Lifetime Data Anal (2017) 23:439–466
DOI 10.1007/s10985-016-9368-x

Semiparametric partially linear varying coefficient models with panel count data

Xin He[1] · **Xuenan Feng**[2] · **Xingwei Tong**[3] · **Xingqiu Zhao**[2]

Received: 19 May 2015 / Accepted: 15 April 2016 / Published online: 27 April 2016

Abstract This paper studies semiparametric regression analysis of panel count data, which arise naturally when recurrent events are considered. Such data frequently occur in medical follow-up studies and reliability experiments, for example. To explore the nonlinear interactions between covariates, we propose a class of partially linear models with possibly varying coefficients for the mean function of the counting processes with panel count data. The functional coefficients are estimated by B-spline function approximations. The estimation procedures are based on maximum pseudo-likelihood and likelihood approaches and they are easy to implement. The asymptotic properties of the resulting estimators are established, and their finite-sample performance is assessed by Monte Carlo simulation studies. We also demonstrate the value of the proposed method by the analysis of a cancer data set, where the new modeling approach provides more comprehensive information than the usual proportional mean model.

Keywords Asymptotic normality · B-spline · Counting process · Maximum likelihood · Maximum pseudo-likelihood · Panel count data · Varying-coefficient

1 Introduction

This paper considers regression analysis of panel count data when certain covariate effects may be much more complex than linear effects. By panel count data, we mean

✉ Xingqiu Zhao
xingqiu.zhao@polyu.edu.hk

[1] University of Maryland, College Park, MD, USA

[2] The Hong Kong Polytechnic University, Hung Hom, Hong Kong

[3] Beijing Normal University, Beijing, China

the data that concern occurrence rates of certain recurrent events and give only the numbers of the events that occur between the observation times, but not their occurrence times. Such data naturally occur in longitudinal follow-up studies on recurrent events in which study subjects can be observed only at discrete time points rather than continuously. Many authors have discussed the analysis of panel count data by using nonparametric and semiparametric methods. For example, Sun and Kalbfleisch (1995), Wellner and Zhang (2000), Zhang and Jamshidian (2003), Lu et al. (2007), and Hu et al. (2009a) studied nonparametric estimation for the mean function of the counting process with panel count data; Thall and Lachin (1988), Sun and Fang (2003), Zhang (2006), and Balakrishnan and Zhao (2009) proposed some nonparametric tests for the problem of nonparametric comparison of treatment groups based on panel count data. Sun and Wei (2000), Cheng and Wei (2000), Hu et al. (2003), and Hua and Zhang (2012) discussed regression analysis of panel count data by the estimating equation-based approaches, while Zhang (2002), Wellner and Zhang (2007), and Lu et al. (2009) presented more efficient inference procedures for joint estimation of parametric and nonparametric components in the proportional mean model by the likelihood-based approaches. In addition, Huang et al. (2006) and Sun et al. (2007) considered the analysis of panel count data with informative observation times.

All these semiparametric regression methods mentioned above have focused on parametric modeling of covariate effects on the recurrent event process. In many applications, a covariate effect may be nonlinear and vary with another covariate. To investigate both linear effects and nonlinear interaction effects between covariates, we propose a class of semiparametric partially linear varying-coefficient models for panel count data. Suppose that $N(t)$ is a counting process arising from a recurrent event study. Let Z be a d-dimensional vector of covariates, and V and W be p-dimensional vectors of covariates. We assume that given (Z, V, W), $N(t)$ is a non-homogeneous Poisson process with the mean function $\Lambda(t|Z, V, W) = E\{N(t)|Z, V, W\}$ having the following form

$$\Lambda(t|Z, V, W) = \Lambda_0(t) \exp\left\{Z'\beta_0 + \sum_{r=1}^{p} V_r \phi_{r0}(W_r)\right\}, \tag{1.1}$$

where $\Lambda_0(\cdot)$ is a completely unknown continuous baseline mean function, $\phi_{r0}(\cdot)$ $(r = 1, \ldots, p)$ are completely unspecified smooth functions, V_r and W_r are the rth components of V and W, and β_0 is a d-dimensional vector of unknown regression parameters. When $V_r = 0$ $(r = 1, \ldots, p)$, the model reduces to linear regression model with panel count data, which has been well studied by Wellner and Zhang (2007) and Lu et al. (2009), among others. When $V_r = 1$ $(r = 1, \ldots, p)$, the model reduces to partly linear regression model for panel count data, which has not been studied in the literature. There are many investigations about nonlinear effects of covariates on response variables for censored data and longitudinal data. For example, Zhang et al. (2014) studied a proportional hazards model with varying coefficients for right-censored and length-biased data; Lindqvist et al. (2015) examined the functional form for covariates in parametric accelerated failure time models with right-censored data by using residual plots; Cheng et al. (2014) provided a simultaneous variable selection and structure identification procedure for ultra-high dimensional longitudinal data. In this article, for notational simplicity, we consider the case with $p = 1$, that is,

$$\Lambda(t|Z, V, W) = \Lambda_0(t)\exp\{Z'\beta_0 + V\phi_0(W)\}. \tag{1.2}$$

For inference about model (1.2), we propose to use likelihood-based methods, where the functional coefficient is estimated by the B-spline function approximation, and the baseline mean function is still directly estimated with parametric components because its B-spline function approximation has some nonlinear restriction that can cause more complicated computing. For this reason, we develop a new algorithm which can be easily implemented.

The remainder of this paper is organized as follows. In Sect. 2, we present two semiparametric methods including maximum pseudo-likelihood and maximum likelihood approaches for joint estimation of parametric and nonparametric components in the model, and also provide corresponding algorithms about computation of the estimates. The asymptotic properties of the resulting estimators are established in Sect. 3, while the proofs are given in Appendix. Section 4 reports some simulation results obtained for assessing the finite sample properties of the proposed estimates and an illustrative example is given in Sect. 5. Some remarks are made in Sect. 6.

2 Semiparametric likelihood approaches

Consider a recurrent event study that consists of n independent subjects and let $N_i(t)$ denote the number of occurrences of the recurrent event of interest before or at time t for subject i. Suppose that for each subject, given covariates (Z_i, V_i, W_i), $N_i(t)$ is a non-homogeneous Poisson process with the mean function given by (1.2), that is,

$$P\{N_i(t) = k|Z_i, V_i, W_i\} = \exp\{\Lambda_i(t|Z_i, V_i, W_i)\}\frac{\{\Lambda_i(t|Z_i, V_i, W_i)\}^k}{k!},$$

where $\Lambda_i(t|Z_i, V_i, W_i) = \Lambda_0(t)\exp\{Z_i'\beta_0 + V_i\phi_0(W_i)\}$. For subject i, suppose that $N_i(\cdot)$ is observed only at finite time points $T_{K_i,1} < \cdots < T_{K_i,K_i} \leq \tau$, where K_i denotes the potential number of observation times, $i = 1, \ldots, n$, and τ is the length of the study. That is, only the values of $N_i(t)$ at these observation times are known and we have panel count data on the $N_i(t)$'s.

In the following, we will assume that given (Z_i, V_i, W_i), $(K_i; T_{K_i,1}, ..., T_{K_i,K_i})$ are independent of the counting processes N_i's. Let $\mathbf{X} = (K, \mathbf{T}, \mathbf{N}, Z, V, W)$, where $\mathbf{T} = (T_{K,1}, ..., T_{K,K})$ and $\mathbf{N} = \big(N(T_{K,1}), \ldots, N(T_{K,K})\big)$. Then $\{\mathbf{X}_i = (K_i, \mathbf{T}_i, \mathbf{N}_i), Z_i, V_i, W_i\ i = 1, ..., n\}$ is a random sample of size n from the distribution of $\mathbf{X}$, where $\mathbf{T}_i = (T_{K_i,1}, ..., T_{K_i,K_i})$ and $\mathbf{N}_i = (N_i(T_{K_i,1}), \ldots, N_i(T_{K_i,K_i}))$.

Without loss of generality, assume that W has support on [0, 1]. For estimation of the smooth function ϕ_0, we use B-spline function approximation. We first introduce some notation (Huang 1999). Let $\mathcal{T} = \{s_i, i = 1, \ldots, m_n + 2l\}$, with

$$0 = s_1 = \cdots = s_l < s_{l+1} < \cdots < s_{m_n+l} < s_{m_n+l+1} = \cdots = s_{m_n+2l} = 1,$$

be a sequence of knots that partition [0, 1] into $m_n + 1$ subintervals $I_i = [s_{l+i}, s_{l+i+1}]$, for $i = 0, 1, \ldots, m_n$. Define Φ_n the class of polynomial splines of order $l \geq 1$ with the knot sequence $\mathcal{T}$. Then Φ_n can be linearly spanned by the normalized B-spline basis functions $\{b_i, i = 1, \ldots, b_{q_n}\}$ with $q_n = m_n + l$ (Schumaker 1981). Let $B_n = (b_1, \cdots, b_{q_n})'$. Then we can approximate ϕ_0 by $\phi_n = B_n'\alpha$, where α is a q_n-dimensional vector of unknown coefficients.

2.1 Maximum pseudo-likelihood approach

The log pseudo-likelihood function for β_0, Λ_0, and ϕ_0 is $l_n^{ps}(\beta_0, \Lambda_0, \phi_0)$ where

$$l_n^{ps}(\beta, \Lambda, \phi) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Big[N_i(T_{K_i,j}) \log \big\{ \Lambda(T_{K_i,j}) \big\} + N_i(T_{K_i,j}) \{ Z_i'\beta + V_i\phi(W_i) \} \\ -\Lambda(T_{K_i,j}) \exp\{ Z_i'\beta + V_i\phi(W_i) \} \Big]$$

after omitting the parts independent of β, Λ, and ϕ.

Let $t_1 < \cdots < t_m$ denote the ordered distinct observation time points in the set of all observation time points $\{T_{K_i,j}, j = 1, \ldots, K_i, i = 1, \ldots, n\}$. Let w_ℓ and $\bar{N}_\ell$ be the number and mean value, respectively, of the observations made at time t_ℓ, $\ell = 1, ..., m$, that is,

$$w_\ell = \sum_{i=1}^{n} \sum_{j=1}^{K_i} I(T_{K_i,j} = t_\ell) \quad \text{and} \quad \bar{N}_\ell = \frac{1}{w_\ell} \sum_{i=1}^{n} \sum_{j=1}^{K_i} N_i(T_{K_i,j}) I(T_{K_i,j} = t_\ell).$$

Define

$$\bar{A}_\ell(\beta, \phi) = \frac{1}{w_\ell} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \exp\{ Z_i'\beta + V_i\phi(W_i) \} I(T_{K_i,j} = t_\ell)$$

and

$$\bar{B}_\ell(\beta, \phi) = \frac{1}{w_\ell} \sum_{i=1}^{n} \sum_{j=1}^{K_i} N_i(T_{K_i,j}) \{ Z_i'\beta + V_i\phi(W_i) \} I(T_{K_i,j} = t_\ell).$$

Then $l_n^{ps}(\beta, \Lambda, \phi)$ can be expressed as

$$l_n^{ps}(\beta, \Lambda, \phi) = \sum_{\ell=1}^{m} w_\ell \{ \bar{N}_\ell \log \Lambda_\ell - \bar{A}_\ell(\beta, \phi)\Lambda_\ell + \bar{B}_\ell(\beta, \phi) \},$$

where $\Lambda_\ell = \Lambda(t_\ell)$, $\ell = 1, \ldots, m$.

Let $\mathcal{R} \subset \mathbb{R}^d$ be a bounded closed set, and let

$$\mathcal{F} = \{\Lambda : \Lambda \text{ is a nondecreasing function over } [0, \tau], \Lambda(0) = 0\},$$

and

$$\Psi_n = \{\phi : \phi = B_n'\alpha \in \Phi_n, ||\phi||_\infty \le M_0\}$$

where τ is the maximum follow-up time of the study and M_0 is a constant. Let $\Theta_n = \mathcal{R} \times \mathcal{F} \times \Psi_n = \{\theta = (\beta, \Lambda, \phi) : \beta \in \mathcal{R}, \Lambda \in \mathcal{F}, \phi \in \Psi_n\}$. Define the estimator $\hat{\theta}_n^{ps} = (\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})$ of $\theta_0 = (\beta_0, \Lambda_0, \phi_0)$ be the value that maximizes $l_n^{ps}(\theta)$ with respect to $\theta \in \Theta_n$. Following Wellner and Zhang (2000, 2007), we define the estimator $\hat{\Lambda}_n^{ps}$ to have jumps only at the observation time points to meet with uniqueness since $l_n^{ps}(\beta, \Lambda, \phi)$ depends on Λ only at the observation time points.

We denote the estimator of ϕ_0 by $\hat{\phi}_n^{ps} = B_n'\hat{\alpha}_n^{ps}$. Following Zhang (2002), one can find the solution of $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\alpha}_n^{ps})$.

Step 1. Choose an initial $(\beta^{(0)}, \alpha^{(0)})$.

Step 2. For given $(\beta^{(k)}, \alpha^{(k)})$ $(k = 0, 1, 2, ...)$, compute

$$\Lambda_\ell^{(k)} = \max_{i \le \ell} \min_{j \ge \ell} \frac{\sum_{i \le r < j} w_r \bar{N}_r}{\sum_{i \le r < j} w_r \bar{A}_r(\beta^{(k)}, \alpha^{(k)})}, \quad \ell = 1, \ldots, m.$$

Step 3. Update (β, α) by finding

$$(\beta^{(k+1)}, \alpha^{(k+1)}) = \text{argmax}_{(\beta,\alpha)\in\mathbb{R}^{d+q_n}} \hat{l}_n^{ps}(\beta, \alpha, \Lambda^{(k)})$$

through the Newton-Raphson algorithm, where

$$\hat{l}_n^{ps}(\beta, \alpha, \Lambda) = \sum_{\ell=1}^{m} w_\ell \{\bar{B}_\ell(\beta, \alpha) - \bar{A}_\ell(\beta, \alpha)\Lambda_\ell\}.$$

Step 4. Repeat Steps 2 and 3 until the convergence is achieved.

2.2 Maximum likelihood approach

The log-likelihood function for β_0, Λ_0, and ϕ_0 is $l_n(\beta_0, \Lambda_0, \phi_0)$ where

$$\begin{aligned} l_n(\beta, \Lambda, \phi) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Big[& \Big\{N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})\Big\} \log\Big\{\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})\Big\} \\ & + \Big\{N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})\Big\}\{Z_i'\beta + V_i\phi(W_i)\} \\ & - \{\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})\}\exp\{Z_i'\beta + V_i\phi(W_i)\}\Big] \end{aligned}$$

after omitting the parts independent of β, Λ, and ϕ, where $T_{K_i,0} = 0$.

Let $(\hat{\beta}_n, \hat{\Lambda}_n, \hat{\phi}_n)$ be the value that maximizes $l_n(\beta, \Lambda, \phi)$ with respect to $(\beta, \Lambda, \phi) \in \Theta_n$. Similarly, the estimator $\hat{\Lambda}_n$ is defined to have jumps only at the observation time points. This estimator can be computed by the algorithm proposed by Wellner and Zhang (2007), but it is computationally expensive. Here, we propose a new algorithm by using the self-consistent algorithm (Hu et al. 2009b).

Define $\lambda_\ell = \Lambda(t_\ell) - \Lambda(t_{\ell-1})$, $\Delta N_i(t_\ell) = N_i(t_\ell) - N_i(t_{\ell-1})$, and $Y_i(t) = I(t \le T_{K_i,K_i})$. Let

$$R_i(t_\ell) = \min\{T_{K_i,j},\ j = 1, \ldots, K_i;\ T_{K_i,j} \ge t_\ell\}$$

and

$$L_i(t_\ell) = \max\{T_{K_i,j},\ j = 1, \ldots, K_i;\ T_{K_i,j} < t_\ell\}$$

denote the most recent observation times of individual i not before and before t_ℓ, respectively. Here $R_i(t_\ell) = t_{m+1} = \infty$ if $t_\ell > T_{K_i,K_i}$. Define $\tilde{\Delta} N_i(t_\ell) = N_i(R_i(t_\ell)) - N_i(L_i(t_\ell))$ and $\tilde{\Delta}\Lambda_i(t_\ell) = \Lambda(R_i(t_\ell)) - \Lambda(L_i(t_\ell))$, that is, $\tilde{\Delta}\Lambda_i(t_\ell) = \sum_{r:\, L_i(t_\ell) < t_r \le R_i(t_\ell)} \lambda_r$. For given β_0 and ϕ_0, we have the following estimating equation for Λ_0:

$$\sum_{i=1}^{n} Y_i(t_\ell)\left[\lambda_\ell \frac{\tilde{\Delta} N_i(t_\ell)}{\tilde{\Delta}\Lambda_i(t_\ell)} - \lambda_\ell \exp\{Z_i'\beta_0 + V_i\phi_0(W_i)\}\right] = 0, \quad \ell = 1, \ldots, m.$$

As Hu et al. (2009b) pointed out, the estimating functions are unbiased and also can be viewed as the expectation of the likelihood estimating functions conditional on panel counts.

We denote the estimators of ϕ_0 by $\hat{\phi}_n = B_n'\hat{\alpha}_n$. To find out the solution of $(\hat{\beta}_n, \hat{\alpha}_n, \hat{\Lambda}_n)$, we propose to implement the following algorithm.

Step 1. Choose the initial $(\beta^{(0)}, \alpha^{(0)}) = (\hat{\beta}_n^{ps}, \hat{\alpha}_n^{ps})$.

Step 2. For given $(\beta^{(k)}, \alpha^{(k)})$, obtain $\lambda_\ell^{(k)}$ $(\ell = 1, \ldots, m)$ by computing

$$\lambda_\ell^{(k,u)} = \frac{\sum_{i=1}^n Y_i(t_\ell)\lambda_\ell^{(k,u-1)} \tilde{\Delta} N_i(t_\ell)/\tilde{\Delta}\Lambda_i^{(k,u-1)}(t_\ell)}{\sum_{i=1}^n Y_i(t_\ell)\exp\{Z_i'\beta^{(k)} + V_i B_n(W_i)'\alpha^{(k)}\}}$$

for $u = 1, 2, \ldots$ until the convergence is achieved. Here we choose $\Lambda^{(0,0)} = \hat{\Lambda}_n^{ps}$ and $\Lambda^{(k,0)} = \Lambda^{(k-1)}$ for $k \ge 1$.

Step 3. Update (β, α) by finding

$$(\beta^{(k+1)}, \alpha^{(k+1)}) = \mathrm{argmax}_{(\beta,\alpha)\in\mathbb{R}^{d+q_n}} \hat{l}_n(\beta, \alpha, \Lambda^{(k)})$$

through the Newton-Raphson algorithm, where

$$\hat{l}_n(\beta, \alpha, \Lambda) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Big[\big\{N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})\big\}\{\beta' Z_i + V_i B_n(W_i)'\alpha\}$$
$$-\{\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})\}\exp\{\beta' Z_i + V_i B_n(W_i)'\alpha\}\Big]$$

Step 4. Repeat Steps 2 and 3 until the convergence is achieved.

3 Asymptotic results

In this section, we study the asymptotic properties of the estimators $\hat{\theta}_n^{ps} = (\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})$ and $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n, \hat{\phi}_n)$ of $\theta_0 = (\beta_0, \Lambda_0, \phi_0)$. Let $\mathcal{B}_{d+2}$ and $\mathcal{B}$ denote the collection of Borel sets in $\mathbb{R}^{d+2}$ and $\mathbb{R}$, respectively, and let $\mathcal{B}_{[0,\tau]} = \{A \cap [0, \tau] : A \in \mathcal{B}\}$. Let $Y = (Z', V, W)'$ with distribution function $F(y)$. Following Wellner and Zhang (2007), define the measures $\mu_1, \mu_2, \nu_1, \nu_2$, and γ as follows: for $A, A_1, A_2 \in \mathcal{B}_{[0,\tau]}$, and $A_3 \in \mathcal{B}_{d+2}$,

$$\nu_1(A \times A_3) = \int_{A_3} \sum_{k=1}^{\infty} P(K = k|Y = y) \sum_{j=1}^{k} P(T_{k,j} \in A|K = k, Y = y)dF(y),$$
$$\mu_1(A) = \nu_1(A \times \mathbb{R}^{d+2}),$$
$$\nu_2(A_1 \times A_2 \times A_3)$$
$$= \int_{A_3} \sum_{k=1}^{\infty}\{P(K = k|Y = y) \sum_{j=1}^{k} P(T_{k,j-1} \in A_1, T_{k,j} \in A_2|K = k, Y = y)\}dF(y),$$
$$\mu_2(A_1 \times A_2) = \nu_2(A_1 \times A_2 \times \mathbb{R}^{d+2}),$$
$$\gamma(A) = \int_{\mathbb{R}^{d+2}} \sum_{k=1}^{\infty} P(K = k|Y = y) \sum_{j=1}^{k} P(T_{k,k} \in A|K = k, Y = y)dF(y).$$

We also define the L_2-metrics d_1 and d_2 as

$$d_1(\theta_1, \theta_2) = \left\{||\beta_1 - \beta_2||^2 + \int |\Lambda_1(t) - \Lambda_2(t)|^2 d\mu_1(t) + E|\phi_1(W) - \phi_2(W)|^2\right\}^{1/2},$$

and

$$d_2(\theta_1, \theta_2) = \Bigg\{||\beta_1 - \beta_2||^2 + \int\int |(\Lambda_1(u) - \Lambda_1(v)) - (\Lambda_2(u) - \Lambda_2(v))|^2 d\mu_2(u, v)$$
$$+E|\phi_1(W) - \phi_2(W)|^2\Bigg\}^{1/2}.$$

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

C1. The maximum spacing of the knots, $\max_{l+1 \le i \le m_n+l+1} |s_i - s_{i-1}| = O(n^{-v})$ with $m_n = O(n^v)$ for $0 < v < 0.5$.
C2. The true parameter $\theta_0 = (\beta_0, \Lambda_0, \phi_0) \in \mathcal{R}^0 \times \mathcal{F} \times \mathcal{F}_r$ with $\Lambda_0(\tau) \le M$ for a constant $M > 0$, and $r = l + a > 0.5$, where

$$\mathcal{F}_r = \{g(\cdot) : |g^{(l)}(w_1) - g^{(l)}(w_2)| \le M_0|w_1 - w_2|^a \text{ for all } 0 \le w_1, w_2 \le 1\}$$

and $g^{(l)}$ is the lth derivative function of g.
C3. The measure $\mu_i \times F$ is absolutely continuous with respect to ν_i, for $i = 1, 2$.
C4. The function M_0^{ps} defined by $M_0^{ps}(X) = \sum_{j=1}^{K} N(T_{K,j}) \log(N(T_{K,j}))$ satisfies $PM_0^{ps}(X) < \infty$.
C5. The function M_0 defined by $M_0(X) = \sum_{j=1}^{K} \Delta N(T_j) \log(\Delta N(T_{K,j}))$ satisfies $PM_0(X) < \infty$.
C6. $\mathcal{C} = \text{supp}(F)$, is a bound set in $\mathbb{R}^{d+2}$. Thus there exist z_0 and v_0 such that $P(|Z| \le z_0) = 1$ and $P(|V| \le v_0) = 1$. That is, the covariates Z and V are uniformly bounded.
C7. If with probability 1, $Z'b + V\psi(W) + \zeta(T_{K,K}) = 0$ for some b, ψ and ζ, then $b = 0$, $\psi = 0$ and $\zeta = 0$.
C8. There exists a positive integer K_0 such that $P(K \le K_0)$=1.

Conditions C1 and C2 are common assumptions in semiparametric estimation problems. Conditions C4-C6 and C8 similar to those required by Wellner and Zhang (2007). Conditions C3 and C7 are needed for identifiability of the model.

Theorem 3.1 (Consistency). *Suppose that conditions C1-C8 hold.*

(i) *If* $\mu_1([b, \tau]) > 0$ *for* $0 < b < \tau$, *then*

$$\lim_{n\to\infty} d_1((\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps} 1_{[0,b]}, \hat{\phi}_n^{ps}), (\beta_0, \Lambda_0 1_{[0,b]}, \phi_0)) = 0 \quad \textit{in Probability}.$$

If $\mu_1(\{\tau\}) > 0$,

$$\lim_{n\to\infty} d_1(\hat{\theta}_n^{ps}, \theta_0) = 0 \quad \textit{in Probability}.$$

(ii) *If* $\gamma([b, \tau]) > 0$ *for* $0 < b < \tau$, *then*

$$\lim_{n\to\infty} d_2((\hat{\beta}_n, \hat{\Lambda}_n 1_{[0,b]}, \hat{\phi}_n), (\beta_0, \Lambda_0 1_{[0,b]}, \phi_0)) = 0 \quad \textit{in Probability}.$$

If $\gamma(\{\tau\}) > 0$, *then*

$$\lim_{n\to\infty} d_2(\hat{\theta}_n, \theta_0) = 0 \quad \textit{in Probability}.$$

To establish the rate of convergence and the asymptotic normality, we need additional conditions.

C9. *For some positive constant* c_0*,* $E[\exp\{c_0 N(\tau)\}] < \infty$.
C10. $P(\cap_{j=1}^{K}\{T_{K,j} \in [\tau_0, \tau]\}) = 1$ *with* $\tau_0 > 0$ *and* $\Lambda_0(\tau_0) > 0$.
C11. *There exists a positive constant* s_0 *such that*

$$P\left(\min_{1\le j\le K}\{T_{K,j} - T_{K,j-1}\} \ge s_0\right) = 1.$$

C12. Λ_0 *is differentiable and the derivative has a positive and finite lower and upper bounds in* $[\tau_0, \tau]$.
C13. *There exists* $\eta_1, \eta_2 \in (0, 1)$ *such that* $a'Var(Z|U, V, W)a \ge \eta_1 a'E(Z'Z|U, V, W)a$ *a.s. for all* $a \in \mathbb{R}^d$*, and* $Var(V|U, W) \ge \eta_2 E(V^2|U, W)$*, where* (U, Y) *has distribution* $\nu_1/\nu_1(\mathbb{R}^+ \times \mathcal{C})$.
C14. *There exists* $\eta_1, \eta_2 \in (0, 1)$ *such that*

$$a'Var(Z|U_1, U_2, Y)a \ge \eta_1 E(Z'Z|U_1, U_2, Y)a, \ a.s.$$

for all $a \in \mathbb{R}^d$*, and* $Var(V|U_1, U_2, W) \ge \eta_2 E(V^2|U_1, U_2, W)$*, where* (U_1, U_2, Y) *has distribution* $\nu_2/\nu_2(\mathbb{R}^{+2} \times \mathcal{C})$.

Conditions C9-C14 and their justifications are similar to those given in Wellner and Zhang (2007).

Theorem 3.2 (Rate of Convergence). *Suppose that conditions C1-C10 hold.*

(i) *If condition C13 holds, then* $n^{\frac{1-v}{3}} d_1(\hat{\theta}_n^{ps}, \theta_0) = O_p(1)$.
(ii) *If conditions C11, C12 and C14 hold, then* $n^{\frac{1-v}{3}} d_2(\hat{\theta}_n, \theta_0) = O_p(1)$.

Theorem 3.3 (Asymptotic Normality). *Suppose that* $\frac{1}{6r-2} < v < \frac{1}{4}$ *with* $r > 1$ *and the conditions C1-C12 hold. Define* $H_1 = \{\mathbf{h}_1 : \boldsymbol{h}_1 \in \mathbb{R}^d, ||\boldsymbol{h}_1|| \le 1\}$ *and* $H_2 = \{h_2 : h_2$ *is a function with bounded total variation in* $[0, \tau], \quad h_2(0) = 0\}$.

(i) *If condition C13 holds, then for* $(\boldsymbol{h}_1, h_2, h_3) \in H_1 \times H_2 \times \mathcal{F}_r$,

$$\begin{aligned}
&\boldsymbol{h}_1'\sqrt{n}(\hat{\beta}_n^{ps} - \beta_0) + \sqrt{n}\int\{\hat{\Lambda}_n^{ps}(t) - \Lambda_0(t)\}dh_2(t) \\
&+\sqrt{n}\int\{\hat{\phi}_n^{ps}(w) - \phi_0(w)\}dh_3(w) \\
&\to_d \ N(0, \sigma_{ps}^2),
\end{aligned}$$

where σ_{ps}^2 *is given in (7.5).*
(ii) *If condition C14 holds, then for* $(\boldsymbol{h}_1, h_2, h_3) \in H_1 \times H_2 \times \mathcal{F}_r$,

$$\begin{aligned}
&\boldsymbol{h}_1'\sqrt{n}(\hat{\beta}_n - \beta_0) + \sqrt{n}\int\{\hat{\Lambda}_n(t) - \Lambda_0(t)\}dh_2(t) \\
&+\sqrt{n}\int\{\hat{\phi}_n(w) - \phi_0(w)\}dh_3(w) \\
&\to_d \ N(0, \sigma^2),
\end{aligned}$$

where σ^2 is given in (7.6).

Remark The proofs of these theorems are given in Appendix. In particular, Theorem 3.3.1 of van der Vaart and Wellner (1996, page 310) cannot be directly applied to prove Theorem 3.3 because the rate of convergence for the proposed estimators is no longer $n^{-1/2}$. We will show the theorem by modifying the conditions required by Theorem 3.3.1 of van der Vaart and Wellner (1996, page 310).

4 Simulation study

To assess the performance of the proposed estimation procedure, we conducted simulation studies under various situations with the focus on the estimation of β_0. In the study, we considered a bivariate covariate $Z = (Z_1, Z_2)'$, where $Z_1 \sim N(1, 1)$ and $Z_2 \sim \text{Uniform}(-1, 1)$. The covariates V and W followed a Bernoulli distribution with success probability 0.5 and a standard uniform distribution over [0, 1]. The follow-up time C_i were generated by $\min(\tilde{C}_i, \tau)$, where $\tilde{C}_i \sim \text{Uniform}(2, 9)$ and $\tau = 8$.

For the observation process, we considered two scenarios. One is to assume that the observation times are independent of covariates and the other is to suppose that the observation process depends on the covariate Z. For the i-th subject, the number of real observation times K_i^* was generated from a discrete uniform distribution between 1 and 5 for the former setup, and it followed a Poisson distribution with mean $\{C_i \exp(Z_{1i} + Z_{2i})/\tau\}$ for the latter one. Furthermore, the observation times $(T_{K_i,1}, \ldots, T_{K_i,K_i^*})$ were the order statistics of a random sample of size K_i^* from the uniform distribution over $(0, C_i)$.

Given K_i^* and $(T_{K_i,1}, \ldots, T_{K_i,K_i^*})$, we generated the panel counts $N_i(T_{K_i,j})$ from

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

for $j = 1, \ldots, K_i^*$ and $i = 1, \ldots, n$. In the above, $N_i(t)$ follows a Poisson distribution with mean $t^2 \exp\{Z_{1i}\beta_{10} + Z_{2i}\beta_{20} + V_i\phi_0(W_i)\}/2$, where $\phi_0(w) = 2\sin(2w + 0.1) + \exp(-0.5w)$. The results given below are based on $n = 100$ or 200, and 500 replications with a bootstrap sample size 100.

Table 1 presents the simulation results by using the proposed maximum pseudo-likelihood and maximum likelihood approaches for the situation where the observation process is independent of covariates and $(\beta_{10}, \beta_{20}) = (1, 1), (1, 0), (1, -1), (0, 1)$, or $(0, 0)$. The table includes the estimated bias (BIAS) given by the averages of the point estimates minus the true value of (β_{10}, β_{20}), the sample standard errors of the estimates (SSE), the means of the bootstrap standard error estimates (BSE), and the empirical 95% coverage probabilities (CP) for (β_{10}, β_{20}). It can be seen that the estimates $(\hat{\beta}_{10}^{ps}, \hat{\beta}_{20}^{ps})$ and $(\hat{\beta}_{10}, \hat{\beta}_{20})$ seem to be unbiased and the two standard error estimates are quite close to each other, indicating that the bootstrap variance estimation procedure provides reasonable estimates. In particular, the maximum likelihood method yields smaller standard error estimates than the maximum pseudo-likelihood approach. Moreover, the empirical coverage probabilities suggest that the normal approximation seems to be appropriate.

Table 1 Simulation results for covariate-independent observation processes

Method	n	(β_{10}, β_{20})	BIAS	SSE	BSE	CP
Maximum pseudo-likelihood	100	(1,1)	(0.0001,0.0005)	(0.0094,0.0202)	(0.0110,0.0233)	(0.966,0.980)
		(1,0)	(0.0003,0.0000)	(0.0099,0.0209)	(0.0114,0.0230)	(0.976,0.968)
		(1,−1)	(0.0007, −0.0007)	(0.0091,0.0201)	(0.0116,0.0232)	(0.974,0.974)
		(0,1)	(0.0013,0.0000)	(0.0185,0.0354)	(0.0205,0.0375)	(0.954,0.948)
		(0,0)	(0.0004,0.0013)	(0.0195,0.0334)	(0.0210,0.0376)	(0.954,0.962)
	200	(1,1)	(0.0003, −0.0004)	(0.0055,0.0118)	(0.0061,0.0129)	(0.970,0.964)
		(1,0)	(−0.0002, 0.0002)	(0.0057,0.0123)	(0.0063,0.0132)	(0.960,0.960)
		(1,−1)	(0.0002,0.0000)	(0.0057,0.0121)	(0.0062,0.0132)	(0.958,0.962)
		(0,1)	(−0.0002, 0.0009)	(0.0112,0.0239)	(0.0122,0.0230)	(0.960,0.930)
		(0,0)	(0.0006,0.0002)	(0.0122,0.0219)	(0.0125,0.0225)	(0.936,0.954)
Maximum likelihood	100	(1,1)	(0.0000, −0.0003)	(0.0069,0.0146)	(0.0079,0.0166)	(0.976,0.970)
		(1,0)	(0.0002, −0.0001)	(0.0071,0.0142)	(0.0082,0.0165)	(0.962,0.970)
		(1,−1)	(0.0000,0.0002)	(0.0070,0.0142)	(0.0078,0.0164)	(0.962,0.972)
		(0,1)	(−0.0002, 0.0021)	(0.0148,0.0285)	(0.0163,0.0305)	(0.964,0.954)
		(0,0)	(−0.0013, 0.0002)	(0.0162,0.0283)	(0.0170,0.0302)	(0.960,0.954)
	200	(1,1)	(0.0000,0.0000)	(0.0044,0.0093)	(0.0047,0.0099)	(0.964,0.952)
		(1,0)	(0.0000, −0.0002)	(0.0047,0.0093)	(0.0050,0.0101)	(0.956,0.962)
		(1,−1)	(−0.0003, −0.0002)	(0.0043,0.0093)	(0.0046,0.0099)	(0.956,0.964)
		(0,1)	(−0.0004, 0.0012)	(0.0097,0.0179)	(0.0100,0.0190)	(0.950,0.972)
		(0,0)	(−0.0007, 0.0002)	(0.0101,0.0172)	(0.0106,0.0191)	(0.954,0.962)

The results for the situation where the observation times depend on the covariates Z are given in Table 2 in which other setups are the same as those in Table 1. As shown in Table 2, the conclusions are similar to those from Table 1 and indicate that the proposed estimation procedure seems to perform well for the scenarios considered here.

Table 3 presents the simulation results of nonparametric estimates for $(\beta_{10}, \beta_{20}) = (1, 1)$ indicating that the estimated $\phi_0(W)$ seems to be unbiased. The conclusions are similar when $(\beta_{10}, \beta_{20}) = (1, 0)$, $(1, -1)$, $(0, 1)$, or $(0, 0)$. Our proposed estimation procedure for ϕ_0 by the B-spline function approximation performs well for all the scenarios in the simulation study.

In addition, we have investigated the computation time of our simulation programs in MATLAB using a PC with Intel Xeon CPU E5520 2.27 GHz. For 500 replications with n=200, it would take about 100 hours for the pseudo-likelihood approach and 15 hours for the likelihood approach.

5 An application

To illustrate the proposed methodology given in the previous sections, we apply it to the bladder cancer study conducted by the Veterans Administration Cooperative

Table 2 Simulation results for covariate-dependent observation processes

Method	n	(β_{10}, β_{20})	BIAS	SSE	BSE	CP
Maximum pseudo-likelihood	100	(1,1)	(0.0003, −0.0010)	(0.0087,0.0177)	(0.0087,0.0201)	(0.940,0.966)
		(1,0)	(0.0004, −0.0002)	(0.0093,0.0196)	(0.0096,0.0217)	(0.946,0.952)
		(1,−1)	(0.0001,0.0009)	(0.0093,0.0208)	(0.0099,0.0225)	(0.956,0.952)
		(0,1)	(0.0003, −0.0041)	(0.0224,0.0406)	(0.0227,0.0417)	(0.946,0.972)
		(0,0)	(−0.0012, −0.0011)	(0.0252,0.0386)	(0.0229,0.0421)	(0.936,0.958)
	200	(1,1)	(0.0001,0.0004)	(0.0054,0.0114)	(0.0052,0.0121)	(0.930,0.958)
		(1,0)	(0.0002,0.0005)	(0.0062,0.0119)	(0.0058,0.0128)	(0.910,0.958)
		(1,−1)	(−0.0001, −0.0001)	(0.0062,0.0140)	(0.0059,0.0134)	(0.922,0.932)
		(0,1)	(−0.0002, 0.0006)	(0.0146,0.0267)	(0.0137,0.0261)	(0.938,0.936)
		(0,0)	(−0.0002, −0.0020)	(0.0163,0.0269)	(0.0145,0.0259)	(0.934,0.936)
Maximum likelihood	100	(1,1)	(−0.0004, 0.0001)	(0.0059,0.0138)	(0.0071,0.0155)	(0.972,0.988)
		(1,0)	(−0.0002, 0.0008)	(0.0066,0.0134)	(0.0075,0.0156)	(0.972,0.964)
		(1,−1)	(−0.0001, 0.0006)	(0.0063,0.0141)	(0.0074,0.0160)	(0.968,0.960)
		(0,1)	(0.0000,0.0010)	(0.0146,0.0278)	(0.0160,0.0311)	(0.964,0.962)
		(0,0)	(−0.0009, −0.0018)	(0.0156,0.0290)	(0.0170,0.0310)	(0.956,0.968)
	200	(1,1)	(0.0001,0.0001)	(0.0036,0.0081)	(0.0041,0.0093)	(0.968,0.966)
		(1,0)	(0.0000,0.0005)	(0.0042,0.0084)	(0.0045,0.0093)	(0.952,0.962)
		(1,−1)	(0.0003,0.0002)	(0.0041,0.0084)	(0.0043,0.0093)	(0.946,0.970)
		(0,1)	(0.0000, −0.0002)	(0.0088,0.0174)	(0.0097,0.0192)	(0.962,0.956)
		(0,0)	(0.0000, −0.0005)	(0.0103,0.0181)	(0.0105,0.0190)	(0.942,0.954)

Urological Research Group (Byar 1980; Andrews and Herzberg 1985; Sun and Wei 2000). In the original study, patients with superficial bladder tumors were randomly divided into three treatment groups (placebo, thiotepa and pyridoxine) and followed for 53 months. At the beginning of the study, two important baseline characteristics, the number of initial bladder tumors and the size of the largest initial tumor, were observed for each patient. After removing all the initial tumors, many patients had multiple recurrences of tumors during the study. At each clinical follow-up visit, the visit time and the number of recurrent tumors between visits were recorded, and then the recurrent tumors were removed. Following Sun and Wei (2000), we will focus on patients in the thiotepa (38) and placebo (47) groups.

For the analysis, we defined Z to be 1 if the patient was given the thiotepa treatment and 0 otherwise. Let V denote the number of initial bladder tumors, and W be the natural logarithm of the size of the largest initial tumor plus 1. Assume that the occurrence process of the bladder tumors can be described by model (1.2). Our model specification regarding $(Z; V; W)$ was based on the previous literature. Both the number of initial tumors and the size of the largest initial tumor have been widely used as important diagnostic factors in cancer studies. Among others, Sun and Wei (2000) and Zhang (2002) concluded that the number of initial bladder tumors is significantly positively related with the tumor recurrence rate but the size of the largest

Table 3 Simulation results of the estimated functional effect $\phi_0(W)$ for $(\beta_{10}, \beta_{20}) = (1, 1)$

W	$\phi_0(W)$	Covariate-independent observation processes				Covariate-dependent observation processes			
		$n = 100$		$n = 200$		$n = 100$		$n = 200$	
		$\hat{\phi}_0^{ps}(W)$	$\hat{\phi}_0(W)$	$\hat{\phi}_0^{ps}(W)$	$\hat{\phi}_0(W)$	$\hat{\phi}_0^{ps}(W)$	$\hat{\phi}_0(W)$	$\hat{\phi}_0^{ps}(W)$	$\hat{\phi}_0(W)$
1/21	1.364	1.366	1.361	1.363	1.364	1.364	1.366	1.361	1.366
2/21	1.526	1.528	1.524	1.525	1.526	1.525	1.526	1.525	1.527
3/21	1.684	1.685	1.681	1.682	1.683	1.682	1.683	1.684	1.684
4/21	1.834	1.835	1.832	1.833	1.835	1.833	1.834	1.835	1.835
5/21	1.977	1.979	1.976	1.976	1.978	1.976	1.977	1.978	1.977
6/21	2.111	2.112	2.110	2.111	2.111	2.110	2.111	2.111	2.111
7/21	2.234	2.235	2.233	2.234	2.234	2.234	2.235	2.234	2.234
8/21	2.345	2.345	2.344	2.345	2.345	2.344	2.345	2.344	2.345
9/21	2.442	2.442	2.441	2.443	2.442	2.441	2.443	2.442	2.442
10/21	2.525	2.525	2.524	2.526	2.525	2.524	2.525	2.525	2.525
11/21	2.593	2.593	2.592	2.593	2.593	2.592	2.593	2.593	2.593
12/21	2.645	2.645	2.645	2.645	2.645	2.643	2.645	2.645	2.645
13/21	2.680	2.680	2.680	2.679	2.680	2.679	2.680	2.680	2.680
14/21	2.698	2.698	2.698	2.697	2.698	2.697	2.698	2.698	2.697
15/21	2.698	2.698	2.698	2.698	2.697	2.697	2.699	2.699	2.698
16/21	2.680	2.680	2.680	2.680	2.680	2.679	2.681	2.681	2.680
17/21	2.645	2.645	2.644	2.645	2.645	2.643	2.646	2.645	2.645
18/21	2.592	2.591	2.591	2.592	2.593	2.591	2.593	2.591	2.592
19/21	2.522	2.520	2.520	2.522	2.524	2.521	2.524	2.521	2.521
20/21	2.436	2.432	2.431	2.436	2.437	2.435	2.437	2.435	2.433

Table 4 Results of the estimated functional effect of the number of the initial tumors on the tumor recurrence rate

Size	Maximum pseudo-likelihood		Maximum likelihood	
	$\hat{\phi}_0^{ps}(W)$	95% pointwise bootstrap CI	$\hat{\phi}_0(W)$	95% pointwise bootstrap CI
1	0.2200*	(0.0526, 0.4227)	0.2325*	(0.0356, 0.4526)
2	0.0296	(−11.4648, 0.4819)	0.0673	(−11.9842, 0.4854)
3	0.2923*	(0.0045, 0.8261)	0.2708*	(0.0540, 0.8248)
4	−0.5134	(−6.5228, 0.9392)	−0.1731	(−8.0308, 1.2608)
5	0.1993	(−5.4838, 1.4465)	0.3540	(−6.5023, 1.4312)
6	−0.2365	(−3.6525, 1.0312)	−0.3196	(−4.0047, 0.6203)
7	−2.6712	(−10.3878, 10.4999)	−2.8510	(−12.2190, 11.4039)

* P-value ≤ 0.05

initial tumor does not have a significant effect. Therefore, we examined the size of the largest initial tumor (W) as a potential moderator (effect modifier) of the association between the tumor recurrence and the number of initial bladder tumors (V). With a bootstrap sample size 1000, the application of the maximum pseudo-likelihood procedure yielded $\hat{\beta}_0^{ps} = -1.2957$ with an estimated standard error of 0.3713, while we obtained $\hat{\beta}_0 = -0.8271$ with an estimated standard error of 0.3828 by applying the maximum likelihood approach. Both results suggest that the thiotepa treatment significantly reduced the recurrence rate of the bladder tumors.

Table 4 presents the estimated $\phi_0(W)$ and its 95% pointwise bootstrap confidence interval for the functional effect of the number of initial tumors on the tumor recurrence rate based on both maximum pseudo-likelihood and maximum likelihood approaches. The results indicate that the number of initial tumors seems to be positively associated with the tumor recurrence rate only when the size of the largest initial tumor is 1 or 3, while the association is insignificant elsewhere. The difference is related to the unbalanced sample sizes for the stratified subgroups defined by W. In particular, the observed sample sizes are $n = 48$ (for $W = 1$), $n = 10$ (for $W = 2$), $n = 16$ (for $W = 3$), $n = 5$ (for $W = 4$), $n = 2$ (for $W = 5$), $n = 3$ (for $W = 6$), and $n = 1$ (for $W = 7$). Therefore, the statistical power to reject $H_0 : \phi_0(W) = 0$ would be relatively low due to the small sample size at most values of W except for $W = 1$ or 3. A 95% bootstrap confidence band could be an alternative approach, but it would be wide and not as informative as the 95% pointwise bootstrap confidence interval due to the small sample size in this application example.

The conclusion is comparable with those given by Sun and Wei (2000), Zhang (2002), Wellner and Zhang (2007) and Lu et al. (2009) among others, but the proposed model reveals more insight on how the effect of the number of initial tumors is moderated by the size of the largest initial tumor. In practice, one may specify $(Z; V; W)$ under a conceptual model according to research questions in which W is a possible moderator (effect modifier) of the association between the recurrent event process and covariate V.

6 Remarks

In this article, we have considered regression analysis of panel count data when certain covariates have nonlinear effects on recurrent events. For estimation of the constant and functional coefficients and the baseline mean function, we have developed spline-based pseudo-likelihood/likelihood approaches that yield the consistency and asymptotical normality of the estimates, and proposed a new algorithm for computing the spline-based maximum likelihood estimators. The proposed inference procedures are robust because the obtained asymptotic results do not rely on the Poisson assumption on the panel counts at all.

It is important to mention that Theorem 3.3 shows not only the asymptotic normality of the parametric estimators but also the asymptotic normality of the functionals of the nonparametric estimators, which can be useful for hypothesis testing problems, while Weller and Zhang (2007) and Lu et al. (2009) focused on the asymptotic distributions of the parametric estimators. Similar to Theorem 3.3, we can establish the asymptotic normality for the functionals of the estimators of the baseline mean function in the proportional mean model proposed by Weller and Zhang (2007) and Lu et al. (2009). In addition, we can also derive the asymptotic normality of the functionals of the spline likelihood-based estimators proposed by Lu et al. (2007), and thus construct a new class of nonparametric tests, which could be more powerful than the existing nonparametric tests for nonparametric comparison of several treat groups with panel count data.

Note that in the foregoing we have assumed that the recurrent event process is independent of the observation times given covariates. To relax this assumption, we can consider the observation history as a covariate in the model and thus, the proposed method can be generalized to the dependent case. Also note that the censoring time has not been considered. Clearly, the proposed estimation procedures work under the independent censoring. However, if the censoring time is informative, a joint modeling approach needs to be developed for further research. In addition, developing an appropriate model-checking procedure for our proposed method is an important direction for future research. Another research direction would involve high-dimensional partially linear proportional mean model with panel count data.

Acknowledgements The authors would like to thank the Editor, Professor Mei-Ling Ting Lee, the Associate Editor and the two reviewers for their constructive and insightful comments and suggestions that greatly improved the paper. Tong's research was partly supported by BCMIIS, NSF China Zhongdian Project 11131002 and NSFC (No. 10971015). Zhao's research was partly supported by the Research Grants Council of Hong Kong (PolyU 504011 and PolyU 503513), the Natural Science Foundation of China (No. 11371299), and The Hong Kong Polytechnic University.

Appendix: Proofs of asymptotic results

In this section we present the proofs of Theorems 3.1, 3.2 and 3.3.

Proof of Theorem 3.1

Here, we only present the proof of part (i) since part (ii) can be verified similarly. Let

$$
\begin{aligned}
m_\theta^{ps}(X) &= \sum_{j=1}^{K} \big[N(T_{K,j}) \log\{\Lambda(T_{K,j}) \exp(\beta' Z + V\phi(W))\} \\
&\quad - \Lambda(T_{K,j}) \exp(\beta' Z + V\phi(W)) \big], \\
M_n^{ps}(\theta) &= P_n m_\theta^{ps}(X), \quad \text{and} \quad M^{ps}(\theta) = P m_\theta^{ps}(X),
\end{aligned}
$$

where P and P_n denote the probability measure and the empirical measure, respectively. Let $h(x) = x \log x - x + 1$. Note that $h(x) \geq (x-1)^2/4$ for $0 \leq x \leq 5$. For any θ in a sufficiently small neighborhood of θ_0,

$$
\begin{aligned}
&M^p(\theta_0) - M^p(\theta) \\
&= \int \Lambda(u) \exp\{Z'\beta + v\phi(w)\} h\left(\frac{\Lambda_0(u)\exp(z'\beta_0 + v\phi_0(w))}{\Lambda(u)\exp(z'\beta + v\phi(w))} \right) d\nu_1(u, z, v, w) \\
&\geq \frac{1}{4} \int \Lambda(u) \exp\{Z'\beta + v\phi(w)\} \left\{ \frac{\Lambda_0(u)\exp(z'\beta_0 + v\phi_0(w))}{\Lambda(u)\exp(z'\beta + v\phi(w))} - 1 \right\}^2 d\nu_1(u, z, v, w).
\end{aligned} \tag{7.1}
$$

Then, using (7.1) and the arguments similar to those in Wellner and Zhang (2007), we can show that $M^{ps}(\theta_0) = M^{ps}(\theta)$ if and only if $\beta = \beta_0$, $\Lambda(t) = \Lambda_0(t)$ a.e. with respect to μ_1, and $\phi = \phi_0$ by C3 and C7.

By the similar arguments as those used in Wellner and Zhang (2007) again, we can also show that $\hat{\Lambda}_n^{ps}(t)$ is uniformly bounded in probability for $t \in [0, b]$ if $\mu_1([b, \tau]) > 0$ for some $0 < b < \tau$ or $t \in [0, \tau]$ if $\mu_1(\{\tau\}) > 0$.

By Helly-Selection Theorem and compactness of Θ_n, it follows that $\hat{\theta}_n^{ps} = (\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})$ has a subsequence $\hat{\theta}_{n_k}^{ps} = (\hat{\beta}_{n_k}^{ps}, \hat{\Lambda}_{n_k}^{ps}, \hat{\phi}_{n_k}^{ps})$ converging to $\theta^+ = (\beta^+, \Lambda^+, \phi^+)$, where Λ^+ is a nondecreasing bound function on $[0, b]$ for $0 < b < \tau$ and it can be defined on $[0, \tau]$ if $\mu_1(\{\tau\}) > 0$.

Note that Θ_n is compact, and the function $m_\theta^{ps}(x)$ is upper semicontinuous in θ for almost all x. Furthermore, $m_\theta^{ps}(X) \leq M_0^{ps}(X) < \infty$ with $PM_0^{ps}(X) < \infty$ by C4. Thus, by Theorem A.1 of Wellner and Zhang (2000), we have

$$
\limsup_{n\to\infty} \sup_{\theta \in \Theta_n} (P_n - P) m_\theta^{ps}(X) \leq 0 \qquad a.s. \tag{7.2}
$$

By the Dominated Convergence Theorem and C4, $M^{ps}(\theta)$ is continuous in θ. Therefore, for any $\varepsilon > 0$, there exists $\phi_0^* \in \Psi_n$ such that

$$
M^{ps}(\beta_0, \Lambda_0, \phi_0) - \varepsilon \leq M^{ps}(\beta_0, \Lambda_0, \phi_0^*) \quad \text{with } ||\phi_0 - \phi_0^*||_\infty = o(1).
$$

Clearly,

$$
M_n^{ps}(\beta_0, \Lambda_0, \phi_0^*) - M^{ps}(\beta_0, \Lambda_0, \phi_0^*) = o_p(1)
$$

and

$$M_n^{ps}(\beta_0, \Lambda_0, \phi_0^*) \le M_n^{ps}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps}).$$

Then, using (7.2) and the arguments similar to those used in Lu et al. (2009), we can show that $M^{ps}(\theta^+) = M^{ps}(\theta_0)$, which yields $\beta^+ = \beta_0$, $\Lambda^+ = \Lambda_0$, a.e., and $\phi^+ = \phi_0$. Therefore, we obtain the weak consistency of $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})$ in the metric d_1.

Proof of Theorem 3.2

To obtain the rate of convergence, we will apply Theorem 3.2.5 of Van der Vaart and Wellner (1996). Let $m_\theta^{ps}(X)$, $M_n^{ps}(\theta)$, and $M^{ps}(\theta)$ be as defined in Appendix A.1. Let $\mu(u, v, w) = \Lambda(u)\exp\{v\phi(w)\}$, $\mu_0(u, v, w) = \Lambda_0(u)\exp\{v\phi_0(w)\}$ and $g(t) = \mu_t(U, Z, V, W)\exp(Z'\beta_t)$, where $(U, Z, V, W) \sim \nu_1$, $\mu_t = t\mu + (1-t)\mu_0$, $\beta_t = t\beta + (1-t)\beta_0$ for $0 \le t \le 1$. Then,

$$\Lambda(U)e^{Z'\beta + V\phi(W)} - \Lambda_0(U)e^{Z'\beta_0 + V\phi_0(W)} = g(1) - g(0).$$

By the mean value theorem, there exits a $0 \le \xi \le 1$ such that $g(1) - g(0) = g'(\xi)$. Since

$$\begin{aligned} g'(\xi) &= \exp(Z'\beta_\xi)[(\mu - \mu_0)(U, V, W) + \{\mu_0 + \xi(\mu - \mu_0)\}(U, V, W)Z'(\beta - \beta_0)] \\ &= \exp(Z'\beta_\xi)[(\mu - \mu_0)(U, V, W)\{1 + \xi Z'(\beta - \beta_0)\} \\ &\quad + \mu_0(U, V, W)Z'(\beta - \beta_0)], \end{aligned}$$

then from (7.1) we have

$$\begin{aligned} &M^{ps}(\theta_0) - M^{ps}(\theta) \\ &\ge c_1 \int \left\{\Lambda(u)\exp(z'\beta + v\phi(w)) - \Lambda_0(u)\exp(z'\beta_0 + v\phi_0(w))\right\}^2 d\nu_1(u, z, v, w) \\ &= c_1 \int [(\mu - \mu_0)(u, v, w)\{1 + \xi z'(\beta - \beta_0)\} \\ &\quad + \mu_0(u, v, w)z'(\beta - \beta_0)]^2 d\nu_1(u, z, v, w) \\ &= c_1\nu_1(g_1 h_1 + g_2)^2 \end{aligned}$$

for a constant c_1, where $g_1(U, Z, V, W) = Z'(\beta - \beta_0)\mu_0(U, V, W)$, $g_2(U, V, W) = (\mu - \mu_0)(U, V, W)$, and $h_1(U, Z, V, W) = 1 + \xi\frac{(\mu - \mu_0)(U, V, W)}{\mu_0(U, V, W)}$ in the notation of Lemma 8.8 of van der Vaart (2002, page 432). To apply van der Vaart's Lemma, we need to show that

$$\{\nu_1(g_1 g_2\}^2 \le c\nu_1(g_1^2)\nu_1(g_2^2) \tag{7.3}$$

for a constant $c < 1$. By the Cauchy-Schwarz inequality and condition (C13), we can show that (7.3) holds for $c = 1 - \eta_1$. Let

$$\Lambda_t = t\Lambda + (1-t)\Lambda_0, \phi_t = t\phi + (1-t)\phi_0, Q(t) = \Lambda_t(U)e^{V\phi_t(W)}.$$

Then

$$g_2(U, V, W) = Q(1) - Q(0) = Q'(\zeta) \quad \text{for } 0 \le \zeta \le 1,$$

and

$$\nu_1(g_2^2) = \nu_1((h_2 g_3 + g_4)^2)$$

where $g_3(U, V, W) = V(\phi(W) - \phi_0(W))\Lambda_0(U)$, $g_4(U) = (\Lambda - \Lambda_0)(U)$, and $h_2(U, V, W) = 1 + \zeta \frac{(\Lambda - \Lambda_0)(U)}{\Lambda_0(U)}$. Similarly, we can show that

$$\{\nu_1(g_3 g_4\}^2 \le (1 - \eta_2)\nu_1(g_3^2)\nu_1(g_4^2).$$

So, by van der Vaart's Lemma, we have

$$\nu_1(g_1 h + g_2)^2 \ge c d_1^2(\theta, \theta_0).$$

To derive the rate of convergence, next we need to find a $\varphi_n(\delta)$ such that

$$E\left\{\sup_{d_1(\theta,\theta_0)<\delta} \sqrt{n}|(P_n - P)(m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X))|\right\} \le c\varphi_n(\delta).$$

Let

$$\mathcal{F}_\delta^{ps} = \left\{m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X) : d_1(\theta, \theta_0) \le \delta\right\}.$$

From the result of Theorem 2.7.5 of Van der Vaart and Wellner (1996) and Lemma A.2 of Huang (1999), for any $\epsilon \le \delta$, we have

$$\log N_{[]}(\epsilon, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B}) \le c\left(\frac{1}{\epsilon} + q_n \log\frac{\delta}{\epsilon}\right),$$

where $||\cdot||_{P,B}$ is the Bernstein Norm defined as $||f||_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$ by van der Vaart and Wellner (1996, page 324). Moreover, we can show that

$$||m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X)||_{P,B}^2 \le c\delta^2,$$

for any $m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X) \in \mathcal{F}_\delta^{ps}$. Therefore, by Lemma 3.4.3 of Van der Vaart and Wellner (1996), we obtain

$$E||n^{1/2}(P_n - P)||_{\mathcal{F}_\delta^{ps}} \le cJ_{[]}(\delta, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B}) \left\{ 1 + \frac{J_{[]}(\delta, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B})}{\delta^2 n^{1/2}} \right\}$$

where

$$\begin{aligned} J_{[]}(\delta, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B}) &= \int_0^\delta \{1 + \log N_{[]}(\epsilon, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B})\}^{1/2} d\epsilon \\ &\le cq_n^{\frac{1}{2}} \int_0^\delta \left(1 + \frac{1}{\epsilon} + \log \frac{\delta}{\epsilon}\right)^{1/2} d\epsilon \le cq_n^{\frac{1}{2}} \delta^{\frac{1}{2}}. \end{aligned}$$

Thus,

$$\varphi_n(\delta) = cq_n^{\frac{1}{2}} \delta^{\frac{1}{2}} \left(1 + \frac{cq_n^{1/2}\delta^{1/2}}{\delta^2 n^{1/2}}\right) = c(q_n^{\frac{1}{2}} \delta^{\frac{1}{2}} + \frac{q_n}{\delta n^{1/2}}).$$

It is easy to see that $\varphi_n(\delta)/\delta$ is decreasing in δ, and

$$r_n^2 \varphi_n\left(\frac{1}{r_n}\right) = r_n^2 \left(q_n^{\frac{1}{2}} r_n^{-\frac{1}{2}} + \frac{q_n}{r_n^{-1} n^{1/2}}\right) = r_n^{\frac{3}{2}} q_n^{\frac{1}{2}} + r_n^3 q_n n^{-\frac{1}{2}} \le cn^{\frac{1}{2}}$$

for $r_n = n^{\frac{1-v}{3}}$ and $0 < v < 1/2$. Hence, it follows from Theorem 3.2.5 of Van der Vaart and Wellner (1996) that $n^{\frac{1-v}{3}} d_1(\hat{\theta}_n^{ps}, \theta_0) = O_p(1)$. Similarly, we can obtain the rate of convergence for $\hat{\theta}_n$.

Proof of Theorem 3.3

First, we prove part (i). Recall that

$$\begin{aligned} l_n^{ps}(\beta, \Lambda, \phi) = \sum_{i=1}^n \sum_{j=1}^{K_i} \Big[& N_i(T_{K_i,j}) \log \left\{\Lambda(T_{K_i,j})\right\} + N_i(T_{K_i,j})\{Z_i'\beta + V_i\phi(W_i)\} \\ & - \Lambda(T_{K_i,j}) \exp\{Z_i'\beta + V_i\phi(W_i)\}\Big]. \end{aligned}$$

We define a sequence of maps S_n^{ps} mapping a neighborhood of $(\beta_0, \Lambda_0, \phi_0)$, denoted by $\mathcal{U}$, in the parameter space for (β, Λ, ϕ) into $l^\infty(H_1 \times H_2 \times \mathcal{F}_r)$ as :

$$\begin{aligned} & S_n^{ps}(\theta)[\mathbf{h}_1, h_2, h_3] \\ & = n^{-1} \frac{d}{d\varepsilon} l_n^{ps}(\beta + \varepsilon \mathbf{h}_1, \Lambda + \varepsilon h_2, \phi + \varepsilon h_3)\Big|_{\varepsilon=0} \\ & = n^{-1} \sum_{i=1}^n \sum_{j=1}^{K_i} \Big[\{N_i(T_{K_i,j}) - \Lambda(T_{K_i,j}) \exp(\beta' Z_i + V_i\phi(W_i))\}\mathbf{h}_1' Z_i \end{aligned}$$

$$
\begin{aligned}
&\quad + \left\{ \frac{N_i(T_{K_i,j})}{\Lambda(T_{K_i,j})} - \exp(\beta' Z_i + V_i \phi(W_i)) \right\} h_2(T_{K_i,j}) \\
&\quad + \{N_i(T_{K_i,j}) - \Lambda(T_{K_i,j}) \exp(\beta' Z_i + V_i \phi(W_i))\} V_i h_3(W_i) \Big] \\
&\equiv A_{n1}^{ps}(\theta)[\mathbf{h}_1] + A_{n2}^{ps}(\theta)[h_2] + A_{n3}^{ps}(\theta)[h_3] \\
&\equiv P_n(\mathbf{h}_1' \dot{l}_\beta^{ps}) + P_n(\dot{l}_\Lambda^{ps}[h_2]) + P_n(\dot{l}_\phi^{ps}[h_3]) \\
&\equiv P_n \psi_{ps}(\theta)[\mathbf{h}_1, h_2, h_3].
\end{aligned}
$$

Correspondingly, we define the limit map $S^{ps} : \mathcal{U} \longrightarrow l^\infty(H_1 \times H_2 \times \mathcal{F}_r)$ as

$$
S^{ps}(\theta)[\mathbf{h}_1, h_2, h_3] = A_1^{ps}(\theta)[\mathbf{h}_1] + A_2^{ps}(\theta)[h_2] + A_3^{ps}(\theta)[h_3],
$$

where

$$
A_1^{ps}(\theta)[\mathbf{h}_1] = P\left[\sum_{j=1}^K \{N(T_{K,j}) - \Lambda(T_{K,j}) \exp(\beta' Z + V\phi(W))\} \mathbf{h}_1' Z \right],
$$

$$
A_2^{ps}(\theta)[h_2] = P\left[\sum_{j=1}^K \left\{ \frac{N(T_{K,j})}{\Lambda(T_{K,j})} - \exp(\beta' Z + V\phi(W)) \right\} h_2(T_{K,j}) \right],
$$

and

$$
A_3^{ps}(\theta)[h_3] = P\left[\sum_{j=1}^K \{N(T_{K,j}) - \Lambda(T_{K,j}) \exp(\beta' Z + V\phi(W))\} V h_3(W) \right].
$$

To derive the asymptotic normality of the estimators $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})$, motivated by the proofs of Theorem 3.3.1 of Van der Vaart and Wellner (1996, page 310) and Theorem 2 of Zeng et al. (2005) , we need to verify the following five conditions.

(a1) $\sqrt{n}(S_n^{ps} - S^{ps})(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps}) - \sqrt{n}(S_n^{ps} - S^{ps})(\beta_0, \Lambda_0, \phi_0) = o_p(1)$.

(a2) $S^{ps}(\beta_0, \Lambda_0, \phi_0) = 0$ and $S_n^{ps}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps}) = o_p(n^{-1/2})$.

(a3) $\sqrt{n}(S_n^{ps} - S^{ps})(\beta_0, \Lambda_0, \phi_0)$ converges in distribution to a tight Gaussian process on $l^\infty(H_1 \times H_2 \times \mathcal{F}_r)$.

(a4) $S^{ps}(\beta, \Lambda, \phi)$ is Fréchet-differentiable at $(\beta_0, \Lambda_0, \phi_0)$ with a continuously invertible derivative $\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)$.

(a5) $S^{ps}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps}) - S^{ps}(\beta_0, \Lambda_0, \phi_0) - \dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\hat{\beta}_n^{ps} - \beta_0, \hat{\Lambda}_n^{ps} - \Lambda_0, \hat{\phi}_n^{ps} - \phi_0) = o_p(n^{-1/2})$.

Condition (a1) holds since

$$
\Big\{ \psi_{ps}(\beta, \Lambda, \phi)[\mathbf{h}_1, h_2, h_3] - \psi_{ps}(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1, h_2, h_3] : \\
d_1((\beta, \Lambda, \phi), (\beta_0, \Lambda_0, \phi_0)) < \delta, (\mathbf{h}_1, h_2, h_3) \in H_1 \times H_2 \times \mathcal{F}_r \Big\}
$$

is a Donsker class for some $\delta > 0$, and that

$$\sup_{(\mathbf{h}_1,h_2,h_3)\in\mathcal{H}} P\left[\psi_{ps}(\beta,\Lambda,\phi)[\mathbf{h}_1,h_2,h_3]-\psi_{ps}(\beta_0,\Lambda_0,\phi_0)[\mathbf{h}_1,h_2,h_3]\right]^2 \longrightarrow 0,$$

as $(\beta,\Lambda,\phi)\longrightarrow(\beta_0,\Lambda_0,\phi_0)$ in d_1.

Clearly, $S^{ps}(\beta_0,\Lambda_0,\phi_0) = 0$. For $h_3 \in \mathcal{F}_r$, let h_{3n} be the B-spline function approximation of h_3 with $||h_3 - h_{3n}||_\infty = O(n^{-vr})$ by Corollary 6.21 of Schumaker (1981, page 227). Then we have $S_n^{ps}(\hat\beta_n^{ps},\hat\Lambda_n^{ps},\hat\phi_n^{ps})[\mathbf{h}_1,h_2,h_{3n}] = 0$. Thus, for $(\mathbf{h}_1,h_2,h_3)\in H_1\times H_2\times\mathcal{F}_r$,

$$\begin{aligned}
&\sqrt{n}\{S_n^{ps}(\hat\beta_n^{ps},\hat\Lambda_n^{ps},\hat\phi_n^{ps})[\mathbf{h}_1,h_2,h_3]\}\\
&\quad=\sqrt{n}P_n\psi_{ps}(\hat\theta_n^{ps})[\mathbf{h}_1,h_2,h_3]-\sqrt{n}P_n\psi_{ps}(\hat\theta_n^{ps})[\mathbf{h}_1,h_2,h_{3n}]\\
&\quad=I_{n1}-I_{n2}+I_{n3}+I_{n4}
\end{aligned}$$

where

$$\begin{aligned}
I_{n1}&=\sqrt{n}(P_n-P)\left\{\psi_{ps}(\hat\theta_n^{ps})[\mathbf{h}_1,h_2,h_3]-\psi_{ps}(\theta_0)[\mathbf{h}_1,h_2,h_3]\right\},\\
I_{n2}&=\sqrt{n}(P_n-P)\left\{\psi_{ps}(\hat\theta_n^{ps})[\mathbf{h}_1,h_2,h_{3n}]-\psi_{ps}(\theta_0)[\mathbf{h}_1,h_2,h_{3n}]\right\},\\
I_{n3}&=\sqrt{n}P_n\left\{\psi_{ps}(\theta_0)[\mathbf{h}_1,h_2,h_3]-\psi_{ps}(\theta_0)[\mathbf{h}_1,h_2,h_{3n}]\right\},
\end{aligned}$$

and

$$I_{n4}=\sqrt{n}P\left\{\psi_{ps}(\hat\theta_n^{ps})[\mathbf{h}_1,h_2,h_3]-\psi_{ps}(\hat\theta_n^{ps})[\mathbf{h}_1,h_2,h_{3n}]\right\}.$$

From (a1), we have $I_{n1} = o_p(1)$ and $I_{n2} = o_p(1)$. Next we need to show $I_{n3} = o_p(1)$ and $I_{n4} = o_p(1)$. Note that

$$E(I_{n3}^2)=P\left\{\psi_{ps}(\theta_0)[\mathbf{h}_1,h_2,h_3]-\psi_{ps}(\theta_0)[\mathbf{h}_1,h_2,h_{3n}]\right\}^2\le c||h_{3n}-h_3||_\infty^2\to 0,$$

and

$$\begin{aligned}
|I_{n4}|&=\Bigg|\sqrt{n}P\Bigg[\sum_{j=1}^{K}\Big\{\Lambda_0(T_{K,j})\exp(Z'\beta_0+V\phi_0(W))\\
&\qquad-\hat\Lambda_n^{ps}(T_{K,j})\exp(Z'\hat\beta_n^{ps}+V\hat\phi_n^{ps}(W))\Big\}V(h_3(W)-h_{3n}(W))\Bigg]\Bigg|\\
&\le c\sqrt{n}d_1(\hat\theta_n^{ps},\theta_0)||h_3-h_{3n}||_\infty\\
&=O(n^{-\frac{1-v}{3}-vr+\frac{1}{2}})
\end{aligned}$$

by Theorem 3.2. Thus (a2) holds for $\frac{1}{6r-2} < v < \frac{1}{2}$.

Condition (a3) holds because $H_1 \times H_2 \times \mathcal{F}_r$ is a Donsker class and the functionals $A_{n1}^{ps}, A_{n2}^{ps}, A_{n3}^{ps}$ are bounded Lipschitz functions with respect to $H_1 \times H_2 \times \mathcal{F}_r$.

For (a4), by the smoothness of $S^{ps}(\beta, \Lambda, \phi)$, the Fréchet differentiability holds and the derivative of $S^{ps}(\beta, \Lambda, \phi)$ at $(\beta_0, \Lambda_0, \phi_0)$, denoted by $\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)$, is a map from the space $\{(\beta-\beta_0, \Lambda-\Lambda_0, \phi-\phi_0) : (\beta, \Lambda, \phi) \in \mathcal{U}\}$ to $l^\infty(H_1 \times H_2 \times \mathcal{F}_r)$ and

$$
\begin{aligned}
&\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] \\
&\quad = \frac{d}{d\varepsilon} \left\{ A_1^{ps}(\theta_0 + \varepsilon(\theta - \theta_0))[\mathbf{h}_1] \right\} \Big|_{\varepsilon=0} + \frac{d}{d\varepsilon} \left\{ A_2^{ps}(\theta_0 + \varepsilon(\theta - \theta_0))[h_2] \right\} \Big|_{\varepsilon=0} \\
&\qquad + \frac{d}{d\varepsilon} \left\{ A_3^{ps}(\theta_0 + \varepsilon(\theta - \theta_0))[h_3] \right\} \Big|_{\varepsilon=0} \\
&\quad = -P \sum_{j=1}^{K} \exp(\beta_0' Z + V\phi_0(W)) \mathbf{h}_1' Z \left[\left\{ \Lambda(T_{K,j}) - \Lambda_0(T_{K,j}) \right\} \right. \\
&\qquad \left. + \Lambda_0(T_{K,j}) \left\{ (\beta - \beta_0)' Z + V(\phi(W) - \phi_0(W)) \right\} \right] \\
&\qquad - P \sum_{j=1}^{K} \exp(\beta_0' Z + V\phi_0(W)) h_2(T_{K,j}) \left[\left\{ \frac{\Lambda(T_{K,j}) - \Lambda_0(T_{K,j})}{\Lambda_0(T_{K,j})} \right. \right. \\
&\qquad \left. \left. + \left\{ (\beta - \beta_0)' Z + V(\phi(W) - \phi_0(W)) \right\} \right] \right. \\
&\qquad - P \sum_{j=1}^{K} \exp(\beta_0' Z + V\phi_0(W)) V h_3(W) \left[\left\{ \Lambda(T_{K,j}) - \Lambda_0(T_{K,j}) \right\} \right. \\
&\qquad \left. + \Lambda_0(T_{K,j}) \left\{ (\beta - \beta_0)' Z + V(\phi(W) - \phi_0(W)) \right\} \right].
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] \\
&\quad = (\beta - \beta_0)' Q_1^{ps}(\mathbf{h}_1, h_2, h_3) + \int (\Lambda(t) - \Lambda_0(t)) d Q_2^{ps}(\mathbf{h}_1, h_2, h_3)(t) \\
&\qquad + \int (\phi(w) - \phi_0(w)) d Q_3^{ps}(\mathbf{h}_1, h_2, h_3)(w)
\end{aligned} \tag{7.4}
$$

where

$$
\begin{aligned}
Q_1^{ps}(\mathbf{h}_1, h_2, h_3) &= -E \left[Z \exp\{\beta_0' Z + V\phi_0(W)\} \sum_{j=1}^{K} \left\{ \Lambda_0(T_{K,j}) \mathbf{h}_1' Z + h_2(T_{K,j}) \right. \right. \\
&\qquad \left. \left. + \Lambda_0(T_{K,j}) V h_3(W) \right\} \right], \\
d Q_2^{ps}(\mathbf{h}_1, h_2, h_3)(t) &= -E \left[\exp\{\beta_0' Z + V\phi_0(W)\} \sum_{j=1}^{K} \frac{1}{\Lambda_0(t)} \left\{ \Lambda_0(t) \mathbf{h}_1' Z + h_2(t) \right. \right. \\
&\qquad \left. \left. + \Lambda_0(t) V h_3(W) \right\} d P(T_{K,j} \le t | K, Y) \right],
\end{aligned}
$$

and

$$
\begin{aligned}
&dQ_3^{ps}(\mathbf{h}_1, h_2, h_3)(w)\\
&\quad = -E\left[V\exp\{\beta_0' Z + V\phi_0(w)\}\sum_{j=1}^{K}\left\{\Lambda_0(T_{K,j})\mathbf{h}_1' Z + h_2(T_{K,j})\right.\right.\\
&\qquad \left.\left. + \Lambda_0(T_{K,j})Vh_3(w)\right\}|W = w\right]dF_W(w)
\end{aligned}
$$

where F_W denotes the cumulative distribution of W.

Next, we show that $Q^{ps} = (Q_1^{ps}, Q_2^{ps}, Q_3^{ps})$ is one-to-one, that is, for $(\mathbf{h}_1, h_2, h_3) \in H_1 \times H_2 \times \mathcal{F}_r$, if $Q^{ps}(\mathbf{h}_1, h_2, h_3) = 0$, then $\mathbf{h}_1 = \mathbf{0}$, $h_2 = 0$, $h_3 = 0$.

Suppose that $Q^{ps}(\mathbf{h}_1, h_2, h_3) = 0$. Then $\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] = 0$ for any (β, Λ, ϕ) in the neighborhood $\mathcal{U}$. In particular, we take $\beta = \beta_0 + \epsilon\mathbf{h}_1$, $\Lambda = \Lambda_0 + \epsilon h_2$, $\phi = \phi_0 + \epsilon h_3$ for a small constant ϵ. Thus we have

$$
\begin{aligned}
0 &= \dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3]\\
&= -\epsilon E\left[\exp\{\beta_0' Z + V\phi_0(W)\}\sum_{j=1}^{K}\Lambda_0(T_{K,j})\left\{\mathbf{h}_1' Z + Vh_3(W) + \frac{h_2(T_{K,j})}{\Lambda_0(T_{K,j})}\right\}^2\right],
\end{aligned}
$$

which yields

$$
\mathbf{h}_1' Z + Vh_3(W) + \frac{h_2(T_{K,j})}{\Lambda_0(T_{K,j})} = 0, \quad j = 1, \ldots, K, \quad a.s.
$$

and so $\mathbf{h}_1 = \mathbf{0}$, $h_2 = 0$, $h_3 = 0$ by C7.

Next we show that (a5) holds. Write

$$
\begin{aligned}
&S^{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1, h_2, h_3] - S^{ps}(\theta_0)[\mathbf{h}_1, h_2, h_3]\\
&\quad - \dot{S}(\beta_0, \Lambda_0, \phi_0)(\hat{\beta}_n^{ps} - \beta_0, \hat{\Lambda}_n^{ps} - \Lambda_0, \hat{\phi}_n^{ps} - \phi_0)[\mathbf{h}_1, h_2, h_3]\\
&= B_{n1} + B_{n2} + B_{n3}
\end{aligned}
$$

where

$$
\begin{aligned}
B_{n1} &= A_1^{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1] - \frac{d}{d\varepsilon}\left\{A_1^{ps}(\theta_0 + \varepsilon(\hat{\theta}_n^{ps} - \theta_0))[\mathbf{h}_1]\right\}\Big|_{\varepsilon=0},\\
B_{n2} &= A_2^{ps}(\hat{\theta}_n^{ps})[h_2] - \frac{d}{d\varepsilon}\left\{A_2^{ps}(\theta_0 + \varepsilon(\hat{\theta}_n^{ps} - \theta_0))[h_2]\right\}\Big|_{\varepsilon=0},
\end{aligned}
$$

and

$$
B_{n3} = A_3^{ps}(\hat{\theta}_n^{ps})[h_3] - \frac{d}{d\varepsilon}\left\{A_3^{ps}(\theta_0 + \varepsilon(\hat{\theta}_n^{ps} - \theta_0))[h_3]\right\}\Big|_{\varepsilon=0}
$$

It is easy to show that $B_{n1} = O_p(d_1^2(\hat{\theta}_n^{ps}, \theta_0))$, $B_{n2} = O_p(d_1^2(\hat{\theta}_n^{ps}, \theta_0))$, and $B_{n3} = O_p(d_1^2(\hat{\theta}_n^{ps}, \theta_0))$. Thus, by Theorem 3.2, (a5) holds for $0 < v < 1/4$.

It follows from (7.4), (a1), (a2) and (a5) that

$$\begin{aligned}
&\sqrt{n}(\hat{\beta}_n^{ps} - \beta_0)' Q_1^{ps}(\mathbf{h}_1, h_2, h_3) + \sqrt{n}\int \{\hat{\Lambda}_n^{ps}(t) - \Lambda_0(t)\} d Q_2^{ps}(\mathbf{h}_1, h_2, h_3)(t) \\
&\quad + \sqrt{n}\int \{\hat{\phi}_n^{ps}(w) - \phi_0(w)\} d Q_3^{ps}(\mathbf{h}_1, h_2, h_3)(w) \\
&\quad = -\sqrt{n}(S_n^{ps} - S^{ps})(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1, h_2, h_3] + o_p(1),
\end{aligned}$$

uniformly in $\mathbf{h}_1$, h_2 and h_3.

For each $(\mathbf{h}_1, h_2, h_3) \in H_1 \times H_2 \times \mathcal{F}_r$, since Q^{ps} is invertible, there exists $(\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}) \in H_1 \times H_2 \times \mathcal{F}_r$ such that

$$Q_1^{ps}(\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}) = \mathbf{h}_1,\ Q_2^{ps}(\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}) = h_2,\ Q_3^{ps}(\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}) = h_3.$$

Therefore, we have

$$\begin{aligned}
&\mathbf{h}_1' \sqrt{n}(\hat{\beta}_n^{ps} - \beta_0) + \sqrt{n}\int \{\hat{\Lambda}_n^{ps}(t) - \Lambda_0(t)\} d h_2(t) \\
&\quad + \sqrt{n}\int \{\hat{\phi}_n^{ps}(w) - \phi_0(w)\} d h_3(w) \\
&\quad = -\sqrt{n}(S_n^{ps} - S^{ps})(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}] + o_p(1) \\
&\quad \to_d N(0, \sigma_{ps}^2),
\end{aligned}$$

where

$$\sigma_{ps}^2 = E\{\psi_{ps}^2(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}]\}. \tag{7.5}$$

To prove part (ii), we define a sequence of maps S_n mapping a neighborhood of $(\beta_0, \Lambda_0, \phi_0), \mathcal{U}$, in the parameter space for (β, Λ, ϕ) into $l^\infty(H_1 \times H_2 \times \mathcal{F}_r)$ as:

$$S_n(\theta)[\mathbf{h}_1, h_2, h_3] = n^{-1}\frac{d}{d\varepsilon} l_n(\beta + \varepsilon \mathbf{h}_1, \Lambda + \varepsilon h_2, \phi + \varepsilon h_3)\Big|_{\varepsilon=0}.$$

Write $\Delta N_i(T_{K_i,j}) = N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})$, $\Delta\Lambda(T_{K_i,j}) = \Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})$, and $\Delta h(T_{K_i,j}) = h(T_{K_i,j}) - h(T_{K_i,j-1})$.

Then, we have

$$\begin{aligned}
&S_n(\theta)[\mathbf{h}_1, h_2, h_3] \\
&\quad = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{K_i} \Big[\{\Delta N_i(T_{K_i,j}) - \Delta\Lambda(T_{K_i,j})\exp(\beta' Z_i + V_i\phi(W_i))\}\mathbf{h}_1' Z_i \\
&\qquad + \left\{\frac{\Delta N_i(T_{K_i,j})}{\Delta\Lambda(T_{K_i,j})} - \exp(\beta' Z_i + V_i\phi(W_i))\right\}\Delta h_2(T_{K_i,j})
\end{aligned}$$

$$
\begin{aligned}
&+\{\Delta N_i(T_{K_i,j}) - \Delta\Lambda(T_{K_i,j})\exp(\beta' Z_i + V_i\phi(W_i))\}V_i h_3(W_i)\big] \\
\equiv\ & A_{n1}(\theta)[\mathbf{h}_1] + A_{n2}(\theta)[h_2] + A_{n3}(\theta)[h_3] \\
\equiv\ & P_n(\mathbf{h}_1'\dot{l}_\beta) + P_n(\dot{l}_\Lambda[h_2]) + P_n(\dot{l}_\phi[h_3]) \\
\equiv\ & P_n\psi(\theta)[\mathbf{h}_1, h_2, h_3].
\end{aligned}
$$

Correspondingly, we define the limit map $S : \mathcal{U} \longrightarrow l^\infty(H_1 \times H_2 \times \mathcal{F}_r)$ as

$$
S(\theta)[\mathbf{h}_1, h_2, h_3] = A_1(\theta)[\mathbf{h}_1] + A_2(\theta)[h_2] + A_3(\theta)[h_3],
$$

where

$$
A_1(\theta)[\mathbf{h}_1] = E\left[\sum_{j=1}^{K}\{\Delta N(T_{K,j}) - \Delta\Lambda(T_{K,j})\exp(\beta' Z + V\phi(W))\}\mathbf{h}_1' Z\right],
$$

$$
A_2(\theta)[h_2] = E\left[\sum_{j=1}^{K}\left\{\frac{\Delta N(T_{K,j})}{\Delta\Lambda(T_{K,j})} - \exp(\beta' Z + V\phi(W))\right\}\Delta h_2(T_{K,j})\right],
$$

and

$$
A_3(\theta)[h_3] = E\left[\sum_{j=1}^{K}\{\Delta N(T_{K,j}) - \Delta\Lambda(T_{K,j})\exp(\beta' Z + V\phi(W))\}V h_3(W)\right].
$$

Furthermore, the derivative of $S(\beta, \Lambda, \phi)$ at $(\beta_0, \Lambda_0, \phi_0)$, denoted by $\dot{S}(\beta_0, \Lambda_0, \phi_0)$, is a map from the space $\{(\beta-\beta_0, \Lambda-\Lambda_0, \phi-\phi_0) : (\beta, \Lambda, \phi) \in \mathcal{U}\}$ to $l^\infty(H_1 \times H_2 \times \mathcal{F}_r)$ and

$$
\begin{aligned}
&\dot{S}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] \\
&\quad = (\beta - \beta_0)' Q_1(\mathbf{h}_1, h_2, h_3) + \int\{\Lambda(t) - \Lambda_0(t)\}dQ_2(\mathbf{h}_1, h_2, h_3)(t) \\
&\qquad + \int\{\phi(w) - \phi_0(w)\}dQ_3(\mathbf{h}_1, h_2, h_3)(w)
\end{aligned}
$$

where

$$
\begin{aligned}
&Q_1(\mathbf{h}_1, h_2, h_3) \\
&\quad = -E\Bigg[Z\exp\{\beta_0' Z + V\phi_0(W)\} \\
&\qquad \times \sum_{j=1}^{K}\left\{\Delta\Lambda_0(T_{K,j})\mathbf{h}_1' Z + \Delta h_2(T_{K,j}) + \Delta\Lambda_0(T_{K,j})V h_3(W)\right\}\Bigg], \\
&dQ_2(\mathbf{h}_1, h_2, h_3)(t)
\end{aligned}
$$

$$
\begin{aligned}
&= -E\Bigg[\exp\{\beta_0' Z + V\phi_0(W)\} \\
&\quad \times \sum_{j=1}^{K}\Bigg\{\Bigg(\mathbf{h}_1' Z + \frac{h_2(t) - h_2(T_{K,j-1})}{\Lambda_0(t) - \Lambda_0(T_{K,j-1})} + V h_3(W)\Bigg) dP(T_{K,j} \le t | K, T_{K,j-1}, Y) \\
&\quad - \Bigg(\mathbf{h}_1' Z + \frac{h_2(T_{K,j}) - h_2(t)}{\Lambda_0(T_{K,j}) - \Lambda_0(t)} + V h_3(W)\Bigg) dP(T_{K,j-1} \le t | K, T_{K,j}, Y)\Bigg\}\Bigg],
\end{aligned}
$$

and

$$
\begin{aligned}
&dQ_3(\mathbf{h}_1, h_2, h_3)(w) \\
&\quad = -E\Big[V \exp\{\beta_0' Z + V\phi_0(w)\} \\
&\quad\quad \times \sum_{j=1}^{K}\left\{\Delta\Lambda_0(T_{K,j})\mathbf{h}_1' Z + \Delta h_2(T_{K,j}) + \Delta\Lambda_0(T_{K,j}) V h_3(w)\right\} | W = w\Bigg] dF_W(w).
\end{aligned}
$$

Next, we show that $Q = (Q_1, Q_2, Q_3)$ is one-to-one, that is, for $(\mathbf{h}_1, h_2, h_3) \in H_1 \times H_2 \times \mathcal{F}_r$, if $Q(\mathbf{h}_1, h_2, h_3) = 0$, then $\mathbf{h}_1 = \mathbf{0}$, $h_2 = 0$, $h_3 = 0$

Suppose that $Q(\mathbf{h}_1, h_2, h_3) = 0$. Then $\dot{S}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] = 0$ for any (β, Λ, ϕ) in the neighborhood $\mathcal{U}$. In particular, we take $\beta = \beta_0 + \epsilon\mathbf{h}_1$, $\Lambda = \Lambda_0 + \epsilon h_2$, $\phi = \phi_0 + \epsilon h_3$ for a small constant ϵ. Thus we have

$$
\begin{aligned}
0 &= \dot{S}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] \\
&= -\epsilon E\left[\exp\{\beta_0' Z + V\phi_0(W)\} \sum_{j=1}^{K} \Delta\Lambda_0(T_{K,j})\left\{\mathbf{h}_1' Z + V h_3(W) + \frac{\Delta h_2(T_{K,j})}{\Delta\Lambda_0(T_{K,j})}\right\}^2\right],
\end{aligned}
$$

which yields

$$
\mathbf{h}_1' Z + V h_3(W) + \frac{\Delta h_2(T_{K,j})}{\Delta\Lambda_0(T_{K,j})} = 0, \;\; j = 1, \ldots, K, \quad a.s.
$$

and so $\mathbf{h}_1 = \mathbf{0}$, $h_2 = 0$, $h_3 = 0$ by C7.

Similarly, we can show that $S(\beta_0, \Lambda_0, \phi_0) = 0$, $S_n(\hat{\beta}_n, \hat{\Lambda}_n, \hat{\phi}_n) = o_p(n^{-1/2})$, and

$$
\begin{aligned}
&S(\hat{\theta}_n)[\mathbf{h}_1, h_2, h_3] - S(\theta_0)[\mathbf{h}_1, h_2, h_3] \\
&\quad = \dot{S}(\beta_0, \Lambda_0, \phi_0)(\hat{\beta}_n - \beta_0, \hat{\Lambda}_n - \Lambda_0, \hat{\phi}_n - \phi_0)[\mathbf{h}_1, h_2, h_3] + O_p(d_2^2(\hat{\theta}_n, \theta_0)) \\
&\quad = \dot{S}(\beta_0, \Lambda_0, \phi_0)(\hat{\beta}_n - \beta_0, \hat{\Lambda}_n - \Lambda_0, \hat{\phi}_n - \phi_0)[\mathbf{h}_1, h_2, h_3] + o_p(n^{-1/2}).
\end{aligned}
$$

for $0 < v < 1/4$. Thus it follows that

$$
\sqrt{n}(\hat{\beta}_n - \beta_0)' Q_1(\mathbf{h}_1, h_2, h_3) + \sqrt{n}\int \{\hat{\Lambda}_n(t) - \Lambda_0(t)\} dQ_2(\mathbf{h}_1, h_2, h_3)(t)
$$

$$+\sqrt{n}\int\{\hat{\phi}_n(w)-\phi_0(w)\}dQ_3(\mathbf{h}_1,h_2,h_3)(w)$$
$$=-\sqrt{n}(S_n-S)(\beta_0,\Lambda_0,\phi_0)[\mathbf{h}_1,h_2,h_3]+o_p(1),$$

uniformly in $\mathbf{h}_1$, h_2 and h_3.

For each $(\mathbf{h}_1, h_2, h_3) \in (\mathbf{h}_1, h_2, h_3)$, since Q is invertible, there exists $(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) \in (\mathbf{h}_1, h_2, h_3)$ such that

$$Q_1^*(\mathbf{h}_1^*, h_2^*, h_3^*) = \mathbf{h}_1,\ Q_2^*(\mathbf{h}_1^*, h_2^*, h_3^*) = 0,\ Q_3^*(\mathbf{h}_1^*, h_2^*, h_3^*) = h_3.$$

Thus, we have

$$\begin{aligned}
&\mathbf{h}_1'\sqrt{n}(\hat{\beta}_n-\beta_0)+\sqrt{n}\int\{\hat{\Lambda}_n(t)-\Lambda_0(t)\}dh_2(t)\\
&+\sqrt{n}\int\{\hat{\phi}_n(w)-\phi_0(w)\}dh_3(w)\\
&=-\sqrt{n}(S_n-S)(\beta_0,\Lambda_0,\phi_0)[\mathbf{h}_1^*,h_2^*,h_3^*]+o_p(1)\\
&\to_d N(0,\sigma^2),
\end{aligned}$$

where

$$\sigma^2 = E\{\psi^2(\beta_0,\Lambda_0,\phi_0)[\mathbf{h}_1^*,h_2^*,h_3^*]\}. \tag{7.6}$$

References

Andrews DF, Herzberg AM (1985) Data: a collection of problems from many fields for the student and research worker. Springer, New York

Balakrishnan N, Zhao X (2009) New multi-sample nonparametric tests for panel count data. Ann Stat 37:1112–1149

Byar DP (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 PH, Edsmyr F (eds) Bladder tumors and other topics in urological oncology. Plenum, New York, pp 363–370

Cheng MY, Honda T, Li J, Peng H (2014) Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal data. Ann Stat 42:1819–1849

Cheng SC, Wei LJ (2000) Inferences for a semiparametric model with panel data. Biometrika 87:89–97

Hu XJ, Lagakos SW, Lockhart RA (2009a) Generalized least squares estimation of the mean function of a counting process based on panel counts. Stat Sinica 19:561–580

Hu XJ, Lagakos SW, Lockhart RA (2009b) Marginal analysis of panel counts through estimating functions. Biometrika 96:445–456

Hu XJ, Sun J, Wei LJ (2003) Regression parameter estimation from panel counts. Scand J Stat 30:25–43

Hua L, Zhang Y (2012) Spline-based semiparametric projected generalized estimating equation method for panel count data. Biostatistics 13:440–454

Huang J (1999) Efficient estimation of the partly linear additive Cox model. Ann Stat 27:1536–1563

Huang CY, Wang MC, Zhang Y (2006) Analysing panel count data with informative observation times. Biometrika 93:763–775

Lindqvist BH, Kvaløy JT, Aaserud S (2015) Residual plots to reveal the functional form for covariates in parametric accelerated failure time models. Lifetime Data Anal 21:353–378

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

Lu M, Zhang Y, Huang J (2009) Semiparametric estimation methods for panel count data using monotone B-splines. J Am Stat Assoc 104:1060–1070
Schumaker L (1981) Spline functions: basic theory. Wiley, New York
Sun J, Fang HB (2003) A nonparametric test for panel count data. Biometrika 90:199–208
Sun J, Kalbfleisch JD (1995) Estimation of the mean function of point processes based on panel count data. Stat Sinica 5:279–290
Sun J, Tong X, He X (2007) Regression analysis of panel count data with dependent observation times. Biometrics 63:1053–1059
Sun J, Wei LJ (2000) Regression analysis of panel count data with covariate-dependent observation and censoring times. J R Stat Soc Ser B 62:293–302
Thall PF, Lachin JM (1988) Analysis of recurrent events: nonparametric methods for random-interval count data. J Am Stat Assoc 83:339–347
Van der Vaart AW (2002) Semiparametric Statistics. In Lectures on Probability Theory and Statistics, Ecole d'Eté de Probabilités de Saint-Flour XXIX-1999. Lecture Notes in Mathematics 1781:330–457
Van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer, New York
Wellner JA, Zhang Y (2000) Two estimators of the mean of a counting process with panel count data. Ann Stat 28:779–814
Wellner JA, Zhang Y (2007) Two likelihood-based semiparametric estimation methods for panel count data with covariates. Ann Stat 35:2106–2142
Zeng D, Lin D, Yin GS (2005) Maximum likelihood estimation for the proportional odds model with random effects. J Am Stat Assoc 100:470–482
Zhang Y (2002) A semiparametric pseudolikelihood estimation method for panel count data. Biometrika 89:39–48
Zhang Y (2006) Nonparametric k-sample tests for panel count data. Biometrika 93:777–790
Zhang F, Chen X, Zhou Y (2014) Proportional hazards model with varying coefficients for length-biased data. Lifetime Data Anal 20:132–157
Zhang Y, Jamshidian M (2003) The gamma-frailty Poisson model for the nonparametric estimation of panel count data. Biometrics 59:1099–1106