Submitted to the Annals of Statistics

TWO LIKELIHOOD-BASED SEMIPARAMETRIC ESTIMATION METHODS FOR PANEL COUNT DATA WITH COVARIATES

BY JON A. WELLNER* AND YING ZHANG†

University of Washington and University of Iowa

We consider estimation in a particular semiparametric regression model for the mean of a counting process with "panel count" data. The basic model assumption is that the conditional mean function of the counting process is of the form $E\{\mathbb{N}(t)|Z\} = \exp(\beta_0^T Z)\Lambda_0(t)$ where Z is a vector of covariates and Λ_0 is the baseline mean function. The "panel count" observation scheme involves observation of the counting process $\mathbb{N}$ for an individual at a random number K of random time points; both the number and the locations of these time points may differ across individuals.

We study semiparametric maximum pseudo-likelihood and maximum likelihood estimators $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$ and $(\hat{\beta}_n, \hat{\Lambda}_n)$ of the unknown parameters (β_0, Λ_0). The pseudo-likelihood estimator is fairly easy to compute, while the maximum likelihood estimator poses more challenges from the computational perspective. We study asymptotic properties of both estimators under the proportional mean model and we also establish asymptotic normality for the estimators of the regression parameter β_0. The methods are validated by some simulation studies and illustrated by an example.

1. Introduction. Suppose that $\mathbb{N} = \{\mathbb{N}(t) : t \geq 0\}$ is a univariate counting process. In many applications, it is important to estimate the ex-

*Research supported in part by National Science Foundation grant DMS-0203320, NIAID grant 2R01 AI291968-04, and an NWO Grant to the Vrije Universiteit, Amsterdam

†Corresponding author

AMS 2000 subject classifications: Primary 60F05, 60F17; secondary 60J65, 60J70

Keywords and phrases: asymptotic distributions, asymptotic efficiency, asymptotic normality, consistency, counting process, empirical processes, information matrix, maximum likelihood, Poisson process, pseudo-likelihood estimators, monotone function

pected number of events $E\{\mathbb{N}(t)|Z\}$ which will occur by the time t, conditionally on a covariate vector Z.

In this paper we consider the proportional mean regression model given by

$$\Lambda(t|Z) \equiv E\{\mathbb{N}(t)|Z\} = e^{\beta_0^T Z}\Lambda_0(t)\,, \tag{1}$$

where the monotone increasing function Λ_0 is the *baseline mean function*. The parameters of primary interest are β_0 and Λ_0.

The observation scheme we want to study is as follows: suppose that we observe the counting process $\mathbb{N}$ at a random number K of random times

$$0 \equiv T_{K,0} < T_{K,1} < \cdots < T_{K,K}.$$

We write $\underline{T}_K \equiv (T_{K,1}, \ldots, T_{K,K})$, and we assume that $(K, \underline{T}_K|Z) \sim G(\cdot|Z)$ is conditionally independent of the counting process $\mathbb{N}$ given the covariate vector Z. We further assume that $Z \sim H$ on $\mathbb{R}^d$ with some mild conditions on H for the identifiability of our semiparametric regression model given in Section 3.

The data for each individual will consist of

$$X = (Z, K, \underline{T}_K, \mathbb{N}(T_{K,1}), \ldots, \mathbb{N}(T_{K,K})) \equiv (Z, K, \underline{T}_K, \underline{\mathbb{N}}_K)\,. \tag{2}$$

This type of data is referred to as *panel count data*. Throughout this manuscript, we will assume that the data consist of $X_1, \ldots, X_n$ i.i.d. as X.

Panel count data arise in many fields including demographic studies, industrial reliability, and clinical trials; see for example Kalbfleisch and Lawless (1985), Gaver and O'Muircheataigh (1987), Thall and Lachin (1988), Thall (1988), Sun and Kalbfleisch (1995), and Wellner and Zhang (2000) where the estimation of either the intensity of event recurrence or the mean function of a counting process with panel count data was studied. Many applications involve covariates whose effects on the underlying counting process are of interest. While there is considerable work on regression modeling for recurrent events based on continuous observations (see, for example Lawless and Nadeau (1995), Cook, Lawless, and Nadeau (1996), and Lin, Wei, Yang, and Ying (2000)), regression analysis with panel count data for counting processes has just started recently. Sun and Wei (2000) and Hu, Sun and Wei (2003) proposed estimating equation methods, while Zhang (1998, 2002) proposed a pseudo-likelihood method for studying the proportional mean model (1) with panel count data.

To derive useful estimators for this model we will often assume, in addition to (1), that the counting process $\mathbb{N}$, conditionally on Z, is a non-homogeneous Poisson process. But our general perspective will be to study the estimators and other procedures when the Poisson assumption *fails to hold* and we assume *only* that the proportional mean assumption (1) holds. Such a program was carried out by Wellner and Zhang (2000) for estimation of Λ_0 without any covariates for this panel count observation model.

The outline of the rest of the paper is as follows: In Section 2, we describe two methods of estimation, namely *maximum pseudo-likelihood* and *maximum likelihood estimators* of (β_0,Λ_0). The basic picture is that the pseudo-likelihood estimator is computationally relatively straightforward and easy to implement, while the maximum likelihood estimators are considerably more difficult, requiring an iterative algorithm in the computation of the profile likelihood. In Section 3, we state the main asymptotic results: strong consistency, rate of convergence and asymptotic normalities of $\hat{\beta}_n^{ps}$ and $\hat{\beta}_n$, for the maximum pseudo-likelihood and maximum likelihood estimators $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$ and $(\hat{\beta}_n, \hat{\Lambda}_n)$ of (β_0, Λ_0) assuming only the proportional mean structure (1), but *not assuming* that $\mathbb{N}$ is a Poisson process. These results are proved in Section 6 by use of tools from empirical process theory. Although pseudo-likelihood methods have been studied in the context of parametric models by Lindsay (1988) and Cox and Reid (2004), not much seems to be known about their behavior in non- and semi-parametric settings such as the one studied here, even assuming that the base model holds.

In Section 4, we present the results of simulation studies to demonstrate the robustness of the methods and compare the relative efficiency of the two methods. An application of our methods to the bladder tumor study is also presented in this section as well. In Section 5, we summarize our findings and present some further problems in this area. Finally, a general theorem concerning asymptotic normality of semiparametric M-estimators, along with some other technical tools upon which the proofs of our main theorems rely, are stated and proved in Appendices A and B.

2. Two Likelihood-Based Semiparametric Estimation Methods.

Maximum Pseudo-likelihood Estimation: The natural pseudo-likelihood estimator for this model uses the marginal distributions of $\mathbb{N}$, conditional on Z,

$$P(\mathbb{N}(t) = k|Z) = \frac{\Lambda(t|Z)^k}{k!} \exp(-\Lambda(t|Z))$$

and *ignores dependence* between $\mathbb{N}(t_1)$, $\mathbb{N}(t_2)$ to obtain the *log-pseudo-likelihood*:

$$l_n^{ps}(\beta,\Lambda) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left\{\mathbb{N}^{(i)}(T_{K_i,j}^{(i)})\log\Lambda(T_{K_i,j}^{(i)}) + \ \mathbb{N}^{(i)}(T_{K_i,j}^{(i)})\beta^T Z_i - e^{\beta^T Z_i}\Lambda(T_{K_i,j}^{(i)})\right\}\,.$$

Let $\mathcal{R} \subset \mathbb{R}^d$ be a bounded and convex set, and let $\mathcal{F}$ be the class of functions

$$\text{(3)}\quad \mathcal{F} \equiv \{\Lambda : \quad [0,\infty) \to [0,\infty) | \Lambda \quad \text{is monotone increasing}, \quad \Lambda(0) = 0\}\,.$$

Then the maximum pseudo-likelihood estimator $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$ of (β_0, Λ_0) is given by

$$(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}) \equiv \text{argmax}_{(\beta,\Lambda)\in\mathcal{R}\times\mathcal{F}}\, l_n^{ps}(\beta,\Lambda)\,.$$

This can be implemented in two steps via the usual (pseudo-) profile likelihood. For each fixed value of β we set

$$\text{(4)}\qquad \hat{\Lambda}_n^{ps}(\cdot,\beta) \equiv \text{argmax}_{\Lambda\in\mathcal{F}}\, l_n^{ps}(\beta,\Lambda)\,,$$

and define

$$l_n^{ps,profile}(\beta) \equiv l_n^{ps}(\beta, \hat{\Lambda}_n^{ps}(\cdot,\beta))\,.$$

Then

$$\hat{\beta}_n^{ps} = \text{argmax}_{\beta \in \mathcal{R}} \, l_n^{ps,profile}(\beta) \,, \qquad \text{and} \qquad \hat{\Lambda}_n^{ps} = \hat{\Lambda}_n^{ps}(\cdot, \hat{\beta}_n^{ps}) \,.$$

Note that $l_n^{ps}(\beta, \Lambda)$ depends on Λ only at the observation time points. By convention, we define our estimator $\hat{\Lambda}_n^{ps}$ to be the one that has jumps only at the observation time points to insure uniqueness.

In fact, the optimization problem in (4) is easily solved and the details of the solution for this optimization problem can be found in Zhang (2002).

Maximum Likelihood Estimation: Under the assumption that $\mathbb{N}$ is (conditionally, given Z) a non-homogeneous Poisson process, the likelihood can be calculated using the (conditional) independence of the increments of $\mathbb{N}$, $\Delta\mathbb{N}(s,t] \equiv \mathbb{N}(t) - \mathbb{N}(s)$, and the Poisson distribution of these increments:

$$P(\Delta\mathbb{N}(s,t] = k|Z) = \frac{[\Delta\Lambda((s,t]|Z)]^k}{k!} \exp(-\Delta\Lambda((s,t]|Z))$$

to obtain the *log-likelihood*:

$$l_n(\beta, \Lambda) \;\; = \;\; \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left\{ \Delta\mathbb{N}_{K_i j}^{(i)} \cdot \log \Delta\Lambda_{K_i j} + \; \Delta\mathbb{N}_{K_i j}^{(i)} \beta^T Z_i - \; e^{\beta^T Z_i} \Delta\Lambda_{K_i j} \right\}$$

where

$$\begin{aligned} \Delta\mathbb{N}_{Kj} &\equiv \mathbb{N}(T_{K,j}) - \mathbb{N}(T_{K,j-1}), \qquad j = 1, \ldots, K \,, \\ \Delta\Lambda_{Kj} &\equiv \Lambda(T_{K,j}) - \Lambda(T_{K,j-1}), \qquad j = 1, \ldots, K \,. \end{aligned}$$

Then

$$(\hat{\beta}_n, \hat{\Lambda}_n) \equiv \text{argmax}_{(\beta,\Lambda) \in \mathcal{R} \times \mathcal{F}} \, l_n(\beta, \Lambda) \,.$$

This maximization can also be carried out in two steps via profile likelihood. For each fixed value of β we set

$$\hat{\Lambda}_n(\cdot, \beta) \equiv \text{argmax}_{\Lambda \in \mathcal{F}} \, l_n(\beta, \Lambda) \,,$$

and define

$$l_n^{profile}(\beta) \equiv l_n(\beta, \hat{\Lambda}_n(\cdot, \beta)) \,.$$

Then

$$\hat{\beta}_n = \text{argmax}_{\beta \in \mathcal{R}} \, l_n^{profile}(\beta) \,, \qquad \text{and} \qquad \hat{\Lambda}_n = \hat{\Lambda}_n(\cdot, \hat{\beta}_n) \,.$$

Similarly, the estimator $\hat{\Lambda}_n$ is defined to have jumps only at the observation time points. To compute the estimate $(\hat{\beta}_n, \hat{\Lambda}_n)$, we adopt a doubly iterative algorithm to update the estimates alternately. The sketch of the algorithm consists of the following steps:

S1. Choose the initial $\beta^{(0)} = \hat{\beta}_n^{ps}$, the maximum pseudo-likelihood estimator.

S2. For given $\beta^{(p)}$ $(p = 0, 1, 2, \cdots)$, the updated estimate of Λ_0, $\Lambda^{(p)}$ is computed by the modified iterative convex minorant algorithm proposed by Jongbloed (1998) on the likelihood function $l_n(\beta^{(p)}, \Lambda)$. Initialize this algorithm using $\Lambda^{(p-1)}$ (in the very first step, we choose the starting value of Λ by interpolating $\hat{\Lambda}_n^{ps}$ linearly between two adjacent jump points to make it monotone increasing and so the likelihood function $l_n(\beta, \Lambda)$ is well defined) and stop the iteration when

$$\left| \frac{l_n(\beta^{(p)}, \Lambda_{new}) - l_n(\beta^{(p)}, \Lambda_{current})}{l_n(\beta^{(p)}, \Lambda_{current})} \right| \leq \eta.$$

S3. For given $\Lambda^{(p)}$, the updated estimate of β, $\beta^{(p+1)}$ is obtained by optimizing $l_n(\beta, \Lambda^{(p)})$ using the Newton-Raphson method. Initialize the algorithm using $\beta^{(p)}$ and stop the iteration when $\|\beta_{new} - \beta_{current}\|_\infty \leq \eta$.

S4. Repeat Steps 2 and 3 until the following convergence criterion is satisfied:

$$\left| \frac{l_n(\beta^{(p+1)}, \Lambda^{(p+1)}) - l_n(\beta^{(p)}, \Lambda^{(p)})}{l_n(\beta^{(p)}, \Lambda^{(p)})} \right| \leq \eta.$$

As in the case of pseudo-likelihood studied in Zhang (2002), it is easy to verify that for any given monotone Λ, the likelihood function $l_n(\beta, \Lambda)$ is a concave function of the regression parameter β with a negative definite Hessian matrix. Because of that, we can also easily show that $l_n(\beta^{(p+1)}, \Lambda^{(p+1)}) -$

$l_n(\beta^{(p)}, \Lambda^{(p)}) \geq 0$ using Taylor expansion. So the value of the (full)-likelihood increases after each iteration.

3. Asymptotic theory: Results.

In this section, we study the properties of the estimators $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$ and $(\hat{\beta}_n, \hat{\Lambda}_n)$. We establish strong consistency and derive the rate of convergence of both estimators in some L_2-metrics related to the observation scheme. We also establish the asymptotic normality of both $\hat{\beta}_n^{ps}$ and $\hat{\beta}_n$ under some mild conditions.

First we give some notation. Let $\mathcal{B}_d$ and $\mathcal{B}$ denote the collection of Borel sets in $\mathbb{R}^d$ and $\mathbb{R}$, respectively, and let $\mathcal{B}_{[0,\tau]} = \{B \cap [0,\tau] : \ B \in \mathcal{B}\}$ and $\mathcal{B}_2[0,\tau] = \mathcal{B}_{[0,\tau]} \times \mathcal{B}_{[0,\tau]}$. On $([0,\tau], \mathcal{B}_{[0,\tau]})$ we define measures μ_1, μ_2, ν_1, ν_2, and γ as follows: for $B, B_1, B_2 \in \mathcal{B}_{[0,\tau]}$ and $C \in \mathcal{B}_d$, set

$$\nu_1(B \times C) = \int_C \sum_{k=1}^{\infty} P(K=k|Z=z) \sum_{j=1}^{k} P(T_{k,j} \in B | K=k, Z=z) dH(z)\,,$$

$$\mu_1(B) = \nu_1(B \times \mathbb{R}^d)\,,$$

$$\nu_2(B_1 \times B_2 \times C)$$

$$= \int_C \sum_{k=1}^{\infty} P(K=k|Z=z) \sum_{j=1}^{k} P(T_{k,j-1} \in B_1, T_{k,j} \in B_2 | K=k, Z=z) dH(z),$$

$$\mu_2(B_1 \times B_2) = \nu_2(B_1 \times B_2 \times \mathbb{R}^d), \quad \text{and}$$

$$\gamma(B) = \int_{\mathbb{R}^d} \sum_{k=1}^{\infty} P(K=k|Z=z) P(T_{k,k} \in B | K=k, Z=z) dH(z)\,.$$

Based on the two measures μ_1 and μ_2, we define the L_2-metrics $d_1(\theta_1, \theta_2)$ and $d_2(\theta_1, \theta_2)$ in the parameter space $\Theta = \mathcal{R} \times \mathcal{F}$ as

$$\begin{aligned} d_1(\theta_1, \theta_2) &= \left\{ |\beta_1 - \beta_2|^2 + \|\Lambda_1 - \Lambda_2\|^2_{L_2(\mu_1)} \right\}^{1/2}, \\ &= \left\{ |\beta_1 - \beta_2|^2 + \int (\Lambda_1(t) - \Lambda_2(t))^2 d\mu_1(t) \right\}^{1/2}, \end{aligned}$$

and

$$d_2(\theta_1, \theta_2) = \left\{ |\beta_1 - \beta_2|^2 + \|\Delta\Lambda_1 - \Delta\Lambda_2\|^2_{L_2(\mu_2)} \right\}^{1/2}$$

$$= \left\{ |\beta_1 - \beta_2|^2 + \int |\Lambda_1(v) - \Lambda_1(u) - (\Lambda_2(v) - \Lambda_2(u))|^2 d\mu_2(u,v) \right\}^{1/2}.$$

To establish consistency, we assume that:

C1. The true parameter $\theta_0 = (\beta_0, \Lambda_0) \in \mathcal{R}^\circ \times \mathcal{F}$ where $\mathcal{R}^\circ$ is the interior of $\mathcal{R}$.

C2. The observation times $T_{K,j}$: for all $j = 1, \ldots, K$, $K = 1, 2, \ldots$, are random variables, taking values in the bounded interval $[0, \tau]$ for some $\tau \in (0, \infty)$, the measure $\mu_j \times H$ on $([0,\tau]^j \times \mathbb{R}^d, \mathcal{B}_j[0,\tau] \times \mathcal{B}_d)$ is absolutely continuous with respect to ν_j for $j = 1, 2$, and $E(K) < \infty$.

C3. The true baseline mean function Λ_0 satisfies $\Lambda_0(\tau) \le M$ for some $M \in (0, \infty)$.

C4. The function M_0^{ps} defined by $M_0^{ps}(X) \equiv \sum_{j=1}^K \mathbb{N}_{Kj} \log(\mathbb{N}_{Kj})$ satisfies $PM_0^{ps}(X) < \infty$.

C5. The function M_0 defined by $M_0(X) \equiv \sum_{j=1}^K \Delta\mathbb{N}_{Kj} \log(\Delta\mathbb{N}_{Kj})$ satisfies $PM_0(X) < \infty$.

C6. $\mathcal{Z} \equiv \mathrm{supp}(H)$, the support of H, is a bounded set in $\mathbb{R}^d$. (Thus there exists $z_0 > 0$ such that $P(|Z| \le z_0) = 1$.)

C7. For all $a \in \mathbb{R}^d$, $a \neq 0$, and $c \in \mathbb{R}$, $P(a^T Z \neq c) > 0$.

Condition 7 is needed together with $\mu_j \times H \ll \nu_j$ from C2 to establish identifiability of the semiparametric model.

Theorem 3.1. *Suppose that Conditions C1-C7 hold and the conditional mean structure of the counting process* $\mathbb{N}$ *is given by (1). Then for every* $b < \tau$ *for which* $\mu_1([b,\tau]) > 0$,

$$d_1\left((\hat\beta_n^{ps}, \hat\Lambda_n^{ps} 1_{[0,b]}), (\beta_0, \Lambda_0 1_{[0,b]}) \right) \to 0 \qquad \text{a.s. as} \quad n \to \infty.$$

In particular, if $\mu_1(\{\tau\}) > 0$, *then*

$$d_1\left((\hat\beta_n^{ps}, \hat\Lambda_n^{ps}), (\beta_0, \Lambda_0) \right) \to 0 \qquad \text{a.s. as} \quad n \to \infty.$$

Moreover, for every $b < \tau$ *for which* $\gamma([b,\tau]) > 0$,

$$d_2\left((\hat\beta_n, \hat\Lambda_n 1_{[0,b]}), (\beta_0, \Lambda_0 1_{[0,b]}) \right) \to 0 \qquad \text{a.s. as} \quad n \to \infty.$$

In particular, if $\gamma(\{\tau\}) > 0$*, then*

$$d_2\left((\hat{\beta}_n, \hat{\Lambda}_n), (\beta_0, \Lambda_0)\right) \to 0 \qquad \textit{a.s. as} \quad n \to \infty.$$

Note that the $j = 2$ part of C2 and C5 are not needed for proving consistency of $\hat{\theta}_n^{ps} = (\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps})$, while the $j = 1$ part of C2 and C4 are not needed for proving consistency of $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n)$.

Remark 3.1. Some condition along the lines of the absolute continuity part of assumption C2 is needed. For example, suppose that $\Lambda_0(t) = t^2$, $\beta_0 = 0$, $\Lambda(t) = t$, and $\beta = 1$. Then if we observe at just one time point T (so $K = 1$ with probability 1), and $T = e^Z$ with probability 1, then $\Lambda_0(T)e^{\beta_0 Z} = \Lambda(T)e^{\beta Z}$ almost surely and the model is not identifiable. C2 holds, in particular, if (K, T_K) is independent of Z.

To derive the rate of convergence, we also assume that:

C8. For some interval $O[T] = [\sigma, \tau]$ with $\sigma > 0$ and $\Lambda_0(\sigma) > 0$, $P(\cap_{j=1}^K \{T_{K,j} \in [\sigma, \tau]\}) = 1$.

C9. $P(K \le k_0) = 1$ for some $k_0 < \infty$.

C10. For some $v_0 \in (0, \infty)$ the function $Z \mapsto E\left(e^{v_0 \mathbb{N}(\tau)} | Z\right)$ is uniformly bounded for $Z \in \mathcal{Z}$.

C11. The observation time points are s_0-separated : i.e. there exists a constant $s_0 > 0$ such that $P(T_{K,j} - T_{K,j-1} \ge s_0$ for all $j = 1, \ldots, K) = 1$. Furthermore, the measure μ_1 is absolutely continuous with respect to Lebesgue measure λ with a derivative $\dot{\mu}_1$ satisfying $\dot{\mu}_1(t) \ge c_0 > 0$ for some positive constant c.

C12. The true baseline mean function Λ_0 is differentiable and the derivative has a positive and finite lower and upper bounds in the observation interval, i.e. there exists a constant $0 < f_0 < \infty$ such that $1/f_0 \le \Lambda_0'(t) \le f_0 < \infty$ for $t \in O[T]$.

C13. For some $\eta \in (0,1)$, $a^T Var(Z|U)a \geq \eta a^T E(ZZ^T|U)a$ a.s. for all $a \in \mathbb{R}^d$ where (U,Z) has distribution $\nu_1/\nu_1(\mathbb{R}^+ \times \mathcal{Z})$.

C14. For some $\eta \in (0,1)$, $a^T Var(Z|U,V)a \geq \eta a^T E(ZZ^T|U,V)a$ a.s. for all $a \in \mathbb{R}^d$ where (U,V,Z) has distribution $\nu_2/\nu_2(\mathbb{R}^{+2} \times \mathcal{Z})$.

Theorem 3.2 *In addition to the conditions required for the consistency, suppose C8, C9, C10, and C13 hold with the constant v_0 in C10 satisfying $v_0 \geq 4k_0(1+\delta_0^{ps})^2$ with $\delta_0^{ps} = \sqrt{c_0\Lambda_0^3(\sigma)/(24\cdot 8f_0)}$ and $\mu_1(\{\tau\}) > 0$. Then*

$$n^{1/3} d_1\left((\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}), (\beta_0, \Lambda_0)\right) = O_p(1).$$

Moreover, if conditions C11, C12, and C14 hold along with the conditions listed above but with the constant v_0 in C10 satisfying $v_0 \geq 4k_0(1+\delta_0)^2$ with $\delta_0 = \sqrt{c_0 s_0^3/(48\cdot 8^2\cdot f_0^4)}$ and $\gamma(\{\tau\}) > 0$, it follows that

$$n^{1/3} d_2\left((\hat{\beta}_n, \hat{\Lambda}_n), (\beta_0, \Lambda_0)\right) = O_p(1).$$

Remark 3.2. Conditions C8, C9, C10 and C12 are sufficient for validity of Theorem 3.2, but they are probably not necessary. Conditions C9 and C10 are mainly used in deriving the rate of convergence when the counting process $\mathbb{N}$ is allowed to be general (but satisfying the mean model (1)). C8 and C9 are generally true in applications. If the counting process is uniformly bounded (which can be justified in many applications) or is (conditionally on covariates) a Poisson process, C10 holds for all $v_0 > 0$. (Thus Theorem 3.2 does give a result under the two different Poisson model assumptions.)

Remark 3.3. It is also worth pointing out that the metrics d_1 and d_2 are closely related. By Lemma 8.1, we can conclude that the two metrics are equivalent under C9 and therefore the consistency and rate of convergence results for the Maximum Likelihood Estimator $(\hat{\beta}_n, \hat{\Lambda}_n)$ hold under the metric d_1 as well.

Remark 3.4. Condition C13 can be justified in many applications. By the Markov inequality, it is easy to see that condition C7 implies that $E(ZZ^T)$

is a positive-definite matrix. If we further assume that, writing E_1 and Var_1 for means and variances under the probability measure $\nu_1/\nu_1(\mathbb{R}^+ \times \mathcal{Z})$, $Var_1(Z|U)$ and $E_1(ZZ^T|U)$ are positive-definite matrices, and if we denote $\lambda_1 = \max\{\text{eigenvalue}(E_1(ZZ^T|U)\}$ and $\lambda_d^* = \min\{\text{eigenvalue}(Var_1(Z|U))\}$, then apparently $\lambda_d^* \leq \lambda_1$. Therefore, for any $a \in \mathbb{R}^d$

$$a^T Var_1(Z|U)a \quad \geq \quad a^T \lambda_d^* a = \frac{\lambda_d^*}{\lambda_1} a^T \lambda_1 a \geq \frac{\lambda_d^*}{\lambda_1} a^T E_1(ZZ^T|U)a.$$

Thus, condition C13 holds by taking $\eta \leq \lambda_d^*/\lambda_1$. Note that both λ_1 and λ_d^* depend on U in general and the argument here works assuming that this ratio has a positive lower bound uniformly in U. We can justify C14 similarly.

Although the overall convergence rate for both the maximum pseudo- and likelihood estimators is of the order $n^{-1/3}$, the rate of convergence for the estimators of the regression parameter, as usual, may still be $n^{-1/2}$. Similar to the results of Huang (1996) for the Cox model with current status data, we can establish asymptotic normality of both $\hat{\beta}_n^{ps}$ and $\hat{\beta}_n$.

Theorem 3.3 *Under the same conditions assumed in Theorem 3.2, the estimators $\hat{\beta}_n^{ps}$ and $\hat{\beta}_n$ are asymptotically normal:*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \to_d Z \sim N_d\left(0, A^{-1}B\left(A^{-1}\right)^T\right), \tag{5}$$

and

$$\sqrt{n}(\hat{\beta}_n^{ps} - \beta_0) \to_d Z^{ps} \sim N_d\left(0, (A^{ps})^{-1}B^{ps}\left((A^{ps})^{-1}\right)^T\right) \tag{6}$$

where

$$\begin{aligned} B &= E\left\{\sum_{j,j'=1}^{K} C_{j,j'}(Z)\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j}, T_{K,j'}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j}, T_{K,j'}\right)}\right]^{\otimes 2}\right\}, \\ A &= E\left\{\sum_{j=1}^{K} \Delta\Lambda_{0Kj} e^{\beta_0^T Z}\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}\right]^{\otimes 2}\right\}, \end{aligned}$$

$$
\begin{aligned}
C_{j,j'}(Z) &= Cov\left[\Delta\mathbb{N}_{Kj}, \Delta\mathbb{N}_{Kj'}|Z, K, \underline{T}_K\right] ,\\
B^{ps} &= E\left\{\sum_{j,j'=1}^{K} C^{ps}_{j,j'}(Z)\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j}\right)}\right]\right.\\
&\qquad\left.\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j'}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j'}\right)}\right]^T\right\},\\
A^{ps} &= E\left\{\sum_{j=1}^{K}\Lambda_{0Kj}e^{\beta_0^T Z}\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j}\right)}\right]^{\otimes 2}\right\},\\
C^{ps}_{j,j'}(Z) &= Cov\left[\mathbb{N}(T_{Kj}), \mathbb{N}(T_{Kj'})|Z, K, T_{K,j}, T_{K,j'}\right] ,\\
\Lambda_{0Kj} &= \Lambda_0(T_{K,j})\,, \qquad \textit{and} \qquad \Delta\Lambda_{0,K,j} = \Lambda_0(T_{K,j}) - \Lambda_0(T_{K,j-1})\,.
\end{aligned}
$$

Note that if the counting process is indeed a conditional Poisson process with the mean function given as specified,

$$
C_{j,j'}(Z) = \begin{cases} \Delta\Lambda_{0Kj}e^{\beta_0^T Z} & , \text{ if } j = j' \\ 0 & , \text{ if } j \neq j'\,. \end{cases}
$$

This yields $B = A = I(\beta_0)$, where $I(\beta_0)$ is the information matrix computed in Wellner, Zhang, and Liu (2004), and thus $A^{-1}B\left(A^{-1}\right)^T = I^{-1}(\beta_0)$. So the estimator $\hat{\beta}_n$ with the conditional Poisson process is efficient. However, under the conditional Poisson process

$$
C^{ps}_{j,j'}(Z) = e^{\beta_0^T Z}\Lambda_{0K(j\wedge j')}
$$

and this yields

$$
\begin{aligned}
B^{ps} &= A^{ps} + 2E_{(K,T_K,Z)}\left\{\sum_{j<j'} e^{\beta_0^T Z}\Lambda_{0Kj}\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j}\right)}\right]\right.\\
&\qquad\left.\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j'}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j'}\right)}\right]^T\right\}\\
&\neq A^{ps}\,.
\end{aligned}
$$

This shows that the semiparametric maximum pseudo-likelihood estimator $\hat{\beta}_n^{ps}$ will not be efficient under the Poisson process assumption.

There is, however, a natural "Poisson regression" model for which the maximum pseudo-likelihood estimator is asymptotically efficient: if we simply assume that the conditional distribution of $(\mathbb{N}(T_{K,1}), \ldots, \mathbb{N}(T_{K,K}))$ given $(K, T_{K,1}, \ldots, T_{K,K}, Z)$ is that of a vector of independent Poisson random variables with means given by $\Lambda(T_{K,j}|Z) = \exp(\beta_0^T Z)\Lambda_0(T_{K,j})$ for $j = 1, \ldots, K$, then

$$C_{j,j'}^{ps}(Z) = \mathrm{Cov}\left[\mathbb{N}(T_{K,j}), \mathbb{N}(T_{K,j'})|Z, K, T_{K,j}, T_{K,j'}\right] = \Lambda(T_{K,j}|Z)1\{j = j'\},$$

so that $B^{ps} = A^{ps} = I_{PoissRegr}(\beta_0)$ and $\hat{\beta}^{ps}$ is asymptotically efficient for this alternative model. In practice, this occurs when $(\mathbb{N}(T_{K,1}), \mathbb{N}(T_{K,2}), \cdots, \mathbb{N}(T_{K,K}))$ consist of cluster Poisson count data in which the counts within a cluster are independent.

4. Numerical Results.

4.1. *Simulation Studies.* We generated data using the same schemes as those given in Zhang (2002). Monte-Carlo bias, standard deviation, and mean squared error of the maximum pseudo-likelihood and maximum likelihood estimates are then compared.

Scenario 1. In this scenario, the data is $\{(Z_i, K_i, \underline{T}_{K_i}^{(i)}, \underline{\mathbb{N}}_{K_i}^{(i)}) : \ i = 1, 2, \ldots, n\}$ with $Z_i = (Z_{i,1}, Z_{i,2}, Z_{i,3})$ where, conditionally on $(Z_i, K_i, \underline{T}_{K_i})$, the counts $\underline{\mathbb{N}}_{K_i}^{(i)}$ were generated from a Poisson process. For each subject, we generate data by the following scheme: $Z_{i,1} \sim Unif(0,1), Z_{i,2} \sim N(0,1), Z_{i,3} \sim Bernoulli(0.5)$; K_i is sampled randomly from the discrete set, $\{1, 2, 3, 4, 5, 6\}$; Given K_i, $\underline{T}_{K_i}^{(i)} = (T_{K_i,1}^{(i)}, T_{K_i,2}^{(i)}, \ldots, T_{K_i,K_i}^{(i)})$ are the order statistics of K_i random observations generated from $Unif(1, 10)$ and rounded to the second decimal point to make the observation times possibly tied; and the panel counts

$$\underline{\mathbb{N}}_{K_i}^{(i)} = (\mathbb{N}^{(i)}(T_{K_i,1}^{(i)}), \mathbb{N}^{(i)}(T_{K_i,2}^{(i)}), \ldots, \mathbb{N}^{(i)}(T_{K_i,K_i}^{(i)}))$$

are generated from the Poisson process with the conditional mean function given by $\Lambda(t|Z_i) = 2t\exp(\beta_0^T Z_i)$, i.e.

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

where $\beta_0 = (\beta_1, \beta_2, \beta_3)^T = (-1.0, 0.5, 1.5)^T$.

For this scenario, we can directly calculate the asymptotic covariance matrices given in Theorem 3.3,

$$\Sigma^{ps} \equiv (A^{ps})^{-1} B^{ps} \left((A^{ps})^{-1}\right)^T = \frac{1582}{17787} W^{-1}$$

and

$$\Sigma = A^{-1} B \left(A^{-1}\right)^T = A^{-1} = \frac{1260}{19179} W^{-1},$$

respectively, where

$$W = E\left\{ e^{\beta_0^T Z} \left[Z - \frac{E\left(Z e^{\beta_0^T Z}\right)}{E\left(e^{\beta_0^T Z}\right)} \right]^{\otimes 2} \right\}.$$

Since it is difficult to calculate the matrix W analytically, we calculated numerically using Mathematica (Wolfram (1966)) to obtain the following numerical results for the asymptotic covariance matrices:

$$\Sigma^{ps} \approx \begin{pmatrix} 0.571104 & 0.000000 & 0.000000 \\ 0.000000 & 0.045304 & 0.000000 \\ 0.000000 & 0.000000 & 0.303752 \end{pmatrix} \tag{7}$$

and

$$\Sigma \approx \begin{pmatrix} 0.421848 & 0.000000 & 0.000000 \\ 0.000000 & 0.033464 & 0.000000 \\ 0.000000 & 0.000000 & 0.224368 \end{pmatrix} \tag{8}$$

We conducted simulation studies with sample sizes of $n = 50$ and $n = 100$, respectively. For each case, the Monte-Carlo sample bias, standard deviation and mean squared error for the semiparametric estimators of the regression parameters and the baseline cumulative mean function at several time points are computed with 1000 repeated samples; see Table 1. We also include the asymptotic standard errors obtained from (7) and (8) in Table 1 to compare with the Monte-Carlo sample standard deviations. The results show that the sample bias for both estimators is small, the standard deviation and mean squared error are smaller for the maximum likelihood method compared to the pseudo-likelihood method and the latter decrease as $n^{-1/2}$ and n^{-1} respectively as sample size increases. Moreover, the standard errors of estimates based on asymptotic theory are all close to the corresponding standard deviations based on the Monte-Carlo simulations, which provides a numerical justification for our asymptotic results in Theorem 3.3.

Table 2 displays the pointwise sample bias, standard deviation and mean squared error for the estimators of the baseline mean function at times, $t = 2, 3, \ldots, 9$. It clearly shows that the bias for both estimators of the mean function, though downward, is negligible compared to the value being estimated. The pointwise standard deviation and mean squared error have the analogous patterns as appeared in Table 1: the maximum likelihood method yields smaller pointwise standard deviations and mean squared errors than the maximum pseudo-likelihood method.

Scenario 2. In this scenario, the data is $\{(Z_i, K_i, \underline{T}^{(i)}_{K_i}, \underline{\mathbb{N}}^{(i)}_{K_i}) : \ i = 1, 2, \ldots, n\}$ with $Z_i = (Z_{i,1}, Z_{i,2}, Z_{i,3})$ and, conditionally on $(Z_i, K_i, \underline{T}^{(i)}_{K_i})$, the counts $\underline{\mathbb{N}}^{(i)}_{K_i}$ were generated from a mixed Poisson process. For each subject, $(Z_i, K_i, \underline{T}^{(i)}_{K_i})$ are generated in exactly the same way as in Scenario 1. The panel counts are, however, generated from a homogeneous Poisson process with a random effect on the intensity: given subject i with covariates Z_i and frailty variable α_i, the counts are generated from the Poisson process with intensity $(\lambda + \alpha_i)\exp(\beta_0^T Z_i)$, where $\lambda = 2.0$ and $\alpha_i \in \{-0.4, 0, 0.4\}$ with probabilities 0.25,0.5,and 0.25, respectively.

In this scenario, the counting process given only the covariates is not a Poisson process. However, the conditional mean function of the counting process given the covariates still satisfies (1) with $\Lambda_0(t) = 2t$ and thus our

TABLE 1
Results of the Monte-Carlo simulation studies for the regression parameters estimates based on 1000 repeated samples for data generated from the conditional Poisson process

	$n = 50$		$n = 100$	
	Pseudo	Full	Pseudo	Full
Estimate of β_1				
BIAS	-0.0052	-0.0032	0.0005	-0.0004
SD	0.1182	0.1048	0.0801	0.0686
ASE	0.1069	0.0919	0.0758	0.0649
MSE$\times 10^2$	1.4005	1.0991	0.6422	0.4703
Estimate of β_2				
BIAS	0.0019	0.0007	-0.0009	-0.0010
SD	0.0352	0.0308	0.0231	0.0199
ASE	0.0301	0.0259	0.0213	0.0183
MSE$\times 10^2$	0.1240	0.0949	0.0533	0.0395
Estimate of β_3				
BIAS	0.0042	0.0017	0.0025	0.0028
SD	0.0811	0.0734	0.0555	0.0467
ASE	0.0779	0.0670	0.0551	0.0474
MSE$\times 10^2$	0.6593	0.5388	0.3087	0.2192

proposed methods are expected to be valid for this case as well. The asymptotic variances given in Theorem 3.3 for this scenario are

$$\Sigma^{ps} = (A^{ps})^{-1} B^{ps} \left((A^{ps})^{-1}\right)^T = \frac{1582}{17787} W^{-1} + \frac{463.12}{17787} W^{-1}\tilde{W}\left(W^{-1}\right)^T$$

and

$$\Sigma = A^{-1}B\left(A^{-1}\right)^T = \frac{1260}{19179} W^{-1} + \frac{7917588}{19179^2} W^{-1}\tilde{W}\left(W^{-1}\right)^T,$$

respectively, where

$$\tilde{W} = E\left\{ e^{2\beta_0^T Z}\left[Z - \frac{E\left(Ze^{\beta_0^T Z}\right)}{E\left(e^{\beta_0^T Z}\right)}\right]^{\otimes 2}\right\}.$$

Using Mathematica (Wolfram (1966)) to compute $\tilde{W}$ numerically we have the following numerical results for the asymptotic covariance matrices:

$$\Sigma^{ps} \approx \begin{pmatrix} 1.172450 & -0.023852 & -0.043178 \\ -0.023852 & 0.108760 & 0.022975 \\ -0.043178 & 0.022975 & 0.448444 \end{pmatrix} \tag{9}$$

TABLE 2
Results of the Monte-Carlo simulation studies for the baseline mean function estimates based on 1000 repeated samples for data generated from the conditional Poisson process

		$n = 50$		$n = 100$	
t	$\Lambda_0(t)$	Pseudo	Full	Pseudo	Full
$t = 2$	4.0				
BIAS$\times 10^2$		-9.5651	-6.0620	-8.5533	-5.3928
SD$\times 10^2$		63.1748	53.7110	48.2530	38.0757
MSE$\times 10^2$		40.8255	29.2161	24.0152	14.7884
$t = 3$	6.0				
BIAS$\times 10^2$		-5.0016	-5.2351	-7.4200	-6.3095
SD$\times 10^2$		81.8728	65.6800	61.3768	46.3566
MSE$\times 10^2$		67.2818	43.4127	38.2216	21.8875
$t = 4$	8.0				
BIAS$\times 10^2$		-9.8907	-5.7329	-6.4369	-6.2300
SD$\times 10^2$		91.0330	75.9852	69.3877	54.1504
MSE$\times 10^2$		83.8484	58.0662	48.5609	29.7110
$t = 5$	10.0				
BIAS$\times 10^2$		-5.6881	-4.5585	-6.3321	-4.2239
SD$\times 10^2$		111.1896	89.6539	82.3000	62.5622
MSE$\times 10^2$		123.9548	80.5861	68.1339	39.3187
$t = 6$	12.0				
BIAS$\times 10^2$		-2.5944	-3.2031	-4.8108	-6.4377
SD$\times 10^2$		124.2285	103.7805	93.3787	72.9975
MSE$\times 10^2$		154.3945	107.8066	87.4272	53.7008
$t = 7$	14.0				
BIAS$\times 10^2$		-1.9150	-3.9224	-5.5356	-4.5218
SD$\times 10^2$		140.3992	118.2790	102.0303	82.1771
MSE$\times 10^2$		197.1559	140.0531	104.4080	67.7352
$t = 8$	16.0				
BIAS$\times 10^2$		-8.1166	-6.8985	-5.8738	-4.9468
SD$\times 10^2$		156.5389	134.1874	117.1169	89.3958
MSE$\times 10^2$		245.7032	180.5385	137.5087	80.1608
$t = 9$	18.0				
BIAS$\times 10^2$		-5.9417	-6.0929	-7.8391	-6.2323
SD$\times 10^2$		178.2617	145.5261	126.4885	100.3230
MSE$\times 10^2$		318.1255	212.1495	160.6079	101.0495

and

$$\Sigma \approx \begin{pmatrix} 0.918986 & -0.019718 & -0.035696 \\ -0.019718 & 0.085924 & 0.018994 \\ -0.035696 & 0.018994 & 0.343985 \end{pmatrix}. \tag{10}$$

As in Scenario 1, we conducted simulation studies with sample sizes of $n = 50$ and $n = 100$, respectively. For each case, the Monte-Carlo sample bias, standard deviation and mean squared error for the semiparametric estimators of the regression parameters and the baseline cumulative mean function at several time points are computed with 1000 repeated samples. The results are shown in Tables 3 and 4, respectively. We observe the same phenomenon as appeared in Scenario 1 that both standard deviation and mean squared error using the maximum likelihood method are smaller than those using the pseudo-likelihood method while the bias is relatively small. We also note that both standard deviation and mean squared error of semiparametric estimators are relatively larger than their counterpart in Scenario 1. This may be caused by violation of the assumption of a conditional Poisson process given only the covariates. We also include the asymptotic standard errors of the regression parameter estimates based on (9) and (10) in Table 3. Again the standard errors derived from the asymptotic theory are all close to the standard deviations based Monte-Carlo simulations.

These simulation studies provided numerical support for the statement that the proposed semiparametric estimation methods are robust against the underlying conditional Poisson process assumption. These methods are valid as long as the proportional mean function model (1) holds. We have also conducted several analytical analyses to compare the semiparametric efficiency between the pseudo-likelihood and maximum likelihood estimation methods. There is considerable evidence that the maximum likelihood method (based on the Poisson process assumption) is more efficient than the pseudo-likelihood method both on and off the Poisson model, with large differences occurring when K is heavy tailed. The detailed analytical results are presented in Wellner, Zhang, and Liu (2004).

TABLE 3

Results of the Monte-Carlo simulation studies for the regression parameter estimates based on 1000 repeated samples for data generated from the mixed Poisson process

	$n = 50$		$n = 100$	
	Pseudo	Full	Pseudo	Full
Estimate of β_1				
BIAS	0.0093	0.0092	0.0053	0.0043
SD	0.1593	0.1428	0.1125	0.1030
ASE	0.1531	0.1356	0.1083	0.0959
MSE$\times 10^2$	2.5460	2.0484	1.2690	1.0635
Estimate of β_2				
BIAS	-0.0025	-0.0012	-0.0003	-0.0006
SD	0.0458	0.0412	0.0302	0.0274
ASE	0.0466	0.0415	0.0330	0.0293
MSE$\times 10^2$	0.2108	0.1698	0.0914	0.0753
Estimate of β_3				
BIAS	0.0030	0.0037	0.0033	0.0006
SD	0.0972	0.0854	0.0697	0.0611
ASE	0.0947	0.0829	0.0670	0.0587
MSE$\times 10^2$	0.9358	0.7308	0.4864	0.3737

4.2. *A Real Data Example.* Using the semiparametric methods proposed in the preceding sections, we analyze the bladder tumor data, extracted from Andrews and Herzberg (1985, pp. 253-260). This data set comes from a bladder tumor study conducted by the Veterans Administration Cooperative Urological Research (Byar et al., 1977). In the study, a randomized clinical trial of three treatments, placebo, pyridoxine pills and thiotepa instillation into the bladder was conducted for patients with superficial bladder tumor when entering the trial. At each follow-up visit, tumors were counted, measured and then removed if observed, and the treatment was continued. The treatment effects, especially the thiotepa instillation, on suppressing the recurrence of bladder tumor have been explored by many authors, for example, Wei, Lin, and Weissfeld (1989), Sun and Wei (2000), Wellner and Zhang (2000) and Zhang (2002).

In this paper, we study the proportional mean model that has been pro-

TABLE 4
Results of the Monte-Carlo simulation studies for the baseline mean function estimates based on 1000 repeated samples for data generated from the mixed Poisson process

		$n = 50$		$n = 100$	
t	$\Lambda_0(t)$	Pseudo	Full	Pseudo	Full
$t = 2$	4.0				
BIAS$\times 10^2$		-9.4431	-7.4610	-7.2387	-4.2480
SD$\times 10^2$		70.2687	59.8600	51.7970	41.8410
MSE$\times 10^2$		50.2686	36.3889	27.3532	17.6872
$t = 3$	6.0				
BIAS$\times 10^2$		-11.2363	-8.0262	-7.9619	-3.6252
SD$\times 10^2$		88.4041	71.9307	67.5299	51.8379
MSE$\times 10^2$		79.4154	52.3844	46.2368	27.0031
$t = 4$	8.0				
BIAS$\times 10^2$		-9.9219	-9.1975	-8.8849	-4.0901
SD$\times 10^2$		111.2754	88.2578	80.9590	62.1734
MSE$\times 10^2$		124.8066	78.7404	66.3330	38.8226
$t = 5$	10.0				
BIAS$\times 10^2$		-12.9604	-14.4471	-7.4010	-4.0248
SD$\times 10^2$		126.0807	102.5314	95.6544	73.2416
MSE$\times 10^2$		160.6432	107.2140	92.0454	53.8053
$t = 6$	12.0				
BIAS$\times 10^2$		-10.9481	-12.2809	-10.2174	-5.5298
SD$\times 10^2$		150.2711	123.3147	110.9413	85.4766
MSE$\times 10^2$		227.0126	153.3269	124.1238	73.3682
$t = 7$	14.0				
BIAS$\times 10^2$		-11.1382	-14.0965	-6.2486	-3.6736
SD$\times 10^2$		170.4857	139.0968	126.9569	100.1551
MSE$\times 10^2$		291.8943	195.4662	161.5709	100.4413
$t = 8$	16.0				
BIAS$\times 10^2$		-19.2660	-15.2879	-8.6318	-4.5795
SD$\times 10^2$		193.3730	159.0769	141.7516	110.1551
MSE$\times 10^2$		377.6430	255.3917	201.6801	121.5513
$t = 9$	18.0				
BIAS$\times 10^2$		-14.8591	-18.7451	-9.7816	-6.7515
SD$\times 10^2$		206.1532	174.6755	151.6876	124.1812
MSE$\times 10^2$		427.1994	308.6290	231.0481	154.6655

TABLE 5
Semiparametric inference for the bladder tumor study based on 100 bootstrap samples from the original data set.

Variable	Method	$\hat{\beta}$	$\hat{sd}(\hat{\beta})$	$\hat{\beta}/\hat{sd}(\hat{\beta})$	p-value
Z_1					
	pseudo	0.1446	0.0518	2.7915	0.0052
	full	0.2069	0.0661	3.1301	0.0017
Z_2					
	pseudo	-0.0450	0.0554	0.8123	0.4166
	full	-0.0355	0.0748	0.4166	0.6351
Z_3					
	pseudo	0.1951	0.3410	0.5721	0.5672
	full	0.0664	0.4438	0.1496	0.8811
Z_4					
	pseudo	-0.6881	0.2854	2.4110	0.0159
	full	-0.7972	0.3150	2.5308	0.0114

posed by Sun and Wei (2000) and Zhang (2002),

$$E\{\mathbb{N}(t)|Z\} = \Lambda_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4), \tag{11}$$

where Z_1 and Z_2 represent the number and size of bladder tumors at the beginning of the trial, and Z_3 and Z_4 are the indicators for the pyridoxine pill and thiotepa instillation treatments, respectively. We choose $\beta^{(0)} = (0,0,0,0)$ to start our iterative algorithm and $\eta = 10^{-10}$ for the convergence criteria to stop the algorithm. Since the asymptotic variances are difficult to estimate, we adopt the bootstrap procedure to estimate the asymptotic standard deviation of the semiparametric estimates of the regression parameters. We used a bootstrap sample size of 100 and evaluated the proposed estimators for each sampled data set. The sample standard deviation of the estimates based on these 100 bootstrap samples is used to estimate the asymptotic standard deviation. The inference based on the bootstrap estimator for asymptotic standard deviation is given in Table 5.

Both methods yield the same conclusion that the baseline number of tumors (the number of tumors observed when entering the trial) significantly affect the recurrence of the tumor (p-value= 0.0052 and 0.0017, respectively, for the pseudo-likelihood and maximum likelihood methods), and the thiotepa instillation treatment appears to reduce the recurrence of tumor significantly. (p-value= 0.0159 and 0.0114, respectively for the pseudo-likelihood and maximum likelihood methods). However, we also notice that the maximum likelihood method, in contrast to what we have observed through both the simulation and analytical studies, yields larger standard deviations compared to the pseudo-likelihood method. Violation of the proportional mean model (11) for this data set could be the explanation for this result. Model diagnostics for modeling of panel count data remains an open problem for future research.

All the numerical experiments in this paper are implemented using R-*Language*. Readers can request the source codes from the second author.

5. Final Remarks and Further Problems. As in the case of panel count data without covariates studied in Sun and Kalbfleisch (1995) and Wellner and Zhang (2000), the pseudo-likelihood estimation method for the semiparametric proportional mean model with panel count data proposed and studied in Zhang (1998, 2002) also has advantages in terms of computational simplicity. However, the maximum pseudo-likelihood estimator is inefficient relative to the maximum likelihood estimator, and could be very inefficient in some cases when the distribution of K is heavy-tailed (Wellner, Zhang, and Liu, 2004). In such cases it is clear that we will want to avoid the pseudo-likelihood estimator, and the computational effort required by the maximum likelihood estimator can be justified by the consequent gain in efficiency.

Our derivation of the asymptotic normality of the maximum likelihood estimator of the regression parameters results in a relatively complicated expression for the asymptotic variance which is difficult to estimate directly. Hence bootstrap inference procedures are an option, and then an efficient algorithm for computing the maximum likelihood estimator is imperative in order to perform the bootstrapping.

In our experiments of simulations, the doubly iterative algorithm using the Newton-Raphson and modified iterative convex minorant methods proposed in this manuscript, although feasible, is not an optimal solution, since the computing time is quite large for the maximum likelihood estimator. It is therefore worth making further efforts toward efficient computing algorithms for semiparametric estimation problems, since it is such a common problem in survival analysis where both the regression parameters and baseline hazard or intensity function need to be estimated.

In view of the efficiency advantages but computational complexity of the maximum likelihood estimator, one may be able to find some compromise or hybrid estimators between the maximum pseudo- and the maximum likelihood estimators which have the computational advantages of the former and the efficiency advantages of the latter.

Profile likelihood inference may also be feasible in this type of problem. Murphy and van der Vaart (1997, 1999, 2000) have developed semiparametric likelihood ratio inference procedures for some related interval censoring models. Extension of the theorems of Murphy and van der Vaart (1997, 1999 and 2000) to semiparametric pseudo-likelihood ratio inference procedures would be useful in making inference about the regression parameter with panel count data and is currently under investigation by the authors.

The extension of the non-standard likelihood ratio procedures and methods of Banerjee and Wellner (2001) for the present model to give tests and confidence intervals for $\Lambda_0(t)$ is also of interest in practice. See Banerjee (2005) for progress in this direction.

One can also explore other models for the mean function replacing the proportional mean model given by (1), such as additive and additive- multiplicative mean model which may also be useful and interesting in practice. Also it is worth exploring if there is computational or efficiency advantages to using the MLE's for one of the class of Mixed Poisson Process $(\mathbb{N}|Z)$, for example the Negative-Binomial model. Further comparisons with the work of Dean and Balshaw (1997), Hougaard, Lee, and Whitmore (1997), and Lawless (1987a,b) would be useful.

6. Asymptotic theory: Proofs.

We use empirical process theory to study the asymptotic properties of the semiparametric maximum pseudo-likelihood and maximum likelihood estimators. The proof of Theorem 3.1 is closely related to the proof of Theorems 4.1 and 4.2 of Wellner and Zhang (2000). The rate of convergence is derived based on the general theorem for the rate of convergence given in Theorem 3.2.5 of van der Vaart and Wellner (1996). The asymptotic normality proofs for both $\hat{\beta}_n^{ps}$ and $\hat{\beta}_n$ are based on the general theorem for M-estimation of regression parameters in the presence of a nonparametric nuisance parameter which is given (and proved) in Appendix A.

First some comments about our miss-specified situation. Because we assume that the true distribution P is not necessarily in either of the Poisson models used to form likelihoods, we are dealing with estimation under a miss-specified model. However, because we assume that the conditional mean model given by (1) is correct, the difficulty is not as serious would be in the general case of complete miss-spefication. Letting $m_\theta^{ps}(X) \equiv \log p_\theta^{ps}(X)$ and $m_\theta(X) \equiv \log p_\theta(X)$ for $P_\theta^{ps} \in \mathcal{P}^{ps}$ and $P_\theta \in \mathcal{P}$ respectively, we have, with P denoting the distribution of X,

$$
\begin{aligned}
P \log\left(\frac{p_{\theta_0}}{p_\theta}\right) &= P_{\theta_0} \log\left(\frac{p_{\theta_0}}{p_\theta}\right) = K(P_{\theta_0}, P_\theta) \\
P \log\left(\frac{p_{\theta_0}^{ps}}{p_\theta^{ps}}\right) &= P_{\theta_0}^{ps} \log\left(\frac{p_{\theta_0}^{ps}}{p_\theta^{ps}}\right) = K(P_{\theta_0}^{ps}, P_\theta^{ps})
\end{aligned}
\tag{12}
$$

where the equalities follow from (1) together with the (natural) exponential family form of the respective Poisson models. Thus $P(m_\theta - m_{\theta_0})$ (and $Pm_\theta^{ps} - m_{\theta_0}^{ps})$) behaves as if the true P is in the model $\mathcal{P}$ (or $\mathcal{P}^{ps}$) respectively. This also implies that $K(P, P_\theta) = P\log(p/p_\theta)$ is minimized over $P_\theta \in \mathcal{P}$ by P_{θ_0}:

assuming that $P\log(p/p_{\theta_0}) < \infty$, it follows that

$$\begin{aligned} P\log\left(\frac{p}{p_{\theta_0}}\right) &= P\log\left(\frac{p}{p_\theta}\right) + P\log\left(\frac{p_\theta}{p_{\theta_0}}\right) \\ &= P\log\left(\frac{p}{p_\theta}\right) - P_{\theta_0}\log\left(\frac{p_{\theta_0}}{p_\theta}\right) \qquad \text{by (12)} \\ &= P\log\left(\frac{p}{p_\theta}\right) - K(P_{\theta_0}, P_\theta) \le P\log\left(\frac{p}{p_\theta}\right) \end{aligned}$$

with equality if $\theta = \theta_0$.

Moreover, for Poisson distributions P_μ and P_ν on $\{0, 1, 2, \ldots\}$ with $\mu, \nu > 0$, we have

$$K(P_\mu, P_\nu) = \nu h(\mu/\nu), \tag{13}$$

where $h(x) \equiv x(\log(x) - 1) + 1$. Since $h(x) \ge (1/4)(x-1)^2$ for $0 \le x \le 5$, it follows that

$$\begin{aligned} K(P_\mu, P_\nu) &= \nu h(\mu/\nu) \ge \nu(\mu/\nu - 1)^2/4 = \frac{1}{4}\frac{(\mu-\nu)^2}{\nu} \\ &\quad \text{if } \ 0 < \mu/\nu \le 5. \end{aligned} \tag{14}$$

Proof of Theorem 3.1: Zhang (2002) has given a proof for the first part of the theorem concerning the semiparametric maximum pseudo-likelihood estimator. Unfortunately however, his proof of theorem 1 on pages 47 and 48 is not correct (and in particular the conditions imposed do not suffice for identifiability as claimed). Here we give proofs for both the maximum pseudo-likelihood and maximum-likelihood estimators.

We first prove the claims concerning the pseudo-likelihood estimators $(\hat\beta_n^{ps}, \hat\Lambda_n^{ps})$. Let $\mathbb{M}_n^{ps}(\theta) = n^{-1} l_n(\beta, \Lambda) = \mathbb{P}_n m_\theta^{ps}(X)$ and $\mathbb{M}^{ps}(\theta) = P m_\theta^{ps}(X)$, where

$$m_\theta^{ps}(X) = \sum_{j=1}^{K} \left\{ \mathbb{N}_{Kj} \log \Lambda_{Kj} + \mathbb{N}_{Kj}\beta^T Z - \Lambda_{Kj}\exp(\beta^T Z) \right\} .$$

First, we show that $\mathbb{M}^{ps}$ has $\theta_0 = (\beta_0, \Lambda_0)$ as its unique maximum point. Computing the expectation conditionally on $(Z, K, \underline{T}_K)$ and using (13) yields

$$\mathbb{M}^{ps}(\theta_0) - \mathbb{M}^{ps}(\theta) = \int \Lambda(u)\exp(\beta^T z) h\left[\frac{\Lambda_0(u)\exp(\beta_0^T z)}{\Lambda(u)\exp(\beta^T z)}\right] d\nu_1(u,z),$$

where $h(x) = x\log(x) - x + 1$. The function $h(x)$ satisfies $h(x) \geq 0$ for $x > 0$ with equality holding only at $x = 1$. Hence $\mathbb{M}^{ps}(\theta_0) \geq \mathbb{M}^{ps}(\theta)$ and $\mathbb{M}^{ps}(\theta_0) = \mathbb{M}^{ps}(\theta)$ if and only if

$$\frac{\Lambda_0(u)\exp(\beta_0^T z)}{\Lambda(u)\exp(\beta^T z)} = 1 \qquad \text{a.e. with respect to} \quad \nu_1. \tag{15}$$

This implies that

$$\beta = \beta_0 \;\text{ and }\; \Lambda(u) = \Lambda_0(u) \qquad \text{a.e. with respect to} \qquad \mu_1 \tag{16}$$

by C2 and C7. Here is a proof of this claim: Let

$$\begin{aligned} f_1(u) &= \Lambda(u) - \Lambda_0(u), \quad & f_2(u) &= \Lambda_0(u)\,, \\ h_1(z) &= \exp(\beta^T z)\,, & h_2(z) &= \exp(\beta^T z) - \exp(\beta_0^T z)\,. \end{aligned}$$

Then (15) implies that

$$\Lambda_0(u)\exp(\beta_0^T z) = \Lambda(u)\exp(\beta^T z)$$

a.e. ν_1, or, equivalently

$$\begin{aligned} 0 &= \{\Lambda(u) - \Lambda_0(u)\}e^{\beta^T z} + \Lambda_0(u)(e^{\beta^T z} - e^{\beta_0^T z}) \\ &= f_1(u)h_1(z) + f_2(u)h_2(z) \qquad \text{a.e. } \nu_1\,. \end{aligned}$$

Since $\mu_1 \times H$ is absolutely continuous with respect to ν_1 by assumption C2, equality holds in the last display a.e. with respect to $\mu_1 \times H$, and it follows (by multiplying across the identity in the last display by ab, integrating with respect to the measure $\mu_1 \times H$, and then applying Fubini's theorem) that

$$0 = \int f_1 a d\mu_1 \int h_1 b dH + \int f_2 a d\mu_1 \int h_2 b dH$$

for all measurable functions $a = a(u)$ and $b = b(z)$. Choosing $a = f_1 1_A$ for $A \in \mathcal{B}_1$ and $b = h_1 1_B$ for $B \in \mathcal{B}_d$ yields

$$0 \;=\; \int f_1^2 1_A d\mu_1 \int h_1^2 1_B dH + \int f_1 f_2 1_A d\mu_1 \int h_1 h_2 1_B dH\,;$$

choosing $a = f_2 1_A$ (for the same $A \in \mathcal{B}_1$) and $b = h_2 1_B$ (again for the same set $B \in \mathcal{B}_d$ yields

$$0 \;=\; \int f_1 f_2 1_A d\mu_1 \int h_1 h_2 1_B dH + \int f_2^2 1_A d\mu_1 \int h_2^2 1_B dH\,.$$

Thus we have

$$\int f_1^2 1_A d\mu_1 \int h_1^2 1_B dH = -\int f_1 f_2 1_A d\mu_1 \int h_1 h_2 1_B dH = \int f_2^2 1_A d\mu_1 \int h_2^2 1_B dH$$

for all $A \in \mathcal{B}_1$ and $B \in \mathcal{B}_d$. By Fubini's theorem this yields

$$\int_{A\times B} f_1^2 h_1^2 d(\mu_1 \times H) = \int_{A\times B} f_2^2 h_2^2 d(\mu_1 \times H)\,.$$

for all such sets A, B. But this implies that the measures γ_1 and γ_2 defined by

$$\gamma_1(A\times B) = \int_{A\times B} f_1^2 h_1^2 d(\mu_1 \times H), \qquad \gamma_2(A\times B) = \int_{A\times B} f_2^2 h_2^2 d(\mu_1 \times H)$$

are equal for all the product sets $A \times B$, and hence by a standard monotone class argument we conclude that $\gamma_1 = \gamma_2$ as measures on $([0,\tau]\times\mathbb{R}^d, \mathcal{B}_1[0,\tau]\times\mathcal{B}_d)$. It follows that

$$f_1^2(u)h_1^2(z) = f_2^2(u)h_2^2(z)$$

a.e. with respect to $\mu_1 \times H$. Thus we conclude that

$$\frac{f_1^2(u)}{f_2^2(u)} = \frac{h_2^2(z)}{h_1^2(z)} \quad \text{a.e. on} \quad \{(u,z):\ f_1^2(u) > 0,\ h_1^2(z) > 0\}\,,$$

or, in other words,

$$\left(\frac{\Lambda(u)}{\Lambda_0(u)} - 1\right)^2 = (1 - \exp((\beta_0 - \beta)^T z))^2$$

a.e. with respect to $\mu_1 \times H$. This implies that (16) holds in view of C7. Integrating across this identity with respect to μ_1 yields

$$\int \left(\frac{\Lambda(u)}{\Lambda_0(u)} - 1\right)^2 d\mu_1(u) = (1 - \exp((\beta_0 - \beta)^T z))^2 \mu_1([0,\tau]) \qquad \text{a.e. } H\,,$$

and hence the right side is a constant a.e. H. But this implies that $\beta = \beta_0$ in view of C7. Combining this with the last display shows that (16) holds.

For any given $\epsilon > 0$, let $\tilde\theta_n^{ps} = (\hat\beta_n^{ps}, (1-\epsilon)\hat\Lambda_n^{ps} + \epsilon\Lambda_0) = (\hat\beta_n^{ps}, \hat\Lambda_n^{ps}) + \epsilon(0, \Lambda_0 - \hat\Lambda_n^{ps})$. Since

$$\mathbb{M}_n^{ps}(\hat\theta_n^{ps}) \geq \mathbb{M}_n^{ps}(\tilde\theta_n^{ps}) = \mathbb{M}_n^{ps}\left(\hat\theta_n^{ps} + \epsilon(0, \Lambda_0 - \hat\Lambda_n^{ps})\right),$$

it follows that

$$\begin{aligned} 0 &\geq \lim_{\epsilon\downarrow 0} \frac{\mathbb{M}_n^{ps}(\hat\theta_n^{ps} + \epsilon(0,\Lambda_0 - \hat\Lambda_n^{ps})) - \mathbb{M}_n^{ps}(\hat\theta_n^{ps})}{\epsilon} \\ &= \mathbb{P}_n\left[\sum_{j=1}^K \left\{\frac{\mathbb{N}_{Kj}}{\hat\Lambda_{nKj}^{ps}\exp(\hat\beta_n^{psT}Z)} - 1\right\}(\Lambda_{0Kj} - \hat\Lambda_{nKj}^{ps})\exp(\hat\beta_n^{psT}Z)\right] \end{aligned}$$

where $\hat\Lambda_{nKj}^{ps} = \hat\Lambda_n^{ps}(T_{K,j})$. This yields

$$\begin{aligned} &\mathbb{P}_n\left[\sum_{j=1}^K \left\{\mathbb{N}_{Kj}\frac{\Lambda_{0Kj}}{\hat\Lambda_{nKj}^{ps}} + \hat\Lambda_{nKj}^{ps}\exp(\hat\beta_n^{psT}Z)\right\}\right] \\ &\quad\leq \mathbb{P}_n\left\{\sum_{j=1}^K \left(\mathbb{N}_{Kj} + \Lambda_{0Kj}\exp(\hat\beta_n^{psT}Z)\right)\right\} \\ &\quad\leq C\mathbb{P}_n\left\{\sum_{j=1}^K (\mathbb{N}_{K,j} + \Lambda_0(T_{K,j}))\right\} \\ &\quad\to_{a.s.} CP\left\{\sum_{j=1}^K (\mathbb{N}_{K,j} + \Lambda_0(T_{K,j}))\right\} < \infty \end{aligned}$$

by C1 - C3 and the strong law of large numbers. (Here C represents a constant. In the sequel the C appearing in different lines may represent different constants.) On the other hand,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}_n & \left[\sum_{j=1}^{K}\left\{\mathbb{N}_{Kj}\frac{\Lambda_{0Kj}}{\hat{\Lambda}^{ps}_{nKj}} + \hat{\Lambda}^{ps}_{nKj}\exp(\hat{\beta}^{psT}_n Z)\right\}\right] \\
& \geq \limsup_{n\to\infty} \mathbb{P}_n \left\{\sum_{j=1}^{K} 1_{[b,\tau]}(T_{K,j})\hat{\Lambda}^{ps}_{nKj}\exp(\hat{\beta}^{psT}_n Z)\right\} \\
& \geq C \limsup_{n\to\infty} \hat{\Lambda}^{ps}_n(b)\mathbb{P}_n\left(\sum_{j=1}^{K} 1_{[b,\tau]}(T_{K,j})\right) \\
& = C \limsup_{n\to\infty} \hat{\Lambda}^{ps}_n(b)\,\mu_1([b,\tau])\,.
\end{aligned}$$

Hence $\hat{\Lambda}^{ps}_n(t)$ is uniformly bounded almost surely for $t \in [0,b]$ if $\mu_1([b,\tau]) > 0$ for some $0 < b < \tau$ or for $t \in [0,\tau]$ if $\mu_1(\{\tau\}) > 0$. Then by the Helly-Selection Theorem and the compactness of $\mathcal{R}\times\mathcal{F}$, it follows that $\hat{\theta}^{ps}_n = \left(\hat{\beta}^{ps}_n, \hat{\Lambda}^{ps}_n\right)$ has a subsequence $\hat{\theta}^{ps}_{n'} = \left(\hat{\beta}^{ps}_{n'}, \hat{\Lambda}^{ps}_{n'}\right)$ converging to $\theta^+ = (\beta^+, \Lambda^+)$, where Λ^+ is an increasing bounded function defined on $[0,b]$ for a $b<\tau$ and it can be defined on $[0,\tau]$ if $\mu_1(\{\tau\}) > 0$. Following the same argument as in proving Theorem 4.1 of Wellner and Zhang (2000), we can show that $\mathbb{M}^{ps}(\theta^+) \geq \mathbb{M}^{ps}(\theta_0)$. Since $\mathbb{M}^{ps}(\theta_0) \geq \mathbb{M}^{ps}(\theta^+)$ by the argument above (15) we conclude that $\mathbb{M}^{ps}(\theta^+) = \mathbb{M}^{ps}(\theta_0)$. Then (16) implies that $\beta^+ = \beta_0$ and $\Lambda^+ = \Lambda_0$ a.e. in μ_1. Finally, the dominated convergence theorem yields the strong consistency of $(\hat{\beta}^{ps}_n, \hat{\Lambda}^{ps}_n)$ in the metric d_1.

Now we turn to the maximum likelihood estimator. Let $\mathbb{M}_n(\theta) = n^{-1}l_n(\beta,\Lambda) = \mathbb{P}_n m_\theta(X)$ and $\mathbb{M}(\theta) = P m_\theta(X)$, where

$$m_\theta(X) = \sum_{j=1}^{K}\left\{\Delta\mathbb{N}_{Kj}\log\Delta\Lambda_{Kj} + \Delta\mathbb{N}_{Kj}\beta^T Z - \Delta\Lambda_{Kj}\exp(\beta^T Z)\right\}\,.$$

First, we show that $\mathbb{M}$ has $\theta_0 = (\beta_0, \Lambda_0)$ as its unique maximum point. Computing the expectation conditionally on $(Z, K, \underline{T}_K)$ and using (13) yields

$$\begin{aligned}
\mathbb{M}(\theta_0) - \mathbb{M}(\theta) \quad = \quad & \int \{\Lambda(v) - \Lambda(u)\}\exp(\beta^T z) \\
& \qquad h\left[\frac{\{\Lambda_0(v) - \Lambda_0(u)\}\exp(\beta_0^T z)}{\{\Lambda(v) - \Lambda(u)\}\exp(\beta^T z)}\right] d\nu_2(u,v,z),
\end{aligned}$$

where $h(x) = x\log(x) - x + 1$. The function $h(x)$ satisfies $h(x) \geq 0$ for $x > 0$ with equality holding only at $x = 1$. Hence $\mathbb{M}(\theta_0) \geq \mathbb{M}(\theta)$ and $\mathbb{M}(\theta_0) = \mathbb{M}(\theta)$ if and only if

$$\text{(17)} \qquad \frac{\{\Lambda_0(v) - \Lambda_0(u)\}\exp(\beta_0^T z)}{\{\Lambda(v) - \Lambda(u)\}\exp(\beta^T z)} = 1 \quad \text{a.e. with respect to} \quad \nu_2.$$

This implies that

$$\text{(18)} \qquad \beta = \beta_0 \ \text{ and } \ \Lambda(v) - \Lambda(u) = \Lambda_0(v) - \Lambda_0(u) \\ \text{a.e. with respect to} \quad \mu_2$$

by C2 and C7. Here is a proof of this claim: Let

$$\begin{aligned} &f_1(u,v) = \Lambda(v) - \Lambda(u) - (\Lambda_0(v) - \Lambda_0(u)), \\ &f_2(u,v) = \Lambda_0(v) - \Lambda_0(u)\,, \\ &h_1(z) = \exp(\beta^T z)\,, \qquad h_2(z) = \exp(\beta^T z) - \exp(\beta_0^T z)\,. \end{aligned}$$

Then (17) implies that

$$\{\Lambda_0(v) - \Lambda_0(u)\}\exp(\beta_0^T z) = \{\Lambda(v) - \Lambda(u)\}\exp(\beta^T z)$$

a.e. ν_2, or, equivalently

$$\begin{aligned} 0 &= \{\Lambda(v) - \Lambda(u) - (\Lambda_0(v) - \Lambda_0(u))\}\, e^{\beta^T z} + (\Lambda_0(v) - \Lambda_0(u))(e^{\beta^T z} - e^{\beta_0^T z}) \\ &= f_1(u,v)h_1(z) + f_2(u,v)h_2(z) \qquad \text{a.e.} \ \ \nu_2\,. \end{aligned}$$

Since $\mu_2 \times H$ is absolutely continuous with respect to ν_2 by assumption C2, it follows by multiplying across the identity in the last display by ab, integrating with respect to the measure $\mu_2 \times H$, and then applying Fubini's theorem that

$$0 \ = \ \int f_1 a d\mu_2 \int h_1 b dH + \int f_2 a d\mu_2 \int h_2 b dH$$

for all measurable functions $a = a(u,v)$ and $b = b(z)$. Choosing $a = f_1 1_A$ for $A \in \mathcal{B}_2[0,\tau]$ and $b = h_1 1_B$ for $B \in \mathcal{B}_d$ yields

$$0 \ = \ \int f_1^2 1_A d\mu_2 \int h_1^2 1_B dH + \int f_1 f_2 1_A d\mu_2 \int h_1 h_2 1_B dH\,;$$

choosing $a = f_2 1_A$ (for the same $A \in \mathcal{B}_2[0,\tau]$) and $b = h_2 1_B$ (again for the same set $B \in \mathcal{B}_d$ yields

$$0 = \int f_1 f_2 1_A d\mu_2 \int h_1 h_2 1_B dH + \int f_2^2 1_A d\mu_2 \int h_2^2 1_B dH .$$

Thus we have

$$\int f_1^2 1_A d\mu_2 \int h_1^2 1_B dH = -\int f_1 f_2 1_A d\mu_2 \int h_1 h_2 1_B dH = \int f_2^2 1_A d\mu_2 \int h_2^2 1_B dH$$

for all $A \in \mathcal{B}_2[0,\tau]$ and $B \in \mathcal{B}_d$. By Fubini's theorem this yields

$$\int_{A\times B} f_1^2 h_1^2 d(\mu_2 \times H) = \int_{A\times B} f_2^2 h_2^2 d(\mu_2 \times H) .$$

for all such sets A, B. But this implies that the measures γ_1 and γ_2 defined by

$$\gamma_1(A \times B) = \int_{A\times B} f_1^2 h_1^2 d(\mu_2 \times H), \qquad \gamma_2(A \times B) = \int_{A\times B} f_2^2 h_2^2 d(\mu_2 \times H)$$

are equal for all the product sets $A \times B$, and hence by a standard monotone class argument we conclude that $\gamma_1 = \gamma_2$ as measures on $([0,\tau]^2 \times \mathbb{R}^d, \mathcal{B}_2[0,\tau] \times \mathcal{B}_d)$.

It follows that

$$f_1^2(u,v) h_1^2(z) = f_2^2(u,v) h_2^2(z)$$

a.e. with respect to $\mu_2 \times H$. Thus we conclude that

$$\frac{f_1^2(u,v)}{f_2^2(u,v)} = \frac{h_2^2(z)}{h_1^2(z)} \quad \text{a.e. on} \quad \{(u,v,z) : \; f_1^2(u,v) > 0, \; h_1^2(z) > 0\} ,$$

or, in other words,

$$\left(\frac{\Lambda(v) - \Lambda(u)}{\Lambda_0(v) - \Lambda_0(u)} - 1 \right)^2 = (1 - \exp((\beta_0 - \beta)^T z))^2$$

a.e. with respect to $\mu_2 \times H$. This implies that (18) holds in view of C7.

Next, we show that $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n)$ is uniformly bounded. We only need to show that $\hat{\Lambda}_n$ is uniformly bounded given that $\hat{\beta}_n$ is bounded since $\hat{\beta}_n$ is in the bounded set $\mathcal{R}$.

For any given $\epsilon > 0$, let $\tilde{\theta}_n = (\hat{\beta}_n, (1-\epsilon)\hat{\Lambda}_n + \epsilon\Lambda_0) = (\hat{\beta}_n, \hat{\Lambda}_n) + \epsilon(0, \Lambda_0 - \hat{\Lambda}_n)$. Since

$$\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\tilde{\theta}_n) = \mathbb{M}_n\left\{\hat{\theta}_n + \epsilon(0, \Lambda_0 - \hat{\Lambda}_n)\right\},$$

it follows that

$$\begin{aligned} 0 &\geq \lim_{\epsilon\downarrow 0} \frac{\mathbb{M}_n(\hat{\theta}_n + \epsilon(0, \Lambda_0 - \hat{\Lambda}_n)) - \mathbb{M}_n(\hat{\theta}_n)}{\epsilon} \\ &= \mathbb{P}_n\left[\sum_{j=1}^{K}\left\{\frac{\Delta\mathbb{N}_{Kj}}{\Delta\hat{\Lambda}_{nKj}\exp(\hat{\beta}_n^T Z)} - 1\right\}(\Delta\Lambda_{0Kj} - \Delta\hat{\Lambda}_{nKj})\exp(\hat{\beta}_n^T Z)\right] \end{aligned}$$

where $\Delta\hat{\Lambda}_{nKj} = \hat{\Lambda}_n(T_{K,j}) - \hat{\Lambda}_n(T_{K,j-1})$. This yields

$$\begin{aligned} &\mathbb{P}_n\left[\sum_{j=1}^{K}\left\{\Delta\mathbb{N}_{Kj}\frac{\Delta\Lambda_{0Kj}}{\Delta\hat{\Lambda}_{nKj}} + \Delta\hat{\Lambda}_{nKj}\exp(\hat{\beta}_n^T Z)\right\}\right] \\ &\quad\leq \mathbb{P}_n\left[\sum_{j=1}^{K}\left\{\Delta\mathbb{N}_{Kj} + \Delta\Lambda_{0Kj}\exp(\hat{\beta}_n^T Z)\right\}\right] \\ &\quad\leq C\mathbb{P}_n\left\{\mathbb{N}_{KK} + \Lambda_0(T_{K,K})\right\} \\ &\quad\to_{a.s.} 2CE\left\{\Lambda_0(T_{K,K})\right\} < \infty, \end{aligned}$$

by C2, C3, C6 and the strong law of large numbers.

On the other hand,

$$\begin{aligned} &\limsup_{n\to\infty} \mathbb{P}_n\left[\sum_{j=1}^{k}\left\{\Delta\mathbb{N}_{Kj}\frac{\Delta\Lambda_{0Kj}}{\Delta\hat{\Lambda}_{nKj}} + \Delta\hat{\Lambda}_{nKj}\exp(\hat{\beta}_n^T Z)\right\}\right] \\ &\quad\geq \limsup_{n\to\infty} \mathbb{P}_n\left\{\sum_{j=1}^{k}\Delta\hat{\Lambda}_{nKj}\exp(\hat{\beta}_n^T Z)\right\} \\ &\quad\geq C\limsup_{n\to\infty} \mathbb{P}_n\left(\sum_{j=1}^{k}\Delta\hat{\Lambda}_{nKj}\right) \\ &\quad= C\mathbb{P}_n\left\{\hat{\Lambda}_n(T_{K,K})\right\} \\ &\quad\geq C\limsup_{n\to\infty} \mathbb{P}_n\left\{1_{[b,\tau]}(T_{K,K})\hat{\Lambda}_n(T_{K,K})\right\} \end{aligned}$$

$$\geq C \limsup_{n\to\infty} \left\{ \hat{\Lambda}_n(b) \mathbb{P}_n \left\{ 1_{[b,\tau]}(T_{K,K}) \right\} \right\}$$
$$\geq C\gamma([b,\tau]) \limsup_{n\to\infty} \hat{\Lambda}_n(b).$$

Hence $\hat{\Lambda}_n(t)$ is uniformly bounded almost surely for $t \in [0,b]$ if $\gamma([b,\tau]) > 0$ for some $0 < b < \tau$ or for $t \in [0,\tau]$ if $\gamma(\{\tau\}) > 0$. Then by the Helly-Selection Theorem and the compactness of $\mathcal{R} \times \mathcal{F}$, it follows that $\hat{\theta}_n = \left(\hat{\beta}_n, \hat{\Lambda}_n\right)$ has a subsequence $\hat{\theta}_{n'} = \left(\hat{\beta}_{n'}, \hat{\Lambda}_{n'}\right)$ converging to $\theta^+ = (\beta^+, \Lambda^+)$, where Λ^+ is an increasing bounded function defined on $[0,b]$ for a $b < \tau$ and it can be defined on $[0,\tau]$ if $\nu(\{\tau\}) > 0$. Following the same argument as in proving Theorem 4.1 of Wellner and Zhang (2000), we can show that $\mathbb{M}(\theta^+) \geq \mathbb{M}(\theta_0)$. This implies that $\beta^+ = \beta_0$ and $\Delta\Lambda^+ = \Delta\Lambda_0$ a.e. in μ_2. Finally, the dominated convergence theorem yields the strong consistency of $\left(\hat{\beta}_n, \hat{\Lambda}_n\right)$ in the metric d_2. □

Proof of Theorem 3.2: We derive the rate of convergence by checking the conditions in Theorem 3.2.5 of van der Vaart and Wellner (1996). Here we give a detailed proof for the first part of the theorem and for the second part, we point out the differences in the proof from the first part. Let

$$\begin{aligned} m_\theta^{ps}(X) &= \sum_{j=1}^K \left\{ \mathbb{N}_{Kj} \log \Lambda(T_{K,j}) + \mathbb{N}_{Kj} \beta^T Z - \Lambda(T_{K,j}) \exp(\beta^T Z) \right\} \\ &= \sum_{j=1}^K \left[\log \left\{ \Lambda(T_{K,j}) \exp(\beta^T Z) \right\}^{\mathbb{N}_{Kj}} - \Lambda(T_{K,j}) \exp(\beta^T Z) \right] \end{aligned}$$

with $\mathbb{N}_{Kj} = \mathbb{N}(T_{K,j})$ and $\mathbb{M}^{ps}(\theta) = P m_\theta^{ps}(X)$. We have

$$\begin{aligned} &\mathbb{M}^{ps}(\theta_0) - \mathbb{M}^{ps}(\theta) \\ &\quad = P\left(\sum_{j=1}^K \left[\mathbb{N}_{Kj} \log \frac{\Lambda_0(T_{K,j}) \exp(\beta_0^T Z)}{\Lambda(T_{K,j}) \exp(\beta^T Z)} \right.\right. \\ &\qquad\qquad \left.\left. - \left\{ \Lambda_0(T_{K,j}) \exp(\beta_0^T Z) - \Lambda(T_{K,j}) \exp(\beta^T Z) \right\} \right] \right) \\ &\quad = E_{(Z,K,\underline{T}_K)} \left[\sum_{j=1}^K \Lambda(T_{K,j}) \exp(\beta^T Z) h\left\{ \frac{\Lambda_0(T_{K,j}) \exp(\beta_0^T Z)}{\Lambda(T_{K,j}) \exp(\beta^T Z)} \right\} \right]. \end{aligned}$$

By (14), for any θ in a sufficiently small neighborhood of θ_0

$$
\begin{aligned}
&\mathbb{M}^{ps}(\theta_0) - \mathbb{M}^{ps}(\theta) \\
&\geq \frac{1}{4} E_{(Z,K,\underline{T}_K)} \left[\sum_{j=1}^{K} \Lambda(T_{K,j}) \exp(\beta^T Z) \left\{ \frac{\Lambda_0(T_{K,j}) \exp(\beta_0^T Z)}{\Lambda(T_{K,j}) \exp(\beta^T Z)} - 1 \right\}^2 \right] \\
&= \frac{1}{4} E_{(Z,K,\underline{T}_K)} \left[\sum_{j=1}^{K} \frac{\left\{ \Lambda_0(T_{K,j}) \exp(\beta_0^T Z) - \Lambda(T_{K,j}) \exp(\beta^T Z) \right\}^2}{\Lambda(T_{K,j}) \exp(\beta^T Z)} \right] \\
&\geq C E_{(Z,K,\underline{T}_K)} \left[\sum_{j=1}^{K} \left\{ \Lambda(T_{K,j}) \exp(\beta^T Z) - \Lambda_0(T_{K,j}) \exp(\beta_0^T Z) \right\}^2 \right] \\
&= C \int \{ \Lambda(u) e^{\beta^T z} - \Lambda_0(u) e^{\beta_0^T z} \}^2 d\nu_1(u, z) \qquad (19)
\end{aligned}
$$

by C1, C2 and C6.

Let $g(t) = \Lambda_t(U) \exp(\beta_t^T Z)$ with $\Lambda_t = t\Lambda + (1-t)\Lambda_0$ and $\beta_t = t\beta + (1-t)\beta_0$ for $0 \leq t \leq 1$ with $(U, Z) \sim \nu_1$. Then $\Lambda(U) \exp(\beta^T Z) - \Lambda_0(U) \exp(\beta_0^T Z) = g(1) - g(0)$, and hence, by the mean value theorem, there exists a $0 \leq \xi \leq 1$ such that $g(1) - g(0) = g'(\xi)$. Since

$$
\begin{aligned}
g'(\xi) &= \exp(\beta_\xi^T Z) \left[(\Lambda - \Lambda_0)(U) + \{ \Lambda_0 + \xi(\Lambda - \Lambda_0) \} (U) (\beta - \beta_0)^T Z \right] \\
&= \exp(\beta_\xi^T Z) \left[(\Lambda - \Lambda_0)(U) \left\{ 1 + \xi(\beta - \beta_0)^T Z \right\} + (\beta - \beta_0)^T Z \Lambda_0(U) \right]
\end{aligned}
$$

from (19) we have

$$
\begin{aligned}
&P \left\{ m_{\theta_0}^{ps}(X) - m_\theta^{ps}(X) \right\} \\
&\geq C \int \left[(\Lambda - \Lambda_0)(u) \left\{ 1 + \xi(\beta - \beta_0)^T z \right\} + (\beta - \beta_0)^T z \Lambda_0(u) \right]^2 d\nu_1(u, z) \\
&= \nu_1 \left[(\Lambda - \Lambda_0)(U) \left\{ 1 + \xi(\beta - \beta_0)^T Z \right\} + (\beta - \beta_0)^T Z \Lambda_0(U) \right]^2 \\
&= \nu_1 \left\{ (\beta - \beta_0)^T Z \Lambda_0(U) \left[1 + \xi(\Lambda - \Lambda_0)(U)/\Lambda_0(U) \right] + (\Lambda - \Lambda_0)(U) \right\}^2 \\
&= \nu_1 \left\{ g_1 h + g_2 \right\}^2
\end{aligned}
$$

where $g_1(U, Z) \equiv (\beta - \beta_0)^T Z \Lambda_0(U)$, $g_2(U) = (\Lambda - \Lambda_0)(U)$, and $h(U, Z) = 1 + \xi(\Lambda - \Lambda_0)(U)/\Lambda_0(U)$ in the notation of Lemma 8.8, page 432, van der

Vaart (2002). To apply van der Vaart's lemma we need to bound $[\nu_1(g_1 g_2)]^2$ by a constant less than one times $\nu_1(g_1^2)\nu_1(g_2^2)$. For the moment we write expectations under ν_1 as E_1. But by the Cauchy-Schwarz inequality and then computing conditionally on U we have

$$
\begin{aligned}
[E_1(g_1 g_2)]^2 &\leq E_1(g_1^2)E_1(g_2^2) = E_1[g_2^2]E_1[E_1(g_1^2|U)] \\
&= E_1\{g_2^2\}E_1\{\Lambda_0^2(U)E_1[(\beta-\beta_0)^T(Z-(Z-E_1(Z|U))^{\otimes 2}(\beta-\beta_0)|U]\} \\
&= E_1\{g_2^2\}E_1\{\Lambda_0^2(U) \\
&\qquad \cdot(\beta-\beta_0)^T\{E_1(ZZ^T|U)-E_1[[Z-E_1(Z|U)]^{\otimes 2}|U]\}(\beta-\beta_0)^T\} \\
&\leq (1-\eta)E_1\{g_2^2\}E_1\{\Lambda_0^2(U)(\beta-\beta_0)^T E_1(ZZ^T|U)(\beta-\beta_0)^T\} \\
&= (1-\eta)E_1\{g_2^2\}E_1\{g_1^2\}
\end{aligned}
$$

where the last inequality follows from C13. Then by van der Vaart's lemma,

$$
\begin{aligned}
\nu_1\{g_1 h + g_2\}^2 &\geq C\{\nu_1(g_1^2)+\nu_1(g_2^2)\} \\
&= C\{|\beta-\beta_0|^2 + \|\Lambda-\Lambda_0\|_{L_2(\mu_1)}^2\} = C d_1^2(\theta,\theta_0)\,.
\end{aligned}
$$

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

$$
E \sup_{d_1(\theta,\theta_0)<\sigma} |\mathbb{G}_n(m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X))| \leq C\phi_n(\sigma).
$$

We let $\mathcal{M}_\delta^1(\theta_0) = \left\{ m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X) : d_1(\theta,\theta_0) < \delta \right\}$ be the class of differences. We shall find an upper bound for the bracketing entropy numbers of this class. We also let $\mathcal{F}_\delta = \{\Lambda \in \mathcal{F} : \ \|\Lambda-\Lambda_0\|_{L_2(\mu_1)} \leq \delta\}$. Since $\mathcal{F}_\delta$ is a class of monotone increasing functions, by Theorem 2.7.5 of van der Vaart and Wellner (1996), for any $\epsilon > 0$, there exists a set of brackets:

$$
[\Lambda_1^l, \Lambda_1^r], \quad [\Lambda_2^l, \Lambda_2^r], \ldots, [\Lambda_q^l, \Lambda_q^r]
$$

with $q \leq \exp(M/\epsilon)$, such that for any $\Lambda \in \mathcal{F}_\delta$, $\Lambda_i^l(t) \leq \Lambda(t) \leq \Lambda_i^r(t)$ for all $t \in O[T]$ and some $1 \leq i \leq q$, and $\int \{\Lambda_i^r(u) - \Lambda_i^l(u)\}^2 d\mu_1(u) \leq \epsilon^2$. (Here we use the fact that μ_1 is a finite measure under our hypotheses, and hence can be normalized to be a probability measure.)

Moreover, we can make these bracketing functions satisfying that $\Lambda_i^r(t) - \Lambda_i^l(t) \le \gamma_1$ and $\Lambda_i^l(t) \ge \gamma_2$ with $\gamma_1, \gamma_2 > 0$ for all $t \in O[T]$ and $1 \le i \le q$ for sufficient small ϵ and δ. Here is the proof for this claim: For any $\Lambda \in \mathcal{F}_\delta$, the result of Lemma 8.2 implies that $\Lambda_0(t) - \epsilon_1 \le \Lambda(t) \le \Lambda_0(t) + \epsilon_1$ for a sufficiently small $\epsilon_1 > 0$ (ϵ_1 can be chosen as $(\delta/C)^{2/3}$ in view of Lemma 8.2) and for all $t \in O[T]$. For any $1 \le i \le q$, there is a $\Lambda \in \mathcal{F}_\delta$ such that $\|\Lambda_i^r - \Lambda\|_{L_2(\mu_1)} \le \epsilon$ and $\|\Lambda - \Lambda_i^l\|_{L_2(\mu_1)} \le \epsilon$, which implies that $\|\Lambda_i^r - \Lambda_0\|_{L_2(\mu_1)} \le \epsilon^*(\epsilon^* = \sqrt{\epsilon^2 + \delta^2})$ and $\|\Lambda_i^l - \Lambda_0\|_{L_2(\mu_1)} \le \epsilon^*$. By Lemma 8.2, this yields that $\Lambda_i^r(t) \le \Lambda_0(t) + \epsilon_2$ and $\Lambda_i^l(t) \ge \Lambda_0(t) - \epsilon_2$ for a sufficient small $\epsilon_2 > 0$. (ϵ_2 can be chosen as $(\epsilon^*/C)^{2/3}$) Therefore our claim is justified by letting $\gamma_1 = 2\epsilon_2$ and $\gamma_2 = \Lambda_0(\sigma) - \epsilon_2$, in view of C8.

Since $\beta \in \mathcal{R}$, a compact set in $\mathbb{R}^d$, we can construct an ϵ-net for $\mathcal{R}$, $\beta_1, \beta_2, \ldots, \beta_p$ with $p = \left[\left(M'/\epsilon^d\right)\right]$ such that for any $\beta \in \mathcal{R}$ there is a s such that

$$|\beta^T Z - \beta_s^T Z| \le \epsilon \quad \text{and} \quad |\exp(\beta^T Z) - \exp(\beta_s^T Z)| \le C\epsilon\,.$$

Therefore we can construct a set of brackets for $\mathcal{M}_\delta^1(\theta_0)$ as follows:

$$[m_{i,s}^{ps^l}(X), m_{i,s}^{ps^r}(X)], \quad \text{for} \quad i = 1, 2, \ldots, q; \quad s = 1, 2, \ldots, p,$$

where

$$\begin{aligned} m_{i,s}^{ps^l}(X) \;&=\; \sum_{j=1}^{K} \Big[\mathbb{N}_{Kj} \log \Lambda_i^l(T_{K,j}) + \mathbb{N}_{Kj}(|\beta_s^T Z| - \epsilon) \\ &\qquad -\Lambda_i^r(T_{K,j})\left\{\exp(\beta_s^T Z) + C\epsilon\right\}\Big] - m_{\theta_0}(X) \end{aligned}$$

and

$$\begin{aligned} m_{i,s}^{ps^r}(X) \;&=\; \sum_{j=1}^{K} \big[\mathbb{N}_{Kj} \log \Lambda_i^r(T_{K,j}) + \mathbb{N}_{Kj}(|\beta_s^T Z| + \epsilon) \\ &\qquad -\Lambda_i^l(T_{K,j})\left\{\exp(\beta_s^T Z) - C\epsilon\right\}\Big] - m_{\theta_0}(X). \end{aligned}$$

In what follows, we show that $\|f_{i,s}(X)\|_{P,B}^2 = \|m_{i,s}^{ps^r}(X) - m_{i,s}^{ps^l}(X)\|_{P,B}^2 \le C\epsilon^2$, where $\|\cdot\|_{P,B}$ is the "Bernstein norm" defined by

$$\|f\|_{P,B} = \left\{2P\left(e^{|f|} - 1 - |f|\right)\right\}^{1/2}$$

(see van der Vaart and Wellner, 1996, page 324). Since $2(e^x - 1 - x) \le x^2 e^x$ for $x \ge 0$, it follows that $\|f\|_{P,B}^2 \le P\left(e^{|f|}|f|^2\right)$. Therefore, $\|f_{i,s}(X)\|_{P,B}^2 \le P\left(e^{|f_{i,s}(X)|}|f_{i,s}(X)|^2\right)$. Note that

$$\begin{aligned} f_{i,s}(X) &= m_{i,s}^{ps^r}(X) - m_{i,s}^{ps^l}(X) \\ &= \sum_{j=1}^{K} \Big\{ \mathbb{N}_{Kj} \left[\log \Lambda_i^r(T_{K,j}) - \log \Lambda_i^l(T_{K,j}) + 2\epsilon\right] \\ &\qquad + \exp(\beta_s^T Z)\left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right) \\ &\qquad + C\epsilon\left(\Lambda_i^r(T_{K,j}) + \Lambda_i^l(T_{K,j})\right)\Big\}. \end{aligned}$$

Hence it follows that

$$\begin{aligned} |f_{i,s}(X)| &\le \mathbb{N}_{KK} \sum_{j=1}^{K} \left| \left(\log \Lambda_i^r(T_{K,j}) - \log \Lambda_i^l(T_{K,j}) + 2\epsilon\right)\right| \\ &\qquad + \exp(\beta_s^T Z) \sum_{j=1}^{K} \left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right) \\ &\qquad + C\epsilon \sum_{j=1}^{K} \left(\Lambda_i^r(T_{K,j}) + \Lambda_i^l(T_{K,j})\right). \end{aligned}$$

Since

$$\begin{aligned} \log y = \log x + (x + \xi(y-x))^{-1}(y-x) \\ \text{for } 0 < x \le y, \quad \text{some } \xi \in [0,1], \end{aligned} \tag{20}$$

we find that

$$\begin{aligned} \log \Lambda_i^r(T_{K,j}) &= \log \Lambda_i^l(T_{K,j}) + \frac{\left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right)}{\Lambda_i^l(T_{K,j}) + \xi\left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right)} \\ &\le \log \Lambda_i^l(T_{K,j}) + \frac{1}{\Lambda_i^l(T_{K,j})}\left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right) \\ &\le \log \Lambda_i^l(T_{K,j}) + \gamma_2^{-1}\left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right) \quad \text{by construction of } \Lambda_i^l \end{aligned}$$

Hence, by C9 and our claim above, we conclude further that

$$\sum_{j=1}^{K} \left| \left(\log \Lambda_i^r(T_{K,j}) - \log \Lambda_i^l(T_{K,j})\right| + 2\epsilon\right),$$

$\sum_{j=1}^K \left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right)$, and $\sum_{j=1}^K \left(\Lambda_i^r(T_{K,j}) + \Lambda_i^l(T_{K,j})\right)$ are all uniformly bounded in $O[T]$, and, more explicitly,

$$\begin{aligned}
&\sum_{j=1}^K (\log \Lambda_i^r(T_{K,j}) - \log \Lambda_i^l(T_{K,j}) + 2\epsilon)^2 \\
&\leq \sum_{j=1}^K \left(\frac{|\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})|}{\gamma_2} + 2\epsilon\right)^2 \\
&\leq \sum_{j=1}^K \left(\frac{2\epsilon_2}{\gamma_2} + 2\epsilon\right)^2 = \sum_{j=1}^K \left(\frac{2\epsilon_2}{\Lambda_0(\sigma) - \epsilon_2} + 2\epsilon\right)^2 \\
&\leq k_0(2+2\delta)^2 \qquad \text{by taking} \quad \epsilon_2 \leq 2^{-1}\Lambda_0(\sigma) \\
&\leq 4k_0(1+\delta_0^{ps})^2
\end{aligned}$$

since the relations $\epsilon_2 = (\epsilon^*/C)^{2/3}$, $\epsilon^* = (\epsilon^2 + \delta^2)^{1/2} \geq \delta$, and $\epsilon_2 \leq 2^{-1}\Lambda_0(\sigma)$ imposes the restriction $\delta \leq C(2^{-1}\Lambda_0(\sigma))^{3/2} \equiv \delta_0^{ps}$ where $C = (c_0/(24f_0))^{1/2}$ from Lemma 8. Therefore, by arguing conditionally on (Z, K, T_K) and using C10,

$$\begin{aligned}
\|f_{i,s}(X)\|_{P,B}^2 &\leq P\left(e^{|f_{i,s}(X)|}|f_{i,s}(X)|^2\right) \\
&\leq CP\left\{e^{v\mathbb{N}_{KK}}\left[\mathbb{N}_{KK}^2 \sum_{j=1}^K \left(\log \Lambda_i^r(T_{K,j}) - \log \Lambda_i^l(T_{K,j}) + 2\epsilon\right)^2\right.\right. \\
&\qquad \left.\left. + \exp(2\beta_s^T Z) \sum_{j=1}^K \left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right)^2 + C\epsilon^2\right]\right\},
\end{aligned}$$

and by C6, C10, and Taylor expansion for $\log \Lambda_i^r(T_{K,j})$ at $\Lambda_i^l(T_{K,j})$ as shown above, we have

$$\begin{aligned}
\|f_{i,s}(X)\|_{P,B}^2 &\leq C\left\{E_{(K,\underline{T}_K)}\left[\sum_{j=1}^K \left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right)^2\right] + \epsilon^2\right\} \\
&\leq C\epsilon^2.
\end{aligned}$$

This shows that the total number of ϵ-brackets for $\mathcal{M}_\delta^1(\theta_0)$ will be of the order $(M/\epsilon)^d\, e^{C(M'/\epsilon)}$ and hence

$$\log N_{[\,]}\left(\epsilon, \mathcal{M}_\delta^1(\theta_0), \|\cdot\|_{P,B}\right) \leq C\,(1/\epsilon)\,.$$

We can similarly verify that $P\left(f_\theta^{ps}(X)\right)^2 \le C\delta^2$ for any $f_\theta^{ps}(X) = m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X) \in \mathcal{M}_\delta^1(\theta_0)$. Hence by Lemma 3.4.3 of van der Vaart and Wellner (1996),

$$E_P^*\|\mathbb{G}_n\|_{\mathcal{M}_\delta^1(\theta_0)} \le C\tilde{J}_{[\,]}\left(\delta, \mathcal{M}_\delta^1(\theta_0), \|\cdot\|_{P,B}\right)\left[1 + \frac{\tilde{J}_{[\,]}\left(\delta, \mathcal{M}_\delta^1(\theta_0), \|\cdot\|_{P,B}\right)}{\delta^2\sqrt{n}}\right],$$

where

$$\begin{aligned}\tilde{J}_{[\,]}\left\{\delta, \mathcal{M}_\delta^1(\theta_0), \|\cdot\|_{P,B}\right\} &= \int_0^\delta \sqrt{1 + \log N_{[\,]}\left(\epsilon, \mathcal{M}_\delta^1(\theta_0), \|\cdot\|_{P,B}\right)}d\epsilon \\ &= C\int_0^\delta \sqrt{1 + \frac{1}{\epsilon}}d\epsilon \le C\int_0^\delta \epsilon^{-1/2}d\epsilon \le C\delta^{1/2}.\end{aligned}$$

Hence $\phi_n(\delta) = \delta^{1/2}(1 + \delta^{1/2}/(\delta^2\sqrt{n})) = \delta^{1/2} + \delta^{-1}/\sqrt{n}$. Then it is easy to see that $\phi_n(\delta)/\delta$ is a decreasing function of δ, and $n^{2/3}\phi_n(n^{-1/3}) = n^{2/3}(n^{-1/6} + n^{1/3}n^{-1/2}) = 2\sqrt{n}$. So it follows by Theorem 3.2.5 of van der Vaart and Wellner (1996) that

$$n^{1/3}d_1\left((\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}), (\beta_0, \Lambda_0)\right) = O_P(1).$$

For the maximum likelihood estimator $(\hat{\beta}_n, \hat{\Lambda}_n)$, similarly, we have

$$\begin{aligned}&\mathbb{M}(\theta_0) - \mathbb{M}(\theta) = P\left\{m_{\theta_0}(X) - m_\theta(X)\right\} \\ &\quad \ge \; CE_{(K,\underline{T}_K,Z)}\left\{\sum_{j=1}^K \left[\Delta\Lambda_0(T_{K,j})\exp(\beta_0^T Z) - \Delta\Lambda(T_{K,j})\exp(\beta^T Z)\right]^2\right\} \\ &\quad = \; C\int\{\Delta\Lambda(u,v)\exp(\beta^T z) - \Delta\Lambda_0(u,v)\exp(\beta_0^T z)\}^2 d\nu_2(u,v,z)\,.\end{aligned}$$

Now we redefine $g(t) = g(t; u, v, z) = \Delta\Lambda_t(u,v)\exp(\beta_t^T z)$ with $\Delta\Lambda_t = t\Delta\Lambda + (1-t)\Delta\Lambda_0$, $\Delta\Lambda(u,v) = \Lambda(v) - \Lambda(u)$, $\Delta\Lambda_0(u,v) = \Lambda_0(v) - \Lambda_0(u)$, and $\beta_t = t\beta + (1-t)\beta_0$. Then by the mean value theorem, we have

$$\begin{aligned}&P\left\{m_{\theta_0}(X) - m_\theta(X)\right\} \\ &\quad \ge \; C\int\Big\{(\Delta\Lambda - \Delta\Lambda_0)(u,v)(1 + \xi(\beta - \beta_0)^T z) \\ &\qquad\qquad + \; (\beta - \beta_0)^T z\Delta\Lambda_0(u,v)\Big\}^2 d\nu_2(u,v,z)\end{aligned}$$

with $0 \le \xi \le 1$. Similar to the argument in the pseudo-likelihood case, we have

$$
\begin{aligned}
& P\{m_{\theta_0}(X) - m_\theta(X)\} \\
\ge\ & C\int \big[(\Delta\Lambda - \Delta\Lambda_0)(u,v)\left\{1+\xi(\beta-\beta_0)^T z\right\} \\
& \qquad + (\beta-\beta_0)^T z\Delta\Lambda_0(u,v)\big]^2\, d\nu_2(u,v,z) \\
=\ & \nu_2\left[(\Delta\Lambda - \Delta\Lambda_0)(U,V)\left\{1+\xi(\beta-\beta_0)^T Z\right\}\right. \\
& \qquad \left. + (\beta-\beta_0)^T Z\Delta\Lambda_0(U)\right]^2 \\
=\ & \nu_2\left\{(\beta-\beta_0)^T z\Delta\Lambda_0(U,V)\left[1+\xi(\Delta\Lambda-\Delta\Lambda_0)(U,V)/\Delta\Lambda_0(U,V)\right]\right. \\
& \qquad \left. + (\Delta\Lambda-\Delta\Lambda_0)(U,V)\right\}^2 \\
=\ & \nu_2\left\{g_1 h + g_2\right\}^2
\end{aligned}
$$

where $g_1(U,V,Z) \equiv (\beta-\beta_0)^T Z\Delta\Lambda_0(U,V)$, $g_2(U,V) = (\Delta\Lambda - \Delta\Lambda_0)(U,V)$, and $h(U,V,Z) = 1 + \xi(\Delta\Lambda - \Delta\Lambda_0)(U,V)/\Delta\Lambda_0(U,V)$ in the notation of Lemma 8.8, page 432, van der Vaart (2002). To apply van der Vaart's lemma we need to bound $[\nu_2(g_1 g_2)]^2$ by a constant less than one times $\nu_2(g_1^2)\nu_2(g_2^2)$. For the moment we write expectations under ν_2 as E_2 and conditional expectations given U,V as $E_2\{\cdot|U,V\}$. But by the Cauchy-Schwarz inequality and then computing conditionally on U,V we have

$$
\begin{aligned}
& [E_2(g_1 g_2)]^2 \\
\le\ & E_2(g_1^2)E_2(g_2^2) = E_2[g_2^2]E_2[E_2(g_1^2|U,V)] \\
=\ & E_2\{g_2^2\}E_2\{\Delta\Lambda_0^2(U,V) \\
& \qquad E_2[(\beta-\beta_0)^T(Z-(Z-E_2(Z|U,V))^{\otimes 2}(\beta-\beta_0)|U,V]\} \\
=\ & E_2\{g_2^2\}E_2\{\Delta\Lambda_0^2(U,V) \\
& \qquad \cdot(\beta-\beta_0)^T\{E_2(ZZ^T|U,V) - E_2[[Z-E_2(Z|U,V)]^{\otimes 2}|U,V]\}(\beta-\beta_0)^T\} \\
\le\ & (1-\eta)E_2\{g_2^2\}E_2\{\Delta\Lambda_0^2(U,V)(\beta-\beta_0)^T E_2(ZZ^T|U,V)(\beta-\beta_0)^T\} \\
=\ & (1-\eta)E_2\{g_2^2\}E_2\{g_1^2\}\,.
\end{aligned}
$$

where the last inequality follows from C14. Then by van der Vaart's lemma

$$
P\{m_{\theta_0}(X) - m_\theta(X)\}
$$

$$
\begin{aligned}
&\geq C\nu_2 \{g_1 h + g_2\}^2 \geq C\{\nu_2(g_1^2) + \nu_2(g_2^2)\} \\
&= C\{|\beta - \beta_0|^2 + \|\Delta\Lambda - \Delta\Lambda_0\|^2_{L_2(\mu_2)}\} = Cd_2^2(\theta, \theta_0)\,.
\end{aligned}
$$

Similarly, to derive the rate of convergence for the maximum likelihood estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n)$, we need to find a $\phi_n^*(\sigma)$ such that

$$
E \sup_{d_2(\theta,\theta_0)<\sigma} |\mathbb{G}_n(m_\theta(X) - m_{\theta_0}(X))| \leq C\phi_n^*(\sigma).
$$

To do so, we define

$$
\mathcal{M}_\delta^2(\theta_0) = \{m_\theta(X) - m_{\theta_0}(X) : \quad d_2(\theta, \theta_0) \leq \delta/k_0\}\,,
$$

Since for any $\theta = (\beta, \Lambda) \in \mathcal{M}_\delta^2(\theta_0)$, $\Lambda \in \mathcal{F}_\delta$ by Lemma 8.1 and C9. Thus we can use the set of brackets:

$$
[\Lambda_1^l, \Lambda_1^r], [\Lambda_2^l, \Lambda_2^r], \cdots, [\Lambda_q^l, \Lambda_q^r]
$$

with $q \leq \exp(M/\epsilon)$ as constructed before. Moreover, by C11 and C12, there exists a constant $a > 0$ such that $\Lambda_0(T_{K,j}) - \Lambda_0(T_{K,j-1}) > 2a$ ($2a = s_0/f_0$ works), and hence we have

$$
\Lambda^l(T_{K,j}) - \Lambda_i^r(T_{K,j-1}) \quad \geq \quad \Lambda_0(T_{K,j}) - \Lambda_0(T_{K,j-1}) - 2\epsilon_2 \geq 2(a - \epsilon_2) > \gamma_3 > 0
$$

for all $1 \leq i \leq q$ and $j = 1, 2, \cdots, K$.

Then using the same arguments as in the pseudo-likelihood case, we can construct a set of brackets for $\mathcal{M}_\delta^2(\theta_0)$ as follows:

$$
\Big\{[m_{i,s}^l(X), m_{i,s}^r(X)], \quad i = 1, 2, \ldots, q; \quad s = 1, 2, \ldots, p\Big\}
$$

where $q \leq \exp(M/\epsilon)$, $p \leq (M'/\epsilon^d)$,

$$
\begin{aligned}
m_{i,s}^l(X) \quad = \quad & \sum_{j=1}^{K} \Big\{\Delta\mathbb{N}_{Kj} \log(\Lambda_i^l(T_{K,j}) - \Lambda_i^r(T_{K,j-1})) + \Delta\mathbb{N}_{Kj}(|\beta_s^T Z| - \epsilon) \\
& -(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j-1}))(\exp(\beta_s^T Z) + C\epsilon)\Big\} - m_{\theta_0}(X)\,,
\end{aligned}
$$

and

$$m^r_{i,s}(X) = \sum_{j=1}^K \Big\{ \Delta\mathbb{N}_{Kj} \log(\Lambda^r_i(T_{K,j}) - \Lambda^l_i(T_{K,j-1})) + \Delta\mathbb{N}_{Kj}(|\beta_s^T Z| + \epsilon) \\ -(\Lambda^l_i(T_{K,j}) - \Lambda^r_i(T_{K,j-1}))(\exp(\beta_s^T Z) - C\epsilon) \Big\} - m_{\theta_0}(X).$$

Let $f_{i,s}(X) = m^r_{i,s}(X) - m^l_{i,s}(X)$. In what follows, we show that $\|f_{i,s}(X)\|^2_{P,B} \leq C\epsilon^2$. Note that

$$\begin{aligned} & f_{i,s}(X) \\ &= m^r_{i,s}(X) - m^l_{i,s}(X) \\ &= \sum_{j=1}^K \Big\{ \Delta\mathbb{N}_{Kj} \Big[\log\Big(\Lambda^r_i(T_{K,j}) - \Lambda^l_i(T_{K,j-1})\Big) - \log\Big(\Lambda^l_i(T_{K,j}) - \Lambda^r_i(T_{K,j-1})\Big) + 2\epsilon \Big] \\ &\quad + \exp(\beta_s^T Z) \Big[\Big(\Lambda^r_i(T_{K,j}) - \Lambda^l_i(T_{K,j})\Big) + \Big(\Lambda^r_i(T_{K,j-1}) - \Lambda^l_i(T_{K,j-1})\Big) \Big] \\ &\quad + C\epsilon \Big[\Big(\Lambda^r_i(T_{K,j}) + \Lambda^l_i(T_{K,j})\Big) - \Big(\Lambda^r_i(T_{K,j-1}) + \Lambda^l_i(T_{K,j-1})\Big) \Big] \Big\}. \end{aligned}$$

By (20) with $y = \Lambda^r_i(T_{K,j}) - \Lambda^l_i(T_{K,j-1})$ and $x = \Lambda^l_i(T_{K,j}) - \Lambda^r_i(T_{K,j-1})$ we find that

$$\begin{aligned} & \log\Big(\Lambda^r_i(T_{K,j}) - \Lambda^l_i(T_{K,j-1})\Big) \\ &\leq \log\Big(\Lambda^l_i(T_{K,j}) - \Lambda^r_i(T_{K,j-1})\Big) \\ &\qquad + \frac{\big[\big(\Lambda^r_i(T_{K,j}) - \Lambda^l_i(T_{K,j})\big) + \big(\Lambda^r_i(T_{K,j-1}) - \Lambda^l_i(T_{K,j-1})\big)\big]}{\Lambda^l_i(T_{K,j}) - \Lambda^r_i(T_{K,j-1})} \\ &\leq \log\Big(\Lambda^l_i(T_{K,j}) - \Lambda^r_i(T_{K,j-1})\Big) \\ &\qquad + \frac{1}{\gamma_3} \Big[\Big(\Lambda^r_i(T_{K,j}) - \Lambda^l_i(T_{K,j})\Big) + \Big(\Lambda^r_i(T_{K,j-1}) - \Lambda^l_i(T_{K,j-1})\Big) \Big]. \end{aligned}$$

Using arguments similar to those for the pseudo-likelihood case, we find that

$$\begin{aligned} & \sum_{j=1}^K \Big[\log\Big(\Lambda^r_i(T_{K,j}) - \Lambda^l_i(T_{K,j-1})\Big) - \log\Big(\Lambda^l_i(T_{K,j}) - \Lambda^r_i(T_{K,j-1})\Big) + 2\epsilon \Big]^2 \\ &\quad \leq 4k_0(1+\delta_0)^2 \end{aligned}$$

with $\delta_0 = \sqrt{c_0 s_0^3/(48 \cdot 8^2 \cdot f_0^4)}$ and therefore

$$\begin{aligned}
\|f_{i,s}(X)\|_{P,B}^2 &\leq P\left(e^{|f_{i,s}(X)|}|f_{i,s}(X)|^2\right) \\
&\leq CP\left\{e^{v\mathbb{N}_{KK}}\left(\mathbb{N}_{KK}^2 \sum_{j=1}^{K}\Big[\log\left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j-1})\right)\right.\right. \\
&\qquad - \log\left(\Lambda_i^l(T_{K,j}) - \Lambda_i^r(T_{K,j-1})\right) + 2\epsilon\Big]^2 \\
&\qquad + \exp(2\beta_s^T Z)\sum_{j=1}^{K}\Big[\left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right) + \left(\Lambda_i^r(T_{K,j-1}) - \Lambda_i^l(T_{K,j-1})\right)\Big]^2 \\
&\qquad \left.\left. + C\epsilon^2\right)\right\} \\
&\leq C\left\{E_{(K,\underline{T}_K)}\left[\sum_{j=1}^{K}\left(\Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right)^2\right] + \epsilon^2\right\} \\
&\leq C\epsilon^2.
\end{aligned}$$

This shows that $\log N_{[\,]}\left(\epsilon, \mathcal{M}_\delta^2(\theta_0), \|\cdot\|_{P,B}\right) \leq C\,(1/\epsilon)$. Similarly, we can easily verify that $P(f(X))^2 \leq C\delta^2$ for any $f(X) = m_\theta(X) - m_{\theta_0}(X) \in \mathcal{M}_\delta^2(\theta_0)$. Then following the same lines as those in getting $\phi_n(\sigma)$ for the pseudo-likelihood case, we find $\phi_n^*(\sigma) = \sigma^{1/2} + \sigma^{-1}/\sqrt{n}$ as well, and this yields

$$n^{1/3} d_2\left((\hat\beta_n, \hat\Lambda_n), (\beta_0, \Lambda_0)\right) = O_P(1).$$

□

Proof of Theorem 3.3: As in the proof of Theorem 3.2, we give a detailed proof for the first part of the theorem, and only outline the differences in the proof for the second part. We prove the theorem by checking the conditions A1-A6 of Theorem 7.1. Note that A1 holds with $\gamma = 1/3$ because of the rate of convergence given in Theorem 3.2. Based on the Poisson model, the pseudo log-likelihood for (β, Λ) with only one observation is given by $m^{ps}(\beta, \Lambda; X) = \sum_{j=1}^{K}\left\{\mathbb{N}_{Kj}\log\Lambda_{Kj} + \mathbb{N}_{Kj}\beta^T Z - e^{\beta^T Z}\Lambda_{Kj}\right\}$, and thus we have

$$m_1^{ps}(\beta, \Lambda; X) = \sum_{j=1}^{K} Z(\mathbb{N}_{Kj} - \Lambda(T_{K,j})\exp(\beta^T Z))$$

$$m_2^{ps}(\beta,\Lambda;X)[h] = \sum_{j=1}^{K}\left(\frac{\mathbb{N}_{Kj}}{\Lambda_{Kj}} - \exp(\beta^T Z)\right) h_{Kj}$$

$$m_{11}^{ps}(\beta,\Lambda;X)[h] = -\sum_{j=1}^{K}\Lambda_{Kj} Z Z^T \exp(\beta^T Z)$$

$$m_{12}^{ps}(\beta,\Lambda;X)[h] = m_{21}^{ps'}(\beta,\Lambda;X)[h] = -\sum_{j=1}^{K} Z\exp(\beta^T Z) h_{Kj}$$

and

$$m_{22}^{ps}(\beta,\Lambda;X)[\mathbf{h},h] = -\sum_{j=1}^{K}\frac{\mathbb{N}_{Kj}}{\Lambda_{Kj}^2}\mathbf{h}_{Kj} h_{Kj},$$

where $\Lambda_{Kj} = \Lambda(T_{K,j})$ and $h_{Kj} = \int_0^{T_{K,j}} h(t)d\Lambda(t)$ for $h \in L_2(\Lambda)$. A2 automatically holds by the model assumption (1). For A3, we need to find a $\mathbf{h}^*$ such that

$$\begin{aligned} &\dot{S}_{12}^{ps}(\beta_0,\Lambda_0)[h] - \dot{S}_{22}^{ps}(\beta_0,\Lambda_0)[\mathbf{h}^*,h] \\ &= P\left\{m_{12}^{ps}(\beta_0,\Lambda_0;X)[h] - m_{22}^{ps}(\beta_0,\Lambda_0;X)[\mathbf{h}^*,h]\right\} = 0, \end{aligned}$$

for all $h \in L_2(\Lambda_0)$. Note that

$$\begin{aligned} &P\left\{m_{12}^{ps}(\beta_0,\Lambda_0;X)[h] - m_{22}^{ps}(\beta_0,\Lambda_0;X)[\mathbf{h}^*,h]\right\} \\ &\quad = -E\left\{\sum_{j=1}^{K}\left[Ze^{\beta_0^T Z} - \frac{\mathbb{N}_{Kj}}{(\Lambda_{0Kj})^2}\mathbf{h}_{Kj}^*\right] h_{Kj}\right\} \\ &\quad = -E_{(K,\underline{T}_K,Z)}\left\{\sum_{j=1}^{K}\left[Ze^{\beta_0^T Z} - \frac{e^{\beta_0^T Z}\mathbf{h}_{Kj}^*}{\Lambda_{0Kj}}\right] h_{Kj}\right\}. \end{aligned}$$

Therefore, an obvious choice of $\mathbf{h}^*$ is

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

Hence

$m^{*ps}(\beta_0,\Lambda_0;X)$

$$
\begin{aligned}
&= m_1^{ps}(\beta_0, \Lambda_0; X) - m_2^{ps}(\beta_0, \Lambda_0; X)[\mathbf{h}^*] \\
&= \sum_{j=1}^{K} \left\{ Z\left(\mathbb{N}_{Kj} - e^{\beta_0^T Z}\Lambda_{0Kj}\right) - \left(\frac{\mathbb{N}_{Kj}}{\Lambda_{0Kj}} - e^{\beta_0^T Z}\right)\Lambda_{0Kj}\frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j}\right)} \right\} \\
&= \sum_{j=1}^{K} \left(\mathbb{N}_{Kj} - e^{\beta_0^T Z}\Lambda_{0Kj}\right) \left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j}\right)} \right],
\end{aligned}
$$

$$
\begin{aligned}
A^{ps} &= -\dot{S}_{11}^{ps}(\beta_0, \Lambda_0) + \dot{S}_{21}^{ps}(\beta_0, \Lambda_0)[\mathbf{h}^*] \\
&= E\left\{ \sum_{j=1}^{K} \left[\Lambda_{0Kj} e^{\beta_0^T Z} Z Z^T - e^{\beta_0^T Z}\Lambda_{0Kj} \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j'}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j'}\right)} Z^T \right] \right\} \\
&= E_{(K, \underline{T}_K, Z)} \left\{ \sum_{j=1}^{K} \Lambda_{0Kj} e^{\beta_0^T Z} \left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j}\right)} \right] Z^T \right\} \\
&= E_{(K, \underline{T}_K, Z)} \left\{ \sum_{j=1}^{K} \Lambda_{0Kj} e^{\beta_0^T Z} \left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j}\right)} \right]^{\otimes 2} \right\},
\end{aligned}
$$

and

$$
\begin{aligned}
B^{ps} &= Em^{*ps}(\beta_0, \Lambda_0; X)^{\otimes 2} \\
&= E_{(K, \underline{T}_K, Z)} \left\{ \sum_{j,j'=1}^{K} C_{j,j'}^{ps}(Z) \left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j}\right)} \right] \right. \\
&\qquad \left. \left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j'}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j'}\right)} \right]^T \right\},
\end{aligned}
$$

with

$$
C_{j,j'}^{ps}(Z) = E\left[\left(\mathbb{N}_{Kj} - e^{\beta_0^T Z}\Lambda_{0Kj}\right) \left(\mathbb{N}_{Kj'} - e^{\beta_0^T Z}\Lambda_{0Kj'}\right) |Z, K, T_{K,j}, T_{K,j'} \right].
$$

To verify A4, we note that the first part automatically holds, because

$$
S_{1n}^{ps}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}) = \mathbb{P}_n m_1^{ps}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}; X) = 0
$$

since $\hat\beta_n^{ps}$ satisfies the pseudo-score equation. Next we shall show that

$$
\begin{aligned}
(21) \qquad & S_{2n}^{ps}(\hat\beta_n^{ps}, \hat\Lambda_n^{ps})[\mathbf{h}^*] \\
&= \mathbb{P}_n \left[\sum_{j=1}^K \frac{1}{\hat\Lambda_{nKj}^{ps}} \left\{ \mathbb{N}_{Kj} - \hat\Lambda_{nKj}^{ps} \exp(\hat\beta_n^{psT} Z) \right\} \right. \\
&\qquad \left. \cdot \Lambda_{0Kj} \frac{E(Z\exp(\beta_0^T Z)|K, T_{K,j})}{E(\exp(\beta_0^T Z)|K, T_{K,j})} \right] \\
&= o_P(n^{-1/2})
\end{aligned}
$$

with $\hat\Lambda_{nKj}^{ps} = \hat\Lambda_n^{ps}(T_{K,j})$. Since $(\hat\beta_n^{ps}, \hat\Lambda_n^{ps})$ maximizes $\mathbb{P}_n m_\theta^{ps}(X)$ over the feasible region, consider a path $\theta_\epsilon = (\hat\beta_n^{ps}, \hat\Lambda_n^{ps} + \epsilon h)$ for $h \in \mathcal{F}$. Then

$$
\lim_{\epsilon \downarrow 0} \frac{d}{d\epsilon} \mathbb{P}_n m_{\theta_\epsilon}^{ps}(X) = \mathbb{P}_n \left[\sum_{j=1}^K \frac{1}{\hat\Lambda_{nKj}^{ps}} \left\{ \mathbb{N}_{Kj} - \hat\Lambda_{nKj}^{ps} \exp(\hat\beta_n^{psT} Z) \right\} h_{Kj} \right] = 0 .
$$

Now choose $h_{Kj} = \hat\Lambda_{nKj}^{ps} E(Z\exp(\beta_0^T Z)|K, T_{K,j}) / E(\exp(\beta_0^T Z)|K, T_{K,j})$. Then to show (21), it suffices to show that

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

where $\alpha_{Kj} = E(Z\exp(\beta_0^T Z)|K, T_{K,j}) / E(\exp(\beta_0^T Z)|K, T_{K,j})$. But I can be decomposed as $I = I_1 - I_2 + I_3$, where

$$
I_1 = (\mathbb{P}_n - P) \left\{ \sum_{j=1}^K \frac{\mathbb{N}_{Kj}}{\hat\Lambda_{nKj}^{ps}} (\Lambda_{0Kj} - \hat\Lambda_{nKj}^{ps}) \alpha_{Kj} \right\},
$$

$$
I_2 = (\mathbb{P}_n - P) \left\{ \sum_{j=1}^K \exp(\hat\beta_n^{psT} Z) (\Lambda_{0Kj} - \hat\Lambda_{nKj}^{ps}) \alpha_{Kj} \right\}
$$

and

$$
I_3 = P \left[\sum_{j=1}^K \frac{1}{\hat\Lambda_{nKj}^{ps}} \left\{ \mathbb{N}_{Kj} - \hat\Lambda_{nKj}^{ps} \exp(\hat\beta_n^{psT} Z) \right\} (\Lambda_{0Kj} - \hat\Lambda_{nKj}^{ps}) \alpha_{Kj} \right].
$$

We show that I_1, I_2 and I_3 are all $o_P(n^{-1/2})$. Let

$$\phi_1(X;\Lambda) = \sum_{j=1}^{K} \frac{\mathbb{N}_{Kj}}{\Lambda_{Kj}}(\Lambda_{0Kj} - \Lambda_{Kj})\alpha_{Kj},$$

$$\phi_2(X;\beta,\Lambda) = \sum_{j=1}^{K} \exp(\beta^T Z)(\Lambda_{0Kj} - \Lambda_{Kj})\alpha_{Kj},$$

and define two classes $\Phi_1(\eta)$ and $\Phi_2(\eta)$ as follows:

$$\Phi_1(\eta) = \left\{\phi_1 : \Lambda \in \mathcal{F} \quad \text{and} \quad \|\Lambda - \Lambda_0\|_{L_2(\mu_1)} \leq \eta\right\}$$

and

$$\Phi_2(\eta) = \{\phi_2 : (\beta,\Lambda) \in \mathcal{R} \times \mathcal{F} \quad \text{and} \quad d_1((\beta,\Lambda),(\beta_0,\Lambda_0)) \leq \eta\}.$$

Using the same bracketing entropy arguments as used in deriving the rate of convergence, it follows that both $\Phi_1(\eta)$ and $\Phi_2(\eta)$ are P-Donsker classes under conditions C1, C6 and C8. Moreover, for the seminorm $\rho_P(f) = \left\{P(f - Pf)^2\right\}^{1/2}$, under conditions C1, C6, C8 and C9, we have $\sup_{\phi_1 \in \Phi_1(\eta)} \rho_P(\phi_1) \to 0$ and $\sup_{\phi_2 \in \Phi_2(\eta)} \rho_P(\phi_2) \to 0$ if $\eta \to 0$. Due to the relationship between P-Donsker and asymptotic equicontinuity (see Corollary 2.3.12 of van der Vaart and Wellner (1996)), this yields $I_1 = o_P(n^{-1/2})$ and $I_2 = o_P(n^{-1/2})$. For I_3, we have

$$\begin{aligned}
I_3 &= P\left[\sum_{j=1}^{K} \frac{\left\{\mathbb{N}_{Kj} - \hat{\Lambda}^{ps}_{nKj}\exp(\hat{\beta}^{ps'}_n Z)\right\}}{\hat{\Lambda}^{ps}_{nKj}}(\Lambda_{0Kj} - \hat{\Lambda}^{ps}_{nKj})\alpha_{Kj}\right] \\
&= E\left[\sum_{j=1}^{K} \frac{\left\{\Lambda_{0Kj}\exp(\beta_0^T Z) - \hat{\Lambda}^{ps}_{nKj}\exp(\hat{\beta}^{psT}_n Z)\right\}}{\hat{\Lambda}^{ps}_{nKj}}(\Lambda_{0Kj} - \hat{\Lambda}^{ps}_{nKj})\alpha_{Kj}\right] \\
&= E\left[\sum_{j=1}^{K} \frac{\left\{(\Lambda_{0Kj} - \hat{\Lambda}^{ps}_{nKj})e^{\beta_0^T Z} + \hat{\Lambda}^{ps}_{nKj}(e^{\beta_0^T Z} - e^{\beta^T Z})\right\}}{\hat{\Lambda}^{ps}_{nKj}}(\Lambda_{0Kj} - \hat{\Lambda}^{ps}_{nKj})\alpha_{Kj}\right] \\
&\leq C d_1^2\left\{(\hat{\beta}^{ps}_n, \hat{\Lambda}^{ps}_n), (\beta_0,\Lambda_0)\right\},
\end{aligned}$$

by performing Taylor expansion of $\exp(\beta^T Z)$ at β_0 along with conditions C1, C6, C8, and the result of Lemma 8.2. Finally the rate of convergence yields $I_3 \leq Cn^{-2/3}$ in probability and thus $I_3 = o_P(n^{-1/2})$.

To verify A5, we note that

$$\sqrt{n}(S_{1n}^{ps} - S_1^{ps})(\beta, \Lambda) - \sqrt{n}(S_{1n}^{ps} - S_1^{ps})(\beta_0, \Lambda_0) \\ = \mathbb{G}_n \left[\sum_{j=1}^{K} Z \left\{ \Lambda_{0Kj} \exp(\beta_0^T Z) - \Lambda_{Kj} \exp(\beta^T Z) \right\} \right] \tag{22}$$

and

$$\sqrt{n}(S_{2n}^{ps} - S_2^{ps})(\beta, \Lambda)[\mathbf{h}^*] - \sqrt{n}(S_{2n}^{ps} - S_2^{ps})(\beta_0, \Lambda_0)[\mathbf{h}^*] \\ = \mathbb{G}_n \left(\sum_{j=1}^{K} \left[\left(\frac{\mathbb{N}_{Kj}}{\Lambda_{Kj}} - \frac{\mathbb{N}_{Kj}}{\Lambda_{0Kj}} \right) - \left\{ e^{\beta^T Z} - e^{\beta_0^T Z} \right\} \right] \Lambda_{0Kj} \alpha_{Kj} \right). \tag{23}$$

Let

$$a(\beta, \Lambda; X) = \sum_{j=1}^{K} Z \left\{ \Lambda_{0Kj} \exp(\beta_0^T Z) - \Lambda_{Kj} \exp(\beta^T Z) \right\}$$

and

$$b(\beta, \Lambda; X) = \sum_{j=1}^{K} \left[\mathbb{N}_{Kj} \left(\frac{1}{\Lambda_{Kj}} - \frac{1}{\Lambda_{0Kj}} \right) - \left\{ \exp(\beta^T Z) - \exp(\beta_0^T Z) \right\} \right] \Lambda_{0Kj} \alpha_{Kj}.$$

For a $\eta > 0$, we define

$$A(\eta) = \{ a(\beta, \Lambda; X) : d_1 \{ (\beta, \Lambda), (\beta_0, \Lambda_0) \} \le \eta, \quad \text{and} \quad (\beta, \Lambda) \in \mathcal{R} \times \mathcal{F} \}$$

and

$$B(\eta) = \{ b(\beta, \Lambda; X) : d_1 \{ (\beta, \Lambda), (\beta_0, \Lambda_0) \} \le \eta, \quad \text{and} \quad (\beta, \Lambda) \in \mathcal{R} \times \mathcal{F} \}.$$

Then by applying the bracketing entropy arguments as in the rate of convergence proof, we can show that both $A(\eta)$ and $B(\eta)$ are P-Donsker classes under conditions C1, C6 and C8 and for a small enough $\eta > 0$. We can also show that $\sup_{a \in A(\eta)} \rho_P \{ a(\beta, \Lambda; X) \} \to 0$ and $\sup_{b \in B(\eta)} \rho_P \{ b(\beta, \Lambda; X) \} \to 0$ if $\eta \to 0$. Then the rate of convergence along with Corollary 2.3.12 of van der Vaart and Wellner (1996) yields that

$$\sup_{|\beta - \beta_0| \le \sigma_n, \|\Lambda - \Lambda_0\| \le C n^{-1/3}} |\mathbb{G}_n a(\beta, \Lambda; X)| = o_P(1)$$

and

$$\sup_{|\beta-\beta_0|\le\sigma_n,\|\Lambda-\Lambda_0\|\le Cn^{-1/3}} |\mathbb{G}_n b(\beta,\Lambda;X)| = o_P(1).$$

Hence A5 holds with $\gamma = 1/3$.

Finally, to verify A6, performing Taylor expansion of $m_1^{ps}(\beta,\Lambda;X)$ at point (β_0,Λ_0), we have

$$\begin{aligned}
m_1^{ps}(\beta,\Lambda:X) &= \sum_{j=1}^K Z\left\{\mathbb{N}_{Kj} - \Lambda_{Kj}\exp(\beta^T Z)\right\}\\
&= m_1^{ps}(\beta_0,\Lambda_0;X) + m_{11}^{ps}(\beta_0,\Lambda_0;X)(\beta-\beta_0) + m_{12}^{ps}(\beta_0,\Lambda_0;X)[\Lambda-\Lambda_0]\\
&\quad -\sum_{j=1}^K \exp(\beta_0^T Z) ZZ^T(\beta-\beta_0)(\Lambda_{Kj}-\Lambda_{0Kj})\\
&\quad -\frac{1}{2}\sum_{j=1}^K Z\exp(\beta_\xi^T Z)\Lambda_{0Kj}(\beta-\beta_0)^T ZZ^T(\beta-\beta_0),
\end{aligned}$$

where $\beta_\xi = \beta_0 + \xi(\beta-\beta_0)$ for some $0<\xi<1$. This yields

$$\begin{aligned}
&|S_1^{ps}(\beta,\Lambda) - S_1^{ps}(\beta_0,\Lambda_0) - \dot S_{11}^{ps}(\beta_0,\Lambda_0)(\beta-\beta_0) - \dot S_{12}^{ps}(\beta_0,\Lambda_0)[\Lambda-\Lambda_0]|\\
&= \left|P\left\{\sum_{j=1}^K \exp(\beta_0^T Z)ZZ^T(\beta-\beta_0)(\Lambda_{Kj}-\Lambda_{0Kj})\right.\right.\\
&\quad \left.\left. +\frac{1}{2}\sum_{j=1}^K Z\exp(\beta_\xi^T Z)\Lambda_{0Kj}(\beta-\beta_0)^T ZZ^T(\beta-\beta_0)\right\}\right|.
\end{aligned} \tag{24}$$

Similarly, we have

$$\begin{aligned}
&|S_2^{ps}(\beta,\Lambda)[\mathbf{h}^*] - S_2^{ps}(\beta_0,\Lambda_0)[\mathbf{h}^*] - \dot S_{21}^{ps}(\beta_0,\Lambda_0)[\mathbf{h}^*](\beta-\beta_0)\\
&\qquad -\dot S_{22}^{ps}(\beta_0,\Lambda_0)[\mathbf{h}^*,\Lambda-\Lambda_0]|\\
&= \frac{1}{2}\left|P\left\{\sum_{j=1}^K \frac{(\Lambda_{Kj}-\Lambda_{0Kj})^2}{\Lambda_{\zeta Kj}}\mathbb{N}_{Kj}{\mathbf{h}^*}_{kj}\right.\right.\\
&\qquad \left.\left. +\sum_{j=1}^K Z\exp(\beta_\zeta^T Z)(\beta-\beta_0)^T ZZ^T(\beta-\beta_0)\right\}\right|,
\end{aligned} \tag{25}$$

where $\beta_\zeta = \beta_0 + \zeta(\beta - \beta_0)$ and $\Lambda_{\zeta Kj} = \Lambda_{0Kj} + \zeta(\Lambda_{Kj} - \Lambda_{0Kj})$ for some $0 \le \zeta \le 1$. Hence by C1, C3, C6, C7 and C8, it follows that

$$(24) \text{ and } (25) \le C\left\{|\beta - \beta_0|^2 + \|\Lambda - \Lambda_0\|^2_{L_2(\mu_1)}\right\},$$

so A6 holds with $\alpha = 2$ and thus the proof for the first part of Theorem 3.3 is complete.

For the second part, first we note that with a single observation, $m(\beta, \Lambda; X) = \sum_{j=1}^K \left\{\Delta\mathbb{N}_{Kj} \log \Delta\Lambda_{Kj} + \Delta\mathbb{N}_{Kj}\beta^T Z - e^{\beta^T Z}\Delta\Lambda_{Kj}\right\}$, and hence

$$\begin{aligned}
m_1(\beta, \Lambda; X) &= \sum_{j=1}^K Z\left[\Delta\mathbb{N}_{Kj} - \Delta\Lambda_{Kj}e^{\beta^T Z}\right],\\
m_2(\beta, \Lambda; X)[h] &= \sum_{j=1}^K \left[\frac{\Delta\mathbb{N}_{Kj}}{\Delta\Lambda_{Kj}} - e^{\beta^T Z}\right]\Delta h_{Kj},\\
m_{11}(\beta, \Lambda; X) &= -\sum_{j=1}^K \Delta\Lambda_{Kj} Z Z^T e^{\beta^T Z},\\
m_{12}(\beta, \Lambda; X)[h] &= m_{21}^T(\beta, \Lambda; X)[h] = -\sum_{j=1}^K Z e^{\beta^T Z}\Delta h_{Kj},\\
m_{22}(\beta, \Lambda; X)[\mathbf{h}, h] &= -\sum_{j=1}^K \frac{\Delta\mathbb{N}_{Kj}}{(\Delta\Lambda_{Kj})^2}\Delta\mathbf{h}_{Kj}\Delta h_{Kj},
\end{aligned}$$

where $\Delta h_{Kj} = \int_{T_{K,j-1}}^{T_{K,j}} h d\Lambda$ for $h \in L_2(\Lambda)$. A1 holds with $\gamma = 1/3$ and the norm $\|\cdot\|$ being $L_2(\mu_2)$ because of the rate of convergence established in Theorem 4.2. A2 holds by the model specification (1.1). For A3, we need to find a $\mathbf{h}^*$ such that

$$\begin{aligned}
&\dot{S}_{12}(\beta_0, \Lambda_0)[h] - \dot{S}_{22}(\beta_0, \Lambda_0)[\mathbf{h}^*, h]\\
&\quad = P\left\{m_{12}(\beta_0, \Lambda_0; X)[h] - m_{22}(\beta_0, \Lambda_0; X)[\mathbf{h}^*, h]\right\} = 0,
\end{aligned}$$

for all $h \in L_2(\Lambda_0)$. Note that

$$P\left\{m_{12}(\beta_0, \Lambda_0; X)[h] - m_{22}(\beta_0, \Lambda_0; X)[\mathbf{h}^*, h]\right\}$$

$$
\begin{aligned}
&= -E\left\{\sum_{j=1}^{K}\left[Ze^{\beta_0' Z} - \frac{\Delta\mathbb{N}_{Kj}}{(\Delta\Lambda_{0Kj})^2}\Delta\mathbf{h}^*_{Kj}\right]\Delta h_{Kj}\right\} \\
&= -E_{(K,\underline{T}_K,Z)}\left\{\sum_{j=1}^{K}\left[Ze^{\beta_0' Z} - \frac{e^{\beta_0' Z}\Delta\mathbf{h}^*_{Kj}}{\Delta\Lambda_{0Kj}}\right]\Delta h_{Kj}\right\}.
\end{aligned}
$$

Therefore, an obvious choice of $\mathbf{h}^*$ satisfies

$$
\Delta\mathbf{h}^*_{Kj} = \Delta\Lambda_{0Kj}\frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}.
$$

Hence

$$
\begin{aligned}
&m^*(\beta_0, \Lambda_0; X) \\
&\quad = m_1(\beta_0, \Lambda_0; X) - m_2(\beta_0, \Lambda_0; X)[\mathbf{h}^*] \\
&\quad = \sum_{j=1}^{K}\left\{Z\left(\Delta\mathbb{N}_{Kj} - e^{\beta_0^T Z}\Delta\Lambda_{0Kj}\right)\right. \\
&\qquad \left. - \left(\frac{\Delta\mathbb{N}_{Kj}}{\Delta\Lambda_{0Kj}} - e^{\beta_0^T Z}\right)\Delta\Lambda_{0Kj}\frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}\right\} \\
&\quad = \sum_{j=1}^{K}\left(\Delta\mathbb{N}_{Kj} - e^{\beta_0^T Z}\Delta\Lambda_{0Kj}\right)\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}\right],
\end{aligned}
$$

$$
\begin{aligned}
A &= -\dot{S}_{11}(\beta_0, \Lambda_0) + \dot{S}_{21}(\beta_0, \Lambda_0)[\mathbf{h}^*] \\
&= E_{(K,\underline{T}_K,Z)}\left\{\sum_{j=1}^{K}\Delta\Lambda_{0Kj}e^{\beta_0^T Z}\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}\right]Z^T\right\} \\
&= E_{(K,\underline{T}_K,Z)}\left\{\sum_{j=1}^{K}\Delta\Lambda_{0Kj}e^{\beta_0^T Z}\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}\right]^{\otimes 2}\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
B &= Em^*(\beta_0, \Lambda_0; X)^{\otimes 2} \\
&= E_{(K,\underline{T}_K,Z)}\left\{\sum_{j,j'=1}^{K}C_{j,j'}(Z)\left[Z - \frac{E\left(Ze^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\beta_0^T Z}|K, T_{K,j-1}, T_{K,j}\right)}\right]^{\otimes 2}\right\}
\end{aligned}
$$

with

$$C_{j,j'}(Z) = E\left[\left(\Delta\mathbb{N}_{Kj} - e^{\beta_0^T Z}\Delta\Lambda_{0Kj}\right)\left(\Delta\mathbb{N}_{Kj'} - e^{\beta_0^T Z}\Delta\Lambda_{0Kj'}\right) |Z, K, T_{K,j-1}, T_{K,j}, T_{K,j'-1}, T_{K,j'}\right].$$

To verify A4, we only need to check the second part, namely to show that

$$\begin{aligned} & S_{2n}(\hat\beta_n, \hat\Lambda_n)[\mathbf{h}^*] \\ = & \mathbb{P}_n\left[\sum_{j=1}^K \frac{\left\{\Delta\mathbb{N}_{Kj} - \Delta\hat\Lambda_{nKj}\exp(\hat\beta_n' Z)\right\}}{\Delta\hat\Lambda_{nKj}} \right. \\ & \left. \Delta\Lambda_{0Kj}\frac{E\left(Z\exp(\beta_0^T Z)|K, T_{K,j-1}, T_{K,j}\right)}{E\left(\exp(\beta_0^T Z)|K, T_{K,j-1}, T_{K,j}\right)}\right] \\ = & o_P(n^{-1/2}). \end{aligned}$$

Denote $\alpha_{Kj}^* = E\left(Z\exp(\beta_0^T Z)|K, T_{K,j-1}, T_{K,j}\right)/E\left(\exp(\beta_0^T Z)|K, T_{K,j-1}, T_{K,j}\right)$ for all $j = 1, 2, \ldots, K$; $K = 1, 2, \ldots$, using the same arguments as those in the case of pseudo-likelihood method, it is equivalent to show that

$$\begin{aligned} I^* &= \mathbb{P}_n\left[\sum_{j=1}^K \frac{1}{\Delta\hat\Lambda_{nKj}}\left\{\Delta\mathbb{N}_{Kj} - \Delta\hat\Lambda_{nKj}\exp(\hat\beta_n' Z)\right\}(\Delta\Lambda_{0Kj} - \Delta\hat\Lambda_{nKj})\alpha_{Kj}^*\right] \\ &= o_P(n^{-1/2}) \end{aligned}$$

Again we rewrite I^* as $I^* = I_1^* - I_2^* + I_3^*$, where

$$I_1^* = (\mathbb{P}_n - P)\left\{\sum_{j=1}^K \frac{\Delta\mathbb{N}_{Kj}}{\Delta\hat\Lambda_{nKj}}(\Delta\Lambda_{0Kj} - \Delta\hat\Lambda_{nKj})\alpha_{Kj}^*\right\},$$

$$I_2^* = (\mathbb{P}_n - P)\left\{\sum_{j=1}^K \exp(\hat\beta_n' Z)(\Delta\Lambda_{0Kj} - \Delta\hat\Lambda_{nKj})\alpha_{Kj}^*\right\}$$

and

$$I_3^* = P\left[\sum_{j=1}^K \frac{1}{\Delta\hat\Lambda_{nKj}}\left\{\Delta\mathbb{N}_{Kj} - \Delta\hat\Lambda_{nKj}\exp(\hat\beta_n' Z)\right\}(\Delta\Lambda_{0Kj} - \Delta\hat\Lambda_{nKj})\alpha_{Kj}^*\right].$$

We shall show that I_1^*, I_2^* and I_3^* are all $o_P(n^{-1/2})$. Let

$$\phi_1^*(\Lambda; X) = \sum_{j=1}^{K} \frac{\Delta\mathbb{N}_{Kj}}{\Delta\Lambda_{Kj}} (\Delta\Lambda_{0Kj} - \Delta\Lambda_{Kj})\alpha_{Kj}^*$$

and

$$\phi_2^*(X; \beta, \Lambda) = \sum_{j=1}^{K} \exp(\beta^T Z)(\Delta\Lambda_{0Kj} - \Delta\Lambda_{Kj})\alpha_{Kj}^*,$$

we define two classes $\Phi_1^*(\eta)$ and $\Phi_2^*(\eta)$ as follows

$$\Phi_1^*(\eta) = \{\phi_1^* : \Lambda \in \mathcal{F} \quad \text{and} \quad \|\Lambda - \Lambda_0\|_{L_2(\mu_2)} \leq \eta\}$$

and

$$\Phi_2^*(\eta) = \{\phi_2^* : (\beta, \Lambda) \in \mathcal{R} \times \mathcal{F} \quad \text{and} \quad d_2\left\{(\beta, \Lambda), (\beta_0, \Lambda_0)\right\} \leq \eta\}.$$

Then by the same bracketing entropy technique used in deriving the rate of convergence of the maximum likelihood estimator, we can show that both $\Phi_1^*(\eta)$ and $\Phi_2^*(\eta)$ are P-Donsker under conditions C1, C6, C8, C11, C12, and results of Lemma 8.2. Moreover, $\sup_{\phi_1^* \in \Phi_1^*(\eta)} \rho_P^2(\phi_1^*) \to 0$ and $\sup_{\phi_2^* \in \Phi_1^*(\eta)} \rho_P^2(\phi_2^*) \to 0$ if $\eta \to 0$. Hence the conclusions $I_1^* = o_P(n^{-1/2})$ and $I_2^* = o_P(n^{-1/2})$ follow from Corollary 2.3.12 of van der Vaart and Wellner (1996). Also as the same as in the first part, the rate of convergence leads to

$$I_3^* \leq C d_2^2 \left((\hat{\beta}_n, \hat{\Lambda}_n), (\beta_0, \Lambda_0) \right) \leq C n^{-2/3} \quad \text{in probability.}$$

This concludes the proof of $I^* = o_P(n^{-1/2})$.

For A5, we note that

$$\begin{aligned} &\sqrt{n}(S_{1n} - S_1)(\beta, \Lambda) - \sqrt{n}(S_{1n} - S_1)(\beta_0, \Lambda_0) \\ &\quad = \mathbb{G}_n \left[\sum_{j=1}^{K} Z \left\{ \Delta\Lambda_{0Kj} \exp(\beta_0^T Z) - \Delta\Lambda_{Kj} \exp(\beta^T Z) \right\} \right] \end{aligned} \tag{26}$$

$$\begin{aligned} &\sqrt{n}(S_{2n} - S_2)(\beta, \Lambda)[\mathbf{h}^*] - \sqrt{n}(S_{2n} - S_2)(\beta_0, \Lambda_0)[\mathbf{h}^*] \\ &\quad = \mathbb{G}_n \left(\sum_{j=1}^{K} \left[\Delta\mathbb{N}_{Kj} \left(\frac{1}{\Delta\Lambda_{Kj}} - \frac{1}{\Delta\Lambda_{0Kj}} \right) \right. \right. \\ &\qquad \left. \left. - \left\{ \exp(\beta^T Z) - \exp(\beta_0^T Z) \right\} \right] \Delta\Lambda_{0Kj} \alpha_{Kj}^* \right). \end{aligned} \tag{27}$$

Let

$$a^*(X;\beta,\Lambda) = \sum_{j=1}^K Z\left\{\Delta\Lambda_{0Kj}\exp(\beta_0^T Z) - \Delta\Lambda_{Kj}\exp(\beta^T Z)\right\}$$

and

$$\begin{aligned} b^*(X;\beta,\Lambda) \;\; = \;\; & \sum_{j=1}^K \Bigg[\Delta\mathbb{N}_{Kj}\left(\frac{1}{\Delta\Lambda_{Kj}} - \frac{1}{\Delta\Lambda_{0Kj}}\right) \\ & - \;\left\{\exp(\beta^T Z) - \exp(\beta_0^T Z)\right\}\Bigg]\Delta\Lambda_{0Kj}\alpha^*_{Kj}. \end{aligned}$$

We define two classes

$$A^*(\eta) = \{a^*(X;\beta,\Lambda):\; d_2\{(\beta,\Lambda),(\beta_0,\Lambda_0)\} \le \eta \quad \text{and} \quad (\beta,\Lambda)\in\mathcal{R}\times\mathcal{F}\}$$

and

$$B^*(\eta) = \{b^*(X;\beta,\Lambda):\; d_2\{(\beta,\Lambda),(\beta_0,\Lambda_0)\} \le \eta \quad \text{and} \quad (\beta,\Lambda)\in\mathcal{R}\times\mathcal{F}\}.$$

Then conditions C1, C3, C6, C7, C8, C9, C11, C12 and result of Lemma 8.2 lead that both $A^*(\eta)$ and $B^*(\eta)$ are P-Donsker classes for a sufficient small $\eta > 0$. Similarly, we can show that $\sup_{a^*\in A^*(\eta)}\rho_P^2(a^*) \to 0$ and $\sup_{b^*\in B^*(\eta)}\rho_P^2(b^*) \to 0$ if $\eta\to 0$. Hence by Corollary 2.3.12 of van der Vaart and Wellner (1996) and the rate of convergence, it follows that

$$\sup_{|\beta-\beta_0|\le\delta_n, \|\Lambda-\Lambda_0\|\le Cn^{-1/3}} |\mathbb{G}_n a^*(X;\beta,\Lambda)| = o_P(1)$$

and

$$\sup_{|\beta-\beta_0|\le\delta_n, \|\Lambda-\Lambda_0\|\le Cn^{-1/3}} |\mathbb{G}_n b^*(X;\beta,\Lambda)| = o_P(1).$$

Finally, performing Taylor expansions of $S_1(\beta,\Lambda)$ and $S_2(\beta,\Lambda)[\mathbf{h}^*]$ at (β_0,Λ_0), we have

$$\begin{aligned} & \left|S_1(\beta,\Lambda) - S_1(\beta_0,\Lambda_0) - \dot S_{11}(\beta_0,\Lambda_0)(\beta-\beta_0) - \dot S_{12}(\beta_0,\Lambda_0)[\Lambda-\Lambda_0]\right| \\ & = \Bigg|P\Bigg\{\sum_{j=1}^K \exp(\beta_0^T Z) Z Z^T(\beta-\beta_0)(\Delta\Lambda_{Kj} - \Delta\Lambda_{0Kj}) \\ & \quad + \frac{1}{2}\sum_{j=1}^K Z\exp(\beta_\xi^T Z)\Delta\Lambda_{0Kj}(\beta-\beta_0)^T Z Z^T(\beta-\beta_0)\Bigg\}\Bigg| \end{aligned} \tag{28}$$

and

$$\begin{aligned}
&\Big| S_2(\beta,\Lambda)[\mathbf{h}^*] - S_2(\beta_0,\Lambda_0)[\mathbf{h}^*] - \dot{S}_{21}(\beta_0,\Lambda_0)[\mathbf{h}^*](\beta-\beta_0) \\
&\qquad - \dot{S}_{22}(\beta_0,\Lambda_0)[\mathbf{h}^*, \Lambda-\Lambda_0] \Big| \\
&= \frac{1}{2}\left| P\left\{ \sum_{j=1}^{K} \frac{(\Delta\Lambda_{Kj})^2}{\Delta\Lambda_{\zeta Kj}} \Delta\mathbb{N}_{Kj}\Delta\mathbf{h}^*{}_{Kj} \right.\right. \\
&\qquad \left.\left. + \sum_{j=1}^{K} \exp(\beta_\zeta^T Z)(\beta-\beta_0)^T Z Z^T (\beta-\beta_0) \right\}\right|, \qquad (29)
\end{aligned}$$

where $\beta_\xi = \beta_0 + \xi(\beta-\beta_0)$, $\beta_\zeta = \beta_0 + \zeta(\beta-\beta_0)$ and $\Delta\Lambda_{\zeta Kj} = \Delta\Lambda_{0Kj} + \zeta(\Delta\Lambda_{Kj} - \Delta\Lambda_{0Kj})$ for $0 \le \xi, \zeta \le 1$. Thus by conditions C1-C3, C6-C9, C11 and C12, we conclude that

$$(28) \text{ and } (29) \le C d_2^2\left((\beta,\Lambda),(\beta_0,\Lambda_0)\right).$$

So A6 holds with $\alpha = 2$, and the asymptotic normality of $\hat{\beta}_n$ is then the direct result of Theorem 7.1. □

7. Appendix A: A general theorem on the asymptotic normality of semiparametric M-estimators. In this section, we present a general theorem dealing with the asymptotic normality of semiparametric M-estimators of regression parameters when the rate of convergence of the estimator for nuisance parameter is slower than $n^{-1/2}$. We consider a general setting of a semiparametric model: given i.i.d. observations $X_1, X_2, \ldots, X_n$, we estimate unknown parameters (β,Λ) by maximizing an objective function $n^{-1}\sum_{i=1}^n m(\beta,\Lambda;X_i) = \mathbb{P}_n m(\beta,\Lambda;X)$, where β is a finite-dimensional parameter and Λ is an infinite-dimensional parameter. If m happens to be the log-likelihood function based on a single observation, then the estimator is simply the semiparametric maximum likelihood estimator. This theorem generalizes the theorem developed by Huang (1996) to accommodate the situation in which data may not come from the proposed model. The notation we use here follows that of Huang (1996).

Let $\beta = (\beta, \Lambda)$, where $\beta \in \mathbb{R}^d$, and Λ is an infinite-dimensional parameter in the class $\mathcal{F}$. Suppose that Λ_η is a parametric path in $\mathcal{F}$ through Λ, i.e. $\Lambda_\eta \in \mathcal{F}$, and $\Lambda_\eta \,|_{\eta=0} = \Lambda$.

Let $\mathbf{H} = \left\{h : \quad h = \frac{\partial \Lambda_\eta}{\partial \eta} \,|_{\eta=0}\right\}$ and for any $h \in \mathbf{H}$, we define

$$m_1(\beta, \Lambda; x) = \nabla_\beta m(\beta, \Lambda; x) \equiv \left(\frac{\partial m(\beta, \Lambda; x)}{\partial \beta_1}, \cdots, \frac{\partial m(\beta, \Lambda; x)}{\partial \beta_d}\right)^T,$$

$$m_2(\beta, \Lambda; x)[h] = \frac{\partial m(\beta, \Lambda_\eta; x)}{\partial \eta} \,|_{\eta=0} ,$$

$$m_{11}(\beta, \Lambda; x) = \nabla^2_\beta m(\beta, \Lambda; x),$$

$$m_{12}(\beta, \Lambda; x)[h] = \frac{\partial m_1(\beta, \Lambda_\eta; x)}{\partial \eta} \,|_{\eta=0} ,$$

$$m_{21}(\beta, \Lambda; x)[h] = \nabla_\beta m_2(\beta, \Lambda; x)[h],$$

$$m_{22}(\beta, \Lambda; x)[h, h] = \frac{\partial^2 m(\beta, \Lambda_\eta; x)}{\partial \eta^2} \,|_{\eta=0} .$$

We also define

$$S_1(\beta, \Lambda) = P m_1(\beta, \Lambda; X),$$

$$S_2(\beta, \Lambda)[h] = P m_2(\beta, \Lambda; X)[h],$$

$$S_{1n}(\beta, \Lambda) = \mathbb{P}_n m_1(\beta, \Lambda; X),$$

$$S_{2n}(\beta, \Lambda)[h] = \mathbb{P}_n m_2(\beta, \Lambda; X)[h],$$

$$\dot{S}_{11}(\beta, \Lambda) = P m_{11}(\beta, \Lambda; X),$$

$$\dot{S}_{12}(\beta, \Lambda)[h] = \dot{S}^T_{21}(\beta, \Lambda)[h] = P m_{12}(\beta, \Lambda; X)[h],$$

$$\dot{S}_{22}(\beta, \Lambda)[h, h] = P m_{22}(\beta, \Lambda; X)[h, h].$$

Furthermore, for $\mathbf{h} = (h_1, h_2, \cdots, h_d)^T \in \mathbf{H}^d$, where $h_j \in \mathbf{H}$ for $j = 1, 2, \cdots, d$, we denote

$$m_2(\beta, \Lambda; x)[\mathbf{h}] = (m_2(\beta, \Lambda; x)[h_1], \cdots, m_2(\beta, \Lambda; x)[h_d])^T ,$$

$$m_{12}(\beta, \Lambda; x)[\mathbf{h}] = (m_{12}(\beta, \Lambda; x)[h_1], \cdots, m_{12}(\beta, \Lambda; x)[h_d]) ,$$

$$m_{21}(\beta, \Lambda; x)[\mathbf{h}] = (m_{21}(\beta, \Lambda; x)[h_1], \cdots, m_{21}(\beta, \Lambda; x)[h_d])^T ,$$

$$m_{22}(\beta,\Lambda;x)[\mathbf{h},h] = \left(m_{22}(\beta,\Lambda;x)[h_1,h],\cdots,m_{22}(\beta,\Lambda;x)[h_d,h]\right)^T,$$

and define

$$S_2(\beta,\Lambda)[\mathbf{h}] = Pm_2(\beta,\Lambda;X)[\mathbf{h}],$$

$$S_{2n}(\beta,\Lambda)[\mathbf{h}] = \mathbb{P}_n m_2(\beta,\Lambda;X)[\mathbf{h}],$$

$$\dot{S}_{12}(\beta,\Lambda)[\mathbf{h}] = Pm_{12}(\beta,\Lambda;X)[\mathbf{h}],$$

$$\dot{S}_{21}(\beta,\Lambda)[\mathbf{h}] = Pm_{21}(\beta,\Lambda;X)[\mathbf{h}],$$

$$\dot{S}_{22}(\beta,\Lambda)[\mathbf{h},h] = Pm_{22}(\beta,\Lambda;X)[\mathbf{h},h].$$

To establish the asymptotic distribution for the M-estimator $\hat{\beta}_n$, we need the following assumptions:

A1. $\left|\hat{\beta}_n - \beta_0\right| = o_p(1)$ and $\left\|\hat{\Lambda}_n - \Lambda_0\right\| = O_p(n^{-\gamma})$ for some $\gamma > 0$ and some norm $\|\cdot\|$.

A2. $S_1(\beta_0,\Lambda_0) = 0$ and $S_2(\beta_0,\Lambda_0)[h] = 0$ for all $h \in \mathbf{H}$.

A3. There exists an $\mathbf{h}^* = (h_1^*,\cdots,h_d^*)^T$, where $h_j^* \in \mathbf{H}$ for $j = 1,\cdots,d$, such that

$$\dot{S}_{12}(\beta_0,\Lambda_0)[h] - \dot{S}_{22}(\beta_0,\Lambda_0)[\mathbf{h}^*,h] = 0,$$

for all $h \in \mathbf{H}$. Moreover, the matrix

$$A = -\dot{S}_{11}(\beta_0,\Lambda_0) + \dot{S}_{21}(\beta_0,\Lambda_0)[\mathbf{h}^*] = -P\left(m_{11}(\beta_0,\Lambda_0;X) - m_{21}(\beta_0,\Lambda_0;X)[\mathbf{h}^*]\right)$$

is non-singular.

A4. The estimator $(\hat{\beta}_n,\hat{\Lambda}_n)$ satisfies

$$S_{1n}(\hat{\beta}_n,\hat{\Lambda}_n) = o_P(n^{-1/2}),$$

and

$$S_{2n}(\hat{\beta}_n,\hat{\Lambda}_n)[\mathbf{h}^*] = o_P(n^{-1/2}).$$

A5. For any $\delta_n \downarrow 0$ and $C > 0$

$$\sup_{|\beta-\beta_0|\le\delta_n,\|\Lambda-\Lambda_0\|\le Cn^{-\gamma}} \left|\sqrt{n}(S_{1n}-S_1)(\beta,\Lambda) - \sqrt{n}(S_{1n}-S_1)(\beta_0,\Lambda_0)\right| = o_P(1),$$

and

$$\sup_{|\beta-\beta_0|\leq\delta_n, \|\Lambda-\Lambda_0\|\leq Cn^{-\gamma}} \left|\sqrt{n}(S_{2n}-S_2)(\beta,\Lambda)[\mathbf{h}^*] - \sqrt{n}(S_{2n}-S_2)(\beta_0,\Lambda_0)[\mathbf{h}^*]\right|$$
$$= o_P(1).$$

A6. For some $\alpha > 1$ satisfying $\alpha\gamma > 1/2$, and for (β, Λ) in a neighborhood of (β_0, Λ_0): $\{(\beta,\Lambda) : |\beta-\beta_0| \leq \delta_n, \quad \|\Lambda-\Lambda_0\| \leq Cn^{-\gamma}\}$,

$$\left|S_1(\beta,\Lambda) - S_1(\beta_0,\Lambda_0) - \dot{S}_{11}(\beta_0,\Lambda_0)(\beta-\beta_0) - \dot{S}_{12}(\beta_0,\Lambda_0)[\Lambda-\Lambda_0]\right|$$
$$= o\left(|\beta-\beta_0|\right) + O\left(\|\Lambda-\Lambda_0\|^\alpha\right),$$

and

$$\left|S_2(\beta,\Lambda)[\mathbf{h}^*] - S_2(\beta_0,\Lambda_0)[\mathbf{h}^*] - \dot{S}_{21}(\beta_0,\Lambda_0)[\mathbf{h}^*](\beta-\beta_0) - \dot{S}_{22}(\beta_0,\Lambda_0)[\mathbf{h}^*,\Lambda-\Lambda_0]\right|$$
$$= o\left(|\beta-\beta_0|\right) + O\left(\|\Lambda-\Lambda_0\|^\alpha\right).$$

Theorem 7.1 *Suppose that Assumptions A1-A6 hold. Then*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = A^{-1}\sqrt{n}\mathbb{P}_n m^*(\beta_0,\Lambda_0;X) + o_{p^*}(1) \to_d N\left(0, A^{-1}B\left(A^{-1}\right)^T\right),$$

where $m^*(\beta_0,\Lambda_0;x) = m_1(\beta_0,\Lambda_0;x) - m_2(\beta_0,\Lambda_0;x)[\mathbf{h}^*]$, $B = Em^*(\beta_0,\Lambda_0;X)^{\otimes 2} = E\left(m^*(\beta_0,\Lambda_0;X)m^*(\beta_0,\Lambda_0;X)^T\right)$, *and* A *is given in Assumption A3.*

Proof: A1 and A5 yield

$$\sqrt{n}(S_{1n}-S_1)(\hat{\beta}_n,\hat{\Lambda}_n) - \sqrt{n}(S_{1n}-S_1)(\beta_0,\Lambda_0) = o_P(1).$$

Since $S_{1n}(\hat{\beta}_n,\hat{\Lambda}_n) = o_{p^*}(n^{-1/2})$ by A4 and $S_1(\beta_0,\Lambda_0) = 0$ by A2, it follows that

$$\sqrt{n}S_1(\hat{\beta}_n,\hat{\Lambda}_n) + \sqrt{n}S_{1n}(\beta_0,\Lambda_0) = o_P(1).$$

Similarly,

$$\sqrt{n}S_2(\hat{\beta}_n,\hat{\Lambda}_n) + \sqrt{n}S_{2n}(\beta_0,\Lambda_0)[\mathbf{h}^*] = o_P(1).$$

Combining these equalities and A6 yields

$$\begin{aligned}&\dot{S}_{11}(\beta_0,\Lambda_0)(\hat{\beta}_n-\beta_0) + \dot{S}_{12}(\beta_0,\Lambda_0)[\hat{\Lambda}_n-\Lambda_0] + S_{1n}(\beta_0,\Lambda_0)\\ &\quad + o(|\hat{\beta}_n-\beta_0|) + O\left(\|\hat{\Lambda}_n-\Lambda_0\|^\alpha\right) = o_P(n^{-1/2}),\end{aligned} \tag{30}$$

and

$$\begin{aligned} &\dot{S}_{21}(\beta_0, \Lambda_0)[\mathbf{h}^*](\hat{\beta}_n - \beta_0) + \dot{S}_{22}(\beta_0, \Lambda_0)[\mathbf{h}^*][\hat{\Lambda}_n - \Lambda_0] \\ &+ S_{2n}(\beta_0, \Lambda_0)[\mathbf{h}^*] + o(|\hat{\beta}_n - \beta_0|) + O\left(\|\hat{\Lambda}_n - \Lambda_0\|^\alpha\right) = o_P(n^{-1/2}). \end{aligned} \tag{31}$$

Because $\alpha\gamma > 1/2$, the rate of convergence assumption A1 implies $\sqrt{n}O\left(\|\hat{\Lambda}_n - \Lambda_0\|^\alpha\right) = o_P(1)$. Thus by A4 and (30) minus (31), it follows that

$$\begin{aligned} &\left(\dot{S}_{11}(\beta_0, \Lambda_0) - \dot{S}_{21}(\beta_0, \Lambda_0)[\mathbf{h}^*]\right)(\hat{\beta}_n - \beta_0) + o(|\hat{\beta}_n - \beta_0|) \\ &\qquad = -\left(S_{1n}(\beta_0, \Lambda_0) - S_{2n}(\beta_0, \Lambda_0)[\mathbf{h}^*]\right) + o_P(n^{-1/2}), \end{aligned}$$

i.e.

$$-\left(A + o(1)\right)(\hat{\beta}_n - \beta_0) = -\mathbb{P}_n m^*(\beta_0, \Lambda_0; X) + o_P(n^{-1/2}).$$

This yields

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta_0) &= \left(A + o(1)\right)^{-1} \sqrt{n}\mathbb{P}_n m^*(\beta_0, \Lambda_0; X) + o_P(1) \\ &\to_d N\left(0, A^{-1}B\left(A^{-1}\right)^T\right). \end{aligned}$$

□

8. Appendix B: Two technical lemmas. **Lemma 8.1** *If $a_0 = b_0$, then for any integer $k \geq 1$,*

$$\sum_{j=1}^{k}(a_j - b_j)^2 \leq k^2 \sum_{j=1}^{k}\left\{(a_j - a_{j-1}) - (b_j - b_{j-1})\right\}^2.$$

Proof: (This is given in Wellner and Zhang (1998), but is included here for completeness.) Note that for $j = 1, 2, \ldots$, since $a_0 = b_0$ we have

$$a_j - b_j = \sum_{j=1}^{j}(a_j - a_{j-1}) - \sum_{j=1}^{j}(b_j - b_{j-1}).$$

Hence it follows by the by the Cauchy-Schwarz inequality that

$$(a_j - b_j)^2 = \left[\sum_{i=1}^{j}\{(a_i - a_{i-1}) - (b_i - b_{i-1})\}\right]^2 \leq j\sum_{i=1}^{j}\{(a_i - a_{i-1}) - (b_i - b_{i-1})\}^2,$$

and thus

$$\begin{aligned}\sum_{j=1}^{k}(a_j - b_j)^2 &\leq \sum_{j=1}^{k} j\sum_{i=1}^{j}[(a_i - a_{i-1}) - (b_i - b_{i-1})]^2 \\ &= \sum_{i=1}^{k}\left\{[(a_i - a_{i-1}) - (b_i - b_{i-1})]^2\sum_{j=i}^{k} j\right\} \\ &\leq k^2\sum_{i=1}^{k}\{(a_i - a_{i-1}) - (b_i - b_{i-1})\}^2.\end{aligned}$$

□

Lemma 8.2 *Suppose that conditions C8, C11, and C12 hold, and that* $\Lambda \in \mathcal{F}$ *satisfies* $\|\Lambda - \Lambda_0\|_{L_2(\mu_1)} \leq \eta$. *Then there exists a constant* C *independent of* Λ *such that*

$$\sup_{t \in O[T]} |\Lambda(t) - \Lambda_0(t)| \leq (\eta/C)^{2/3}.$$

Proof. Suppose that $t_0 \in O[T]$ satisfies

$$|\Lambda(t_0) - \Lambda_0(t_0)| \geq (1/2)\sup_{t \in O[T]} |\Lambda(t) - \Lambda_0(t)| \equiv \xi/2.$$

Then either $\Lambda(t_0) \geq \Lambda_0(t_0) + \xi/2$, or $\Lambda_0(t_0) \geq \Lambda(t_0) + \xi/2$; i.e. $\Lambda(t_0) \leq \Lambda_0(t_0) - \xi/2$. In the first case we have

$$\begin{aligned}\eta^2 &\geq \int\{\Lambda(t) - \Lambda_0(t)\}^2 d\mu_1(t) \\ &\geq \int_{t_0}^{\Lambda_0^{-1}(\xi/2+\Lambda_0(t_0))}\{\Lambda(t) - \Lambda_0(t)\}^2\,\dot{\mu}_1(t)dt \\ &\geq \int_{t_0}^{\Lambda_0^{-1}(\xi/2+\Lambda_0(t_0))}\{\Lambda_0(t_0) + \xi/2 - \Lambda_0(t)\}^2\,\dot{\mu}_1(t)dt\end{aligned}$$

$$
\begin{aligned}
&\geq \int_{\Lambda_0(t_0)}^{\xi/2+\Lambda_0(t_0)} \{\Lambda_0(t_0)+\xi/2-x\}^2 \dot{\mu}_1(\Lambda_0^{-1}(x)) \frac{1}{\Lambda_0'\{\Lambda_0^{-1}(x)\}} dx \\
&\geq (c_0/f_0) \int_{\Lambda_0(t_0)}^{\xi/2+\Lambda_0(t_0)} \{\Lambda_0(t_0)+\xi/2-x\}^2 dx \geq \frac{c_0}{24 f_0}\xi^3 \\
&= \frac{c_0}{24 f_0} \left(\sup_{t \in O[T]} |\Lambda(t)-\Lambda_0(t)| \right)^3 .
\end{aligned}
$$

This yields the stated conclusion with $C \equiv \sqrt{c_0/(24 f_0)}$.
In the second case the same conclusion holds by a similar argument. □

The result of Lemma 8.2 can be extended to the interval $S[T] = (0, \tau)$ as long as C12 is valid on $S[T]$ and $\dot{\mu}_1(t)$ is uniformly bounded away from zero for $t \in S[T]$.

References.

ANDREWS, D. F. AND HERZBERG, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer, New York.

BANERJEE, M. AND WELLNER, J. A. (2001). "Likelihood ratio tests for monotone functions," *Ann. Statist.*, **29**, 1699 - 1731.

BANERJEE, M. (2005). Likelihood based inference for monotone functions: towards a general theory. *Technical Report*, Department of Statistics, University of Michigan.

BYAR, D. P., BLACKARD, C., AND THE VACURG. (1977). "Comparisons of placebo, pyridoxine, and topical thiotepa in preventing stage I bladder cancer." *Urology* **10**, 556 - 561.

COOK, R. J., LAWLESS, J. F., AND NADEAU, C. (1996). "Robust tests for treatment comparisons based on recurrent event responses," *Biometrics*, **52**, 557 - 571.

COX, D. R., AND REID, N. (2004). "A note on pseudo-likelihood constructed from marginal densities," *Biometrika* **91**, 729 - 737.

DEAN, C. B. AND BALSHAW, R. (1997). "Efficiency lost by analyzing counts rather than event times in Poisson and overdispersed Poisson regression models," *J. Amer. Statist. Assoc.*, **92**, 1387-1398.

GAVER, D. P., AND O'MUIRCHEARTAIGH, I. G. (1987). "Robust empirical Bayes analysis of event rates," *Technometrics*, **29**, 1-15.

HOUGAARD, P., LEE, M.T., AND WHITMORE, G. A. (1997). "Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes," *Biometrics*, **53**, 1225 - 1238.

HU, X.J., SUN, J., AND WEI, L.J. (2003). "Regression parameter estimation from panel counts." *Scandinavian Journal of Statistics* **30**, 25-43.

HUANG, J. (1996). "Efficient estimation for the Cox model with interval censoring," *Ann. Statist.*, **24**, 540-568.

JONGBLOED, G. (1998). "The iterative convex minorant algorithm for nonparametric estimation," *Journal of Computation and Graphical Statistics*, **7**, 310-321.

KALBFLEISCH, J. D. AND LAWLESS, J. F. (1985). "The analysis of panel count data under a Markov assumption," *J. Amer. Statist. Assoc.*, **80**, 863 - 871.

LAWLESS, J. F. (1987a). "Regression methods for Poisson process data," *J. Amer. Statist. Assoc.*, **82**, 808-815.

LAWLESS, J. F. (1987b). "Negative binomial and mixed Poisson regression," *Canad. J. Statist.*, **15**, 209-225.

LAWLESS, J. F. AND NADEAU, C. (1995). "Some simple robust methods for the analysis of recurrent events," *Technometrics*, **37**, 158 - 168.

LIN, D. Y., WEI, L. J., YANG, I., AND YING, Z. (2000). "Semiparametric regression for the mean and rate functions of recurrent events," *J. Roy. Statist. Soc. B*, 711-730.

LINDSAY, B. G. (1988). "Composite likelihood methods." In *Statistical Inference from Stochastic Processes*, Ed. N. U. Prahbu, pp. 221 - 239. American Mathematical Society, Providence.

MURPHY, S. AND VAN DER VAART, A. W. (1997). "Semiparametric likelihood ratio inference," *Ann. Statist.*, **25**, 1471 - 1509.

MURPHY, S. AND VAN DER VAART, A. W. (1999). "Observed information in semi-parametric models," *Bernoulli*, **5**, 381–412.

MURPHY, S. AND VAN DER VAART, A. W. (2000). "On profile likelihood," *J. Amer. Statist. Assoc.*, **95**, 449 - 485.

SUN, J. AND KALBFLEISCH, J. D. (1995). "Estimation of the mean function of point processes based on panel count data," *Statistica Sinica*, **5**, 279 - 290.

SUN, J. AND WEI, L. J. (2000). "Regression analysis of panel count data with covariate-dependent observation and censoring times," *J. R. Stat. Soc. Ser. B*, **62**, 293 - 302.

THALL, P. F., AND LACHIN, J. M. (1988). "Analysis of Recurrent Events: Nonparametric Methods for Random-Interval Count Data," *J. Amer. Statist. Assoc.*, **83**, 339-347.

THALL, P. F. (1988). "Mixed Poisson likelihood regression models for longitudinal interval count data," *Biometrics*, **44**, 197-209.

VAN DER VAART, A.W. (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, A.W., AND WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

WEI, L. J., LIN, D. Y., AND WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065 - 1073.

WELLNER, J. A. AND ZHANG, Y. (2000). "Two estimators of the mean of a counting process with panel count data," *Ann. Statist.*, **28**, 779 - 814.

WELLNER, J. A. AND ZHANG, Y. (1998). "Two estimators of the mean of a counting process with panel count data," Technical Report 341, Dept. of Statistics, Univ. Washington. www.stat.washington.edu/www/research/reports/.

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.* (Eds. Lin, D.Y. and P.J., Heagerty). Springer-Verlag, New York, 143-174.

WOLFRAM, S. (1996). *The Mathematica Book.* Third Edition. Wolfram Research, Champaign.

ZHANG, Y. (1998). *Estimation for Counting Processes Based on Incomplete Data*, Unpublished Ph.D. dissertation, University of Washington.

ZHANG, Y. (2002). "A semiparametric pseudo likelihood estimation method for panel count data," *Biometrika*, **89**, 39-48.

DEPARTMENT OF STATISTICS
BOX 354322
UNIVERSITY OF WASHINGTON
SEATTLE, WA 98195-4322
E-MAIL: jaw@stat.washington.edu

UNIVERSITY OF IOWA
BIOSTATISTICS
200 HAWKINS DRIVE, C22 GH
IOWA CITY, IOWA 52242-1009
E-MAIL: ying-j-zhang@uiowa.edu