},t)\leq g_s^r(\mathbf{x},\boldsymbol{z},t)$, for all $\mathbf{x}$, $t\in[0,\tau]$ and $\boldsymbol{z}\in\mathcal{G}$, and $\|g_s^r-g_s^l\|_2^2\leq\varepsilon^2$.

Next, consider the class $\mathcal{M}_{n,\eta}\equiv\{m_g(\mathbf{O})-m_{g_n}(\mathbf{O}):g\in\mathcal{F}_{n,\eta}\}$, where

$$m_g(\mathbf{O})=\sum_{j=1}^{K}[N(T_{K,j})-g(\mathbf{X},Z(T_{K,j},W),T_{K,j})]^2\xi(T_{K,j}).$$

For $i=1,\cdots,(\frac{\eta}{\varepsilon})^{c_4q_n}$, define

$$\begin{aligned}m_i^l(\mathbf{O})=\sum_{j=1}^{K}\Big[&\left\{|g_i^l|\vee|g_i^r|\right\}^2(\mathbf{X},Z(T_{K,j},W),T_{K,j})-2N(T_{K,j})g_i^r(\mathbf{X},Z(T_{K,j},W),T_{K,j})+\\&+2N(T_{K,j})g_n(\mathbf{X},Z(T_{K,j},W),T_{K,j})-g_n^2(\mathbf{X},Z(T_{K,j},W),T_{K,j})\Big]\xi(T_{K,j}),\end{aligned}$$

$$m_i^r(\mathbf{O}) = \sum_{j=1}^{K} \Big[\big\{ |g_i^l| \wedge |g_i^r| \big\}^2 (\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - 2N(T_{K,j}) g_i^l(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) +$$
$$+ 2N(T_{K,j}) g_n(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - g_n^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \Big] \xi(T_{K,j}),$$

where $a \vee b = \min\{a, b\}$ and $a \wedge b = \max\{a, b\}$. Then, it is easy to show that $\|m_i^r(\mathbf{O}) - m_i^l(\mathbf{O})\|_{P,B}^2 \leq c_5\varepsilon^2$ with a positive constant c_5, where $\|\cdot\|_{P,B}$ is the "Bernstein norm" defined by $\|f\|_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$ (see van der Vaart and Wellner, 1996, p. 324). Thus $\{[m_i^l(\mathbf{O}), m_i^r(\mathbf{O})], i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_4 q_n}\}$ is the set of brackets for $\mathcal{M}_{n,\eta}$, which implies that

$$\log N_{[]}(\varepsilon, \mathcal{M}_{n,\eta}, \|\cdot\|_{P,B}) \leq c_4 q_n \log(\eta/\varepsilon).$$

Moreover, by some calculations, we can verify that $\|m_g(\mathbf{O}) - m_{g_n}(\mathbf{O})\|_{P,B}^2 \leq c_6\eta^2$ for any $g \in \mathcal{F}_{n,\eta}$ by C3 and C4. Therefore, by Lemma 3.4.3 of van der Vaart and Wellner (1996), we obtain

$$E\|n^{1/2}(\mathbb{P} - P)\|_{\mathcal{M}_{n,\eta}} \leq c_7 \tilde{J}_{[]}(\eta, \mathcal{M}_{n,\eta}, \|\cdot\|_{P,B}) \left\{ 1 + \frac{\tilde{J}_{[]}(\eta, \mathcal{M}_{n,\eta}, \|\cdot\|_{P,B})}{\eta^2 n^{1/2}} M_3 \right\}, \tag{A1}$$

where M_3 is a constant, $\|n^{1/2}(\mathbb{P} - P)\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |n^{1/2}(\mathbb{P} - P)f|$, and

$$\tilde{J}_{[]}(\eta, \mathcal{M}_{n,\eta}, \|\cdot\|_{P,B}) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{M}_{n,\eta}, \|\cdot\|_{P,B})\}^{1/2} d\varepsilon \leq c_8 q_n^{1/2} \eta.$$

The right-hand side of Eq. (A1) yields $\varphi_n(\eta) = c_9(q_n^{1/2}\eta + q_n/n^{1/2})$. It is easy to see that $\varphi_n(\eta)/\eta$ is decreasing in η, and

$$r_n^2 \varphi\left(\frac{1}{r_n}\right) = r_n q_n^{1/2} + r_n^2 q_n / n^{1/2} \leq 2n^{1/2},$$

for $r_n = n^{\frac{1-\nu}{2}}$ and $0 < \nu < 1/2$.

Note that

$$\begin{aligned}
& P[m_g(\mathbf{O}) - m_{g_n}(\mathbf{O})] = \\
&= P\left[\int_0^\tau \left\{ [N(t) - g(\mathbf{X}, Z(t, W), t)]^2 - [N(t) - g_n(\mathbf{X}, Z(t, W), t)]^2 \right\} \xi(t) dH(t) \right] = \\
&= E\left[\int_0^\tau (g_n - g)(\mathbf{X}, Z(t, W), t)[2N(t) - (g + g_n)(\mathbf{X}, Z(t, W), t)] \xi(t) dH(t) \right] \geq \\
&\geq E\left[\int_0^\tau (g - g_n)^2 (\mathbf{X}, Z(t, W), t) \xi(t) dH(t) \right] \geq \\
&\geq \|g - g_n\|_2^2.
\end{aligned}$$

Furthermore, in the proof of Theorem 4.1, we have that $\|\hat{g}_n - g_n\|_2$ converges to zero in probability. Thus, by Theorem 3.4.1 of van der Vaart and Wellner (1996),

$$n^{\frac{1-\nu}{2}}\|\hat{g}_n - g_n\|_2 = O_p(1).$$

Since $\|\tilde{\mu}_n - \tilde{\mu}_0\|_\infty = O(n^{-\nu r})$, then $\|g_n - g_0\|_2 = O_p(n^{-\nu r})$, and so

$$\|\hat{g}_n - g_0\|_2 = O_p(n^{-\nu r}) + O_p(n^{-\frac{1-\nu}{2}}) = O_p(n^{-\min\{\nu r, \frac{1-\nu}{2}\}}).$$

Therefore by the similar arguments as those used in the proof of consistency of $\hat{\boldsymbol{\beta}}_n$, $\hat{\boldsymbol{\alpha}}_n$, and $\hat{\tilde{\mu}}_n$, we can get the rate of convergence of $\hat{\tilde{\mu}}_n$, $\hat{\boldsymbol{\beta}}_n$, and $\hat{\boldsymbol{\alpha}}_n$ as stated in Theorem 4.2. The choice of $\nu = 1/(1+2r)$ yields the rate of convergence of $r/(1+2r)$, which completes the proof. □

Proof of Theorem 3. Let $\mathcal{H} \equiv \{(\mathbf{h}_1, \mathbf{h}_2, h_3) : (\mathbf{h}_1', \mathbf{h}_2')' \in \mathcal{R}, h_3 \in \mathcal{F}_r, \|\mathbf{h}_1\| \le 1, \|\mathbf{h}_2\| \le 1, \|h_3\|_\infty \le 1\}$. We define a sequence of maps S_n mapping a neighborhood of $\tilde{\boldsymbol{\theta}}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tilde{\mu}_0)$, denoted by $\mathcal{U}$, in the parameter space for $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \tilde{\mu})$ into $l^\infty(\mathcal{H})$ as

$$\begin{aligned} S_n(\tilde{\theta})[\mathbf{h}_1, \mathbf{h}_2, h_3] &\equiv n^{-1}\frac{d}{d\varepsilon}L_n(\boldsymbol{\beta} + \varepsilon\mathbf{h}_1, \boldsymbol{\alpha} + \varepsilon\mathbf{h}_2, \tilde{\mu} + \varepsilon h_3)\Big|_{\varepsilon=0} = \\ &= -\frac{2}{n}\sum_{i=1}^{n}\int_0^\tau [N_i(t) - \exp\{\beta'\mathbf{X}_i + \alpha'\phi(\mathcal{F}_{it}, W_i) + \tilde{\mu}(t)\}] \times \\ &\quad \times \exp\{\beta'\mathbf{X}_i + \alpha'\phi(\mathcal{F}_{it}, W_i) + \tilde{\mu}(t)\} \times \\ &\quad \times [\mathbf{h}_1'\mathbf{X}_i + \mathbf{h}_2'\phi(\mathcal{F}_{it}, W_i) + h_3(t)]d\tilde{H}_i(t) \equiv \\ &\equiv \mathbb{P}_n\psi(\tilde{\theta}; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3], \end{aligned}$$

where

$$\begin{aligned} \psi(\tilde{\theta}; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] = -2\int_0^\tau &[N(t) - \exp\{\beta'\mathbf{X} + \alpha'\phi(\mathcal{F}_t, W) + \tilde{\mu}(t)\} \times \\ &\times \exp\{\beta'\mathbf{X} + \alpha'\phi(\mathcal{F}_t, W) + \tilde{\mu}(t)\} \times \\ &\times [\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_t, W) + h_3(t)]d\tilde{H}(t). \end{aligned}$$

Correspondingly, we define the limit map $S : \mathcal{U} \longrightarrow l^\infty(\mathcal{H})$ as

$$S(\tilde{\theta})[\mathbf{h}_1, \mathbf{h}_2, h_3] = P\psi(\tilde{\theta}; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3],$$

where $l^\infty(\mathcal{H})$ is the space of bounded functionals on $\mathcal{H}$ under the supermum norm $\|f\|_\infty = \sup_{h\in\mathcal{H}}|f(h)|$.

To derive the asymptotic normality of the estimator $\hat{\tilde{\boldsymbol{\theta}}}_n = (\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\alpha}}_n, \hat{\tilde{\mu}}_n)$, motivated by the proof of Theorem 3.3.1 of van der Vaart and Wellner (1996), we first need to verify the following four conditions.

(i) $\sqrt{n}(S_n - S)(\hat{\tilde{\boldsymbol{\theta}}}_n) - \sqrt{n}(S_n - S)(\tilde{\boldsymbol{\theta}}_0) = o_p(1)$.
(ii) $\sqrt{n}(S_n - S)(\tilde{\boldsymbol{\theta}}_0)$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H})$.
(iii) $S(\tilde{\boldsymbol{\theta}}_0) = 0$ and $S_n(\hat{\tilde{\boldsymbol{\theta}}}_n) = o_p(n^{-1/2})$.

(iv) $\tilde{\boldsymbol{\theta}} \longmapsto S(\tilde{\boldsymbol{\theta}})$ is Fréchet-differentiable at $\tilde{\boldsymbol{\theta}}_0$ with a continuous derivative $\dot{S}(\tilde{\boldsymbol{\theta}}_0)$ and

$$\sqrt{n}\Big(S(\hat{\tilde{\boldsymbol{\theta}}}_n) - S(\tilde{\boldsymbol{\theta}}_0)\Big) - \sqrt{n}\dot{S}(\tilde{\boldsymbol{\theta}}_0)\Big(\hat{\tilde{\boldsymbol{\theta}}}_n - \tilde{\boldsymbol{\theta}}_0\Big) = o_p(1).$$

Note that

$$\begin{aligned}&\sqrt{n}(S_n - S)(\hat{\tilde{\boldsymbol{\theta}}}_n)[\mathbf{h}_1, \mathbf{h}_2, h_3] - \sqrt{n}(S_n - S)(\tilde{\boldsymbol{\theta}}_0)[\mathbf{h}_1, \mathbf{h}_2, h_3] = \\ &= \sqrt{n}(\mathbb{P}_n - P)\Big(\psi(\hat{\tilde{\boldsymbol{\theta}}}_n; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] - \psi(\tilde{\boldsymbol{\theta}}_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3]\Big).\end{aligned}$$

For some $\delta > 0$, define

$$\mathcal{F}_\delta = \Big\{\psi(\tilde{\theta}; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] - \psi(\tilde{\theta}_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] : \rho\big(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}_0\big) < \delta, (\mathbf{h}_1, \mathbf{h}_2, h_3) \in \mathcal{H}\Big\}.$$

It is easy to see that $\mathcal{H}$ is a Donsker class base on the argument in pages 154–157 of van der Vaart and Wellner (1996). Since

$$\begin{aligned}&\big|\psi(\tilde{\theta}; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3]\big| = \\ &= \Big| -2\int_0^\tau [N(t) - \exp\{\boldsymbol{\beta}'\mathbf{X} + \boldsymbol{\alpha}'\phi(\mathcal{F}_t, W) + \tilde{\mu}(t)\}]\exp\{\boldsymbol{\beta}'\mathbf{X} + \boldsymbol{\alpha}'\phi(\mathcal{F}_t, W) + \tilde{\mu}(t)\} \times \\ &\quad \times [\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_t, W) + h_3(t)]d\tilde{H}(t)\Big| \le \\ &\le M_1\|\mathbf{h}_1\| + M_2\|\mathbf{h}_2\| + M_3\|h_3\|_\infty,\end{aligned}$$

for constants M_1, M_2, M_3, which means that $\psi(\tilde{\theta}; \mathbf{O})$ is a bounded Lipschitz functional with respect to $\mathcal{H}$; thus $\mathcal{F}_\delta$ is a Donsker class for $\delta > 0$. For any $\tilde{\boldsymbol{\theta}}_1 = (\boldsymbol{\beta}_1, \boldsymbol{\alpha}_1, \tilde{\mu}_1)$ and $\tilde{\boldsymbol{\theta}}_2 = (\boldsymbol{\beta}_2, \boldsymbol{\alpha}_2, \tilde{\mu}_2)$, it is easy to verify that

$$P\Big|\psi(\tilde{\boldsymbol{\theta}}_1; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] - \psi(\tilde{\boldsymbol{\theta}}_2; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3]\Big|^2 \le c_{10}\rho^2\big(\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2\big),$$

for a constant c_{10}. Thus condition (i) holds by Lemma 13.3 of Kosorok (2008).

Condition (ii) is also satisfied since $\{\psi(\tilde{\boldsymbol{\theta}}_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] : (\mathbf{h}_1, \mathbf{h}_2, h_3) \in \mathcal{H}\}$ is a Donsker class.

Clearly, $S(\tilde{\boldsymbol{\theta}}_0) = 0$. For $h_3 \in \mathcal{F}_r$, let h_{3n} be the B-spline function approximation of h_3 with $\|h_{3n} - h_3\|_\infty = O(n^{-\nu r})$, then we have $S_n(\hat{\tilde{\boldsymbol{\theta}}}_n)[\mathbf{h}_1, \mathbf{h}_2, h_{3n}] = 0$.

Thus, for $(\mathbf{h}_1, \mathbf{h}_2, h_3) \in \mathcal{H}$,

$$\begin{aligned}&n^{\frac{1}{2}}S_n(\hat{\tilde{\theta}}_n)[\mathbf{h}_1, \mathbf{h}_2, h_3] = \\ &= n^{\frac{1}{2}}\big\{S_n(\hat{\tilde{\boldsymbol{\theta}}}_n)[\mathbf{h}_1, \mathbf{h}_2, h_3] - S_n(\hat{\tilde{\boldsymbol{\theta}}}_n)[\mathbf{h}_1, \mathbf{h}_2, h_{3n}]\big\} = \\ &= n^{\frac{1}{2}}(\mathbb{P}_n - P)\psi(\hat{\tilde{\boldsymbol{\theta}}}_n; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] - n^{\frac{1}{2}}(\mathbb{P}_n - P)\psi(\tilde{\boldsymbol{\theta}}_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] + \\ &\quad - \Big\{n^{\frac{1}{2}}(\mathbb{P}_n - P)\psi(\hat{\tilde{\boldsymbol{\theta}}}_n; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_{3n}] - n^{\frac{1}{2}}(\mathbb{P}_n - P)\psi(\tilde{\boldsymbol{\theta}}_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_{3n}]\Big\} + \\ &\quad + n^{\frac{1}{2}}\mathbb{P}_n\Big\{\psi(\tilde{\boldsymbol{\theta}}_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] - \psi(\tilde{\boldsymbol{\theta}}_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_{3n}]\Big\} + \\ &\quad + n^{\frac{1}{2}}P\Big\{\psi(\hat{\tilde{\boldsymbol{\theta}}}_n; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] - \psi(\hat{\tilde{\boldsymbol{\theta}}}_n; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_{3n}]\Big\} \equiv \\ &\equiv Q_{1n} - Q_{2n} + Q_{3n} + Q_{4n}.\end{aligned}$$

It follows from condition (i) that both Q_{1n} and Q_{2n} are $o_p(1)$. Note that

$$\begin{aligned}
|Q_{4n}| &= \Big|2n^{\frac{1}{2}}P\int_0^\tau [\exp\{\hat{\boldsymbol{\beta}}_n'\mathbf{X}+\hat{\boldsymbol{\alpha}}_n'\phi(\mathcal{F}_t,W)+\hat{\tilde{\mu}}_n(t)\}-N(t)]\times \\
&\quad \times \exp\{\hat{\boldsymbol{\beta}}_n'\mathbf{X}+\hat{\boldsymbol{\alpha}}_n'\phi(\mathcal{F}_t,W)+\hat{\tilde{\mu}}_n(t)\}(h_3-h_{3n})(t)d\tilde{H}(t)\Big| \le \\
&\le c_{11}\Big|n^{\frac{1}{2}}P\int_0^\tau [\exp\{\hat{\boldsymbol{\beta}}_n'\mathbf{X}+\hat{\boldsymbol{\alpha}}_n'\phi(\mathcal{F}_t,W)+\hat{\tilde{\mu}}_n(t)\}- \\
&\quad -\exp\{\boldsymbol{\beta}_0'\mathbf{X}+\boldsymbol{\alpha}_0'\phi(\mathcal{F}_t,W)+\tilde{\mu}_0(t)\}](h_3-h_{3n})(t)d\tilde{H}(t)\Big| = \\
&= c_{11}\Big|n^{\frac{1}{2}}P\int_0^\tau \exp\{f_n^*\}\{\hat{f}_n-f_0\}(h_3-h_{3n})(t)d\tilde{H}(t)\Big| \le \\
&\le c_{12}n^{\frac{1}{2}}\rho\big(\hat{\tilde{\theta}}_n,\tilde{\theta}_0\big)\|h_{3n}-h_3\|_\infty \le \\
&\le n^{\frac{1}{2}}O(n^{-\frac{1-\nu}{2}})O(n^{-\nu r}) = o_p(1)
\end{aligned}$$

for constants c_{11} and c_{12}, where $f_0=\boldsymbol{\beta}_0'\mathbf{X}+\boldsymbol{\alpha}_0'\phi(\mathcal{F}_t,W)+\tilde{\mu}_0$, $\hat{f}_n=\hat{\boldsymbol{\beta}}_n'\mathbf{X}+\hat{\boldsymbol{\alpha}}_n'\phi(\mathcal{F}_t,W)+\hat{\tilde{\mu}}_n$, and $f_n^*=(1-\xi)f_0+\xi\hat{f}_n$ with $0\le\xi\le1$. Furthermore, Q_{3n} is also $o_p(1)$ since

$$\begin{aligned}
&P\Big[\psi(\tilde{\boldsymbol{\theta}}_0;\mathbf{O})[\mathbf{h}_1,\mathbf{h}_2,h_3]-\psi(\tilde{\boldsymbol{\theta}}_0;\mathbf{O})[\mathbf{h}_1,\mathbf{h}_2,h_{3n}]\Big]^2 = \\
&= P\Big[2\int_0^\tau [N(t)-\exp\{\boldsymbol{\beta}_0'\mathbf{X}+\boldsymbol{\alpha}_0'\phi(\mathcal{F}_t,W)+\tilde{\mu}_0(t)\}]\times \\
&\quad \times \exp\{\boldsymbol{\beta}_0'\mathbf{X}+\boldsymbol{\alpha}_0'\phi(\mathcal{F}_t,W)+\tilde{\mu}_0(t)\}(h_{3n}-h_3)(t)d\tilde{H}(t)\Big]^2 \le \\
&\le c_{13}\|h_{3n}-h_3\|_\infty^2 \longrightarrow 0, \qquad n\longrightarrow\infty,
\end{aligned}$$

for a constant c_{13}. Thus $S_n(\hat{\tilde{\boldsymbol{\theta}}}_n)=o_p(n^{-1/2})$.

For the proof of condition (iv), by the smoothness of $S(\tilde{\boldsymbol{\theta}})$, the Fréchet differentiability holds and the derivative of $S(\tilde{\boldsymbol{\theta}})$ at $(\tilde{\boldsymbol{\theta}}_0)$, denoted by $\dot{S}(\tilde{\boldsymbol{\theta}}_0)$, is a map from the space $\{(\tilde{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}_0):\tilde{\boldsymbol{\theta}}\in\mathcal{U}\}$ to $l^\infty(\mathcal{H})$ and

$$\begin{aligned}
&\dot{S}(\tilde{\boldsymbol{\theta}}_0)(\tilde{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}_0)[\mathbf{h}_1,\mathbf{h}_2,h_3] = \\
&= \frac{dS(\boldsymbol{\beta}_0+\varepsilon(\boldsymbol{\beta}-\boldsymbol{\beta}_0),\boldsymbol{\alpha}_0+\varepsilon(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0),\tilde{\mu}+\varepsilon(\tilde{\mu}-\tilde{\mu}_0))[\mathbf{h}_1,\mathbf{h}_2,h_3]}{d\varepsilon}\Big|_{\varepsilon=0} = \\
&= 2P\int_0^\tau [2\exp\{\boldsymbol{\beta}_0'\mathbf{X}+\boldsymbol{\alpha}_0'\phi(\mathcal{F}_t,W)+\tilde{\mu}_0(t)\}-N(t)]\exp\{\boldsymbol{\beta}_0'\mathbf{X}+\boldsymbol{\alpha}_0'\phi(\mathcal{F}_t,W)+\tilde{\mu}_0(t)\}\times \\
&\quad \times[(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{X}+(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)'\phi(\mathcal{F}_t,W)+(\tilde{\mu}-\tilde{\mu}_0)(t)]\times \\
&\quad \times[\mathbf{h}_1'\mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]d\tilde{H}(t) = \\
&= 2P\int_0^\tau \exp\{2(\boldsymbol{\beta}_0'\mathbf{X}+\boldsymbol{\alpha}_0'\phi(\mathcal{F}_t,W)+\tilde{\mu}_0(t))\}\times \\
&\quad \times[(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{X}+(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)'\phi(\mathcal{F}_t,W)+(\tilde{\mu}-\tilde{\mu}_0)(t)]\times \\
&\quad \times[\mathbf{h}_1'\mathbf{X}+\mathbf{h}_2'\phi(\mathcal{F}_t,W)+h_3(t)]d\tilde{H}(t) \equiv \\
&\equiv \sigma_1(\mathbf{h}_1,\mathbf{h}_2,h_3)'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\sigma_2(\mathbf{h}_1,\mathbf{h}_2,h_3)'(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)+\int_0^\tau(\tilde{\mu}-\tilde{\mu}_0)(t)d\sigma_3(\mathbf{h}_1,\mathbf{h}_2,h_3),
\end{aligned}$$

where

$$\sigma_1(\mathbf{h}_1, \mathbf{h}_2, h_3) = 2P\left\{\int_0^\tau g_0^2(\boldsymbol{X}, Z(t, W), t)[\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_t, W) + h_3(t)]\mathbf{X}d\tilde{H}(t)\right\},$$

$$\sigma_2(\mathbf{h}_1, \mathbf{h}_2, h_3) = 2P\left\{\int_0^\tau g_0^2(\boldsymbol{X}, Z(t, W), t)[\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_t, W) + h_3(t)]\phi(\mathcal{F}_t, W)d\tilde{H}(t)\right\},$$

and

$$\sigma_3(\mathbf{h}_1, \mathbf{h}_2, h_3)(t) = 2P\left\{\int_0^t g_0^2(\boldsymbol{X}, Z(s, W), s)[\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_s, W) + h_3(s)]d\tilde{H}(s)\right\},$$

with $g_0(\mathbf{X}, Z(t, W), t) = \exp\{\boldsymbol{\beta}_0'\mathbf{X} + \boldsymbol{\alpha}_0'\phi(\mathcal{F}_t, W) + \tilde{\mu}_0(t)\}$.

Thus, condition (iv) follows from

$$\begin{aligned}
&\left|\sqrt{n}\Big[S(\hat{\tilde{\boldsymbol{\theta}}}_n) - S(\tilde{\boldsymbol{\theta}}_0) - \dot{S}(\tilde{\boldsymbol{\theta}}_0)(\hat{\tilde{\boldsymbol{\theta}}}_n - \tilde{\boldsymbol{\theta}}_0)\Big][\mathbf{h}_1, \mathbf{h}_2, h_3]\right| = \\
&= \bigg|2\sqrt{n}P\int_0^\tau \Big\{(\exp\{\hat{f}_n\} - \exp\{f_0\})\exp\{\hat{f}_n\} - \exp\{2f_0\}(\hat{f}_n - f_0)\Big\} \times \\
&\quad \times [\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_t, W) + h_3(t)]d\tilde{H}(t)\bigg| = \\
&= \bigg|2\sqrt{n}P\int_0^\tau \Big\{\Big[\exp\{f_0\}(\hat{f}_n - f_0) + \frac{\exp\{f_0\}}{2}(\hat{f}_n - f_0)^2 + o_p\big(\|\hat{f}_n - f_0\|_2^2\big)\Big]\exp\{\hat{f}_n\} - \\
&\quad - \exp\{2f_0\}(\hat{f}_n - f_0)\Big\}[\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_t, W) + h_3(t)]d\tilde{H}(t)\bigg| = \\
&= \bigg|2\sqrt{n}P\int_0^\tau \Big\{\exp\{f_0 + f_n^*\}(\hat{f}_n - f_0)^2 + \\
&\quad + \left[\frac{\exp\{f_0\}}{2}(\hat{f}_n - f_0)^2 + o_p\big(\|\hat{f}_n - f_0\|_2^2\big)\right]\exp\{\hat{f}_n\}\Big\} \times \\
&\quad \times [\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_t, W) + h_3(t)]d\tilde{H}(t)\bigg| \le \\
&\le c_{14}\sqrt{n}\Big[\rho^2\big(\hat{\tilde{\boldsymbol{\theta}}}_n, \tilde{\boldsymbol{\theta}}_0\big) + o_p\Big(\rho^2\big(\hat{\tilde{\boldsymbol{\theta}}}_n, \tilde{\boldsymbol{\theta}}_0\big)\Big)\Big] = \\
&= O_p(n^{\frac{1}{2}-(1-\nu)}) + o_p(n^{\frac{1}{2}-(1-\nu)}) = o_p(1),
\end{aligned}$$

for a constant c_{14}. Therefore, by conditions (i) to (iv), we have

$$\begin{aligned}
&\sqrt{n}\dot{S}(\tilde{\boldsymbol{\theta}}_0)(\hat{\tilde{\boldsymbol{\theta}}}_n - \tilde{\boldsymbol{\theta}}_0)[\mathbf{h}_1, \mathbf{h}_2, h_3] = \\
&= \sigma_1(\mathbf{h}_1, \mathbf{h}_2, h_3)'\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + \sigma_2(\mathbf{h}_1, \mathbf{h}_2, h_3)'\sqrt{n}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) + \\
&\quad + \int_0^\tau \sqrt{n}(\hat{\mu}_n - \mu_0)(t)d\sigma_3(\mathbf{h}_1, \mathbf{h}_2, h_3)(t) = \\
&= -\sqrt{n}(S_n - S)(\tilde{\boldsymbol{\theta}}_0)[\mathbf{h}_1, \mathbf{h}_2, h_3] + o_p(1). \qquad \text{(A2)}
\end{aligned}$$

Next, we will derive the asymptotic normality of estimators from Eq. (A2).
Rewrite

$$\begin{aligned}
&\sigma_3(\mathbf{h}_1, \mathbf{h}_2, h_3)(t) = \\
&= 2E\left\{\sum_{j=1}^{K} \xi(T_{K,j}) I(T_{K,j} \le t) g_0^2(\boldsymbol{X}, Z(T_{K,j}, W), T_{K,j})[\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_{T_{K,j}}, W) + h_3(T_{K,j})]\right\} = \\
&= 2E\left\{\sum_{j=1}^{K} I(T_{K,j} \le t)\Big[E\Big\{\xi(T_{K,j}) g_0^2(\boldsymbol{X}, Z(T_{K,j}, W), T_{K,j})\Big| K, T_{K,j}\Big\} h_3(T_{K,j}) + \right.\\
&\quad \left. + E\Big\{\xi(T_{K,j}) g_0^2(\boldsymbol{X}, Z(T_{K,j}, W), T_{K,j})(\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_{T_{K,j}}, W))\Big| K, T_{K,j}\Big\}\Big]\right\}.
\end{aligned}$$

Thus, we can take

$$\begin{aligned}
h_3(T_{K,j}) &= -a^{-1}(K, T_{K,j}) E\Big\{\xi(T_{K,j}) g_0^2(\boldsymbol{X}, Z(T_{K,j}, W), T_{K,j})(\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2'\phi(\mathcal{F}_{T_{K,j}}, W))\Big| K, T_{K,j}\Big\} = \\
&= -\begin{pmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{pmatrix}' a^{-1}(K, T_{K,j}) E\left\{g_0^2(\boldsymbol{X}, Z(T_{K,j}, W), T_{K,j})\xi(T_{K,j})\begin{pmatrix}\mathbf{X}\\ \phi(\mathcal{F}_{T_{K,j}}, W)\end{pmatrix}\Big| K, T_{K,j}\right\}
\end{aligned}$$

such that $\sigma_3(\mathbf{h}_1, \mathbf{h}_2, h_3)(t) \equiv 0$. Furthermore, for this h_3, we have

$$\big(\sigma_1(\mathbf{h}_1, \mathbf{h}_2, h_3)\sigma_2(\mathbf{h}_1, \mathbf{h}_2, h_3)\big) = 2A\begin{pmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{pmatrix}.$$

Thus, it follows from Eq. (A1) that

$$\begin{pmatrix}\sigma_1(\mathbf{h}_1, \mathbf{h}_2, h_3)\\ \sigma_2(\mathbf{h}_1, \mathbf{h}_2, h_3)\end{pmatrix}' \sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\\ \hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\end{pmatrix} = -\sqrt{n}(\mathbb{P}_n - P)\psi(\tilde{\boldsymbol{\theta}}_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3] + o_p(1).$$

Thus, $2\begin{pmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{pmatrix}' A\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\\ \hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\end{pmatrix} \longrightarrow N(0, \sigma^2)$ with

$$\sigma^2 = \{\psi(\tilde{\boldsymbol{\theta}}_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3]\}^2 = 4\begin{pmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{pmatrix}' B\begin{pmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{pmatrix}.$$

Therefore,

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\\ \hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\end{pmatrix} \longrightarrow N(0, A^{-1}B(A^{-1})').$$

This completes the proof of Theorem 3. □

References

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine and topical thiotepa. In: Pavane-Macaluso, M., Smith, P. H. and Edsmyr, F. (Eds.), *Bladder Tumors and Other Topics in Urological Oncology*. Plenum, New York, NY, pp. 363–370.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analysis of event rates. *Technometrics* **29**, 1–15.

He, X., Tong, X., Sun, J. and Cook, R. (2008). Regression analysis of multivariate panel Count data. *Biostatistics* **9**, 234–248.

Hu, X. J., Lagakos, S. W. and Lockhart, R. A. (2009). Generalized least squares estimation of the mean function of a counting process based on panel counts. *Statistica Sinica* **19**, 561–580.

Hu, X. J., Sun, J. and Wei, L. J. (2003). Regression parameter estimation from panel counts. *Scandinavian Journal of Statistics* **30**, 25–43.

Huang, J. (1999), Efficient estimation of the partly linear additive Cox model. *Journal of the American Statistical Assocication* **27**, 1536–1563.

Jamshidian, M. (2004). On algorithms for restricted maximum likelihood estimation. *Computational Statistics & Data Analysis* **45**, 137–157.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel count data under a Markov assumption. *Journal of the American Statistical Association* **80**, 863–871.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, NY.

Li, N., Sun, L. and Sun, J. (2010). Semiparametric transformation models for panel count data with dependent observation process. *Statistics in Biosciences* **2**, 191–210.

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.

Lu, M., Zhang, Y. and Huang, J. (2009). Semiparametric estimation methods for panel count using monotone polynomial splines. *Journal of the American Statistical Association* **104**, 1060–1070.

Schumaker, L. (1981). *Spline Functions: Basic Theory*. Wiley, New York, NY.

Stakgold, I. (1998). *Green's Functions and Boundary Value Problems*. John Wiley and Sons, New York, NY.

Staniswalis, J. G., Thall, P. F. and Salch, J. (1997). Semiaprametric regression analysis for recurrent event interval counts. *Biometrics* **53**, 1334–1353.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics* **8**, 1048–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics* **10**, 1040–1053.

Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics* **13**, 689–705.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Annals of Statistics* **14**, 590–606.

Sun, J. and Kalbleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica* **5**, 279–290.

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society. Series B* **62**, 293–302.

Sun, J. and Zhao, X. (2013). *Statistical Analysis of Panel Count Data*. Springer, New York, NY.

Thall, P. F. and Lachin J. M., (1988). Analysis of recurrent events: nonparametric methods for random-interval count data. *Journal of the American Statistical Association* **83**, 339–347.

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

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Annals of Statistics* **28**, 779–814.

Wellner, J. A. and Zhang, Y. (2007). Two likelihood based semiparametric estimation methods for panel count data with covariates. *Annals of Statistics* **35**, 2106–2142.

Zhang, Y. (2002). A semiparametric pseudolikehood estimation method for panel count data. *Biometrika* **89**, 39–48.

Zhang, Y. and Jamshidian, M. (2004). On algorithms for nonparametric maximum likelihood estimation of the failure function with censored data. *Journal of Computational and Graphical Statistics* **13**, 123–140.

Zhao, X., and Tong, X. (2011). Semiparametric regression analysis of panel count data with informative observation times. *Computational Statistics & Data Analysis* **55**, 291–300.

Zhao, X., Tong, X. and Sun, J. (2013). Robust estimation for panel count data with informative observation times. *Computational Statistics & Data Analysis* **57**, 33–40.

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Annals of Statistics* **26**, 1760–1782.

