Semiparametric Estimation Methods for Panel Count Data Using Monotone Polynomial Splines

Minggen Lu, Ying Zhang, and Jian Huang

We study semiparametric likelihood-based methods for panel count data with proportional mean model $E[\mathbb{N}(t)|Z] = \Lambda_0(t)\exp(\beta_0^T Z)$, where Z is a vector of covariates and $\Lambda_0(t)$ is the baseline mean function. We propose to estimate $\Lambda_0(t)$ and β_0 jointly with $\Lambda_0(t)$ approximated by monotone polynomial B-splines and to compute the estimators using the generalized Rosen algorithm utilized in Zhang and Jamshidian (2004) for nonparametric maximum likelihood estimation problems. We show that the proposed spline-based likelihood estimators of $\Lambda_0(t)$ are consistent with a possibly better than $n^{1/3}$ convergence rate. The normality of estimators of the β_0 is also established. Comparisons between the proposed estimators and their alternatives studied in Wellner and Zhang (2007) are made through simulations studies, regarding their finite sample performance and computational complexity. A real example from a bladder tumor clinical trial is used to illustrate the methods.

KEY WORDS: Counting process; Empirical process; Generalized Rosen Algorithm; Maximum pseudo-likelihood method; Maximum likelihood method; Monotone polynomial splines; Monte Carlo.

1. INTRODUCTION

In many long-term clinical trials or epidemiological studies, the subjects are observed at several discrete time-points during the study. Only the number of recurrent events occurred before each observation is observed. The number of observations and observation times may vary from individual to individual. This kind of data is referred to as panel count data.

A motivating example of panel count data is the bladder tumor randomized clinical trial conducted by the Veterans Administration Cooperative Urological Research Group. In this study, all patients had superficial bladder tumors when they entered the trial and they were randomly assigned to one of the three arms: placebo, pyridoxine pill or thiotepa instillation. Many patients had multiple recurrences of the tumor and new tumors were removed at each visit. One of the primary goals of this study was to determine the effect of treatment on suppressing the tumor recurrence; see for example, Byar (1980), Wei, Lin, and Weissfeld (1989), Wellner and Zhang (2000, 2007), Sun and Wei (2000), and Lu, Zhang, and Huang (2007).

Panel count data can be treated as incomplete observations on counting processes. The methods for estimating the mean function of a counting process with this type of data have been explored in literatures; see for example, Kalbfleisch and Lawless (1985), Thall and Lachin (1988), Sun and Kalbfleisch (1995), and Wellner and Zhang (2000). In many applications, the effects of covariates on the underlying counting process are of primary interest. Several studies have considered semiparametric regression methods with the proportional mean model for panel count data, namely,

$$E[\mathbb{N}(t)|Z] = \Lambda_0(t)\exp(\beta_0^T Z), \tag{1}$$

where the true baseline mean function $\Lambda_0(t)$ is left completely unspecified and β_0 is a vector of regression parameters. Sun and Wei (2000) proposed some procedures for estimating regression parameters using generalized estimating equations techniques. Wellner and Zhang (2007) studied both semiparametric maximum pseudo-likelihood estimator and semiparametric maximum likelihood estimator with the proportional mean model (1), assuming the underlying counting process is nonhomogeneous Poisson conditional on covariates. Their methods are shown to be robust against possible misspecification of the underlying counting process, as long as the proportional mean model (1) holds. The semiparametric maximum pseudo-likelihood estimator is fairly easy to compute, but it can be very inefficient when the distribution of the number of observations K is heavily tailed as shown in an example given by Wellner, Zhang, and Liu (2004). The semiparametric maximum likelihood estimator is more efficient than semiparametric maximum pseudo-likelihood estimator, but its computation is very intensive. Although the asymptotic normality was established for both the pseudo-likelihood and likelihood estimators of β_0, there are in general no explicit forms of the asymptotic variances, which make it difficult to estimate the standard errors of these estimators. A bootstrap semiparametric inference procedure proposed in Wellner and Zhang (2007) is a common procedure in practice. However, this procedure often requires a substantial amount of computing effort. It is, therefore, desirable to develop methods that not only maintain good statistical properties but also are less computationally intensive.

The spline estimation of an unknown function such as a hazard function or a survival function has been studied by many researchers. For example, in the context of nonparametric estimation with right censored data, Kooperberg, Stone, and Truong (1995) reparameterized the likelihood function in terms of the hazard function and approximated the logarithm hazard function using polynomial splines. Although the nonnegativity of hazard function is guaranteed, their estimation procedure incurred additional computing effort when the estimate of survival function is desired. Lu, Zhang, and Huang (2007) studied nonparametric likelihood-based estimators of the mean function with panel count data using monotone polynomial splines. The spline likelihood estimators outperform the estimators proposed by Wellner and Zhang (2000) in terms of rate of convergence and mean square errors. Moreover, the spline likelihood estimators are much less computationally demanding than its alternative. These advantages motivate us to use monotone polynomial splines in semiparametric estimation for panel count data.

In this article, we use the monotone cubic B-splines (Schumaker, 1981) to directly approximate the logarithm of true baseline mean function $\log \Lambda_0(t)$, i.e.

$$\log \Lambda_0(t) \approx \sum_{j=1}^{q_n} \alpha_j B_j(t),$$

subject to $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{q_n}$. The monotonicity of the resulted spline function is guaranteed by imposing nondecreasing constraints on the coefficients $\alpha_j, j = 1, \ldots, q_n$ (Schumaker, 1981). Reparameterizaton of semiparametric proportional mean model (1)

$$E\{\mathbb{N}(t)|Z\} = \Lambda_0(t) \exp(\beta_0^T Z) \approx \exp\left\{\sum_{j=1}^{q_n} \alpha_j B_j(t) + \beta_0^T Z\right\}$$

leads to simultaneous estimation of the spline coefficients $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_{q_n})$ and regression parameters β_0. Since the number of basis B-splines, q_n, is often taken much smaller than the sample size, the dimension of the estimation problem is greatly reduced. Therefore, semiparametric spline estimations are expected to be less computationally expansive than the semiparametric estimators of Wellner and Zhang (2007). The generalized Rosen algorithm utilized by Zhang and Jamshidian (2004) for nonparametric estimation problems is modified for computing the estimates of $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_{q_n})$ and β_0 jointly.

The rest of the paper is organized as follows. Section 2 characterizes the spline pseudo-likelihood $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$ and spline likelihood estimators $(\hat{\beta}_n, \hat{\Lambda}_n)$ and describes the generalized Rosen algorithm for computing these estimators. Section 3 states the asymptotic properties of the spline estimators, including consistency, rate of convergence, and asymptotic normality. Section 4 carries out two sets of simulation studies to evaluate finite sample performance of the spline estimators and compare computational complexities between the spline estimators and their alternatives. Section 5 applies the spline methods to the bladder tumor example.

Section 6 summarizes our findings and discusses some related problems. Finally, the proofs of the asymptotic properties are given in the Appendix.

2. TWO LIKELIHOOD-BASED SPLINE SEMIPARAMETRIC METHODS

Let $\{\mathbb{N}(t) : t \geq 0\}$ be a univariate counting process with the conditional mean function given by (1), where Z is a time-independent covariate vector with distribution F on $\mathbb{R}^d$. K is the total number of observations on the counting process and $T = (T_{K,1}, \cdots, T_{K,K})$ is a sequence of random observation times with $0 < T_{K,1} < \ldots < T_{K,K}$. The cumulative numbers of recurrent events up to these times, $\mathbb{N} = \{\mathbb{N}(T_{K,1}), \ldots, \mathbb{N}(T_{K,K})\}$ with $0 \leq \mathbb{N}(T_{K,1}) \leq \cdots \leq \mathbb{N}(T_{K,K})$, are observed accordingly. We assume that (K, T_K) is conditionally independent of $\mathbb{N}$, given the covariate vector Z. The panel count data of the counting process consist of $X = (K, T, \mathbb{N}, Z)$. In the sequel, we denote the observed data consisting of independently and identically distributed random vectors by $X_1, \cdots, X_n$, where $X_i = (K_i, T_i, \mathbb{N}^{(i)}, Z_i)$ with $T_i = (T^{(i)}_{K_i,1}, \cdots, T^{(i)}_{K_i,K_i})$ and $\mathbb{N}^{(i)} = \{\mathbb{N}^{(i)}(T^{(i)}_{K_i,1}), \cdots, \mathbb{N}^{(i)}(T^{(i)}_{K_i,K_i})\}$, for $i = 1, \cdots, n$.

Zhang (2002) and Wellner and Zhang (2007) proposed a semiparametric pseudo-likelihood method for estimating (β_0, Λ_0) under a working assumption that the underlying counting process given the covariates is a nonhomogeneous Poisson process with the conditional mean function given by (1). By ignoring the dependence of the cumulative counts within a subject, they established the log pseudo-likelihood for (β, Λ),

$$l_n^{ps}(\beta, \Lambda) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Big[\mathbb{N}^{(i)}(T^{(i)}_{K_i,j}) \log \Lambda(T^{(i)}_{K_i,j}) + \mathbb{N}^{(i)}(T^{(i)}_{K_i,j}) \beta^T Z_i - \exp(\beta^T Z_i) \Lambda(T^{(i)}_{K_i,j}) \Big], \quad (2)$$

after omitting the additive terms that are independent of (β_0, Λ_0).

Assuming that $\mathbb{N}(t)$ is (conditionally, given Z) a nonhomogeneous Poisson process and using the conditional independence of the increments of $\mathbb{N}(t)$, Wellner and Zhang (2007) also established the log likelihood for (β, Λ),

$$l_n(\beta, \Lambda) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left[\Delta \mathbb{N}_{K_i,j}^{(i)} \log \Delta \Lambda_{K_i,j} + \Delta \mathbb{N}_{K_i,j}^{(i)} \beta^T Z_i - \exp(\beta^T Z_i) \Delta \Lambda_{K_i,j} \right], \tag{3}$$

where

$$\begin{aligned} \Delta \mathbb{N}_{K,j}^{(i)} &= \mathbb{N}^{(i)}(T_{K,j}^{(i)}) - \mathbb{N}^{(i)}(T_{K,j-1}^{(i)}), \\ \Delta \Lambda_{K,j} &= \Lambda(T_{K,j}^{(i)}) - \Lambda(T_{K,j-1}^{(i)}), \end{aligned}$$

for $j = 1, 2, \cdots, K$.

As described in Zhang (2002) and Wellner and Zhang (2007), the semiparametric maximum pseudo-likelihood/likelihood estimation can be implemented in two steps based on the profile likelihood method in which estimators of $\Lambda_0(t)$ are defined as nondecreasing step functions with jumps only possibly occurring at the observation times. It is easy to compute the semiparametric maximum pseudo-likelihood estimator, but the computation of the semiparametric maximum likelihood method is very time-consuming, especially when the sample is large. Regardless of the underlying counting process, both likelihood-based semiparametric approaches given in Wellner and Zhang (2007) are shown to be consistent and the asymptotic normality of the estimators of β_0 is established. The semiparametric maximum likelihood method in general tends to be more efficient than its pseudo-likelihood counterpart, even when the model is misspecified.

In this manuscript, we propose to estimate the baseline mean function using a monotone polynomial spline instead of the step function, in order to alleviate the computation burden and to achieve better rate of convergence in estimating the baseline function. The approach follows the idea of method of sieves in nonparametric maximum likelihood estimation, originally proposed by Geman and Hwang (1982).

For a finite closed interval $[a, b]$, let $\mathcal{T} = \{t_i\}_1^{m_n+2l}$, with

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

be a sequence of knots that partition $[a, b]$ into $m_n + 1$ subintervals $J_i = [t_{l+i}, t_{l+i+1}]$, for $i = 0, \ldots, m_n$. Denote $\varphi_{l,t}$ the class of polynomial splines of order $l \geq 1$ with the knot sequence $\mathcal{T}$. The class $\varphi_{l,t}$ can be linearly spanned by the B-spline basis functions $\{B_i, 1 \leq i \leq q_n\}$ with $q_n = m_n + l$ (Schumaker, 1981). Now, we define a subclass of $\varphi_{l,t}$,

$$\psi_{l,t} = \left\{ \sum_{i=1}^{q_n} \alpha_i B_i : \alpha_1 \leq \cdots \leq \alpha_{q_n} \right\}.$$

According to Theorem 5.9 of Schumaker (1981), $\psi_{l,t}$ is a class of monotone nondecreasing splines since the monotonicity of the B-splines is guaranteed by the nondecreasing order of coefficients. We approximate the logarithm of smooth monotone baseline mean function $\log \Lambda_0(t)$ by $\sum_{i=1}^{q_n} \alpha_i B_i(t)$ and estimate the coefficients $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{q_n})$ and regression parameters β jointly through maximizing the approximated pseudo-likelihood and likelihood subject to nondecreasing constraints, respectively.

Let $\hat{\alpha}^{ps} = (\hat{\alpha}_1^{ps}, \hat{\alpha}_1^{ps}, \cdots, \hat{\alpha}_{q_n}^{ps})$ and $\hat{\beta}_n^{ps}$ be the values that maximize spline pseudo-likelihood,

$$l_n^{ps}(\beta, \alpha) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left[\mathbb{N}^{(i)}(T_{K_i,j}^{(i)}) \beta^T Z_i + \mathbb{N}^{(i)}(T_{K_i,j}^{(i)}) \sum_{l=1}^{q_n} \alpha_l B_l(T_{K_i,j}^{(i)}) - \exp\left\{ \beta^T Z_i + \sum_{l=1}^{q_n} \alpha_l B_l(T_{K_i,j}^{(i)}) \right\} \right], \tag{4}$$

under constraints $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{q_n}$.

Similarly, let $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_1, \cdots, \hat{\alpha}_{q_n})$ and $\hat{\beta}_n$ be the values that maximize spline likelihood,

$$l_n(\beta, \alpha) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left[\Delta \mathbb{N}_{K_i,j}^{(i)} \log \Delta \Lambda_{K_i,j} + \Delta \mathbb{N}_{K_i,j}^{(i)} \beta^T Z_i - \exp(\beta^T Z_i) \Delta \Lambda_{K_i,j} \right], \tag{5}$$

where

$$\Delta \Lambda_{K_i,j} = \exp\left(\sum_{l=1}^{q_n} \alpha_l B_l(T_{K_i,j}^{(i)}) \right) - \exp\left(\sum_{l=1}^{q_n} \alpha_l B_l(T_{K_i,j-1}^{(i)}) \right),$$

subject to the same constraints as above. We denote the spline pseudo-likelihood estimator of $\Lambda_0(t)$ by $\hat{\Lambda}_n^{ps}(t) = \exp(\sum_{j=1}^{q_n} \hat{\alpha}_j^{ps} B_j(t))$ and the spline likelihood estimator of $\Lambda_0(t)$ by $\hat{\Lambda}_n(t) = \exp(\sum_{j=1}^{q_n} \hat{\alpha}_j B_j(t))$, respectively.

We note that both the spline pseudo-likelihood and spline likelihood functions are concave with respect to the unknown coefficients $\alpha_1, \alpha_2. \cdots, \alpha_{q_n}$ and regression parameters β. So the spline likelihood-based estimation problem is equivalent to a nonlinear convex programming problem subject to linear inequality constraints. Specifically, the spline estimation problems (4) and (5) can be formulated as the linear inequality constrained maximization problem

$$\max_{\theta \in \Theta_\alpha \times \mathbb{R}^d} l(\theta | X), \tag{6}$$

where $\theta = (\alpha, \beta)$ with $\alpha \in \Theta_\alpha = \{\alpha : \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{q_n}\}$. Jamshidian (2004) proposed a generalized gradient projection method for optimizing a nonlinear objective function with linear inequality constraints, based on the generalized Euclidean metric $\|x\| = x^T W x$ with W being a positive definite matrix and possibly varying from iteration to iteration. This method essentially generalizes the algorithm proposed by Rosen (1960) and will be referred to as the generalized Rosen algorithm throughout this manuscript. Zhang and Jamshidian (2004) applied the generalized Rosen algorithm to large-scale nonparametric maximum likelihood estimation problems by choosing $W = D_H$, the matrix containing only the diagonal elements of the negative Hessian matrix H, in order to avoid the storage problem in updating H. However, this will increase the number of iterations and thereby the computing time.

We use $W = H$ directly because the dimension of unknown parameter space is usually small in our applications due to the use of polynomial splines. The use of the full Hessian matrix substantially reduces the number of iterations. As a result, the spline estimators are expected to be much less computationally demanding than their alternatives proposed by Wellner and Zhang (2007). We now describe the algorithm that was used in computing the proposed spline estimators.

Let $\dot{\ell}(\theta)$ and W be the gradient and negative Hessian matrix of the log pseudo-likelihood or log likelihood with respect to θ, respectively. Let $\mathcal{A} = \{i_1, i_2, \cdots, i_m\}$ denote the index set of active constraints, i.e. $\alpha_{i_j} = \alpha_{i_j+1}$, for $j = 1, 2, \cdots, m$, during the numerical computation. We define a working matrix corresponding to this set, given as follows:

$$A = \begin{bmatrix} 0 & \cdots & -1 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & -1 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 & \cdots & 0 \end{bmatrix}_{m \times (q_n + d)},$$

where d is the dimension of regression parameter β. The generalized Rosen algorithm is implemented in the following steps.

S0: (**Computing the feasible search direction**)

$$\underline{d} = \left(I - W^{-1} A^T \left(A W^{-1} A^T\right)^{-1} A\right) W^{-1} \dot{\ell}(\theta).$$

S1: (**Forcing the updated θ fulfill the constraints**) If the resulted direction $\underline{d}$ is not nondecreasing in its components, compute

$$\gamma = \min_{i \notin \mathcal{A} \text{ and } d_i > d_{i+1}} \left(-\frac{\alpha_{i+1} - \alpha_i}{d_{i+1} - d_i} \right).$$

Doing so guarantees that $\alpha_{i+1} + \gamma d_{i+1} \geq \alpha_i + \gamma d_i$, for $i = 1, 2, \cdots, q_n$.

S2: (**Step-Halving line search**) Looking for a smallest integer k starting from 0 such that

$$\ell\left(\theta + (1/2)^k \underline{d}\right) > \ell(\theta).$$

S3: (**Updating the solution**) If $\gamma > (1/2)^k$, replace θ by $\tilde{\theta} = \theta + (1/2)^k \underline{d}$ and check the stopping criterion (S5).

S4: (**Updating the active constraint set**) If $\gamma \leq (1/2)^k$, in addition to replace θ by $\tilde{\theta} = \theta + \gamma \underline{d}$, 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 $\|\underline{d}\| \geq \varepsilon$ for a small $\varepsilon > 0$, go to S0. Otherwise, compute $\lambda = \left(AW^{-1}A^T\right)^{-1} AW^{-1}\dot{\ell}(\theta)$.

i. If $\lambda_i \leq 0$ for all $i \in \mathcal{A}$, set $\hat{\theta} = \theta$ 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 A and go to S0.

To initialize the algorithm, we choose $\alpha = (0, 0, 1, \cdots, q_n - 2)^T_{1\times q_n}$ and $\beta = (0, \cdots, 0)^T_{1\times d}$. With this choice, $A = (-1, 1, 0, \cdots, 0)_{1\times(q_n+d)}$. In our experience, the active constraint set is usually identified in the first few iterations and no updates for A are needed thereafter.

3. ASYMPTOTIC RESULTS

In this section, we study the asymptotic properties of the spline pseudo-likelihood estimator $(\hat{\beta}^{ps}_n, \hat{\Lambda}^{ps}_n)$ and the spline likelihood estimator $(\hat{\beta}_n, \hat{\Lambda}_n)$. Let $\mathcal{B}_d$ and $\mathcal{B}$ denote the collection of Borel sets in $\mathbb{R}^d$ and $\mathbb{R}$, respectively, and let $\mathcal{B}_{[0,\tau]} = \{B \cap [0,\tau] : B \in \mathcal{B}\}$. Following Wellner and Zhang (2007), define the measures $\mu_1, \mu_2, \nu_1, \nu_2$, and γ as follows: for $B, B_1, B_2 \in \mathcal{B}_{[0,\tau]}$, and $C \in \mathcal{B}_d$,

$$\begin{aligned}
\nu_1(B \times C) &= \int_C \sum_{k=1}^{\infty} P(K=k|Z=z) \sum_{j=1}^{k} P(T_{k,j} \in B | K=k, Z=z) dF(z), \\
\mu_1(B) &= \nu_1(B \times \mathbb{R}^d), \\
\nu_2(B_1 \times B_2 \times C) &= \int_C \sum_{k=1}^{\infty} P(K=k|Z=z) \sum_{j=1}^{k} P(T_{k,j-1} \in B_1, T_{k,j} \in B_2 | K=k, Z=z) dF(z), \\
\mu_2(B_1 \times B_2) &= \nu_2(B_1 \times B_2 \times \mathbb{R}^d), \\
\gamma(B) &= \int_{\mathbb{R}^d} \sum_{k=1}^{\infty} P(K=k|Z=z) P(T_{k,k} \in B | K=k, Z=z) dF(z).
\end{aligned}$$

We study the consistency and the rate of convergence in the L_2-metrics d_1 and d_2, given by

$$\begin{aligned}
d_1(\vartheta_1, \vartheta_2) &= \left\{ |\beta_2 - \beta_1|^2 + \int |\Lambda_2(t) - \Lambda_1(t)|^2 d\mu_1(t) \right\}^{1/2}, \\
d_2(\vartheta_1, \vartheta_2) &= \left\{ |\beta_2 - \beta_1|^2 + \int\int |\Lambda_1(u) - \Lambda_1(v) - (\Lambda_2(u) - \Lambda_2(v))|^2 d\mu_2(u,v) \right\}^{1/2},
\end{aligned}$$

where $\vartheta_i = (\beta_i, \Lambda_i)$, for i =1 and 2, with $\Lambda_1, \Lambda_2 \in \mathcal{F} = \{\Lambda : \Lambda$ is monotone nondecreasing, $\Lambda(0) = 0\}$.

Let

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

be a sequence of knots with $m_n = O(n^\nu)$, for $0 < \nu < 1/2$. To study the asymptotic properties of the spline estimators, we need to allocate the knots properly and assume the smoothness of the true baseline mean function.

C1: The maximum spacing of the knots, $\Delta \equiv \max_{l+1 \le i \le m_n+l+1} |t_i - t_{i-1}| = O(n^{-\nu})$. Moreover, there exists a constant $M > 0$ such that $\Delta/\delta \leq M$ uniformly in n, where $\delta = \min_{l+1 \leq i \leq m_n+l+1} |t_i - t_{i-1}|$.

C2: The true baseline mean function Λ_0 has a bounded rth derivative in $[0, \tau]$ with $r \geq 1$. Moreover, the first derivative has a positive lower bound in $[0, \tau]$. That is, there exists a constant $C_0 > 0$ such that $\Lambda_0'(t) \geq C_0$, for $t \in [0, \tau]$.

Some regularity conditions for observation schemes and underlying counting process provided in Wellner and Zhang (2007) are also needed in this project.

C3: The parameter space of β, $\mathcal{R}$, is bounded and convex on $\mathbb{R}^d$ and the true parameter $\vartheta_0 = (\beta_0, \Lambda_0) \in \mathcal{R}^o \times \mathcal{F}$, where $\mathcal{R}^o$ is the interior of $\mathcal{R}$.

C4: The measure $\mu_i \times F$ is absolutely continuous with respect to ν_i, for $i = 1, 2$, and $E(K) < \infty$.

C5: There exists a z_0 such that $P(|Z| \leq z_0) = 1$. That is, the covariate vector is uniformly bounded.

C6: For all $a \in \mathbb{R}^d, a \neq 0$, and $c \in \mathbb{R}, P(a^T Z \neq c) > 0$.

C7: (a) The function $M_0^{ps}(X) = \sum_{j=1}^{K} \mathbb{N}_{Kj} \log(\mathbb{N}_{Kj})$ satisfies $PM_0^{ps}(X) < \infty$. (b) The function $M_0(X) = \sum_{j=1}^{K} \Delta\mathbb{N}_{Kj} \log(\Delta\mathbb{N}_{Kj})$ satisfies $PM_0(X) < \infty$.

C8: There exists a positive integer M_0 such that $P(K \leq M_0) = 1$. That is, the number of observations is finite.

C9: $E\left\{e^{C\mathbb{N}(t)}\right\}$ is uniformly bounded in $[0, \tau]$. The τ can be viewed as the termination time in a follow-up study.

C10: The observation time points are s_0-separated. That is, there exists a constant s_0 such that $P(T_{K,j} - T_{K,j-1} \geq s_0) = 1$, for all $j = 1, 2, \cdots, K$.

Remark 1. C1 is similar to those required by Stone (1986) and Zhou, Shen, and Wolfe (1998). C2 is standard in the literature of nonparametric smoothing estimation. C3 is a common assumption in semiparametric estimation problems. C4 and C6 are needed to establish the identifiability of the semiparametric model. The conditions related to the observation schemes, C5, C8 and C10 are mild and are easily justified in many applications. C9 holds if the underlying counting process is uniformly bounded which is often true in real life problems, or if it is a Poisson or mixed Poisson process conditional on covariates.

Remark 2. C7(a) and C7(b) are used in the proof of consistency for the pseudo-likelihood estimator $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$ and the likelihood estimator $(\hat{\beta}_n, \hat{\Lambda}_n)$, respectively. C10 is only needed for likelihood estimator.

Theorem 1 (Consistency). Suppose that C1 - C9 hold and the counting process $\mathbb{N}$ satisfies the proportional mean regression model (1). Then, for every $0 < b < \tau$ for which $\mu_1([b, \tau]) > 0$,

$$\lim_{n\to\infty} d_1((\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps} 1_{[0,b]}), (\beta_0, \Lambda_0 1_{[0,b]})) = 0,$$

in probability. In particular, if $\mu_1(\{\tau\}) > 0$, then

$$\lim_{n\to\infty} d_1((\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}), (\beta_0, \Lambda_0)) = 0,$$

in probability. In addition, if C10 holds, then, for every $0 < b < \tau$ for which $\gamma([b, \tau]) > 0$,

$$\lim_{n\to\infty} d_2((\hat{\beta}_n, \hat{\Lambda}_n 1_{[0,b]}), (\beta_0, \Lambda_0 1_{[0,b]})) = 0,$$

in probability. In particular, if $\gamma(\{\tau\}) > 0$, then

$$\lim_{n\to\infty} d_2((\hat{\beta}_n, \hat{\Lambda}_n), (\beta_0, \Lambda_0)) = 0,$$

in probability.

Remark 3. The metrics d_1 and d_2 are closely related. By Lemma 8.1 of Wellner and Zhang (2000), the two metrics are equivalent under C8, and therefore the consistency and rate of convergence results (shown below) for maximum likelihood estimator $(\hat{\beta}_n, \hat{\Lambda}_n)$ hold for metric d_1 as well.

To derive the rate of convergence and asymptotic normality of the estimators of regression parameters, we need to assume the following three additional conditions.

C11: There exists an interval $O[T] = [\sigma, \tau]$ with $\sigma > 0$ such that $\Lambda_0(\sigma) > 0$ and $P\left(\cap_{j=1}^{K}\{T_{Kj} \in [\sigma, \tau]\}\right) = 1$.

C12: There exists $\eta \in (0, 1)$ such that $a^T Var(Z|U)a \geq \eta a^T E(ZZ^T|U)a$ a.s., for all $a \in \mathbb{R}^d$.

C13: There exists $\eta \in (0, 1)$ such that $a^T Var(Z|U, V)a \geq \eta a^T E(ZZ^T|U, V)a$ a.s., for all $a \in \mathbb{R}^d$.

Theorem 2 (Rate of Convergence). In addition to C1-C9, suppose C11 and C12 hold. If ν is chosen to be $1/(1 + 2r)$, then

$$n^{r/(1+2r)} d_1((\hat{\Lambda}_n^{ps}, \hat{\beta}_n^{ps}), (\Lambda_0, \beta_0)) = O_p(1).$$

Moreover, under C1-C11 and C13, it follows that

$$n^{r/(1+2r)} d_2((\hat{\Lambda}_n, \hat{\beta}_n), (\Lambda_0, \beta_0)) = O_p(1).$$

Remark 4. The justifications for these three additional conditions were given in Wellner and Zhang (2007). Theorem 2 indicates that the spline estimators can have a higher rate of convergence than the semiparametric estimators studied in Wellner and Zhang (2007), if the baseline mean function is sufficiently smooth, because $r/(1+2r) \geq 1/3$ when $r \geq 1$.

Theorem 3 (Asymptotic Normality). Under the conditions listed in Theorem 2, the estimators $\hat{\beta}_n^{ps}$ and $\hat{\beta}_n$ are asymptotically normal and

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_n^{ps} - \beta_0) &\to_d N(0, (A^{ps})^{-1} B^{ps} (A^{psT})^{-1}) \equiv_d N(0, \Sigma^{ps}), \\
\sqrt{n}(\hat{\beta}_n - \beta_0) &\to_d N(0, A^{-1} B (A^T)^{-1}) \equiv_d N(0, \Sigma),
\end{aligned}$$

where

$$\begin{aligned}
A^{ps} &= E\left\{ \sum_{j=1}^{K} \Lambda_0(T_{Kj}) e^{\beta_0^T Z} \left[Z - \frac{E(Z e^{\beta_0^T Z} | K, T_{K,j})}{E(e^{\beta_0^T Z} | K, T_{K,j})} \right]^{\otimes 2} \right\}, \\
B^{ps} &= E\left\{ \sum_{j,j'=1}^{K} C_{j,j'}^{ps}(Z) \left[Z - \frac{E(Z e^{\beta_0^T Z} | K, T_{K,j})}{E(e^{\beta_0^T Z} | K, T_{K,j})} \right] \left[Z - \frac{E(Z e^{\beta_0^T Z} | K, T_{K,j'})}{E(e^{\beta_0^T Z} | K, T_{K,j'})} \right]^T \right\}, \\
A &= E\left\{ \sum_{j=1}^{K} \Delta\Lambda_0(T_{Kj}) e^{\beta_0^T Z} \left[Z - \frac{E(Z e^{\beta_0^T Z} | K, T_{K,j-1}, T_{K,j})}{E(e^{\beta_0^T Z} | K, T_{K,j-1}, T_{K,j})} \right]^{\otimes 2} \right\}, \\
B &= E\left\{ \sum_{j,j'=1}^{K} C_{j,j'}(Z) \left[Z - \frac{E(Z e^{\beta_0^T Z} | K, T_{K,j}, T_{K,j'})}{E(e^{\beta_0^T Z} | K, T_{K,j}, T_{K,j'})} \right]^{\otimes 2} \right\},
\end{aligned}$$

with

$$C_{j,j'}^{ps}(Z) = Cov(\mathbb{N}(T_{Kj}), \mathbb{N}(T_{Kj'}) | Z, K, T_{K,j}, T_{K,j'})$$

and

$$C_{j,j'}(Z) = Cov(\Delta\mathbb{N}(T_{Kj}), \Delta\mathbb{N}(T_{Kj'}) | Z, K, T_K).$$

Remark 5. Although the overall rate of convergence for the pseudo-likelihood estimator $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$ and likelihood estimator $(\hat{\beta}_n, \hat{\Lambda}_n)$ are $n^{r/(1+2r)} < n^{1/2}$, the rate of convergence for $\hat{\beta}_n^{ps}$ and $\hat{\beta}_n$ are still $n^{1/2}$. Theorem 3 shows that the spline estimators of β_0 and their alternatives proposed by Wellner and Zhang (2007) are asymptotically equivalent. The proofs of these theorems are sketched in the Appendix.

4. SIMULATION STUDIES

Two Monte Carlo simulation studies are carried out to compare the statistical properties and computational complexity between the spline estimators and their alternatives studied in Wellner and Zhang (2007), and to demonstrate the robustness of the proposed methods as well. The data are simulated in the same manner as in Zhang (2002). In each simulation study, we generate n independently and identically distributed observations $\{(K_i, T_i, \mathbb{N}^{(i)}, Z_i) : i = 1, 2, \ldots, n\}$ with $Z_i = (Z_{i1}, Z_{i2}, Z_{i3})$. For each subject i, the data are generated by the following schemes: $Z_{i1} \sim$ Uniform(0,1), $Z_{i2} \sim N(0,1)$, and $Z_{i3} \sim$ Bernoulli(0.5); K_i is randomly sampled from a discrete uniform distribution $\{1, 2, 3, 4, 5, 6\}$; Given K_i, the random panel observation times $T_i = (T^{(i)}_{K_i,1}, \ldots, T^{(i)}_{K_i,K_i})$ are K_i ordered random draws from Uniform(0,10) and rounded to the second decimal point to make the observation times possibly tied. The two simulations differ in the methods of generating the panel counts $\mathbb{N}^{(i)} = \{\mathbb{N}^{(i)}(T^{(i)}_{K_i,1}), \ldots, \mathbb{N}^{(i)}(T^{(i)}_{K_i,K_i})\}$, given (K_i, T_i). They are described as follows:

Simulation 1. The panel counts are generated from a Poison process with conditional mean function $E(\mathbb{N}(t)|Z) = 2te^{\beta_0^T Z}$. That is,

$$\mathbb{N}^{(i)}(T^{(i)}_{K_i,j}) - \mathbb{N}^{(i)}(T^{(i)}_{K_i,j-1}) \sim \text{Poisson}\{2(T^{(i)}_{K_i,j} - T^{(i)}_{K_i,j-1}) \exp(\beta_0^T Z_i)\},$$

for $j = 1, 2, \ldots, K_i$, where $\beta_0 = (\beta_1, \beta_2, \beta_3)^T = (-1.0, 0.5, 1.5)^T$. Under this simulation setting, we can directly compute the asymptotic covariance matrices stated in Theorem 3,

$$\Sigma^{ps} = (A^{ps})^{-1} B^{ps} ((A^{ps})^{-1})^T = \frac{29}{315} \Omega^{-1}$$

and

$$\Sigma = A^{-1}B(A^{-1})^T = A^{-1} = \frac{42}{617}\Omega^{-1},$$

where $\Omega = E\{e^{\beta_0^T Z}[Z - E(Ze^{\beta_0^T Z})/E(e^{\beta_0^T Z})]^{\otimes 2}\}$. A direct calculation of Ω results in

$$\Sigma^{ps} = \begin{pmatrix} 0.591151 & 0.000000 & 0.000000 \\ 0.000000 & 0.046894 & 0.000000 \\ 0.000000 & 0.000000 & 0.314416 \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} 0.437094 & 0.000000 & 0.000000 \\ 0.000000 & 0.034673 & 0.000000 \\ 0.000000 & 0.000000 & 0.232478 \end{pmatrix}.$$

Simulation 2. The panel counts are generated from a mixed Poisson process. We first generate a random sample $\gamma_1, \gamma_2, \ldots, \gamma_n \sim \{-0.4, 0, 0.4\}$ with $\mathrm{pr}(\gamma_i = -0.4) = \mathrm{pr}(\gamma_i = 0.4) = 1/4$ and $\mathrm{pr}(\gamma_i = 0) = 1/2$, for $i = 1, 2, \ldots, n$. Given γ_i, the panel counts for the ith subject are generated according to Poisson$\{(2+\gamma_i)t\exp(\beta_0^T Z_i)\}$. That is,

$$\mathbb{N}^{(i)}(T_{K_i,j}^{(i)}) - \mathbb{N}^{(i)}(T_{K_i,j-1}^{(i)})|\gamma_i \sim \text{Poisson}\{(2+\gamma_i)(T_{K_i,j}^{(i)} - T_{K_i,j-1}^{(i)})\exp(\beta_0^T Z_i)\},$$

for $j = 1, 2, \ldots, K_i$. This counting process given only the covariates is not a Poisson process. However, the conditional mean given covariates still satisfy the proportional mean model (1) with $\Lambda_0(t) = 2t$ and thus the proposed method is expected to be valid in this scenario as well. The asymptotic covariance matrices in Theorem 3 for this simulation setting are given by

$$\Sigma^{ps} = (A^{ps})^{-1}B^{ps}((A^{ps})^{-1})^T = \frac{29}{315}\Omega^{-1} + \frac{2}{75}\Omega^{-1}\tilde{\Omega}(\Omega^{-1})^T$$

and

$$\Sigma = A^{-1}B(A^{-1})^T = A^{-1} = \frac{42}{617}\Omega^{-1} + \frac{8383.2}{617^2}\Omega^{-1}\tilde{\Omega}(\Omega^{-1})^T,$$

where $\tilde{\Omega} = E\{e^{2\beta_0^T Z}[Z - E(Ze^{\beta_0^T Z})/E(e^{\beta_0^T Z})]^{\otimes 2}\}$. A direct calculation of $\tilde{\Omega}$ along with the Ω

calculated above leads to

$$\Sigma^{ps} = \begin{pmatrix} 1.207044 & -0.024428 & -0.044223 \\ -0.024428 & 0.111885 & 0.023531 \\ -0.044223 & 0.023531 & 0.462607 \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} 0.945694 & -0.020173 & -0.036519 \\ -0.020173 & 0.088343 & 0.019432 \\ -0.036519 & 0.019432 & 0.354853 \end{pmatrix}.$$

In these two simulations, the cubic B-splines are used in computing the spline estimators. Let $T_{\min}$ and $T_{\max}$ be the respective minimum and maximum values of the collection of total distinct observation times in the data. The interval $[T_{\min}, T_{\max}]$ is equally divided into m_n+1 subintervals, in which m_n is selected as the cubic root of the number of distinct observation times plus 1. Hence, the spacing of the knots $\Delta_i = t_i - t_{i-1}$ is proportional to $n^{-1/3}$, for $i = l+1, l+2, \ldots, m_n+l+1$, and C1 listed in Section 3 automatically satisfies. In our studies, we generate 1000 Monte Carlo samples with $n = 50$ and 100, respectively, for each scenario.

To compare the estimators for $\Lambda_0(t)$ in detail, we calculate the estimates of $\Lambda_0(t)$ at the time-points $t = 1.5, 2.0, 2.5, \ldots, 9.5$. For Simulation 1, the square of pointwise biases and the pointwise mean squared errors at these time-points for $n = 50$ and 100, respectively, are plotted in Fig.1. From the graph, we can see that the biases of the four estimators are clearly negligible compared to the mean squared errors, and the pointwise mean squared errors of the spline estimators are smaller than their alternatives. The spline likelihood estimator appears to be the most efficient one among the four. When sample size doubles, the pointwise mean squared errors drop substantially, which indicates the consistency of these estimators. The results of the Monte Carlo study for the regression parameters are summarized in Table 1, we find that the biases for all estimators of β_0 are small, and the Monte Carlo standard deviations of the spline estimators are almost identical to those of their alternatives, which are consistent with the asymptotic results given in Theorem 3. Similar to their alternatives proposed in Wellner and Zhang (2007), the spline likelihood estimators of β_0 are more efficient than the spline pseudo-likelihood estimators. Moreover, the standard errors calculated based on Theorem 3 are close to the Monte Carlo standard deviations for all estimators and hence the use of asymptotic results in applications with moderate sample size is justified.

For Simulation 2, the finite-sample study is similarly conducted. The results are displayed in Table 2 and Fig.2. The same patterns as those in Simulation 1 are observed. The Monte Carlo standard deviations of the likelihood estimators of both $\Lambda_0(t)$ and β_0 are smaller than those of the pseudo-likelihood estimators, and the biases for all estimators are negligible. Again, the spline estimators for Λ_0 have smaller mean square errors than their alternatives studied in Wellner and Zhang (2007), and the spline estimators of β_0 behave the same as their alternatives asymptotically. The standard errors calculated based on the asymptotic theory are all close to the Monte Carlo standard deviations. This simulation also reinforces the conclusion made in Wellner and Zhang (2007) that the likelihood method based on Poisson process is robust again the underlying counting process. However, the mean squared errors of the estimators of both Λ_0 and β_0 increase when the Poisson process model is misspecified for the true underlying counting process.

We also compare the computing time among the four estimators and summarize the results in Table 3. The computational advantage of the spline estimators over their alternatives in Wellner and Zhang (2007) is remarkable, especially for the case of likelihood estimation. This advantage makes the bootstrap procedure for estimating the standard errors of the spline-based maximum likelihood estimates of β_0 feasible in practice.

To assess the inference performance, the coverage probabilities of 95% confidence intervals for β_0 with both Poisson and mixed Poisson processes are obtained with 1000 Monte Carlo samples. For each Monte Carlo sample, the standard errors of the estimates of β_0 are estimated by the bootstrap standard deviation with 1000 replications, and then the Wald-95% confidence intervals are constructed. Table 4 exhibits the right coverage probabilities of 95% confidence intervals with sample 50 and 100, respectively. Although we used 1000 replications in bootstrap for this study, we actually found out that the bootstrap method with only 100 replications yields reasonable estimates of the standard errors for this application. This finding supports the statement about the number of bootstrap samples needed to yield a valid estimate of standard error given by Efron and Tibshirani (1993).

As a concluding remark, the proposed spline semiparametric bootstrap inference procedure is a sound practical method for applications with moderate sample size.

5. A REAL EXAMPLE: BLADDER TUMOR TRAIL

The proposed methods are illustrated using the bladder tumor data described in the introduction (Andrews and Herzberg, 1985, p.250-60). In this randomized clinical trial, a total of 116 patients were randomly assigned into one of three treatment groups, 40 to placebo, 31 to pyridoxine, and 38 to thiotepa. The number of follow-ups and follow-up times varied greatly from patient to patient.

To investigate the efficacy of the two treatments (pyridoxine pill and thiotepa installation) on suppressing the recurrences of bladder tumors, we use the proportional mean model proposed by Sun and Wei (2000) and Wellner and Zhang (2007),

$$E\{\mathbb{N}(t)|Z\} = \Lambda_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4),$$

where Z_1 and Z_2 represent number and largest size of bladder tumors at beginning of the trial; and Z_3 and Z_4 are the indicators for pyridoxine pill and thiotepa installation, respectively.

For this example only the spline estimators of regression parameters are calculated. The bootstrap procedure is applied to estimate the standard errors. The results of inference based on 1000 bootstrap samples are summarized in Table 5.

Both the spline pseudo-likelihood method and spline likelihood method yield the same conclusion that the individuals with more tumors at baseline tend to have more recurrent tumors with p-values 0.012 and 0.01, respectively. On average, for each additional tumor at the baseline, the number of recurrent tumors will increase by 15.6% or 23% using the pseudo-likelihood method or the likelihood method, respectively. The thiotepa treatment significantly suppresses the recurrence of tumors with p-values 0.015 and 0.019 based on the pseudo-likelihood method and likelihood method, respectively. On average, the number of recurrent tumors for thiotepa group is 50% or 45% of that for control group by the pseudo-likelihood method or likelihood method, adjusting for the baseline tumor number and size. These conclusions are consistent with those made in Wellner and Zhang (2007).

6. FINAL REMARKS AND FURTHER PROBLEMS

The monotone spline semiparametric methods have the advantage over the methods proposed by Wellner and Zhang (2007) in terms of the computing efficiency and the convergence rate of the estimators of the infinite dimensional parameter $\Lambda_0(t)$. Meanwhile, the estimates of regression parameters β_0 in the spline methods have the same asymptotic distributions as their alternatives proposed in Wellner and Zhang (2007). Furthermore, the ease of computing spline estimators makes the statistical inference based on the bootstrap procedure feasible in practice. The proposed spline method provides a practical approach for semiparametric regression analysis with panel count in which joint estimation of the nonparametric component and parametric regression parameters is a challenging task.

We have used the pre-specified partition for monotone polynomial splines. It would be preferable to adaptively select the number and spacings of the knots such as using the penalized likelihood method. One may also explore other models such as additive and additive-multiplicative mean model instead of proportional mean model (1). In this manuscript, we assume the number of observation and observation times (K, T) given covariates Z are independent of the underlying counting process $\mathbb{N}$. This assumption may not be realistic in some applications, since patients with rapid disease progression may tend to visit the clinics more often. Therefore, it is worthwhile to extend our methods to such applications.

APPENDIX: PROOFS AND TECHNICAL DETAILS

The proofs for asymptotic results are sketched here. The empirical process theory is the major technical tool to prove the asymptotic results. The notations used in this section follow those given in van der Vaart and Wellner (1996), Huang (1996, 1999), and Wellner and Zhang (2007). Here we only sketch the proofs for the spline pseudo-likelihood estimator, since the proofs for the spline likelihood estimator are basically parallel.

Let $N_{[\,]}(\epsilon, \phi_{l,t}, L_2(\mu_i)), i = 1, 2$, be the bracketing number of ε-brackets with metric $L_2(\mu_i)$ needed to cover the class $\phi_{l,t}$ which is defined in van der Vaart and Wellenr (1996). Using the bracketing number theorem developed in Shen and Wong (1994, p.597), we have the following technical lemma which will be used extensively in our proofs.

Lemma A1. Assume $\psi_{l,t}$ is the set of all monotone polynomial splines with order l and sequence $\mathcal{T} = \{t_i\}_1^{m_n+2l}$. Then, for any $\eta > 0$ and $\varepsilon \le \eta$,

$$\log N_{[\ ]}(\varepsilon, \psi_{l,t}, L_2(\mu_i)) \le c q_n \log(\eta/\varepsilon),$$

for $i = 1, 2$ and a positive constant c, where $q_n = m_n + l$ is the number of spline basis functions.

Proof of Theorem 1 (Consistency)

Let $X_i = (K_i, T_i, \mathbb{N}^{(i)}, Z_i),\ i = 1, 2, \cdots, n,$ be n independently and identically distributed copies of $X = (K, T, \mathbb{N}, Z)$. The pseudo-likelihood for $\vartheta = (\beta, \Lambda)$ can be rewritten as

$$m_n^{ps}(\vartheta) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left[\mathbb{N}^{(i)}(T_{K_i,j}^{(i)})\log\{\Lambda(T_{K_i,j}^{(i)})\exp(\beta^T Z_i)\} - \Lambda(T_{K_i,j}^{(i)})\exp(\beta^T Z_i)\right].$$

Let $\mathbb{M}_n^{ps}(\vartheta) = \mathbb{P}_n m_\vartheta^{ps}(X) = \frac{1}{n} m_n^{ps}(\vartheta)$ and $\mathbb{M}^{ps}(\vartheta) = P m_\vartheta^{ps}(X)$, where

$$m_\vartheta^{ps}(X) = \sum_{j=1}^{K}\{\mathbb{N}(T_{K,j})\log\Lambda(T_{K,j}) + \mathbb{N}(T_{K,j})\beta^T Z - \Lambda(T_{K,j})\exp(\beta^T Z)\}.$$

Wellner and Zhang (2007) showed that $\mathbb{M}^{ps}(\vartheta_0) \ge \mathbb{M}^{ps}(\vartheta)$ and $\mathbb{M}^{ps}(\vartheta_0) = \mathbb{M}^{ps}(\vartheta)$ if and only if $\beta = \beta_0$ and $\Lambda(u) = \Lambda_0(u)$ a.e. with respect to μ_1 for true parameters $\vartheta_0 = (\beta_0, \Lambda_0)$ by C2 and C6.

Let $\mathcal{T} = \{t_i\}_{i=1}^{m_n+2l}$ with

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

be a sequence of knots with $q_n = m_n + l = O(n^\nu)$, for $0 < \nu < 1/2$. There exists a monotone spline $\Lambda_n \in \psi_{l,t}$ with order $l \ge r + 2$ and knots $\mathcal{T}$ such that $\|\Lambda_n - \Lambda_0\|_\infty = \sup_{t\in[\sigma,\tau]} |\Lambda_n(t) - \Lambda_0(t)| = O(n^{-\nu r})$ by Lemma A1 of Lu, Zhang, and Huang (2007).

Dominated Convergence Theorem and C7(a) yield that $\mathbb{M}^{ps}(\vartheta)$ is continuous in ϑ. Therefore, for any arbitrary $\varepsilon > 0$, there exists $\Lambda_0^* \in \psi_{l,t}$ such that $\mathbb{M}^{ps}(\beta_0, \Lambda_0) - \varepsilon \leq \mathbb{M}^{ps}(\beta_0, \Lambda_0^*)$ with $\|\Lambda_0 - \Lambda_0^*\|_\infty = o(1)$.

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

Lemma A1 guarantees that $\psi_{l,t}$ is compact. Then by Helly-Selection Theorem and compactness of $\mathcal{R} \times \psi_{l,t}$, it concludes that $\hat{\vartheta}_n^{ps} = (\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$ has a subsequence $\hat{\vartheta}_{n_k}^{ps} = (\hat{\beta}_{n_k}^{ps}, \hat{\Lambda}_{n_k}^{ps})$ converging to $\vartheta^+ = (\beta^+, \Lambda^+)$, where Λ^+ is a nondecreasing function on $[\sigma, \tau]$.

Note that $\psi_{l,t}$ is compact, and the function $\vartheta \mapsto m_\vartheta^{ps}(x)$ is upper semicontinuous in ϑ for almost all x. Furthermore, $m_\vartheta^{ps}(X) \leq M_0^{ps}(X)$ with $PM_0^{ps}(X) < \infty$ by C7(a). Thus, Lemma A.1 (One-sided Gilvenko-Cantelli Theorem) of Wellner and Zhang (2000) yields

$$\limsup_{n\to\infty} \sup_{\vartheta \in \mathcal{R}\times\psi_{l,t}} (\mathbb{P}_n - P) m_\vartheta^{ps}(x) \leq 0 \quad \text{a.s.} \tag{A1}$$

Next, we show that $\mathbb{M}_n^{ps}(\beta_0, \Lambda_0^*) - \mathbb{M}^{ps}(\beta_0, \Lambda_0^*) = o_p(1)$. Define the class

$$\mathcal{N}_\eta^{ps} = \{m_{(\beta_0,\Lambda)}^{ps}(X) - m_{(\beta_0,\Lambda_0)}^{ps}(X) : \ \Lambda \in \psi_{l,t} \text{ and } \|\Lambda - \Lambda_0\|_{L_2(\mu_1)} \leq \eta\}.$$

It easy to show that $\mathcal{N}_\eta^{ps}$ is a Donsker class by C2, C8, C9, and Lemma A1. Cauchy-Schwarz inequality and C5, C8, and C9 yield

$$P\left(m_{(\beta_0,\Lambda)}^{ps}(X) - m_{(\beta_0,\Lambda_0)}^{ps}(X)\right)^2 \leq CP\sum_{j=1}^{K}(\Lambda(T_{K,j}) - \Lambda_0(T_{K,j}))^2 \leq c\eta^2,$$

for $\Lambda \in \psi_{l,t}$ and $\|\Lambda - \Lambda_0\|_{L_2(\mu_1)} \leq \eta$.

It is followed that, for the seminorm $\rho_P(f) = \{P(f - Pf)^2\}^{1/2}$,

$$\sup_{f \in \mathcal{N}_\eta^{ps}} \rho_P(f) \leq \sup_{f \in \mathcal{N}_\eta^{ps}} \{Pf^2\}^{1/2} \leq c\eta \to 0,$$

as $\eta \to 0$. Due to Corollary 2.3.13 of van der Vaart and Wellner (1996) (the relationship between $P-$Donsker and asymptotic equicontinuity), we have

$$(\mathbb{P}_n - P)(m^{ps}_{(\beta_0,\Lambda^*)}(X) - m^{ps}_{(\beta_0,\Lambda_0)}(X)) = o_p(n^{-1/2}).$$

Furthermore, $(\mathbb{P}_n - P)m^{ps}_{(\beta_0,\Lambda_0)}(X) = o_p(1)$ due to the Central Limit Theorem. Thus,

$$\mathbb{M}_n^{ps}(\beta_0, \Lambda_0^*) - \mathbb{M}^{ps}(\beta_0, \Lambda_0^*) = o_p(1).$$

Moreover, $\mathbb{M}_n^{ps}(\beta_0, \Lambda_0^*) \leq \mathbb{M}_n^{ps}(\hat\beta_n^{ps}, \hat\Lambda_n^{ps})$, and it follows that

$$\mathbb{M}^{ps}(\beta_0, \Lambda_0) - \varepsilon \leq \mathbb{M}^{ps}(\beta_0, \Lambda_0^*) = \mathbb{M}_n^{ps}(\beta_0, \Lambda_0^*) + o_p(1) \leq \mathbb{M}_n^{ps}(\hat\beta_n^{ps}, \hat\Lambda_n^{ps}) + o_p(1).$$

Inequality (A1) yields

$$\limsup_{n_k \to \infty} (\mathbb{P}_n - P) m^{ps}_{\hat\vartheta^{ps}_{n_k}}(X) \leq \limsup_{n_k \to \infty} \sup_{\vartheta \in \mathcal{R} \times \psi_{l,t}} (\mathbb{P}_n - P) m^{ps}_\vartheta(X) \leq 0.$$

It follows that

$$\mathbb{M}^{ps}(\beta_0, \Lambda_0) - \varepsilon \leq \limsup_{n_k \to \infty} \mathbb{M}_n^{ps}(\hat\vartheta^{ps}_{n_k}) \quad \leq \quad \limsup_{n_k \to \infty} \mathbb{M}^{ps}(\hat\vartheta^{ps}_{n_k}) = \mathbb{M}^{ps}(\vartheta^+)$$

in probability. The continuity of $\mathbb{M}^{ps}$, $\hat\vartheta^{ps}_{n_k} \to \vartheta^+$, and Dominated Convergence Theorem yield the last limit. Hence, for any $\varepsilon > 0$, $-\varepsilon \leq \mathbb{M}^{ps}(\vartheta^+) - \mathbb{M}^{ps}(\vartheta_0) \leq 0$. This implies that $\mathbb{M}^{ps}(\vartheta^+) = \mathbb{M}^{ps}(\vartheta_0)$. It follows that $\beta_0 = \beta^+$ and $\Lambda_0(u) = \Lambda^+(u)$ a.e. Since this is true for any convergent subsequence, we conclude that all the limits of subsequence of $\hat\vartheta^{ps}_n$ are equal to ϑ_0 a.e. This implies that $\lim_{n\to\infty} \hat\Lambda^{ps}_n(u) = \Lambda_0(u)$ in probability. Since $\hat\Lambda^{ps}_n$ is uniformly bounded for $t \in [0, b]$ if $\mu_1[b, \tau] > 0$ for some $0 < b < \tau$ or for $t \in [0, \tau]$ if $\mu_1(\tau) > 0$, the Dominated Convergence Theorem yields the weak consistency of $(\hat\beta^{ps}_n, \hat\Lambda^{ps}_n)$ in the metric d_1.

Proof of Theorem 2 (Rate of Convergence)

The proof of the rate of convergence is based on Theorem 3.2.5 of van der Vaart and Wellner (1996). For any ϑ in a sufficiently small neighborhood of ϑ_0, Wellner and Zhang (2007) showed that $\mathbb{M}^{ps}(\vartheta_0) - \mathbb{M}^{ps}(\vartheta) \geq C d_1^2(\vartheta, \vartheta_0)$, for $C > 0$. Therefore, the first condition of Theorem 3.2.5 of van der Vaart and Wellner (1996) holds.

For any $\eta > 0$, define the class

$$\mathcal{F}_\eta^{ps} = \{\Lambda | \Lambda \in \psi_{l,t},\ \|\Lambda - \Lambda_0\|_{L_2(\mu_1)} \leq \eta\}.$$

By Theorem 1, $\hat{\Lambda}_n^{ps} \in \mathcal{F}_\eta^{ps}$, for any $\eta > 0$ and sufficiently large n.

Next, define the class

$$\mathcal{M}_\eta^{ps} = \{m_\vartheta^{ps}(X) - m_{\vartheta_0}^{ps}(X) :\ \Lambda \in \mathcal{F}_\eta^{ps} \text{ and } d_1(\vartheta, \vartheta_0) \leq \eta\}.$$

With C2, C4, C5, C8, C9, and the result of Lemma A1, for any $\varepsilon \leq \eta$, we can have

$$\log N_{[\ ]}(\varepsilon, \mathcal{M}_\eta^{ps}, ||\cdot||_{P,B}) \leq c q_n \log(\eta/\varepsilon),$$

where $\|\cdot\|_{P,B}$ is the Bernstein Norm defined as $\|f\|_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$ by van der Vaart and Wellner (1996, p. 324). Moreover, some algebraic calculations lead to $||\mathbb{M}^{ps}(\vartheta) - \mathbb{M}^{ps}(\vartheta_0)||_{P,B}^2 \leq C\eta^2$, for any $m_\vartheta^{ps}(X) - m_{\vartheta_0}^{ps}(X) \in \mathcal{M}_\eta^{ps}$. Therefore, by Lemma 3.4.3 of van der Vaart and Wellner (1996), we obtain

$$E_P||n^{1/2}(\mathbb{P} - P)||_{\mathcal{M}_\eta^{ps}} \leq C J_{[\ ]}(\eta, \mathcal{M}_\eta^{ps}, ||\cdot||_{P,B}) \left\{1 + \frac{J_{[\ ]}(\eta, \mathcal{M}_\eta^{ps}, ||\cdot||_{P,B})}{\eta^2 n^{1/2}}\right\}, \qquad \text{(A2)}$$

where

$$J_{[\ ]}(\eta, \mathcal{M}_\eta^{ps}, ||\cdot||_{P,B}) = \int_0^\eta \{1 + \log N_{[\ ]}(\varepsilon, \mathcal{M}_\eta^{ps}, ||\cdot||_{P,B})\}^{1/2} d\varepsilon \leq c_0 {q_n}^{1/2} \eta.$$

The right hand side of (A2) yields $\phi_n(\eta) = C({q_n}^{1/2}\eta + q_n/n^{1/2})$. It is easy to see that $\phi_n(\eta)/\eta$ is decreasing in η, and

$$r_n^2 \phi_n(\frac{1}{r_n}) = r_n {q_n}^{1/2} + r_n^2 q_n/n^{1/2} \le 2n^{1/2},$$

for $r_n = n^{(1-\nu)/2}$ and $0 < \nu < 1/2$. Hence, $n^{(1-\nu)/2} d_1(\hat{\vartheta}_n^{ps}, \vartheta_0) = O_p(1)$ by Theorem 3.2.5 of van der Vaart and Wellner (1996). The choice of $\nu = 1/(1+2r)$ yields the rate of convergence of $n^{r/(1+2r)}$ which completes the proof.

Proof of Theorem 3 (Asymptotic Normality)

We show the normality by checking the assumptions A1 - A6 of Theorem 6.1 of Wellner and Zhang (2007). This theorem is a generalization of a result of Huang (1996). The rate of convergence given in Theorem 2 guarantees the assumption A1 holds with $\gamma > 1/3$. Assumptions A2, A3, A5 and A6 only depend on the model and verification of these assumptions in the likelihood-based methods for panel count data has been done in Wellner and Zhang (2007). Here, we only need to verify Assumption A4 for the spline pseudo-likelihood method, that is,

$$\begin{aligned} S_{1n}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}) &= \mathbb{P}_n \sum_{j=1}^{K} Z(\mathbb{N}_{Kj} - \hat{\Lambda}_{nKj}^{ps} \exp(\hat{\beta}_n^{psT} Z)) = o_p(n^{-1/2}), \\ S_{2n}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})[\mathbf{h}^*] &= \mathbb{P}_n \sum_{j=1}^{K} \left(\frac{\mathbb{N}_{Kj}}{\hat{\Lambda}_{nKj}^{ps}} - \exp(\hat{\beta}_n^{psT} Z) \right) \mathbf{h}_{Kj}^* = o_p(n^{-1/2}), \end{aligned}$$

where

$$\mathbf{h}_{Kj}^* = \Lambda_{0Kj} \frac{E(Z e^{\beta_0^T Z} | K, T_{K,j})}{E(e^{\beta_0^T Z} | K, T_{K,j})},$$

with $\Lambda_{0Kj} = \Lambda_0(T_{Kj})$ and $\hat{\Lambda}_{nKj}^{ps} = \hat{\Lambda}_n^{ps}(T_{Kj})$.

Since $\hat{\beta}_n^{ps}$ satisfies the pseudo-score equation, it follows that

$$S_{1n}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}) = \mathbb{P}_n \sum_{j=1}^{K} Z(\mathbb{N}_{Kj} - \hat{\Lambda}_{nKj}^{ps} \exp(\hat{\beta}_n^{psT} Z)) = 0.$$

The first part of A4 holds. Therefore, we only need to verify

$$S_{2n}(\beta_n^{ps}, \Lambda_n^{ps})[\mathbf{h}^*] = o_p(n^{-1/2}).$$

Since $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$ maximizes $\mathbb{P}_n m_{\vartheta}^{ps}(x)$ over the feasible region, for $\vartheta_\varepsilon = (\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps} + \varepsilon h)$ with any $h \in \psi_{l,t}$, we have

$$\begin{aligned}
& \lim_{\varepsilon \downarrow 0} \frac{d}{d\varepsilon} \mathbb{P}_n m_{\vartheta_\varepsilon}^{ps}(X) \\
= \quad & \mathbb{P}_n \sum_{j=1}^{K} \lim_{\varepsilon \downarrow 0} \frac{d}{d\varepsilon} \left[\mathbb{N}_{Kj} \log(\hat{\Lambda}_{nKj} + \varepsilon h) + \mathbb{N}_{Kj} \hat{\beta}_n^{psT} Z - \exp(\hat{\beta}_n^{psT} Z)(\hat{\Lambda}_{nKj}^{ps} + \varepsilon h) \right] \\
= \quad & \mathbb{P}_n \sum_{j=1}^{K} \left[\frac{\mathbb{N}_{Kj}}{\hat{\Lambda}_{nKj}^{ps}} - \exp(\hat{\beta}_n^{psT} Z) \right] h = 0.
\end{aligned}$$

Take $h = \hat{\Lambda}_{nKj}^{ps} \alpha_{Kj}$ with $\alpha_{Kj} = \dfrac{E(Z \exp(\beta_0^T Z)|K, T_{Kj})}{E(\exp(\beta_0^T Z)|K, T_{Kj})}$. We have

$$\mathbb{P}_n \left[\sum_{j=1}^{K} \frac{1}{\hat{\Lambda}_{nKj}^{ps}} \left\{ \mathbb{N}_{Kj} - \hat{\Lambda}_{nKj}^{ps} \exp(\hat{\beta}_n^{psT} Z) \right\} \hat{\Lambda}_{nKj}^{ps} \alpha_{Kj} \right] = 0.$$

Thus, to show

$$S_{2n}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})[\mathbf{h}^*] = \mathbb{P}_n \left[\sum_{j=1}^{K} \frac{1}{\hat{\Lambda}_{nKj}^{ps}} \{ \mathbb{N}_{Kj} - \hat{\Lambda}_{nKj}^{ps} \exp(\hat{\beta}_n^{psT} Z) \} \Lambda_{0Kj} \alpha_{Kj} \right] = o_p(n^{-1/2}),$$

it is equivalent to show that

$$\begin{aligned}
I \quad = \quad & \mathbb{P}_n \left[\sum_{j=1}^{K} \frac{1}{\hat{\Lambda}_{nKj}^{ps}} \{ \mathbb{N}_{Kj} - \hat{\Lambda}_{nKj}^{ps} \exp(\hat{\beta}_n^{psT} Z) \} (\Lambda_{0Kj} - \hat{\Lambda}_{nKj}^{ps}) \alpha_{Kj} \right] \\
= \quad & I_1 - I_2 + I_3 = o_p(n^{-1/2}),
\end{aligned}$$

where

$$\begin{aligned} I_1 &= (\mathbb{P}_n - P)\left[\sum_{j=1}^{K} \frac{\mathbb{N}_{Kj}}{\hat{\Lambda}^{ps}_{nKj}}(\Lambda_{0Kj} - \hat{\Lambda}^{ps}_{nKj})\alpha_{Kj}\right], \\ I_2 &= (\mathbb{P}_n - P)\left[\sum_{j=1}^{K} \exp(\hat{\beta}_n^{psT} Z)(\Lambda_{0Kj} - \hat{\Lambda}^{ps}_{nKj})\alpha_{Kj}\right], \\ I_3 &= P\left[\sum_{j=1}^{K} \frac{1}{\hat{\Lambda}^{ps}_{nKj}}\{\mathbb{N}_{Kj} - \hat{\Lambda}^{ps}_{nKj}\exp(\hat{\beta}_n^{psT} Z)\}(\Lambda_{0Kj} - \hat{\Lambda}^{ps}_{nKj})\alpha_{Kj}\right]. \end{aligned}$$

Next we show that I_1, I_2, and I_3 are all $o_p(n^{-1/2})$.

Let

$$\Phi_1^{ps}(\eta) = \left\{\sum_{j=1}^{K} \frac{\mathbb{N}_{Kj}}{\Lambda_{Kj}}(\Lambda_{0Kj} - \Lambda_{Kj})\alpha_{Kj} : \ \Lambda \in \psi_{l,t} \ \text{ and } \ \|\Lambda - \Lambda_0\|_{L_2(\mu_1)} \le \eta\right\}.$$

It is easy to show that $\Phi_1^{ps}(\eta)$ is a Donsker class by C2, C8, C9, and C11. Cauchy-Schwarz inequality in addition to C2, C8, C9, and C11 yield

$$P\left(\sum_{j=1}^{K} \frac{\mathbb{N}_{Kj}}{\Lambda_{Kj}}(\Lambda_{0Kj} - \Lambda_{Kj})\alpha_{Kj}\right)^2 < CP\sum_{j=1}^{K}(\Lambda_{0Kj} - \Lambda_{Kj})^2 < C\eta^2,$$

for $\Lambda \in \psi_{l,t}$ and $\|\Lambda - \Lambda_0\|_{L_2(\mu_1)} \le \eta$. It is followed that, for the seminorm $\rho_P(f) = \{P(f - Pf)^2\}^{1/2}$,

$$\sup_{f \in \Phi_1^{ps}(\eta)} \rho_P(f) \le \sup_{f \in \Phi_1^{ps}(\eta)} \{Pf^2\}^{1/2} \le c\eta \to 0$$

as $\eta \to 0$. Due to Corollary 2.3.13 of van der Vaart and Wellner (1996), we have $I_1 = o_p(n^{-1/2})$.

Next, Let

$$\Phi_2^{ps}(\eta) = \left\{\sum_{j=1}^{K} \exp(\beta^T Z)(\Lambda_{0Kj} - \Lambda_{Kj})\alpha_{Kj} : \ \Lambda \in \psi_{l,t} \ \text{ and } \ d_1(\vartheta, \vartheta_0) \le \eta\right\}.$$

C5, C8, and C11 along with the assumption that $\mathcal{R}$ is compact yield $\Phi_2^{ps}(\eta)$ is a Donsker class. Moreover, by conditions C5, C8, C11, and Cauchy-Schwarz inequality, we have

$$P\left(\sum_{j=1}^{K}\exp(\beta^T Z)(\Lambda_{0Kj}-\Lambda_{Kj})\alpha_{Kj}\right)^2 < CP\sum_{j=1}^{K}(\Lambda_{0Kj}-\Lambda_{Kj})^2 < C\eta^2,$$

for $\Lambda \in \psi_{l,t}$ and $d_1(\vartheta,\vartheta_0) \le \eta$. Therefore, due to Corollary 2.3.13 of van der Vaart and Wellner (1996), $I_2 = O_p(n^{-1/2})$.

The quantity I_3 equals

$$\begin{aligned}
& P\sum_{j=1}^{K}\frac{1}{\hat{\Lambda}^{ps}_{nKj}}\{\Lambda_{0Kj}\exp(\beta_0^T Z) - \hat{\Lambda}^{ps}_{nKj}\exp(\hat{\beta}_n^{psT}Z)\}(\Lambda_{0Kj}-\hat{\Lambda}^{ps}_{nKj})\alpha_{Kj} \\
= \;& P\sum_{j=1}^{K}\frac{1}{\hat{\Lambda}^{ps}_{nKj}}(\Lambda_{0Kj}-\hat{\Lambda}^{ps}_{nKj})^2\exp(\beta_0^T Z)\alpha_{Kj} + P\sum_{j=1}^{K}(\exp(\beta_0^T Z)-\exp(\hat{\beta}_n^{psT}Z))(\Lambda_{0Kj}-\hat{\Lambda}^{ps}_{nKj})\alpha_{Kj}.
\end{aligned}$$

Because $\hat{\Lambda}^{ps}_{nKj}$ is uniformly bounded in probability as shown in the proof for consistency, by C5, C8, and C11, we have

$$P\sum_{j=1}^{K}\frac{1}{\hat{\Lambda}^{ps}_{nKJ}}(\Lambda_{0Kj}-\hat{\Lambda}^{ps}_{nKj})^2\exp(\beta_0^T Z)\alpha_{Kj} \le C_1\sum_{j=1}^{K}(\Lambda_{0Kj}-\hat{\Lambda}^{ps}_{nKj})^2.$$

Performing Taylor expansion of $\exp(\hat{\beta}_n^{psT}Z)$ at $\beta=\beta_0$, we have

$$\exp(\hat{\beta}_n^{psT}Z)-\exp(\beta_0^T Z) = \exp(\xi^T Z)(\hat{\beta}_n^{psT}Z-\beta_0^T Z),$$

where ξ is between $\hat{\beta}_n^{ps}$ and β_0^T. By Conditions C5, C8, and C11, we have

$$\begin{aligned}
& P\sum_{j=1}^{K}(\exp(\beta_0^T Z)-\exp(\hat{\beta}_n^{psT}Z))(\Lambda_{0Kj}-\hat{\Lambda}^{ps}_{nKj})\alpha_{Kj} \\
\le \;& CP\sum_{j=1}^{K}(\beta_0^T Z-\hat{\beta}_n^{psT}Z)(\Lambda_{0Kj}-\hat{\Lambda}^{ps}_{nKj}) \\
\le \;& C\sum_{j=1}^{K}P(\hat{\beta}_n^{ps}-\beta_0)^T Z^T Z(\hat{\beta}_n^{ps}-\beta_0) + CP\sum_{j=1}^{K}(\Lambda_{0Kj}-\hat{\Lambda}^{ps}_{nKj})^2.
\end{aligned}$$

The last inequality holds due to Cauchy-Schwarz inequality.

Hence,

$$I_3 \leq C|\hat{\beta}_n^{ps} - \beta_0|^2 + CE\sum_{j=1}^{K}(\Lambda_{0Kj} - \hat{\Lambda}_{nKj}^{ps})^2 = Cd_1^2((\hat{\Lambda}_n^{ps}, \hat{\beta}_n^{ps}), (\Lambda_0, \beta_0)).$$

Finally, the rate of convergence yields $I_3 = o_p(n^{-1/2})$.

REFERENCES

Andrews, D. F., and Herzberg, A. M. (1985), *Data; A Collection of Problems from Many Fields for the Students and Research Works,* New York: Springer-Verlag.

Byar, D. P. (1980), "The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumors: Comparisons of Placebo, Pyridoxine, and Topical Thiotepa," In *Bladder Tumors and Other Topics in Urological Oncology*, 363 - 370, edited by M. Macaluso, P. H. Smith, and F. Edsmyr. New York: Plenum.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap,* London: Chapman and Hall.

Geman, A. and Hwang, C. R. (1982), "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," *The Annals of Statistics,* 10, 401-414.

Huang, J. (1996), "Efficient Estimation for the Proportional Hazards Model with Interval Censored Data," *The Annals of Statistics,* 24, 540-568.

Huang, J. (1999), "Efficient Estimation of the Partly Linear Additive Cox Model," *The Annals of Statistics,* 27, 1536-1563.

Jamshidian, M. (2004), "On Algorithms for Restricted Maximum Likelihood Estimation," *Computational Statistics and 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 American Statistical Association,* 80, 863-871.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995), “Hazard Regression,” *Journal of American Statistical Association,* 90, 78-94.

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.

Rosen, J. B. (1960), “The Gradient Projection Method for Nonlinear Programming,” *Journal of the Society for Industrial and Applied Mathematics,* 8, 181-217.

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

Shen, X., and Wong, W. H. (1994), “Convergence Rate of Sieve Estimates,” *The Annals of Statistics,* 22, 580-615.

Sun, J., and Kalbfleisch, J. D. (1995), “Estimation of the Mean Function of Point Processes Based on Panel Count Data,” *Statistical Sinica,* 5, 279-290.

Stone, C. J. (1986), “The Dimensionality Reduction Principal for Generalized Additive Models,” *The Annals of Statistics,* 14, 590-606.

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,* Ser. B 62, 293 -302.

Thall, P. F., and Lachin, J. M. (1988), “Analysis of Recurrent Events: Nonparametric Methods for Random-Interval Count Data,” *Journal of American Statistical Association,* 83, 339-347.

Van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes,* New York: Springer-Verlag.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989), "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions," *Journal of American Statistical Association,* 84, 1065-1073.

Wellner, J. A., and Zhang, Y. (2000), "Two Estimators of the Mean of A Counting Process with Panel Count Data," *The Annals of Statistics,* 28, 779-814.

Wellner, J. A., and Zhang, Y. (2007), "Likelihood-Based Semiparametric Estimation Methods for Panel Count Data with Covariates," *The Annals of Statistics,* 35, 2106-2142.

Wellner, J. A., Zhang, Y., and Liu, H. (2004), " A semiparametric regression model for panel count data: When do pseudo-likelihood estimators become badly inefficient?" *Proceedings of the Second Seattle Biostatistical Symposium: Analysis of Correlated Data. Lecture Notes in Statistics.* 179, 143-174. Springer, New York.

Zhang, Y. (2002), "A Semiparametric Pseudolikelihood Estimation Method for Panel Count Data," *Biometrika,* 89, 39-48.

Zhang, Y., and Jamshidian, M. (2004), "On Algorithms for NPMLE of the Failure Function with Censored Data," *The Journal of Computational and Graphical Statistics,* 13, 123-140.

Zhou, S., Shen, X., and Wolfe, D. A. (1998), "Local Asymptotics for Regression Splines and Confidence Region," *The Annals of Statistics,* 26, 1760-1782.

Table 1. Comparison of Bias and Standard Deviation for β_0, Based on the Data Generated from Poisson Process, n = 50 or 100.

	n = 50				n = 100			
	a	b	c	d	a	b	c	d
Estimate of β_1								
BIAS	-0.0050	-0.0045	-0.0025	-0.0035	0.0013	0.0010	0.0016	0.0023
SD	0.1200	0.1196	0.1055	0.1042	0.0854	0.0839	0.0754	0.0742
ASE	0.1087	0.1087	0.0935	0.0935	0.0769	0.0769	0.0661	0.0661
Estimate of β_2								
BIAS	0.0007	0.0011	-0.0008	-0.0010	0.0007	0.0007	0.0005	0.0006
SD	0.0338	0.0330	0.0296	0.0294	0.0238	0.0232	0.0208	0.0204
ASE	0.0306	0.0306	0.0263	0.0263	0.0217	0.0217	0.0186	0.0186
Estimate of β_3								
BIAS	-0.0007	-0.0001	0.0004	-0.0002	-0.0021	-0.0023	-0.0013	-0.0016
SD	0.0850	0.0837	0.0734	0.0724	0.0600	0.0594	0.0509	0.0508
ASE	0.0793	0.0793	0.0682	0.0682	0.0561	0.0561	0.0482	0.0482

a. Maximum Pseudo-likelihood Estimator

b. Spline Pseudo-likelihood Estimator

c. Maximum Likelihood Estimator

d. Spline Likelihood Estimator

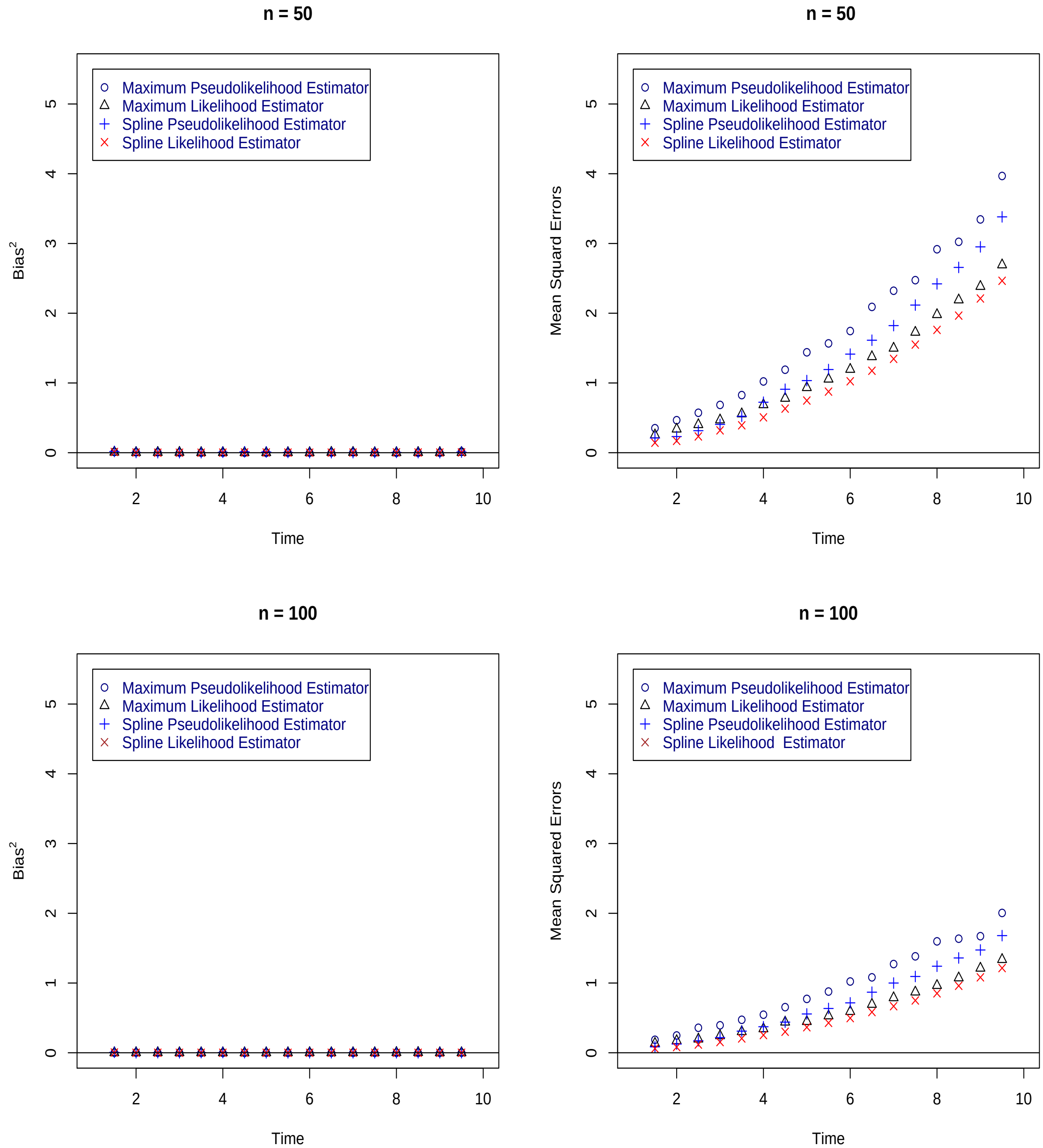

Figure 1. Simulation 1, with Data from Poisson Process and $\Lambda_0(t) = 2t$.

Table2. Comparison of Bias and Standard Deviation for β_0, Based on the Data Generated from the Mixed Poisson Process, n = 50 or 100.

	n = 50				*n = 100*			
	a	b	c	d	a	b	c	d
Estimate of β_1								
BIAS	0.0003	0.0006	-0.0010	0.0004	0.0028	0.0034	0.0050	0.0052
SD	0.1535	0.1548	0.1376	0.1377	0.1095	0.1103	0.1007	0.1007
ASE	0.1554	0.1554	0.1375	0.1375	0.1099	0.1099	0.0972	0.0972
Estimate of β_2								
BIAS	-0.0012	-0.0019	-0.0016	-0.0019	-0.0001	-0.0002	-0.0001	-0.0001
SD	0.0467	0.0466	0.0426	0.0423	0.0319	0.0323	0.0294	0.0295
ASE	0.0473	0.0473	0.0420	0.0420	0.0334	0.0334	0.0297	0.0297
Estimate of β_3								
BIAS	0.0016	0.0013	0.0022	0.0013	0.0004	0.0002	0.0010	0.0008
SD	0.1050	0.1048	0.0924	0.0916	0.0693	0.0693	0.0614	0.0614
ASE	0.0962	0.0962	0.0842	0.0842	0.0680	0.0680	0.0596	0.0596

a. Maximum Pseudo-likelihood Estimator

b. Spline Pseudo-likelihood Estimator

c. Maximum Likelihood Estimator

d. Spline Likelihood Estimator

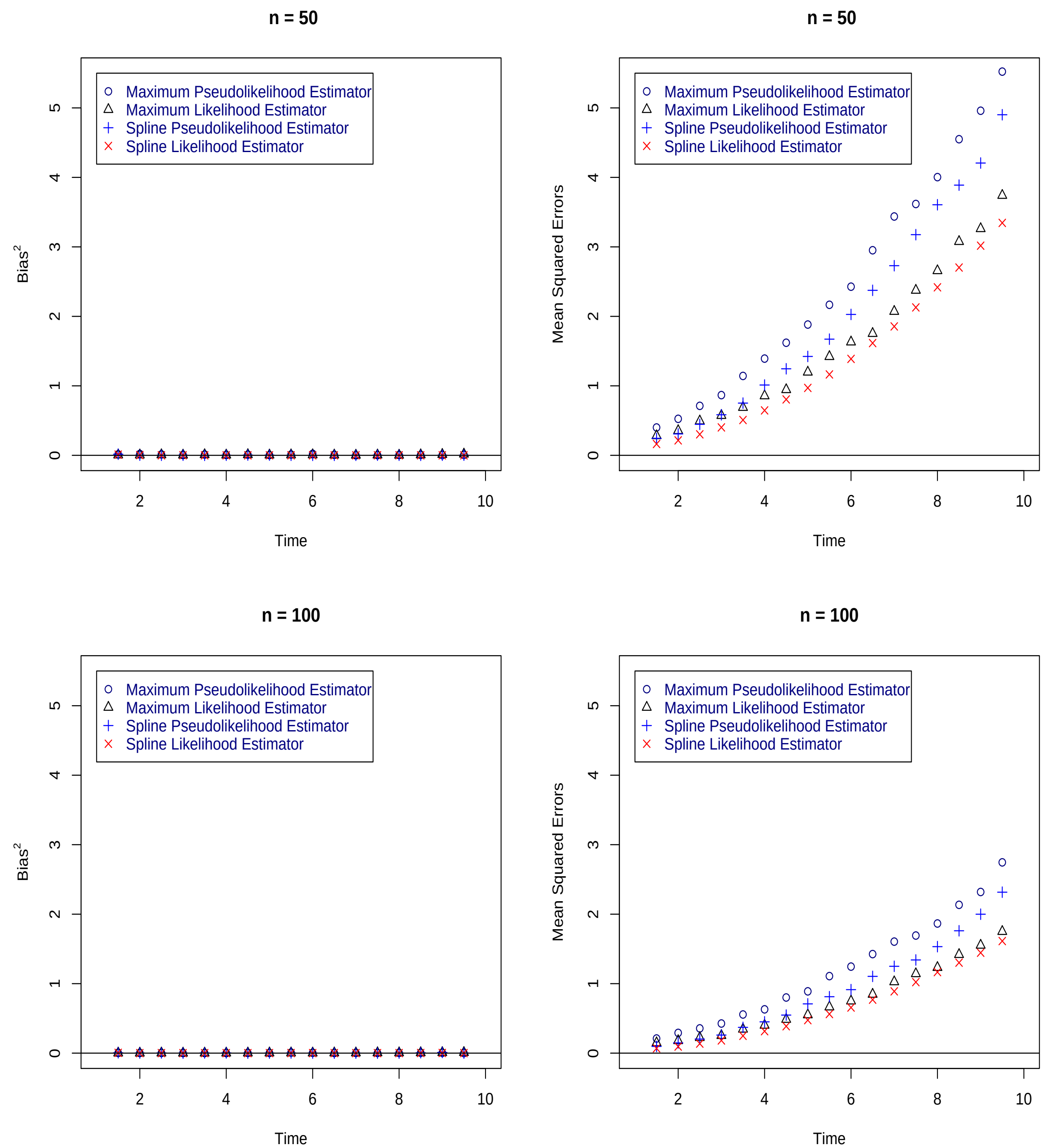

Figure 2. Simulation 2, with Data from the Mixed Poisson Process and $\Lambda_0(t) = 2t$.

Table 3. Comparison of Computing Time in Seconds among Four Estimators, Based on the Data Generated from the Poisson Process or Mixed Poisson Process with Sample Size 50 or 100.

	Poisson Process		*Mixed Poisson Process*	
Estimators	*n = 50*	*n = 100*	*n = 50*	*n = 100*
Maximum Pseudo-likelihood	9.2	17.2	9.5	20.7
Spline Pseudo-likelihood	3.0	10.4	3.2	7.8
Maximum Likelihood	603.4	1534.1	662.5	1564.6
Spline Likelihood	14.2	42.1	13.7	43.2

Table 4. Coverage Probabilities of 95% Confidence Intervals, Based on 1000 Bootstrap Samples and Data Generated from the Poisson Process or Mixed Poisson Process with Sample Size 50 or 100.

	Poisson Process			*Mixed Poisson Process*		
Estimators	β_1	β_2	β_3	β_1	β_2	β_3
n = 50						
Spline Pseudo-likelihood	95.73%	95.28%	88.50%	95.02%	93.82%	93.42%
Spline Likelihood	94.99%	93.96%	94.55%	94.62%	94.62%	92.72%
n = 100						
Spline Pseudo-likelihood	94.88%	95.27%	93.91%	93.19%	94.18%	94.97%
Spline Likelihood	94.67%	96.52%	94.98%	93.58%	94.77%	95.76%

Table 5. Spline Inference for the Bladder Tumor Study, Based on 1000 Bootstrap Samples from the Original Data.

	Spline Pseudo-likelihood				*Spline Likelihood*			
Variable	$\hat{\beta}$	$\hat{sd}(\hat{\beta})$	$\hat{\beta}/\hat{sd}(\hat{\beta})$	p-value	$\hat{\beta}$	$\hat{sd}(\hat{\beta})$	$\hat{\beta}/\hat{sd}(\hat{\beta})$	p-value
Z_1	0.145	0.057	2.535	0.012	0.208	0.082	2.515	0.010
Z_2	-0.049	0.061	-0.731	0.475	-0.035	0.082	-0.429	0.672
Z_3	0.191	0.280	0.694	0.507	0.064	0.395	0.167	0.870
Z_4	-0.688	0.276	-2.486	0.015	-0.797	0.331	-2.403	0.019