*Submitted to Bernoulli*

# Nonparametric Inference for Reversed Mean Models with Panel Count Data

LI LIU[1] WEN SU[2] GUOSHENG YIN[2] XINGQIU ZHAO[3] and YING ZHANG[4]

[1]*School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, 430072, China*
*E-mail:* lliu.math@whu.edu.cn
[2]*Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong*
*E-mail:* jenna.wen.su@connect.hku.hk
*E-mail:* gyin@hku.hk; Corresponding author
[3]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong*
*E-mail:* xingqiu.zhao@polyu.edu.hk
[4]*Department of Biostatistics, University of Nebraska Medical Center, Omaha, NE, USA*
*E-mail:* ying.zhang@unmc.edu

Panel count data typically refer to data arising from studies with recurrent events, in which subjects are observed only at discrete time points rather than under continuous observations. We investigate a general situation where a recurrent event process is eventually truncated by an informative terminal event and we are particularly interested in behaviors of the recurrent event process near the terminal event. We propose a reversed mean model for estimating the mean function of the recurrent event process. We develop a two-stage sieve likelihood-based method to estimate the mean function, which overcomes the computational difficulties arising from a nuisance functional parameter involved in the likelihood. The consistency and the convergence rate of the two-stage estimator are established. Allowing for the convergence rate slower than the standard rate, we develop the general weak convergence theory of $M$-estimators with a nuisance functional parameter, and then apply it to the proposed estimator for deriving the asymptotic normality. Furthermore, a class of two-sample tests is developed. The proposed methods are evaluated with extensive simulation studies and illustrated with panel count data from the Chinese Longitudinal Healthy Longevity Study.

*MSC2020 subject classifications:* Primary 62G05, 62G10, 62G20
*Keywords:* Nonparametric tests; Recurrent events; Reversed mean model; Terminal event

## 1. Introduction

Panel count data often arise in many applied fields, for example, econometrics, epidemiology, medicine, public health, and reliability. As a special type of recurrent event data, the event process of panel count data is observed at finite distinct time points. Only the number of recurrent events that have occurred between the observation times is known and the times of event occurrences are not recorded. In many panel studies, the follow-up of subjects may be truncated by a terminal event which is likely to be associated with accumulation of recurrent events. For example, death is the terminal outcome of the worsening of serious illness which is often manifested as disease recurrences, and panel count data of this format are abundant in biomedical research. In a bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group of U.S.A. (Andrews and Herzberg, 1985), observed data include the clinical visit times or observation times and the numbers of recurrent tumors between clinical visits for all patients in the three treatment groups, i.e., placebo, thiotepa,

[*]The first two authors contribute equally to this work.

and pyridoxine. The follow-up of patients was truncated by death, which led to panel count data with the terminal event of death. As another example, the Chinese Longitudinal Healthy Longevity Study (CLHLS) collected information on health status and quality of elderly individuals from 22 provinces in China; refer to Zeng et al. (2017) for a detailed description. One of the study objectives was to identify the major factors contributing to human health and longevity. The CLHLS conducted face-to-face interviews with 9093 individuals aged 77 or older on random days in 1998 as the baseline wave, with follow-up waves in 2000, 2002, 2005, 2008, 2011 and 2014, respectively. During the study period, more than 90% of the participants experienced serious illness at least once and some died or were lost to follow-up. Taking the occurrences of serious illness as the outcome of interest, only the numbers of such events between the waves were observed for each individual, while the exact times of these event occurrences were not tracked. Due to the lack of information on the precise timing of serious illness events, the event counts could be viewed as panel count data truncated by death. Investigators were particularly interested in learning the progression of serious illness prior to death.

For the analysis of panel count data with no terminal event, many methods have been developed, including parametric approaches (e.g., Kalbfeisch and Lawless, 1985; Hinde, 1982; Breslow, 1984; and Thall, 1988), nonparametric estimation procedures (e.g., Sun and Kalbfeisch, 1995; Wellner and Zhang, 2000; Hu et al., 2009; Zhang and Jamshidian, 2003; Huang et al., 2006; Lu et al., 2007), nonparametric testing methods (e.g., Thall and Lachin, 1988; Sun and Fang, 2003; Zhang, 2006; Balakrishnan and Zhao, 2009; Zhao and Zhang, 2017), and semiparametric and nonparametric regression models (e.g., Cheng and Wei, 2000; Sun and Wei, 2000; Hu et al., 2003; Huang et al., 2006; Wellner and Zhang, 2007; Lu et al., 2009; Zhao et al., 2019; Ma and Sundaram, 2018). However, the aforementioned methods cannot handle a more complicated but realistic situation in biomedical studies where a recurrent event process is truncated by a terminal event. There exists limited research to take into account the effect of a terminal event in the panel count data. In particular, two classes of methods have been proposed: joint modeling through frailty variables (e.g., Huang and Wang, 2004; Zeng and Cai, 2010; Sun et al., 2012; Zhou et al., 2017; Diao, Zeng et al., 2017) and marginal modeling by the inverse probability weighting technique (e.g., Zhao et al., 2011; Zhao et al., 2013). The first approach is robust but cannot clearly reveal the relationship between the recurrent event process and the terminal event through shared unknown latent variables, which thus does not emphasize on the behavior of the recurrent event process near the informative terminal event. The second one may be inappropriate when a terminal event such as death truncates the recurrent event process.

There has been limited research studying the stochastic process behavior near a terminal event with application to ascertaining the quality of life and medical cost near death (e.g. Chan and Wang, 2010; Li et al., 2013). To the best of our knowledge, no method has been proposed to model the recurrent event process near a terminal event for studying the direct relationship between the terminal event and the recurrent event process based on panel count data. Our goal is to develop new inference procedures for panel count data truncated by a terminal event by modeling recurrent events reversely from the terminal event, which allows us to focus more on the behavior of recurrent events near the terminal event. For this purpose, we propose a reversed mean model which can intuitively and clearly explain the relationship between the recurrent event process and the terminal event.

The main contributions of this work are fourfold. First, we propose a reversed mean model anchoring at a terminal event to characterize the explicit relationship between a recurrent event process and a terminal event. Second, we develop a two-stage nonparametric sieve likelihood-based estimation procedure by treating the survival function of the terminal event as a nuisance functional parameter. Third, we establish the asymptotic properties of the proposed estimator and, in particular, we develop a general theorem for the asymptotic normality of nonparametric M-estimators with a nuisance functional parameter when estimators have a convergence rate slower than the standard rate $n^{-1/2}$. Fourth, we propose a class of nonparametric tests for comparing mean functions of recurrent event processes with panel count data in the presence of an informative terminal event.

The rest of this paper is organized as follows. In Section 2, we propose a reversed mean model anchoring at a terminal event for a recurrent event process, and develop a two-stage estimation procedure. In Section 3, we establish the asymptotic properties including the consistency, the convergence rate, and the asymptotic normality of nonparametric $M$-estimators with a nuisance functional parameter. A class of nonparametric two-sample tests are developed for nonparametric comparison of mean functions in Section 4. In Section 5, we conduct simulation studies to assess the performance of the proposed estimators and testing procedure. Section 6 reports the analysis results of the CLHLS data, and Section 7 concludes with some remarks. The lemmas and proofs of theorems are given in Section 8, while the proofs of the lemmas are provided in the supplementary materials.

# 2. Methodology

## 2.1. Model setting

In a study involving recurrent events, suppose that each subject yields a counting process $N(t)$, denoting the total number of occurrences of the event of interest up to time $t$, $0 \leq t \leq \tau$, where $\tau$ is a known constant time point. In general, not every subject can be followed up to $\tau$ due to censoring. In a more complicated but common situation, for each subject there exists a terminal event such as death which may truncate the follow-up prior to $\tau$. Let $U$ denote the time of the terminal event which can truncate the counting process $N(\cdot)$, and let $C$ denote the censoring time for $(U, N(\cdot))$ after which neither the counting process nor the terminal event is observable. Suppose that $N(\cdot)$ can only be observed at discrete observation times $0 < T_{K,1} < \cdots < T_{K,K}$, where the total number of observations $K$ is an integer-valued random variable. Define $Y = U \wedge C$, where $a \wedge b = \min(a, b)$. Let $\Delta = 1_{\{U \leq C\}}$ be a censoring indicator, let $\underline{T} = (T_{K,1}, \ldots, T_{K,K})$ be panel observation times on the counting process, and let $\underline{N} = (N_1, \ldots, N_K) = \big(N(T_{K,1}), \ldots, N(T_{K,K})\big)$ denote the cumulative event counts corresponding to $\underline{T}$. The observed data consist of i.i.d. copies of $D = (Y, \Delta, \underline{N}, \underline{T}, K)$.

Chan and Wang (2010, 2017) proposed backward stochastic models to study the terminal behavior of recurrent event processes. Kong et al. (2018) treated a terminal event time as a covariate in regression analysis of longitudinal data. Motivated by the aforementioned work, we propose a reversed mean model for nonparametric inference based on panel count data with an informative terminal event. Our model can directly reflect the effect of a terminal event on a recurrent event process, which has not been investigated in the literature of panel count data.

For ascertaining the behavior of the counting process near the terminal event time $U$, we study the reversed counting process $\widetilde{N}(t; U)$, the event count from time $t$ to $U$, with the reversed mean model,

$$E(\widetilde{N}(t; U) | U = u) = \Lambda(u - t), \tag{1}$$

where $\Lambda(\cdot)$ is a nondecreasing function with $\Lambda(0) = 0$. To gain more insight into the model, we plot in Figure 1 the reversed conditional mean functions $\Lambda(U - t)$ with $\Lambda(t) = 8\{1 - \exp(-t)\}$ for three randomly chosen times of the terminal event assuming $U \sim \mathrm{Uniform}(2, 4)$. It shows that the expected number of recurrent events from a fixed time $t$ to the terminal event increases as it prolongs.

Because $\widetilde{N}(t; U)$ may not be observable at some $t$ due to censoring, model (1) is not directly applicable for estimating $\Lambda(\cdot)$ using the observed panel count data. Noting that $N(t) = \widetilde{N}(0; U) - \widetilde{N}(t; U)$, we have

$$E(N(t) | U = u) = \Lambda(u) - \Lambda(u - t)$$

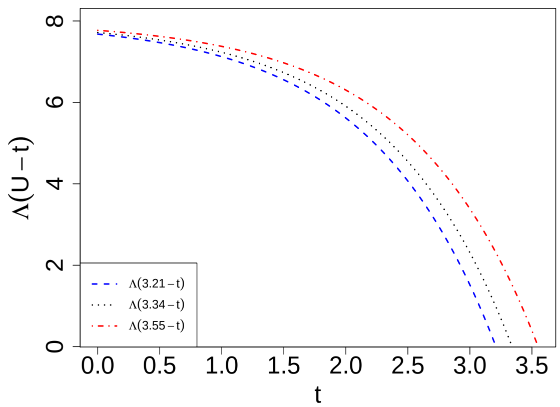


**Figure 1**. Plots of three trajectories of the reversed conditional mean function $\Lambda(U-t)$ in (1).

and this model facilitates a nonparametric maximum likelihood estimation method to study the reversed mean function with panel count data truncated by a terminal event following Wellner and Zhang (2000) and Lu et al. (2007).

To develop a valid estimation procedure, we assume: (i) $U$ and $C$ are independent; (ii) The distribution of the censoring time $C$ is non-informative to $\Lambda$; (iii) The distribution of $(K, \underline{T})$ is also non-informative to $\Lambda$. Model (1) automatically implies that $\widetilde{N}(\cdot)$ and $U$ are not independent. The goal of our proposed model is to directly quantify the behavior of the recurrent events near the informative terminal event.

## 2.2. Two-stage sieve likelihood-based estimation procedure

Let $\triangle N_j = N(T_{K,j}) - N(T_{K,j-1})$, $\underline{\triangle N} = (\triangle N_1, \ldots, \triangle N_K)$ with realization $\underline{\triangle n} = (\triangle n_1, \ldots, \triangle n_K)$, where $\triangle n_j = n_j - n_{j-1}$, and $\underline{t} = (t_{K,1}, \ldots, t_{K,K})$. Assume that our working model for $N(t)$ is a non-homogeneous Poisson process with the mean function $\Lambda(u) - \Lambda(u-t)$ given the process is truncated by the terminal event occurred at $U = u$, then

$$P(\underline{\triangle N} = \underline{\triangle n} | U = u, K, \underline{T} = \underline{t}) = \prod_{j=1}^{K} \frac{\exp(-\triangle\Lambda_j(u))(\triangle\Lambda_j(u))^{\triangle n_j}}{(\triangle n_j)!},$$

where $t_{K,0} \equiv 0$, $n_0 \equiv 0$, and for $j \geq 1$, $\triangle\Lambda_j(u) = \Lambda(u - t_{K,j-1}) - \Lambda(u - t_{K,j})$. Since the terminal event is subject to censoring, this conditional probability cannot be immediately used for constructing the likelihood function of the observed data $D$. The censored terminal event time, however, can be integrated out from the likelihood if we facilitate a distribution for $U$ (e.g. Kong et al., 2018). Under this working model, the likelihood of $(\Lambda, F)$ for the i.i.d. sample $\underline{D} = \{D_i : i = 1, \ldots, n\}$ is

$$L_n(\Lambda, F; \underline{D}) = \prod_{i=1}^{n} \left( \prod_{j=1}^{K_i} \frac{\left(\triangle\Lambda_{i,j}(Y_i)\right)^{\triangle N_{i,j}} \exp\left(-\triangle\Lambda_{i,j}(Y_i)\right)}{\triangle N_{i,j}!} f(Y_i) \right)^{\Delta_i}$$

$$\times \left( \int_{Y_i}^{\infty} \prod_{j=1}^{K_i} \frac{\left(\triangle\Lambda_{i,j}(u)\right)^{\triangle N_{i,j}} \exp\left(-\triangle\Lambda_{i,j}(u)\right)}{\triangle N_{i,j}!} dF(u) \right)^{1-\Delta_i},$$

where $T_{K_i,0} \equiv 0$, $\triangle N_{i,j} = N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})$, and $\triangle\Lambda_{i,j}(u) = \Lambda(u - T_{K_i,j-1}) - \Lambda(u - T_{K_i,j})$. The log-likelihood of $(\Lambda, F)$ is

$$l_n(\Lambda, F; \underline{D}) = \frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i \sum_{j=1}^{K_i}\left\{\triangle N_{i,j}\log(\triangle\Lambda_{i,j}(Y_i)) - \triangle\Lambda_{i,j}(Y_i)\right\}\right.$$
$$\left. +(1-\Delta_i)\log\left\{\int_{Y_i}^{\infty}\prod_{j=1}^{K_i}\frac{\left(\triangle\Lambda_{i,j}(u)\right)^{\triangle N_{i,j}}\exp\left(-\triangle\Lambda_{i,j}(u)\right)}{\triangle N_{i,j}!}dF(u)\right\}\right],$$

after omitting the term $\log f(Y_i)$ and the part unrelated to $(\Lambda, F)$.

We denote the true values of $\Lambda(u)$ and $\triangle\Lambda_j(u)$ as $\Lambda_0(u)$ and $\triangle\Lambda_{0j}(u)$, respectively. And let $f_0$ and $F_0$ denote the density and cumulative distribution functions of $U$, respectively. Clearly, joint estimation of $\Lambda_0$ and $F_0$ is a daunting problem in view of the complicated likelihood structure. We consider a two-stage estimation procedure in the spirit of pseudo-likelihood estimation as described below.

Stage 1: Obtain the Kaplan–Meier estimator of $F_0$, $\widehat{F}_n(t)$, based on $\{(Y_i, \Delta_i), i = 1, \ldots, n\}$.
Stage 2: Derive the log pseudo-likelihood of $\Lambda$ as $l_n(\Lambda, \widehat{F}_n; \underline{D}) := l_n(\Lambda; \underline{D})$.

Considering the complexity of $l_n(\Lambda; \underline{D})$, we propose to estimate the smooth function $\Lambda_0$ using B-spline function approximation (Lu, Zhang and Huang, 2007). Let $\mathcal{T} = \{t_i, i = 1, \ldots, m_n + 2l\}$, with $0 = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m_n+l} < t_{m_n+l+1} = \cdots = t_{m_n+2l} = \tau$, be a sequence of knots that partition $[0, \tau]$ into $m_n + 1$ subintervals $I_i = [t_{l+i}, t_{l+i+1}]$, for $i = 0, 1, \ldots, m_n$. Let $\Phi_n$ be a class of polynomial splines of order $l \geq 1$ with the knot sequence $\mathcal{T}$, and thus $\Phi_n$ can be linearly spanned by the normalized B-spline basis functions $\{B_i, i = 1, \ldots, q_n\}$ with $q_n = m_n + l$ (Schumaker, 2007). Define a subclass of $\Phi_n$: $\Psi_n = \left\{\sum_{i=1}^{q_n}\alpha_i B_i : 0 \leq \alpha_1 \leq \cdots \leq \alpha_{q_n}, \sum_{i=1}^{q_n}\alpha_i B_i(0) = 0\right\}$. These constraints sufficiently warrant that $\Lambda(t) = \sum_{i=1}^{q_n}\alpha_i B_i(t)$ is non-decreasing and $\Lambda(0) = 0$. The constraints $0 \leq \alpha_1 \leq \ldots \leq \alpha_n$ are required to guarantee the non-decreasing property of $\Lambda$ and the constraint $\sum_{i=1}^{q_n}\alpha_i B_i(0) = 0$ implies $\alpha_1 = 0$ as $(B_1(0), \ldots, B_{q_n}(0)) = (1, 0, \ldots, 0)$.

The estimator $\widehat{\Lambda}_n$ of $\Lambda_0$ maximizes $l_n(\Lambda; \underline{D})$ over $\Lambda \in \Psi_n$, and the spline pseudo-likelihood estimator of $\Lambda_0$ is denoted by $\widehat{\Lambda}_n = \sum_{i=1}^{q_n}\widehat{\alpha}_i B_i$.

# 3. Asymptotic Results

## 3.1. Consistency and convergence rate

Let $g^{(r)}$ denote the $r$-th derivative function of $g$, and let $M_j, j = 0, 1, \ldots, 5$, and $c$ denote different constants throughout the paper. Define $\mathcal{F} = \{F : F \text{ is a distribution function on } [0, \infty)\}$,

$$\mathcal{H}_r = \{g : |g^{(r-1)}(s) - g^{(r-1)}(t)| \leq c|s - t|, \text{ any } 0 \leq s < t < \infty, r \geq 1\},$$

$$\text{and} \quad \Psi = \{\Lambda : \Lambda \text{ is an increasing function on } [0,\tau], \Lambda(0) = 0, \Lambda \in \mathcal{H}_r\}.$$

Let $\mathcal{B}$ denote a collection of Borel sets in $\mathcal{R}$, and let $\mathcal{B}_{[0,\tau]} = \{B \bigcap [0,\tau] : B \in \mathcal{B}\}$. Define the measure $\mu_1$ on $([0,\tau]^2, \mathcal{B}^2_{[0,\tau]})$ by

$$\begin{aligned} &\mu_1(B_1 \times B_2) \\ &= \int_0^\infty \left[ \sum_{k=1}^\infty P(K = k | U = u) \sum_{j=1}^k P(u - T_{k,j-1} \in B_1, u - T_{k,j} \in B_2 | K = k, U = u) \right] dF_0(u), \end{aligned}$$

for $B_1, B_2 \in \mathcal{B}_{[0,\tau]}$, and the metric $d_1$ on $\Psi$ by

$$d_1^2(\Lambda_1, \Lambda_2) = \iint \left| \big(\Lambda_1(t) - \Lambda_1(s)\big) - \big(\Lambda_2(t) - \Lambda_2(s)\big) \right|^2 d\mu_1(s,t)$$

for $\Lambda_1, \Lambda_2 \in \Psi$. Also define the measure $\mu_2$ on $([0,\tau], \mathcal{B}_{[0,\tau]})$ as

$$\mu_2(B) = \int_0^\infty \sum_{k=1}^\infty P(K = k | U = u) \sum_{j=1}^k P(u - T_{k,j} \in B | K = k, U = u) dF_0(u), \quad \text{for } B \in \mathcal{B}_{[0,\tau]}.$$

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

(C1) $\Lambda_0 \in \Psi$ with $0 < \Lambda_0(\tau) < \infty$; $P\Big( \bigcap_{j=1}^K \big\{ U - T_{K,j} \in [0,\tau] \big\} \Big) = 1$.

(C2) $P(Y < U) = M_1$ for $0 < M_1 < 1$.

(C3) There is a positve constant $c$ such that $E\big(e^{cN(\tau)}\big) < \infty$.

(C4) There is some $0 < u_0 < \tau$ such that the support of the terminal event time $U$ is $[u_0, \tau]$. The time has a density $f_0$ with respect to Lebesgue measure which has a version which is continuous on $[u_0, \tau]$ and there is a constant $M_2 > 0$ such that for all $u \in [u_0, \tau]$ we have $f_0(u) \geq M_2$.

(C5) There exists a constant $M_3 > 0$ such that $P(C \geq \tau) = M_3$.

(C6) $q_n = O(n^\nu)$ with $0 < \nu < 1/2$, $\max_i |t_i - t_{i-1}| = O(n^{-\nu})$ and $\dfrac{\max_i |t_i - t_{i-1}|}{\min_i |t_i - t_{i-1}|} \leq M_4$ uniformly for $n$, where $t_i, i = l, \ldots, m_n + l + 1$ are the partition knots of interval $[0,\tau]$.

(C7) The observation time points are $s_0$-separated, i.e., there exists a constant $s_0 > 0$ such that $P(T_{K,j} - T_{K,j-1} \geq s_0 \text{ for all } j = 1, \ldots, K) = 1$. Furthermore, $\mu_2$ is absolutely continuous with respect to the Lebesgue measure with a derivative $\dot{\mu}_2(t) \geq M_5 > 0$ for some positive constant $M_5$.

(C8) The true baseline function $\Lambda_0$ is differentiable and its derivative has positive and finite lower and upper bounds in the observation interval, i.e., there exists a constant $1 < M_0 < \infty$ such that $1/M_0 \leq \dot{\Lambda}_0(t) \leq M_0$ for $t \in [0,\tau]$.

Condition (C1) indicates that the observation times are separated from the time origin and the residual time $U - T$ falls into an observed time interval. (C2) is a regular one for survival data meaning that the censoring rate falls between 0 and 1. Condition (C3) can be satisfied if the process $N(t)$ is uniformly bounded or it is a Poisson or mixed Poisson process. Condition (C4) on the distribution of the terminal event time can be satisfied by most continuous random variables. Condition (C5) is mild technical conditions for the proof. Condition (C6) is required to ensure the approximation for the monotone function by Lu, Zhang and Huang (2007, 2009). Conditions (C7) and (C8) are technical ones similar to (C11) and (C12) of Wellner and Zhang (2007). The first part of (C7) and (C8) can yield $P(c_1 \leq \Lambda_0(Y - T_{K,j-1}) - \Lambda_0(Y - T_{K,j}) \leq c_2, j = 1, \ldots, K) = 1$ for two positive constants $c_1$ and $c_2$, which is needed to derive the asymptotic properties of the proposed estimator. The second part of (C7) is used in the proofs of Lemma 3 and Theorems 1 and 2. Serving as the technical conditions for proving the theorems, they are quite reasonable in view of real data applications. Condition (C7) simply implies that the adjacent observations should be minimally separated in time and Condition (C8) implies that the process $\widetilde{N}(\cdot)$ has a jump with a non-zero probability at any time prior to the terminal event.

Based on the asymptotic properties of the Kaplan–Meier (KM) estimator, we can derive the consistency and the convergence rate of the two-stage estimator $\widehat{\Lambda}_n$ by extending the empirical process theories for $M$-estimators to those with a nuisance functional parameter.

**Theorem 1** *(Consistency of $\widehat{\Lambda}_n$) Under Conditions (C1)–(C8), it holds that $d_1(\widehat{\Lambda}_n, \Lambda_0) \to 0$ almost surely.*

**Theorem 2** *(Convergence rate of $\widehat{\Lambda}_n$) Under Conditions (C1)–(C8), for any $r > 1$, we have $d_1(\widehat{\Lambda}_n, \Lambda_0) = O_p(n^{-r/(1+2r)})$.*

Theorem 2 shows that the convergence rate of the nonparametric two-stage estimator $\widehat{\Lambda}_n$ reaches $n^{-r/(1+2r)}$ for any $r > 1$, which is slower than $n^{-1/2}$ but faster than $n^{-1/3}$.

## 3.2. Asymptotic theory of nonparametric M-estimators with a nuisance functional parameter

Suppose that $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample from the distribution of $X$, and $l_n(\Lambda, F; \mathbf{X}) = \sum_{i=1}^{n} m(\Lambda, F; X_i)$ is an objective function based on $\mathbf{X}$, where $\Lambda$ is an unknown function in the class $\Psi$, and $F$ is a nuisance functional parameter. Let $\Psi_n$ be the sieve parameter space satisfying $\Psi_n \subseteq \Psi_{n+1} \subseteq \cdots \subseteq \Psi$, for $n \geq 1$. Assume that $\widehat{F}_n$ is a consistent estimator of $F$ and $\widehat{\Lambda}_n$ is the estimator of $\Lambda_0$ by maximizing $l_n(\Lambda, \widehat{F}_n; \mathbf{X})$ in the sieve parameter space $\Psi_n$.

Let $\mathcal{H}$ be a space containing $\Psi$, and $l^\infty(\mathcal{H})$ be the space of bounded functionals on $\mathcal{H}$ under the supremum norm $\|f\|_\infty = \sup\limits_{h \in \mathcal{H}} |f(h)|$. For $h \in \mathcal{H}$, we define a sequence of maps $G_n$ in the parameter space for $(\Lambda, F)$ to $l^\infty(\mathcal{H})$ as the derivative of $n^{-1} l_n(\Lambda, F; \mathbf{X})$ with respect to $\Lambda$ in the direction $h$:

$$
\begin{aligned}
G_n(\Lambda, F)[h] =& n^{-1} \lim_{\delta \to 0} \frac{l_n(\Lambda + \delta h, F; \mathbf{X}) - l_n(\Lambda, F; \mathbf{X})}{\delta} \\
=& n^{-1} \sum_{i=1}^{n} \lim_{\delta \to 0} \frac{m(\Lambda + \delta h, F; X_i) - m(\Lambda, F; X_i)}{\delta} \\
=& \mathbb{P}_n m_1(\Lambda, F; X)[h]
\end{aligned}
$$

and define $G(\Lambda, F)[h] = P m_1(\Lambda, F; X)[h]$, where $P$ and $\mathbb{P}_n$ denote the probability measure and empirical measure with $Pf = \int f dP$ and $\mathbb{P}_n f = n^{-1} \sum_{i=1}^{n} f(X_i)$, respectively.

To establish the asymptotic normality, we need the following conditions.

(B1) $\sqrt{n}(G_n - G)(\widehat{\Lambda}_n, \widehat{F}_n)[h] - \sqrt{n}(G_n - G)(\Lambda_0, F_0)[h] = o_p(1)$;

(B2) $G(\Lambda_0, F_0)[h] = 0$ and $G_n(\widehat{\Lambda}_n, \widehat{F}_n)[h] = o_p(n^{-1/2})$;

(B3) $G(\Lambda, F)[h]$ is Fréchet-differentiable with respect to $\Lambda$ and $F$ with the continuous derivatives $\dot{G}_{1,\Lambda,F}[h]$ and $\dot{G}_{2,\Lambda,F}[h]$, respectively;

(B4) $G(\widehat{\Lambda}_n, \widehat{F}_n)[h] - G(\Lambda_0, F_0)[h] - \dot{G}_{1,\Lambda_0,\widehat{F}_n}(\widehat{\Lambda}_n - \Lambda_0)[h] - \dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h] = o_p(n^{-1/2})$;

(B5) $\sqrt{n}(G_n - G)(\Lambda_0, F_0)[h] + \sqrt{n}\dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h]$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H})$.

**Remark 1** *Conditions (B2), (B3) and (B5) are analogous to the analytical conditions in Theorem 3.3.1 of van der Varrt and Wellner (1996); (B1) and (B4) mean the remainders of the Taylor expansions are negligible, which are weaker than those in van der Varrt and Wellner (1996).*

**Theorem 3** *(General functional asymptotic normality) Under Assumptions (B1)–(B5), for any $h \in \mathcal{H}$, we have*

$$-\sqrt{n}\dot{G}_{1,\Lambda_0,\widehat{F}_n}(\widehat{\Lambda}_n - \Lambda_0)[h] = \sqrt{n}\dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h] + \sqrt{n}(G_n - G)(\Lambda_0, F_0)[h] + o_p(1),$$

*and $-\sqrt{n}\dot{G}_{1,\Lambda_0,\widehat{F}_n}(\widehat{\Lambda}_n - \Lambda_0)[h]$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H})$.*

**Remark 2** *Theorem 3.3.1 of van der Varrt and Wellner (1996) implies that the estimator has a convergence rate of $n^{-1/2}$, while Theorem 3 allows the order of $\widehat{\Lambda}_n$ to be slower than $n^{-1/2}$. Theorem 3 states a general functional asymptotic normality property for the nonparametric $M$-estimator with a nuisance parameter, which extends Theorem 1 in Zhao and Zhang (2017) to more practical cases. As an application, we derive the asymptotic normality of the two-stage functional estimator for panel count data truncated by a terminal event.*

## 3.3. Asymptotic normality of functionals of two-stage nonparametric M-estimators for panel count data

The next theorem is obtained by applying Theorem 3 to the proposed estimator.

**Theorem 4** *Under Conditions (C1)–(C8), for any $h \in \mathcal{H}_r$, we have*

$$\begin{aligned}
\sqrt{n}P \Bigg[ &\Delta \sum_{j=1}^{K} \frac{\Delta\widehat{\Lambda}_{nj}(Y) - \triangle\Lambda_{0j}(Y)}{\triangle\Lambda_{0j}(Y)} h_j(Y) + (1-\Delta)\Big\{ \int_Y^\tau S_0(u) d\widehat{F}_n(u) \Big\}^{-2} \\
&\times \int_Y^\tau S_0(u) \sum_{j=1}^{K} \left\{ \frac{\triangle N_j}{\triangle\Lambda_j(u)} - 1 \right\} h_j(u) d\widehat{F}_n(u) \\
&\times \int_Y^\tau S_0(u) \sum_{j=1}^{K} \left\{ \frac{\triangle N_j}{\triangle\Lambda_j(u)} - 1 \right\} \{\Delta\widehat{\Lambda}_{nj}(u) - \triangle\Lambda_{0j}(u)\} d\widehat{F}_n(u) \Bigg] \xrightarrow{d} N(0, \sigma_1^2[h]),
\end{aligned}$$

*where $\Delta\widehat{\Lambda}_{nj}(u) = \widehat{\Lambda}_n(u - T_{j-1}) - \widehat{\Lambda}_n(u - T_j)$, $\Delta\Lambda_{0j}(u) = \Lambda_0(u - T_{j-1}) - \Lambda_0(u - T_j)$, $S_0(u) = \prod_{j=1}^{K} \{\triangle\Lambda_{0j}(u)\}^{\triangle N_j} \exp\{-\triangle\Lambda_{0j}(u)\}/\{\triangle N_j!\}$, and $\sigma_1^2[h]$ is given in the proof of Theorem 4.*

Theorem 4 plays a key role for constructing new statistics for multi-sample nonparametric comparison of panel count data truncated by a terminal event.

# 4. Nonparametric Tests

Consider $n$ independent subjects from $J$ groups with $n_1 + \cdots + n_J = n$, where $n_r$ is the sample size of the $r$th group. Let $\widetilde{N}^{(r)}(t;U)$ denote the reversed counting process from a terminal event time $U$ for group $r$, $r = 1, \ldots, J$, that subjects to panel observations. Suppose that $E(\widetilde{N}^{(r)}(t;U)|U = u) = \Lambda_r(u-t)$ with $\Lambda_r(0) = 0$, $r = 1, \ldots, J$. The problem of interest is to test the null hypothesis $H_0 : \Lambda_1 = \Lambda_2 = \cdots = \Lambda_J = \Lambda_0$. We illustrate the test for two-sample comparison which can be extended to $J$-sample comparison by using similar ideas in Zhang (2006) and Balakrishnan and Zhao (2009).

Let $\widehat{\Lambda}_r$ denote the two-stage sieve likelihood-based estimator of $\Lambda_r$ based on the data in the $r$th group for $r = 1, 2$, and let $\widehat{\Lambda}_0$ be the two-stage sieve likelihood-based estimator of $\Lambda_0$ based on the pooled data. Using the same notation as in Section 2.2, we denote $\triangle\Lambda_{r,i,j}(u) = \Lambda_r(u - T_{K_i,j-1}) - \Lambda_r(u - T_{K_i,j})$ for $r = 0, 1, 2$, and let $\triangle\widehat{\Lambda}_{r,i,j}(u)$ be the corresponding two-stage estimators. For testing the hypothesis $H_0$, our method is motivated by the asymptotic results in Theorem 4 and an idea commonly used in survival analysis, such as Anderson et al. (1993), Zhang (2006), Balakrishnan and Zhao (2009) and Zhao and Zhang (2017). More specifically, we propose the test statistic as

$$
\begin{aligned}
Q_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Bigg[ & \Delta_i \sum_{j=1}^{K_i} h_{n,i,j}(Y_i) \frac{\triangle\widehat{\Lambda}_{1,i,j}(Y_i) - \triangle\widehat{\Lambda}_{2,i,j}(Y_i)}{\triangle\widehat{\Lambda}_{0,i,j}(Y_i)} \\
& + (1-\Delta_i) \Big\{ \int_{Y_i}^{\tau} \widehat{S}_0(u) d\widehat{F}_n(u) \Big\}^{-2} \int_{Y_i}^{\tau} \widehat{S}_0(u) \sum_{j=1}^{K_i} h_{n,i,j}(u) \Big\{ \frac{\triangle N_{i,j}}{\triangle\widehat{\Lambda}_{0,i,j}(u)} - 1 \Big\} d\widehat{F}_n(u) \\
& \times \int_{Y_i}^{\tau} \widehat{S}_0(u) \sum_{j=1}^{K_i} \Big\{ \frac{\triangle N_{i,j}}{\triangle\widehat{\Lambda}_{0,i,j}(u)} - 1 \Big\} \Big\{ \triangle\widehat{\Lambda}_{1,i,j}(u) - \triangle\widehat{\Lambda}_{2,i,j}(u) \Big\} d\widehat{F}_n(u) \Bigg],
\end{aligned}
$$

where $\widehat{S}_0(u)$ represents the value of $S_0(u)$ with $\Lambda_0$ replaced by $\widehat{\Lambda}_0$, and $h_n(u)$'s are bounded weight processes with $h_{n,i,j}(u) = h_n(u - T_{K_i,j-1}) - h_n(u - T_{K_i,j})$. This statistic can be viewed as a numerical integral of the weighted relative difference between $\triangle\widehat{\Lambda}_1$ and $\triangle\widehat{\Lambda}_2$.

**Theorem 5** *(Asymptotic distribution of the test statistic) Suppose the conditions in Theorem 4 hold, and there exists a bounded function $h_0(u)$ such that $h_0 \in \mathcal{H}_r$ and*

$$
E\left[ \sum_{j=1}^{K} \left\{ h_{n,j}(U) - h_{0,j}(U) \right\}^2 \right] = o(n^{-1/(1+2r)}),
$$

*where $h_{n,j}(u) = h_n(u - T_{K,j-1}) - h_n(u - T_{K,j})$ and $h_{0,j}(u) = h_0(u - T_{K,j-1}) - h_0(u - T_{K,j})$. If $n_1/n \to p$ as $n \to \infty$ with $0 < p < 1$, then under $H_0 : \Lambda_1 = \Lambda_2 = \Lambda_0$,*

*(i) $Q_n$ has an asymptotic distribution $N(0, \sigma_2^2)$, where $\sigma_2^2 = \left(\frac{1}{p} + \frac{1}{1-p}\right)\sigma_0^2$ with $\sigma_0^2 = E\{m_1^2(\Lambda_0, F_0; D)[h_0]\}$;*

*(ii) $\sigma_0^2$ can be consistently estimated by*

$$
\widehat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^{n} \left[ \Delta_i \left\{ \sum_{j=1}^{K_i} \frac{\triangle N_{i,j} - \triangle\widehat{\Lambda}_{0,i,j}(Y_i)}{\triangle\widehat{\Lambda}_{0,i,j}(Y_i)} h_{n,i,j}(Y_i) \right\}^2 \right.
$$

$$+(1-\Delta_i)\frac{\left\{\sum_{j=1}^{K_i}\int_{Y_i}^{\tau}\widehat{S}_0(u)\frac{\triangle N_{i,j}-\triangle\widehat{\Lambda}_{0,i,j}(u)}{\triangle\widehat{\Lambda}_{0,i,j}(u)}h_{n,i,j}(u)d\widehat{F}_n(u)\right\}^2}{\left\{\int_{Y_i}^{\tau}\widehat{S}_0(u)d\widehat{F}_n(u)\right\}^2}\Bigg].$$

Theorem 5 presents the asymptotic normality of the test statistic $Q_n$ as well as a consistent estimator for the asymptotic variance of $Q_n$. We can choose different weight processes to conduct hypothesis testing, for which we provide typical choices of weight processes in simulation studies. In particular, we relax the condition of a monotone weight process required in Zhang (2006) and Balakrishnan and Zhao (2009) by a bounded weight process, and thus the choice of a weight process for our test statistic is more general.

Using Theorem 5, we can construct the test statistic $Q_n/\{(1/p+1/(1-p))\hat{\sigma}_0\}$, which follows the standard normal distribution asymptotically, to test $H_0$ for two-sample comparison. In a more general setting for $J$-sample comparison, we can use the techniques in Balakrishna and Zhao (2009) to construct a test statistic which follows a $\chi^2$ distribution asymptotically.

## 5. Simulation Studies

We conducted simulation studies to evaluate the finite-sample performance of the proposed two-stage estimators and nonparametric tests. We first generated the number of observations $K_i$ from the uniform distribution with an equal probability of $1/8$ on $\{3,4,5,6,7,8,9,10\}$. We then generated censoring time $C_i$ from Uniform$(\tau/4,\tau)$ and the terminal event time $U_i$ from Uniform$(0,4)$+Exp$(1)$, where $\tau$ was chosen such that the censoring rate was around 0.2 and 0.4 respectively. We obtained $Y_i=\min\{U_i,C_i\}$ and $\Delta_i=I_{(U_i\le C_i)}$. For given $Y_i$ and $K_i$, we took observation times $T_{ij}, j=1,\cdots,K_i$, to be the order statistics from Uniform$(0,Y_i)$. Finally, the $N_{ij}$'s were generated from a Poisson process with mean function $E(N_i(t)|U_i)=\Lambda(U_i)-\Lambda(U_i-t)$. We considered two mean functions:

CASE 1: $\Lambda_1(u|\gamma_i)=8\gamma_i\{1-\exp(-u)\},\ \Lambda_2(u|\gamma_i)=8\gamma_i\beta\{1-\exp(-u)\}$ with $\beta=1,1.3,1.6$, respectively;

CASE 2: $\Lambda_1(u|\gamma_i)=3\gamma_i u/2,\ \Lambda_2(u|\gamma_i)=\gamma_i\sqrt{\beta u}$ with $\beta=2,5,10$, respectively.

For each case, we took $\gamma_i=1$ and $\gamma_i\sim$Gamma$(2,1/2)$ corresponding to Poisson and mixed Poisson processes. The true mean functions in Case 1 do not overlap, while those in Case 2 cross over. For the purpose of comparison, we considered three types of weights under case 1 with $h_n(u)=\widehat{\Lambda}_0(u)W_k(u)$ for $k=1,2,3$, where $W_1(u)=1$, $W_2(u)=n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{K_i}I(u\le Y_i-T_{K_i,j})$ and $W_3(u)=n^{-1}\sum_{i=1}^{n}I(u\le Y_i)$, and two additional types of weights for Case 2 with $W_4(u)=n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{K_i}I(u>Y_i-T_{K_i,j})$ and $W_5(u)=n^{-1}\sum_{i=1}^{n}I(u>Y_i)$. We adopted cubic spline basis functions and took $q_n=7$ with the order of $O(n^{1/4})$ in the simulation. The sample sizes were set as $(n_1,n_2)=(40,60)$, $(60,100)$ and $(120,120)$, and 500 Monte Carlo replications were carried out.

The simulation results are summarized in Figures 2 and 3 and Tables 1 and 2. Figure 2 shows the plots of the estimates for $\Lambda_l, l=1,2$ under Case 1 with $(n_1,n_2)=(120,120)$ and censoring rate 0.2, and the plots for other cases are similar. It can be seen that the estimates are close to the true functions, which indicates the functional estimates are consistent. Besides, the pointwise 1.96 standard

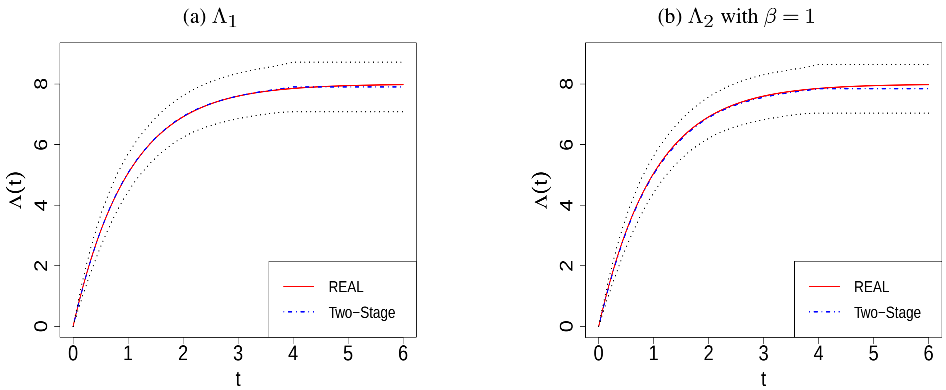


**Figure 2**. Plots of the estimates for $\Lambda_1$ and $\Lambda_2$ in Case 1 with a censoring rate of $20\%$ and $n_1 = n_2 = 120$. The solid line is the true function, the dot–dashed line is the pointwise mean estimate, and the dotted lines describe the pointwise 1.96 standard deviation (1.96-SD) error-bars of the estimates.

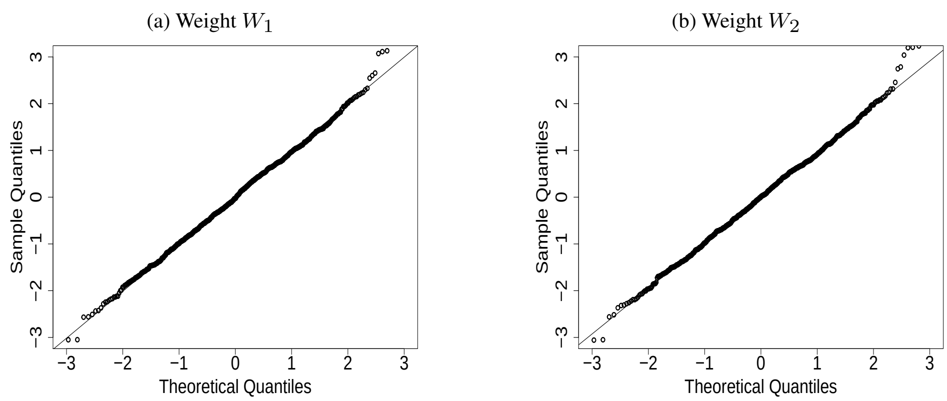


**Figure 3**. Normal quantile-quantile plots of the test statistics under two different weights in Case 1 with a censoring rate of $20\%$ and $n_1 = n_2 = 120$.

deviation (1.96-SD) error bars based on the sampling standard deviation depicts the degree of the variability of the functional estimator in the 500 Monte Carlo simulations. To evaluate the asymptotic normality in Theorem 5, we provide the quantile-quantile (QQ) plots of the test statistics against the standard normal. Figure 3 presents the normal QQ plots under two different weights in Case 1 with $\beta = 1$, $(n_1, n_2) = (120, 120)$, and censoring rate 0.2. The results reveal that the asymptotic normality is justified in finite samples with moderate size. Under Case 1, it is obvious that $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$ are close to each other for $\beta = 1$, and there are some differences between the two estimates when $\beta$ takes other values. Under Case 2, the estimates of $\Lambda_1$ and $\Lambda_2$ cross over at different points for different values of $\beta$.

Table 1 shows that the sizes of the proposed tests are all around the nominal value 0.05, and the power values are closer to 1 when $\beta$ is much larger than the null value 1 under Case 1 with weights $W_1, W_2$ and $W_3$. Table 2 presents the power values of the proposed tests under Case 2. When the

**Table 1.** Estimated size and power of the tests under Case 1 in the simulation

| CR | $\beta$ | $\gamma_i = 1$ $W_1$ | $W_2$ | $W_3$ | $\gamma_i \sim \text{Gamma}(2, 1/2)$ $W_1$ | $W_2$ | $W_3$ |
|---|---|---|---|---|---|---|---|
| | | $n_1 = 40, n_2 = 60$ | | | $n_1 = 40, n_2 = 60$ | | |
| 0.4 | 1 | 0.064 | 0.058 | 0.060 | 0.051 | 0.057 | 0.055 |
| | 1.3 | 0.591 | 0.528 | 0.583 | 0.222 | 0.202 | 0.214 |
| | 1.6 | 0.966 | 0.938 | 0.953 | 0.551 | 0.505 | 0.545 |
| | | $n_1 = 60, n_2 = 100$ | | | $n_1 = 60, n_2 = 100$ | | |
| | 1 | 0.054 | 0.054 | 0.054 | 0.053 | 0.048 | 0.048 |
| | 1.3 | 0.752 | 0.687 | 0.743 | 0.296 | 0.278 | 0.284 |
| | 1.6 | 0.996 | 0.991 | 0.997 | 0.728 | 0.675 | 0.711 |
| | | $n_1 = 120, n_2 = 120$ | | | $n_1 = 120, n_2 = 120$ | | |
| | 1 | 0.043 | 0.049 | 0.044 | 0.045 | 0.047 | 0.046 |
| | 1.3 | 0.914 | 0.887 | 0.915 | 0.431 | 0.409 | 0.431 |
| | 1.6 | 1.000 | 1.000 | 1.000 | 0.868 | 0.882 | 0.897 |
| | | $n_1 = 40, n_2 = 60$ | | | $n_1 = 40, n_2 = 60$ | | |
| 0.2 | 1 | 0.062 | 0.052 | 0.057 | 0.048 | 0.054 | 0.048 |
| | 1.3 | 0.710 | 0.649 | 0.699 | 0.233 | 0.220 | 0.237 |
| | 1.6 | 0.991 | 0.986 | 0.991 | 0.626 | 0.580 | 0.614 |
| | | $n_1 = 60, n_2 = 100$ | | | $n_1 = 60, n_2 = 100$ | | |
| | 1 | 0.050 | 0.047 | 0.047 | 0.049 | 0.050 | 0.048 |
| | 1.3 | 0.867 | 0.812 | 0.855 | 0.351 | 0.322 | 0.337 |
| | 1.6 | 1.000 | 1.000 | 1.000 | 0.815 | 0.793 | 0.807 |
| | | $n_1 = 120, n_2 = 120$ | | | $n_1 = 120, n_2 = 120$ | | |
| | 1 | 0.050 | 0.046 | 0.049 | 0.055 | 0.057 | 0.053 |
| | 1.3 | 0.972 | 0.954 | 0.967 | 0.535 | 0.505 | 0.527 |
| | 1.6 | 1.000 | 1.000 | 1.000 | 0.959 | 0.946 | 0.957 |

Note: CR represents the censoring rate, and $W_j$'s represent the weights used in the test statistics.

two mean functions cross over, the power relies on the choice of the weight process, and the tests with weights $W_4$ and $W_5$ perform the best. As explained in Balakrishnan and Zhao (2009), this phenomenon is caused by the fact that the differences before and after the intersection point of two mean functions have opposite signs, and the difference after the intersection point dominates that before the point. Thus, the tests with weights $W_4$ and $W_5$ yield higher power, because they weigh the difference at later time points more than earlier ones in comparison with weights $W_1 - -W_3$. In addition, both tables show that the power increases as the sample size increases or the censoring rate decreases. In summary, the simulation studies numerically justify the asymptotic properties of the proposed estimator and two-sample test.

## 6. Application

We applied the proposed method to the panel count data of occurrences of serious illness truncated by death from the CLHLS study. Focusing on the individuals who entered the longitudinal study in 1998 and were followed up to 2014, we identified 3050 elderly people after removing 6043 individuals with missing or erroneous records for the analysis. By using month as the time metric to record the recurrent serious illness starting from the baseline survey, we let $N_i(t)$ be the observed number of serious illnesses occurring up to month $t$ for the $i$th individual, and let $T_{ij}$ be the observation time

**Table 2.** Estimated size and power of the tests under Case 2 in the simulation

| | | $\gamma_i = 1$ | | | | | $\gamma_i \sim \mathrm{Gamma}(2, 1/2)$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CR | $\beta$ | $W_1$ | $W_2$ | $W_3$ | $W_4$ | $W_5$ | $W_1$ | $W_2$ | $W_3$ | $W_4$ | $W_5$ |
| | | $n_1 = 40, n_2 = 60$ | | | | | $n_1 = 40, n_2 = 60$ | | | | |
| 0.4 | 2 | 0.932 | 0.542 | 0.842 | 0.944 | 0.810 | 0.760 | 0.376 | 0.606 | 0.858 | 0.754 |
| | 5 | 0.408 | 0.056 | 0.210 | 0.660 | 0.556 | 0.338 | 0.078 | 0.142 | 0.556 | 0.540 |
| | 8 | 0.130 | 0.074 | 0.040 | 0.342 | 0.382 | 0.134 | 0.062 | 0.038 | 0.328 | 0.392 |
| | | $n_1 = 60, n_2 = 100$ | | | | | $n_1 = 60, n_2 = 100$ | | | | |
| | 2 | 0.984 | 0.656 | 0.922 | 0.998 | 0.976 | 0.880 | 0.444 | 0.726 | 0.950 | 0.882 |
| | 5 | 0.488 | 0.050 | 0.214 | 0.872 | 0.778 | 0.352 | 0.058 | 0.148 | 0.668 | 0.674 |
| | 8 | 0.084 | 0.120 | 0.036 | 0.450 | 0.506 | 0.106 | 0.080 | 0.046 | 0.408 | 0.500 |
| | | $n_1 = 120, n_2 = 120$ | | | | | $n_1 = 120, n_2 = 120$ | | | | |
| | 2 | 1.000 | 0.862 | 0.994 | 0.998 | 0.988 | 0.978 | 0.620 | 0.890 | 0.994 | 0.978 |
| | 5 | 0.632 | 0.052 | 0.298 | 0.970 | 0.950 | 0.392 | 0.050 | 0.162 | 0.834 | 0.840 |
| | 8 | 0.072 | 0.194 | 0.032 | 0.606 | 0.692 | 0.080 | 0.118 | 0.038 | 0.416 | 0.578 |
| | | $n_1 = 40, n_2 = 60$ | | | | | $n_1 = 40, n_2 = 60$ | | | | |
| 0.2 | 2 | 0.990 | 0.844 | 0.974 | 0.992 | 0.968 | 0.856 | 0.560 | 0.744 | 0.920 | 0.858 |
| | 5 | 0.652 | 0.124 | 0.370 | 0.886 | 0.814 | 0.358 | 0.062 | 0.190 | 0.626 | 0.636 |
| | 8 | 0.132 | 0.066 | 0.038 | 0.534 | 0.548 | 0.078 | 0.050 | 0.034 | 0.330 | 0.412 |
| | | $n_1 = 60, n_2 = 100$ | | | | | $n_1 = 60, n_2 = 100$ | | | | |
| | 2 | 0.998 | 0.938 | 0.996 | 1.000 | 0.996 | 0.964 | 0.710 | 0.902 | 0.990 | 0.966 |
| | 5 | 0.772 | 0.142 | 0.476 | 0.988 | 0.974 | 0.476 | 0.104 | 0.246 | 0.790 | 0.784 |
| | 8 | 0.150 | 0.108 | 0.042 | 0.752 | 0.812 | 0.110 | 0.068 | 0.046 | 0.454 | 0.578 |
| | | $n_1 = 120, n_2 = 120$ | | | | | $n_1 = 120, n_2 = 120$ | | | | |
| | 2 | 1.000 | 0.994 | 1.000 | 1.000 | 0.998 | 0.996 | 0.908 | 0.982 | 0.998 | 0.992 |
| | 5 | 0.930 | 0.204 | 0.714 | 0.994 | 0.992 | 0.630 | 0.108 | 0.328 | 0.924 | 0.914 |
| | 8 | 0.192 | 0.144 | 0.038 | 0.912 | 0.954 | 0.110 | 0.084 | 0.041 | 0.567 | 0.702 |

Note: CR represents the censoring rate, and $W_j$'s represent the weights used in the test statistics.

points for the $i$th individual, $j = 1, \ldots, K_i$, where $K_i$ is the number of observation times for individual $i$ during the whole survey process. Death as the terminal event was possibly censored by lost of follow-up or alive by the end of the study. In this study, the longest follow-up time was $\tau = 197$ and the censoring rate was $27\%$.

Our goal is to determine whether there was a difference between urban and rural populations on serious illness occurrences during their life time. Let $\Lambda_U$ and $\Lambda_R$ denote the mean functions of recurrent event processes for the urban and rural populations respectively, and we are interested in testing the null hypothesis: $H_0 : \Lambda_U = \Lambda_R = \Lambda_0$. The sample sizes of the two study groups are $(n_U, n_R) = (1489, 1561)$. We obtained the two-stage likelihood-based estimates $\widehat{\Lambda}_U$, $\widehat{\Lambda}_R$, and $\widehat{\Lambda}_0$ based on each group and the pooled data respectively, as shown in Figure 5. Because this study mainly involved elderly people, for large $u$, the estimate of $\Lambda(u)$ was largely based on the subjects with longer survival; for example, $u = 160$ months corresponded to those who lived 91 years old or beyond. Therefore, it is reasonable to expect these subjects had suffered much more serious diseases during their life time. In addition, it appears that for the elders with longevity, the serious diseases might mostly occur when they were around 70, which contributed to the sharp rising in the estimate of the cumulative event counts for a larger value of $u$. It can be seen that people living in urban areas seemed to experience more serious illnesses during their life time compared with those in rural areas.

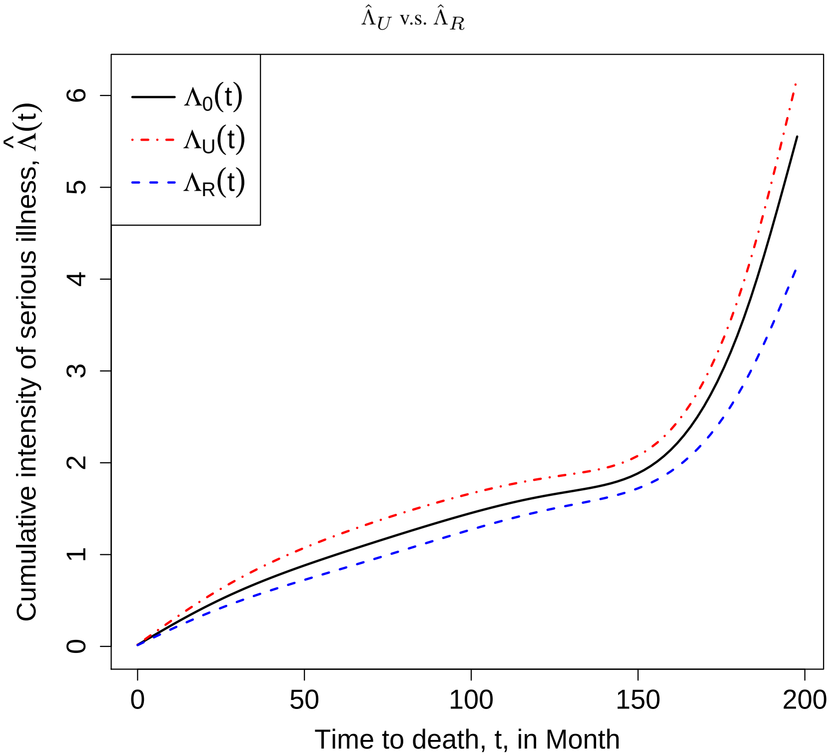


**Figure 4**. Plots of the estimates of the mean functions $\Lambda$'s for the urban, rural and the pooled group based on the CLHLS data. The solid line denotes the estimate for the common mean function based on the pooled data, the dot–dashed line denotes that for $\Lambda_U$, and the dashed line denotes that for $\Lambda_R$.

We further applied the proposed test in Section 4 to the hypothesis $H_0$. We chose weights $W_1$, $W_2$ and $W_3$ based on the patterns of the estimated mean functions, and adopted cubic spline basis functions with $q_n = 7$ which is in the order of $O(n^{1/4})$. As shown in Table 3, the life-time experience of serious illness was significantly different between people in urban and rural areas regardless the choice of weight processes. People living in urban areas had a significantly higher serious disease rate compared to those in rural areas, while the caution in interpreting the difference needs to be exercised. A relatively more stressful life style in urban areas might have contributed to a higher disease rate in people's life time, but incomprehensive health care systems and limited access to medical facilities in rural areas that resulted in fewer disease diagnoses might also be a contributing factor to explain the difference.

To examine whether our model has a reasonable fit to the CLHLS data, we estimated $\Lambda$ for subgroups defined by $U$. We divided the CLHLS data into two subsets according to the median value of $U$ (about 49 months from the study enrollment), where Group 1 consists of the individuals with $U$ values smaller than 49 months and Group 2 with $U \geq 49$. The sample sizes for two groups are 1000

**Table 3.** Two-sample test results of $H_0 : \Lambda_U = \Lambda_R$ with different weights for the CLHLS data

| | $W_1$ | $W_2$ | $W_3$ |
|---|---|---|---|
| $Q_n$ | 8.790 | 4.789 | 7.314 |
| SE | 2.211 | 1.256 | 1.837 |
| $p$-value | $< 10^{-3}$ | $< 10^{-3}$ | $< 10^{-3}$ |

Note: $W_j$'s represent the weights used in the test statistics, $Q_n$ represents the observed statistic, and SE represents the estimated standard error of $Q_n$.

and 1218, respectively. The estimates of mean function $\Lambda(t)$ for different groups are given in Figure 5. It shows that the mean function of Group 1 is similar to that of Group 2 for $t \in [0, 49]$, indicating the homogeneous temporal model is a reasonable assumption.

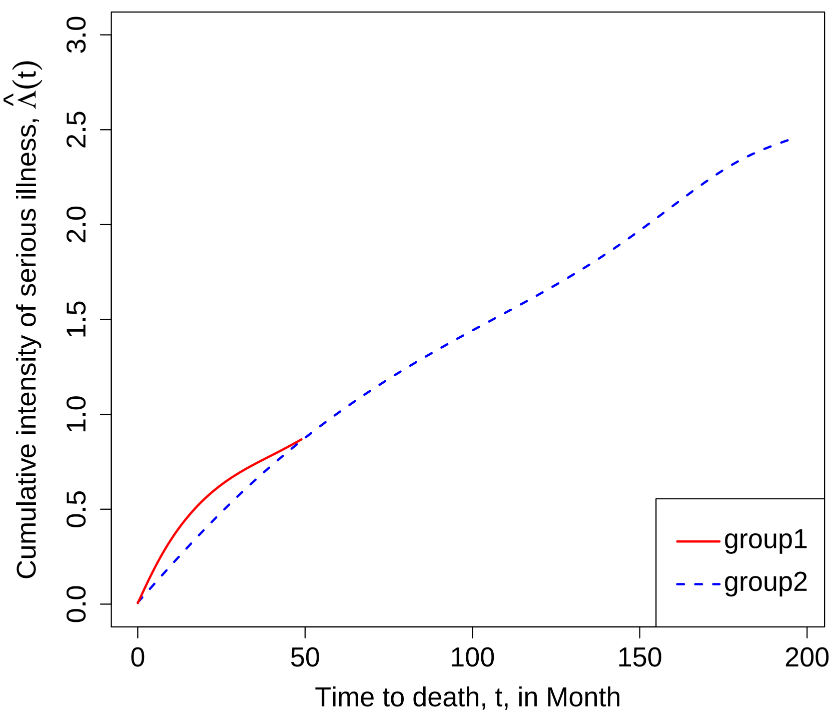


**Figure 5**. Plots of the estimates for the mean functions $\Lambda$'s from two groups divided by values of the terminal event time $U$ in the CLHLS data. The solid line denotes the estimate for the mean function based on the data with terminal event time smaller than 49 months from the study enrollment, and the dashed line denotes that with terminal event time greater than 49 months.

# 7. Concluding Remarks

We have proposed a flexible and intuitive nonparametric reversed conditional mean model to characterize the behavior of a recurrent event process near an informative terminal event that truncates the process in panel count data. The asymptotic properties of the two-stage estimator are thoroughly established, which are further applied to construct a class of new tests for nonparametric comparison of recurrent event processes. In the development of the two-stage estimation procedure, we base a Poisson process as the working model to obtain the likelihood function, while the asymptotic results do not depend on the Poisson process assumption, implying the proposed method is robust against the underlying stochastic mechanism of counting process $N(t)$.

Our work focuses on comparing the recurrent event processes truncated by a terminal event under the assumption that the distribution of the terminal event is the same across groups. It can be easily extended to comparing the whole disease processes that include both the recurrent event process and terminal event process by testing the null hypothesis, $H_0 : \Lambda_1 = \Lambda_2$ and $F_1 = F_2$, for two populations.

While the nonparametric inference is the focal point of this work, the reversed mean model for panel count data subject to an informative terminal event can be extended to semiparametric regression analysis that studies the effects of covariates on the recurrent event process prior to the terminal event. This extension is critical to analyze the CLHLS data in order to investigate the effects of socioeconomic status, family lifestyle, and demographic profile on the health of the aging population, which warrants for future research.

In general, the conditional expectation for the reversed count process $\widetilde{N}(t;U)$ can be expressed by

$$E(\widetilde{N}(t;U)|U=u) = \Lambda_{\widetilde{N}}(t|u), \quad 0 \le t \le u,$$

where $\Lambda_{\widetilde{N}}(\cdot|u)$ is a nonincreasing function for each fixed $u$ with $\Lambda_{\widetilde{N}}(u|u) = 0$. In this paper, we have focused on a specific reversed mean model, $\Lambda_{\widetilde{N}}(t|u) = \Lambda(u-t)$, which can be viewed as a homogeneous temporal model as the reversed event count only depends on the time to the terminal event $u-t$ but not the terminal event time $u$. This model is suitable for the CLHLS data as demonstrated in the previous section. Further research can be focused on other forms of $\Lambda_{\widetilde{N}}(t|u)$ or other models on $\widetilde{N}(t;U)$ such as a model with a latent acceleration parameter suggested by a reviewer.

# 8. Proofs of Theorems

## 8.1. Lemmas

First, we introduce more notation. Define the metric $d_2$ on $\Psi$ as

$$d_2^2(\Lambda_1, \Lambda_2) = \int \left|\Lambda_1(t) - \Lambda_2(t)\right|^2 d\mu_2(t) \qquad \text{for} \quad \Lambda_1, \Lambda_2 \in \Psi,$$

$\mathcal{F}_\eta = \{F : \|F - F_0\|_\infty \le \eta, F \in \mathcal{F}\}$, $\quad \Psi_{n\delta} = \{\Lambda : d_1(\Lambda, \Lambda_0) \le \delta, \Lambda \in \Psi_n\}$, and $\Psi_\delta^0 = \{\Lambda : d_1(\Lambda, \Lambda_0) \le \delta, \Lambda \in \Psi\}$. Let $\overline{F}(u) = 1 - F(u)$,

$$A_j(u) = \frac{\triangle N_j}{\triangle \Lambda_j(u)} - 1, \quad \text{and} \quad S(u) = \prod_{j=1}^{K} \frac{[\triangle \Lambda_j(u)]^{\triangle N_j} \exp(-\triangle \Lambda_j(u))}{\triangle N_j!}.$$

For the two quantities above, we use the notation $\widehat{A}_{nj}(u)$ and $\widehat{S}(u)$ when $\Lambda(u) = \widehat{\Lambda}_n(u)$, and $A_{0j}(u)$ and $S_0(u)$ when $\Lambda(u) = \Lambda_0(u)$. Define

$$m(\Lambda, F; D) = \Delta \sum_{j=1}^{K} \left\{ \triangle N_j \log(\triangle\Lambda_j(Y)) - \triangle\Lambda_j(Y) \right\} + (1-\Delta) \log \left\{ \int_Y^{\tau} S(u) dF(u) \right\},$$

$$m_1(\Lambda, F; D)[h] = \Delta \sum_{j=1}^{K} A_j(Y) h_j(Y) + (1-\Delta) \frac{\int_Y^\infty S(u) \sum_{j=1}^{K} A_j(u) h_j(u) dF(u)}{\int_Y^\infty S(u) dF(u)},$$

$$\begin{aligned} m_{11}(\Lambda, F; D)[h_1, h_2] = & -\Delta \sum_{j=1}^{K} \frac{\triangle N_j}{\triangle\Lambda_j^2(Y)} h_{1j}(Y) h_{2j}(Y) + (1-\Delta) \Big( \int_Y^\infty S(u) dF(u) \Big)^{-1} \\ & \times \int_Y^\infty S(u) \Bigg( \left[ \sum_{j=1}^{K} A_j(u) h_{1j}(u) \right] \cdot \left[ \sum_{j=1}^{K} A_j(u) h_{2j}(u) \right] \\ & - \sum_{j=1}^{K} \frac{\triangle N_j}{(\triangle\Lambda_j(u))^2} h_{1j}(u) h_{2j}(u) \Bigg) dF(u) - (1-\Delta) \Big( \int_Y^\infty S(u) dF(u) \Big)^{-2} \\ & \times \int_Y^\infty S(u) \sum_{j=1}^{K} A_j(u) h_{1j}(u) dF(u) \int_Y^\infty S(u) \sum_{j=1}^{K} A_j(u) h_{2j}(u) dF(u), \end{aligned}$$

$$\begin{aligned} m_{12}(\Lambda, F; D)[h_1, h_3] = & (1-\Delta) \Big( \int_Y^\infty S(u) dF(u) \Big)^{-1} \int_Y^\infty S(u) \sum_{j=1}^{K} A_j(u) h_{1j}(u) dh_3(u) \\ & - (1-\Delta) \Big( \int_Y^\infty S(u) dF(u) \Big)^{-2} \int_Y^\infty S(u) \sum_{j=1}^{K} A_j(u) h_{1j}(u) dF(u) \int_Y^\infty S(u) dh_3(u) \end{aligned}$$

with $h_{lj}(u) = h_l(u - T_{K,j-1}) - h_l(u - T_{K,j})$ for $l = 1, 2$, $G_n(\Lambda, F)[h] = \mathbb{P}_n m_1(\Lambda, F; D)[h]$,

$$\dot{G}_{1,\Lambda_0,F_0}(\Lambda - \Lambda_0)[h] = P(m_{11}(\Lambda_0, F_0; D)[h, \Lambda - \Lambda_0]), \quad \text{and}$$
$$\dot{G}_{2,\Lambda_0,F_0}(F - F_0)[h] = P(m_{12}(\Lambda_0, F_0; D)[h, F - F_0]).$$

Write $L(\Lambda, F; D) = \exp\{m(\Lambda, F; D)\}$.

**Lemma 1** *Under Conditions (C1)–(C3), $Pm(\Lambda, F_0; D)$ has a unique maximizer $\Lambda_0$.*

**Lemma 2** *Under Conditions (C1)–(C3) and (C7)–(C8), the class of functions $\{m(\Lambda, F; D) : \Lambda \in \Psi, F \in \mathcal{F}, \Lambda$ is uniformly bounded$\}$ is Donsker.*

**Lemma 3** *Let $\mathcal{G} = \{h : [0, \tau] \to [0, M]\}$ for a positive constant $M$. Define*

$$\mathcal{G}_\delta(F) = \{m(\Lambda, F; D) - m(\Lambda_0, F; D) : \Lambda \in \Psi_{n\delta}\},$$

$$\mathcal{G}_{1,\delta}(F)[h] = \{m_1(\Lambda, F; D)[h] - m_1(\Lambda_0, F; D)[h] : \Lambda \in \Psi_{n\delta}\},$$
$$\mathcal{G}_{2,\delta}(F)[h] = \Big\{V_n\Big(\Lambda, \Lambda_0, h, F\Big) : \Lambda \in \Psi_{n\delta}\Big\}$$

*for $F \in \mathcal{F}_\eta$ and $h \in \mathcal{G}$. Then under Conditions (C1)–(C3) and (C6)–(C8), we have for any $0 < \epsilon < \delta$,*

$$\log N_{[]}(\epsilon, \mathcal{G}_\delta(F), \|\cdot\|_{P,B}) \le cq_n \log(\delta/\epsilon), \tag{2}$$
$$\log N_{[]}(\epsilon, \mathcal{G}_{1,\delta}(F)[h], \|\cdot\|_{P,B}) \le cq_n \log(\delta/\epsilon), \tag{3}$$
$$\log N_{[]}(\epsilon, \mathcal{G}_{2,\delta}(F)[h], \|\cdot\|_{P,B}) \le cq_n \log(\delta/\epsilon), \tag{4}$$

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

**Lemma 4** *(i) Under Condition (C4), for any integrable function $\phi(x)$ on $[0, \infty)$,*

$$\left(\int_y^\infty \phi(x) d[F(x) - F_0(x)]\right)^2 \lesssim \int_y^\infty \dot{\phi}^2(x) dF_0(x) \cdot \|F - F_0\|_\infty^2 + \phi^2(y)\|F - F_0\|_\infty^2$$

*for any $F \in \mathcal{F}$ and any $y \in [0, \infty)$.*

*(ii) Under Conditions (C1)–(C4), we have for small enough $\delta$ and $\Lambda \in \Psi_\delta^0$,*

$$\Big|P\Big(m(\Lambda, F; D) - m(\Lambda, F_0; D) - (1-\Delta)\frac{F_0(Y) - F(Y)}{\overline{F}_0(Y)}\Big)\Big| \lesssim d_1(\Lambda, \Lambda_0)\|F - F_0\|_\infty + \|F - F_0\|_\infty^2. \tag{5}$$

*Particularly, $\big|P\big(m(\Lambda_0, F; D) - m(\Lambda_0, F_0; D)\big)\big| \lesssim \|F - F_0\|_\infty$.*

**Lemma 5** *Assume that for given $F \in \mathcal{F}_\eta$, and for arbitrary function $\phi_n : (0, \infty) \to R$ such that $\delta \to \phi_n(\delta)/\delta^\beta$ is decreasing for some $0 < \beta < 2$, every $\delta > 0$, every $\Lambda \in \Psi_n$,*

$$P(m(\Lambda, F; D) - m(\Lambda_0, F; D)) \lesssim -d_1^2(\Lambda, \Lambda_0) + d^2(F, F_0) + d_1(\Lambda, \Lambda_0)d(F, F_0),$$
$$E\left[\sup_{d_1(\Lambda,\Lambda_0)<\delta, \Lambda\in\Psi_n} |\sqrt{n}(\mathbb{P}_n - P)(m(\Lambda, F; D) - m(\Lambda_0, F; D))|\right] \le \phi_n(\delta),$$

*where $d(F, F_0)$ represents the distance metric between $F$ and $F_0$. Let $r_n > 0$ satisfy $\phi_n(r_n) \le \sqrt{n} r_n^2$ for every $n \in \mathbb{N}$. If $\hat{\Lambda}_n \in \Psi_n$ is a consistent estimator for $\Lambda$ satisfying $\mathbb{P}_n m(\hat{\Lambda}_n, F; D) \ge \mathbb{P}_n m(\Lambda_0, F; D) - O_p(r_n^2)$, then $d_1(\hat{\Lambda}_n, \Lambda_0) = O_p(r_n + d(F, F_0))$.*

According to Lemma 3 in the supplementary material of Kong and Nan (2016), we have the following conclusion on the consistency of the Kaplan–Meier (KM) estimator.

**Lemma 6** *(Consistency of KM estimator) Suppose Condition (C5) holds, then $\|\widehat{F}_n(t) - F_0(t)\|_\infty = O_p(n^{-1/2})$.*

Following Propositions 3, 4 and Theorem 5 of Akritas (2000), we can obtain the central limit theorem for integral of the KM estimator.

**Lemma 7** *(Asymptotic normality of KM estimator) Let*

$$\tilde{\phi}(u) = \overline{F}_0(u-)\Big[\phi(u) - \frac{1}{F_0(u)}\int_u^\tau \phi(t) dF_0(t)\Big]$$

*and* $M(u;D_i) = I_{(Y_i \le u, \Delta_i = 1)} - \int_{-\infty}^{u} I_{(Y_i \ge s)} dH_0(s)$ *with* $G_0$ *being the common distribution of* $C_i$, $\overline{F}_0(u) = 1 - F_0(u)$ *and* $H_0 = 1 - (1 - F_0)(1 - G_0)$. *Then under Condition (C2), we have for any integrable function* $\phi(u)$ *satisfying* $\displaystyle\int_0^\tau \frac{\phi^2(u)}{1 - G_0(u-)} dF_0(u) < \infty$,

$$\int_0^\tau \phi(u) d[\widehat{F}_n(u) - F_0(u)] = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \frac{\tilde{\phi}(u)}{1 - H_0(u-)} dM(u;D_i),$$

*and*

$$n^{1/2} \int_0^\tau \phi(u) d(\widehat{F}_n(u) - F_0(u)) \xrightarrow{d} N(0, \sigma^2),$$

*where* $\sigma^2$ *is defined as in (4) of Akritas (2000).*

## 8.2. Proof of Theorem 1

According to Lemma A1 in Lu et al. (2007), for $\Lambda_0 \in \mathcal{H}_r$, there exists $\Lambda_n \in \Psi_n$ such that $\|\Lambda_n - \Lambda_0\|_\infty = O(n^{-r\nu})$. Choose a positive monotone $h_n \in \Psi_n$ such that $\|h_n\|^2_{L_2(\mu_1)} = O(n^{-\nu r} + n^{-(1-\nu)/2})$. Then for any $\alpha > 0$, $\|\Lambda_n - \Lambda_0 + \alpha h_n\|^2_{L_2(\mu_1)} = O(n^{-\nu r} + n^{-(1-\nu)/2})$ and $\inf_{t-s \ge s_0} (\triangle\Lambda_n(t,s) - \triangle\Lambda_0(t,s) + \alpha\triangle h_n(t,s)) > 0$ for sufficiently large $n$, where $\triangle h_n(t,s) = h_n(t) - h_n(s)$ and $s_0$ is defined as in (C7). Let $H_n(\alpha) = m(\Lambda_n + \alpha h_n, \hat{F}_n; D)$. Then the first derivative of $H_n$ is

$$\begin{aligned}\dot{H}_n(\alpha) =& \Delta \sum_{j=1}^{K} \Big\{ \frac{\triangle N_j}{\triangle\Lambda_{nj}(Y) + \alpha h_{n,j}(Y)} - 1 \Big\} h_{n,j}(Y) + \frac{1 - \Delta}{\int_Y^\infty S_n(u) d\hat{F}_n(u)} \\ & \times \int_Y^\infty S_n(u) \sum_{j=1}^{K} \Big\{ \frac{\triangle N_j}{\triangle\Lambda_{nj}(u) + \alpha h_{n,j}(u)} - 1 \Big\} h_{n,j}(u) d\hat{F}_n(u),\end{aligned}$$

where $S_n(u)$ represents $S(u)$ taking value at $\triangle\Lambda_j(u) = \triangle\Lambda_{nj}(u) + \alpha h_{n,j}(u)$. We claim that $P(d_1(\hat{\Lambda}_n, \Lambda_n) \le \alpha_0 \|h_n\|_{L_2(\mu_1)}) \to 1$ by showing that $\mathbb{P}_n \dot{H}_n(\alpha_0) < 0$ and $\mathbb{P}_n \dot{H}_n(-\alpha_0) > 0$ for any $\alpha_0 > 0$. Note that $\mathbb{P}_n \dot{H}_n(\alpha_0) = (\mathbb{P}_n - P)\dot{H}_n(\alpha_0) + P\dot{H}_n(\alpha_0) := I_1 + I_2$. Similar to Lemma 3, using Conditions (C7) and (C8), we can show that the class $\mathcal{L}_{1\delta}(F) = \{m_1(\Lambda, F; D)[\Lambda - \Lambda_n] : \Lambda \in \Psi_{n\delta}, F \in \mathcal{F}_\eta\}$ with $\delta^2 = n^{-\nu r} + n^{-(1-\nu)/2}$ is Donsker. Hence, $I_1 = O_p(n^{-1/2})$. Since $\|\Lambda_n - \Lambda_0 + \alpha_0 h_n\|^2_{L_2(\mu_1)} = O(n^{-\nu r} + n^{-(1-\nu)/2}) = o(1)$ and $\|\hat{F}_n - F_0\|_\infty = O_p(n^{-1/2})$, then by Lemma 4, we have

$$\begin{aligned}I_2 \lesssim & P\Bigg( \Delta \sum_{j=1}^{K} \Big( \frac{\triangle\Lambda_{0j}(Y)}{\triangle\Lambda_{nj}(Y) + \alpha_0 h_{n,j}(Y)} - 1 \Big) h_{n,j}(Y) \\ & + \frac{1 - \Delta}{\overline{F}_0(Y)} \int_Y^\infty \sum_{j=1}^{K} \Big( \frac{\triangle\Lambda_{0j}(u)}{\triangle\Lambda_{nj}(u) + \alpha_0 h_{n,j}(u)} - 1 \Big) h_{n,j}(u) dF_0(u) \Bigg) + \|\hat{F}_n - F_0\|_\infty.\end{aligned}$$

Define $a_{nj} = \triangle\Lambda_{nj} - \triangle\Lambda_{0j} + \alpha_0 h_{n,j}$ and $b(s) = \triangle\Lambda_{0j}/(\triangle\Lambda_{0j} + sa_{nj})$ for $0 \leq s \leq 1$. Then

$$b(s) = 1 + \big(-\frac{a_{nj}}{\triangle\Lambda_{0j}}\big)s + \frac{\triangle\Lambda_{0j}a_{nj}^2}{(\triangle\Lambda_{0j} + \xi a_{nj})^2}s^2,$$

for some $\xi \in (0,1)$. Since $\triangle\Lambda_{0j}$ and $a_{nj}$ are bounded on $[0,\tau]$, there exists constants $c_1$ and $c_2$ such that

$$c_1 E\left\{\sum_{j=1}^{K} a_{nj}^2(U)\right\} \leq E\left\{\sum_{j=1}^{K} \frac{\triangle\Lambda_{0j}(U)a_{nj}^2(U)}{(\triangle\Lambda_{0j}(U) + \xi a_{nj}(U))^2}\right\} \leq c_2 E\left\{\sum_{j=1}^{K} a_{nj}^2(U)\right\}.$$

Therefore,

$$E\left\{\sum_{j=1}^{K} \frac{\triangle\Lambda_{0j}(U)a_{nj}^2(U)}{(\triangle\Lambda_{0j}(U) + \xi a_{nj}(U))^2}\right\} = O(n^{-\nu r} + n^{-(1-\nu)/2}),$$

and so

$$I_2 \leq E\left\{\sum_{j=1}^{K}(-c_1 a_{nj} + c_2 a_{nj}^2)h_{n,j} + c_3 n^{-1/2}\right\} \leq -\frac{c_1}{2}E\left\{\sum_{j=1}^{K} a_{nj} + c_3 n^{-1/2}\right\} = -c(n^{-\nu r} + n^{-(1-\nu)/2})$$

for some constant $c > 0$. Noting that $n^{-1/2} = o(n^{-\nu r} + n^{-(1-\nu)/2})$ for $0 < \nu < 1/2$, we have $\mathbb{P}_n \dot{H}_n(\alpha_0) < 0$ except on an event with probability converging to zero. The same arguments can show that $\mathbb{P}_n \dot{H}_n(-\alpha_0) > 0$ with probability converging to 1 as $n \to \infty$. Thus, we get that $P(d_1(\widehat{\Lambda}_n, \Lambda_0) > \varepsilon) \to 0$ for any $\varepsilon > 0$. So there exits a measurable set $\Xi$ with $P(\Xi) \geq 1 - \epsilon$ for any $\epsilon > 0$ such that $\widehat{\Lambda}_n$ defined on $[0,\tau]$ is uniformly bounded on $\Xi$.

We now restrict us on the measurable set $\Xi$. Recalling that $\|\Lambda_n - \Lambda_0\|_\infty = O(q_n^{-r})$, we have

$$\begin{aligned}
&\mathbb{P}_n m(\Lambda_n, \widehat{F}_n; D) - \mathbb{P}_n m(\Lambda_0, \widehat{F}_n; D)\\
=&(\mathbb{P}_n - P)\big(m(\Lambda_n, \widehat{F}_n; D) - m(\Lambda_0, \widehat{F}_n; D)\big) + P\big(m(\Lambda_n, \widehat{F}_n; D) - m(\Lambda_0, \widehat{F}_n; D)\big) := I_1 + I_2.
\end{aligned}$$

Since a Donsker class is also Glivanko-Cantilli class, then by Lemma 7, we have $|I_1| = o_p(1)$. In addition, $|I_2| \leq cE(N(T_{K,K}) + 1)\|\Lambda_n - \Lambda_0\|_\infty = O(q_n^{-r}) = o(1)$. Thus, we have

$$\mathbb{P}_n m(\Lambda_n, \widehat{F}_n; D) = \mathbb{P}_n m(\Lambda_0, \widehat{F}_n; D) + o_p(1).$$

Then the definition of $\widehat{\Lambda}_n$ yields that

$$\mathbb{P}_n m(\widehat{\Lambda}_n, \widehat{F}_n; D) \geq \sup_{\Lambda_n \in \Psi_n} \mathbb{P}_n m(\Lambda_n, \widehat{F}_n; D) \geq \mathbb{P}_n m(\Lambda_0, \widehat{F}_n; D) + o_p(1) = \mathbb{P}_n m(\Lambda_0, F_0; D) + o_p(1), \tag{6}$$

where the last equality is from Lemma 6 and Lemma 4. Using Lemma 2, we have

$$0 \leq P(m(\Lambda_0, F_0; D) - m(\widehat{\Lambda}_n, F_0; D)) = o_p(1), \tag{7}$$

where the second inequality is obtained from (6) and the last equality is from Lemma 4. Besides, we have by Lemma 1 that for any $\delta > 0$, $\sup_{d_1(\Lambda,\Lambda_0) > \delta} Pm(\Lambda, F_0; D) < Pm(\Lambda_0, F_0; D)$. This shows that

$$\{d_1(\widehat{\Lambda}_n, \Lambda_0) > \delta\} \subset \{Pm(\widehat{\Lambda}_n, F_0; D) < Pm(\Lambda_0, F_0; D)\},$$

with the sequence of the events on the right going to a null event in view of inequality (7). Since this relation holds on the measurable set $\Xi$ with $P(\Xi) \geq 1-\epsilon$, it implies $d_1(\widehat{\Lambda}_n, \Lambda_0) \to 0$ almost uniformly. So we conclude the almost sure convergence of $\widehat{\Lambda}_n$ using the arguments of Lemma 1.9.2 in van der Vaart and Wellner (1996).

## 8.3. Proof of Theorem 2

We prove the theorem by verifying the conditions in Lemma 5. Using the triangle inequality, we have

$$\begin{aligned}
Pm(\Lambda, \widehat{F}_n; D) - Pm(\Lambda_0, \widehat{F}_n; D) \leq & \Big|P\Big(m(\Lambda, \widehat{F}_n; D) - m(\Lambda, F_0; D) - (1-\Delta)\frac{F_0(Y) - F(Y)}{\overline{F}_0(Y)}\Big)\Big| \\
& + \Big|P\Big(m(\Lambda_0, \widehat{F}_n; D) - m(\Lambda_0, F_0; D) - (1-\Delta)\frac{F_0(Y) - F(Y)}{\overline{F}_0(Y)}\Big)\Big| \\
& + P(m(\Lambda, F_0; D) - m(\Lambda_0, F_0; D)) := I_1 + I_2 + I_3.
\end{aligned}$$

From Lemma 4, it follows that for $\Lambda \in \Psi_\delta^0$, $I_1 + I_2 \lesssim d_1(\Lambda, \Lambda_0)\|\widehat{F}_n - F_0\|_\infty + \|\widehat{F}_n - F_0\|_\infty^2$. For $I_3$,

$$\begin{aligned}
I_3 = & P\left(\Delta \sum_{j=1}^K \left\{\triangle N_j \log \frac{\triangle\Lambda_j(Y)}{\triangle\Lambda_{0j}(Y)} - (\triangle\Lambda_j(Y) - \triangle\Lambda_{0j}(Y))\right\}\right) + P\left((1-\Delta)\log \frac{\int_Y^\infty S(u)dF_0(u)}{\int_Y^\infty S_0(u)dF_0(u)}\right) \\
:= & I_{31} + I_{32}.
\end{aligned}$$

For $I_{31}$ and $I_{32}$, we can show that $I_{31} \lesssim -P\Big(\Delta \sum_{j=1}^K \Big(\triangle\Lambda_j(Y) - \triangle\Lambda_{0j}(Y)\Big)^2\Big)$ and

$$P(I_{32}) \lesssim -P\left(\frac{1-\Delta}{\overline{F}_0(Y)}\int_Y^\infty \sum_{j=1}^K (\triangle\Lambda_j(u) - \triangle\Lambda_{0j}(u))^2 dF_0(u)\right).$$

Thus, we have $I_3 \lesssim -d_1^2(\Lambda, \Lambda_0)$. This gives that

$$P(m(\Lambda, \widehat{F}_n; D) - m(\Lambda_0, \widehat{F}_n; D)) \lesssim -d_1^2(\Lambda, \Lambda_0) + d_1(\Lambda, \Lambda_0)d(\widehat{F}_n, F_0) + d^2(\widehat{F}_n, F_0). \tag{8}$$

Moreover, similar to the proof of part (ii) in Lemma 3, we can show that

$$\|m(\Lambda, \widehat{F}_n; D) - m(\Lambda_0, \widehat{F}_n; D)\|_{P,B}^2 \lesssim d_1(\Lambda, \Lambda_0)d(\widehat{F}_n, F_0) + d_1^2(\Lambda, \Lambda_0) \lesssim \delta^2$$

for $n^{-1/2} \lesssim O(\delta)$ and $d_1(\Lambda, \Lambda_0) \lesssim O(\delta)$. Therefore, Lemma 3 gives that

$$J_{[]}(\delta, \mathcal{G}_\delta(\widehat{F}_n), \|\cdot\|_{P,B}) = \int_0^\delta (1 + \log N_{[]}(\epsilon, \mathcal{G}_\delta(\widehat{F}_n), \|\cdot\|_{P,B}))^{1/2} d\epsilon \leq c q_n^{1/2}\delta.$$

Then by Lemma 3.4.3 of van der Vaart and Wellner (1996), we obtain

$$\begin{aligned}
& E\left\{\sup_{d_1(\Lambda,\Lambda_0)<\delta, \Lambda\in\Psi_n} \sqrt{n}|(\mathbb{P}_n - P)(m(\Lambda, \widehat{F}_n; D) - m(\Lambda_0, \widehat{F}_n; D))|\right\} \\
\leq & J_{[]}(\delta, \mathcal{G}_\delta(\widehat{F}_n), \|\cdot\|_{P,B})(1 + J_{[]}(\delta, \mathcal{G}_\delta(\widehat{F}_n), \|\cdot\|_{P,B})/(\delta^2 n^{1/2})) \lesssim q_n^{1/2}\delta + q_n n^{-1/2}.
\end{aligned}$$

Setting $\phi_n(\delta) = q_n^{1/2}\delta + q_n n^{-1/2}$ and $r_n = n^{-(1-\nu)/2}$, it is easy to see that $\phi_n(\delta)/\delta$ is a decreasing function of $\delta$ and $\phi_n(r_n)/r_n^2 = O(n^{1/2})$.

According to Lemma A1 in Lu et al. (2007), there exists $\Lambda_n \in \Psi_n$ with order $l \geq r+2$ such that $\|\Lambda_n - \Lambda_0\|_\infty = O(n^{-r\nu})$ for $0 < \nu < 1/2$. Then

$$\begin{aligned}
&\mathbb{P}_n m(\widehat{\Lambda}_n, \widehat{F}_n; D) - \mathbb{P}_n m(\Lambda_0, \widehat{F}_n; D) \\
=&\big(\mathbb{P}_n m(\widehat{\Lambda}_n, \widehat{F}_n; D) - \mathbb{P}_n m(\Lambda_n, \widehat{F}_n; D)\big) + (\mathbb{P}_n - P)\big(m(\Lambda_n, \widehat{F}_n; D) - m(\Lambda_0, \widehat{F}_n; D)\big) \\
&+ P\big(m(\Lambda_n, \widehat{F}_n; D) - m(\Lambda_0, \widehat{F}_n; D)\big) \\
:=&I_1 + I_2 + I_3,
\end{aligned}$$

where $I_1 \geq 0$ by the definition of $\widehat{\Lambda}_n$. For $I_2$, we set

$$\tilde{\mathcal{G}}(F) = \Big\{ \frac{m(\Lambda, F; D) - m(\Lambda_0, F; D)}{n^{-r\nu+\epsilon}} : \Lambda \in \Psi_n, \|\Lambda - \Lambda_0\|_\infty = O(n^{-r\nu}) \Big\}$$

for any $0 < \epsilon < 1/2 - r\nu$, and it can be argued that $\tilde{\mathcal{G}}(\widehat{F}_n)$ is P-Donsker by (C1)–(C3) and (C6)–(C8). In addition, we can obtain that $Pf^2 \to 0$ as $n \to \infty$ for any $f \in \tilde{\mathcal{G}}(\widehat{F}_n)$. Thus, by Corollary 2.3.12 of van der Varrt and Wellner (1996), we have

$$I_2 = n^{-r\nu+\epsilon}(\mathbb{P}_n - P)\Big(\frac{m(\Lambda_n, \widehat{F}_n; D) - m(\Lambda_0, \widehat{F}_n; D)}{n^{-r\nu+\epsilon}}\Big) = O_p(n^{-r\nu+\epsilon}n^{-1/2}) = o(n^{-2r\nu}).$$

For $I_3$, we have $I_3 \geq -O_p(n^{-2r\nu})$ by following the deduction for (8). Hence,

$$\mathbb{P}_n m(\widehat{\Lambda}_n, \widehat{F}_n; D) - \mathbb{P}_n m(\Lambda_0, \widehat{F}_n; D) \geq -O_p(n^{-2r\nu}).$$

To satisfy the conditions in Lemma 5, we need $n^{-2r\nu} = O(n^{-(1-\nu)})$. The choice of $\nu = 1/(1+2r)$ yields the convergence rate of $n^{r/(1+2r)}$. This completes the proof of Theorem 2.

## 8.4. Proof of Theorem 3

From Assumptions (B2)–(B4), we have

$$-\sqrt{n}\dot{G}_{1,\Lambda_0,\widehat{F}_n}(\widehat{\Lambda}_n - \Lambda_0)[h] - \sqrt{n}\dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h] = -\sqrt{n}G(\widehat{\Lambda}_n, \widehat{F}_n)[h] + o_p(1). \quad (9)$$

Assumptions (B1) and (B2) give

$$-\sqrt{n}G(\widehat{\Lambda}_n, \widehat{F}_n)[h] = \sqrt{n}(G_n - G)(\Lambda_0, F_0)[h] + o_p(1). \quad (10)$$

Then it follows from (9) and (10) that

$$-\sqrt{n}\dot{G}_{1,\Lambda_0,\widehat{F}_n}(\widehat{\Lambda}_n - \Lambda_0)[h] = \sqrt{n}(G_n - G)(\Lambda_0, F_0)[h] + \sqrt{n}\dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h] + o_p(1).$$

Thus, the theorem is proved by Assumption (B5).

## 8.5. Proof of Theorem 4

To prove the theorem, we need to verify Assumptions (B1)–(B5) by Theorem 3.

(a) We show (B1). Noting that $d_1(\widehat{\Lambda}_n, \Lambda_0) = O_p(n^{-r/(1+2r)})$ by Theorem 2, we have $\lim_{n\to\infty} P(\widehat{\Lambda}_n \in \Psi_{n\delta}) = 1$ with $\delta = O(n^{-r/(1+2r)})$. Hence, we only need to prove

$$\sup_{\Lambda\in\Psi_{n\delta}} |\sqrt{n}(G_n - G)(\Lambda, \widehat{F}_n)[h] - \sqrt{n}(G_n - G)(\Lambda_0, F_0)[h]| = o_p(1) \tag{11}$$

uniformly in $h \in \mathcal{H}_r$. Lemma 3 gives that $J_{[]}(\delta, g_{1,\delta}(\widehat{F}_n)[h], \|\cdot\|_{P,B}) \lesssim q_n^{1/2}\delta$. Applying Lemma 3.4.3 of van de Vaart and Wellner (1996), we obtain that

$$\begin{aligned}
&E\left\{\sup_{\Lambda\in\Psi_{n\delta}} |\sqrt{n}(\mathbb{P}_n - P)(m_1(\Lambda, \widehat{F}_n; D)[h] - m_1(\Lambda_0, \widehat{F}_n; D)[h])|\right\}\\
&\lesssim J_{[]}(\delta, g_{1,\delta}(\widehat{F}_n)[h], \|\cdot\|_{P,B})\big(1 + J_{[]}(\delta, g_{1,\delta}(\widehat{F}_n)[h], \|\cdot\|_{P,B})/(\delta^2 n^{1/2})\big) = o(1).
\end{aligned}$$

Then by Markov's inequality, we have

$$\sup_{\Lambda\in\Psi_{n\delta}} |\sqrt{n}(G_n - G)(\Lambda, \widehat{F}_n)[h] - \sqrt{n}(G_n - G)(\Lambda_0, \widehat{F}_n)[h]| = o_p(1).$$

In addition, Lemma 1 shows that

$$|\sqrt{n}(G_n - G)(\Lambda_0, \widehat{F}_n)[h] - \sqrt{n}(G_n - G)(\Lambda_0, F_0)[h]| = o_p(1).$$

Thus, by the triangle inequality, we obtain (11).

(b) We show (B2). A simple calculation yields that $G(\Lambda_0, F_0)[h] = Pm_1(\Lambda_0, F_0; D)[h] = 0$ for any $h \in \mathcal{H}_r$. It remains to show $G_n(\widehat{\Lambda}_n, \widehat{F}_n)[h] = \mathbb{P}_n m_1(\widehat{\Lambda}_n, \widehat{F}_n; D)[h] = o_p(n^{-1/2})$ for $h \in \mathcal{H}_r$. We first note that $\widehat{\Lambda}_n = \sum_{l=1}^{q_n} \widehat{\alpha}_l B_l$ satisfies the following equation $\frac{\partial \mathbb{P}_n m(\Lambda, \widehat{F}_n; D)}{\partial \alpha_l}\Big|_{\alpha_l = \widehat{\alpha}_l} = 0$, which means that for any $h_n \in \Psi_n$, $G_n(\widehat{\Lambda}_n, \widehat{F}_n)[h_n] = 0$. Besides, for any $h \in \mathcal{H}_r$, there exists $h_n \in \Psi_n$ such that $\|h^{(r)} - h_n^{(r)}\|_\infty = O(n^{-r\nu}), r = 0, 1$ by Lemma A1 in Lu et al. (2007). We now only need to show that $G_n(\widehat{\Lambda}_n, \widehat{F}_n)[h - h_n] = o_p(n^{-1/2})$. To this end, we write

$$\begin{aligned}
G_n(\widehat{\Lambda}_n, \widehat{F}_n)[h - h_n] &= \Big(G_n(\widehat{\Lambda}_n, \widehat{F}_n)[h - h_n] - G_n(\Lambda_0, \widehat{F}_n)[h - h_n]\Big)\\
&\quad + \Big(G_n(\Lambda_0, \widehat{F}_n)[h - h_n] - G_n(\Lambda_0, F_0)[h - h_n]\Big) + G_n(\Lambda_0, F_0)[h - h_n]\\
&\quad := I_1 + I_2 + I_3.
\end{aligned}$$

We first note

$$\begin{aligned}
&P|I_1|\\
&\le P\left(\Delta \sum_{j=1}^{K} \Big|\frac{\triangle\widehat{\Lambda}_{nj}(Y) - \triangle\Lambda_{0j}(Y)}{\triangle\widehat{\Lambda}_{nj}(Y)}\Big|\right)\|h - h_n\|_\infty
\end{aligned}$$

$$
\begin{aligned}
&+P\Bigg((1-\Delta)\Bigg|\frac{\int_Y^\infty \widehat{S}_n(u)\sum_{j=1}^K \widehat{A}_{nj}(u)(h_j-h_{n,j})(u)dF_0(u)}{\int_Y^\infty \widehat{S}_n(u)dF_0(u)}-\frac{\int_Y^\infty S_0(u)\sum_{j=1}^K A_{0j}(u)(h_j-h_{n,j})(u)dF_0(u)}{\int_Y^\infty S_0(u)dF_0(u)}\Bigg|\Bigg)\\
&+P\Bigg((1-\Delta)\Bigg|\frac{\int_Y^\infty \widehat{S}_n(u)\sum_{j=1}^K \widehat{A}_{nj}(u)(h_j-h_{n,j})(u)d\widehat{F}_n(u)}{\int_Y^\infty \widehat{S}_n(u)d\widehat{F}_n(u)}-\frac{\int_Y^\infty \widehat{S}_n(u)\sum_{j=1}^K \widehat{A}_{nj}(u)(h_j-h_{n,j})(u)dF_0(u)}{\int_Y^\infty \widehat{S}_n(u)dF_0(u)}\Bigg|\Bigg)\\
&+P\Bigg((1-\Delta)\Bigg|\frac{\int_Y^\infty S_0(u)\sum_{j=1}^K A_{0j}(u)(h_j-h_{n,j})(u)d\widehat{F}_n(u)}{\int_Y^\infty S_0(u)d\widehat{F}_n(u)}-\frac{\int_Y^\infty S_0(u)\sum_{j=1}^K A_{0j}(u)(h_j-h_{n,j})(u)dF_0(u)}{\int_Y^\infty S_0(u)dF_0(u)}\Bigg|\Bigg)\\
:=&I_{11}+I_{12}+I_{13}+I_{14}.
\end{aligned}
$$

It can be seen that $I_{11}\le cP\Big(\Delta\sum_{j=1}^K\big|\triangle\widehat{\Lambda}_{nj}(Y)-\triangle\Lambda_{0j}(Y)\big|\Big)\cdot\|h-h_n\|_\infty$. For $I_{12}$, we take $h_1=\triangle\widehat{\Lambda}_n-\triangle\Lambda_0$ and

$$
\begin{aligned}
g(t)=&\Bigg(\int_Y^\infty \exp\Big(\sum_{j=1}^K\Big[\triangle N_j\log(\triangle\Lambda_{0j}+th_{1j})(u)\Big]-(\triangle\Lambda_{0j}+th_{1j})(u)-\log\triangle N_j!\Big)dF_0(u)\Bigg)^{-1}\\
&\times\int_Y^\infty \exp\Big(\sum_{j=1}^K\Big[\triangle N_j\log(\triangle\Lambda_{0j}+th_{1j})(u)-(\triangle\Lambda_{0j}+th_{1j})(u)-\log\triangle N_j!\Big]\Big)\\
&\times\sum_{j=1}^K\Big(\frac{\triangle N_j}{(\triangle\Lambda_{0j}+th_{1j})(u)}-1\Big)(h_j-h_{n,j})(u)dF_0(u).
\end{aligned}
$$

Then we have

$$
\begin{aligned}
I_{12}=&P\big((1-\Delta)|g(1)-g(0)|\big)=P\big((1-\Delta)|\dot{g}(\xi)|\big)\\
\le&c\Bigg[P\Bigg(\frac{1-\Delta}{\overline{F}_0(Y)}\int_Y^\infty\sum_{j=1}^K h_{1j}^2(u)dF_0(u)\Bigg)\Bigg]^{1/2}\|h-h_n\|_\infty.
\end{aligned}
$$

Moreover, by Lemma 4

$$
\begin{aligned}
I_{13}=&P\Bigg((1-\Delta)\Bigg|\frac{\int_Y^\infty \widehat{S}_n(u)\sum_{j=1}^K \widehat{A}_{nj}(u)(h_j-h_{n,j})(u)d(\widehat{F}_n-F_0)(u)}{\int_Y^\infty \widehat{S}_n(u)dF^\xi(u)}\\
&-\frac{\int_Y^\infty \widehat{S}_n(u)\sum_{j=1}^K \widehat{A}_{nj}(u)(h_j-h_{n,j})(u)dF^\xi(u)\cdot\int_Y^\infty \widehat{S}_n(u)d(\widehat{F}_n-F_0)(u)}{\Big(\int_Y^\infty \widehat{S}_n(u)dF^\xi(u)\Big)^2}\Bigg|\Bigg)\\
\lesssim&\|\widehat{F}_n-F_0\|_\infty\cdot\|h-h_n\|_\infty,
\end{aligned}
$$

where $F^{\xi} = F_0 + \xi(\widehat{F}_n - F_0)$ for some $\xi \in (0,1)$ and $I_{14} \lesssim \|\widehat{F}_n - F_0\|_\infty \cdot \|h - h_n\|_\infty$. Therefore,

$$P|I_1| = O_p\big(d_1(\widehat{\Lambda}_n, \Lambda_0) \cdot \|h - h_n\|_\infty + \|\widehat{F}_n - F_0\|_\infty \cdot \|h - h_n\|_\infty\big) = o_p(n^{-1/2}).$$

Similarly, we can obtain that $P|I_2| \lesssim \|\widehat{F}_n - F_0\|_\infty \|h - h_n\|_\infty = o_p(n^{-1/2})$, and

$$\begin{aligned}
PI_3^2 \lesssim & \frac{1}{n} P\left(\Delta \sum_{j=1}^K A_{0j}(Y)(h_j - h_{n,j})(u) + (1-\Delta) \frac{\int_Y^\infty S_0(u) \sum_{j=1}^K A_{0j}(u)(h_j - h_{n,j})(u) dF_0(u)}{\int_Y^\infty S_0(u) dF_0(u)}\right)^2 \\
\lesssim & \frac{1}{n}\|h - h_n\|_\infty^2.
\end{aligned}$$

Thus, we have $G_n(\widehat{\Lambda}_n, \widehat{F}_n)[h - h_n] = o_p(n^{-1/2})$.

(c) We show (B3). By the smoothness of $G_n(\Lambda, F)[h]$ with respect to $\Lambda$ and $F$, we claim that $G(\Lambda, F)[h]$ is Fréchet-differentiable with respect to $\Lambda$ and $F$ at $(\Lambda_0, F_0)$. Moreover, their Fréchet derivatives are

$$\begin{aligned}
&\dot{G}_{1,\Lambda_0,F_0}(\Lambda - \Lambda_0)[h] = P(m_{11}(\Lambda_0, F_0)[h, \Lambda - \Lambda_0]) \quad \text{and} \\
&\dot{G}_{2,\Lambda_0,F_0}(F - F_0)[h] = P(m_{12}(\Lambda_0, F_0)[h, F - F_0]).
\end{aligned}$$

(d) We show (B4). Since $G(\Lambda_0, F_0)[h] = 0$, we have

$$\begin{aligned}
&G(\widehat{\Lambda}_n, \widehat{F}_n)[h] - G(\Lambda_0, F_0)[h] - \dot{G}_{1,\Lambda_0,\widehat{F}_n}(\widehat{\Lambda}_n - \Lambda_0)[h] - \dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h] \\
=&\big(\dot{G}_{1,\Lambda^\xi,\widehat{F}_n} - \dot{G}_{1,\Lambda_0,\widehat{F}_n}\big)(\widehat{\Lambda}_n - \Lambda_0)[h] + (\dot{G}_{2,\Lambda_0,F^\xi} - \dot{G}_{2,\Lambda_0,F_0})(\widehat{F}_n - F_0)[h] := I_1 + I_2,
\end{aligned}$$

where $F^{\xi} = F_0 + \xi(\widehat{F}_n - F_0)$ for some $\xi \in (0,1)$. We recall that $h_1 = \triangle\widehat{\Lambda}_n - \triangle\Lambda_0$ and split $I_1$ into $I_{11} - I_{12} - I_{13}$, where

$$\begin{aligned}
|I_{11}| =& \Bigg| P\left(\Delta \sum_{j=1}^K \Big(\frac{\triangle N_j}{(\triangle\Lambda_j^\xi(Y))^2} - \frac{\triangle N_j}{\triangle\Lambda_{0j}^2(Y)}\Big) h_{1j}(Y) h_j(Y)\right) \\
&+ P\Bigg((1-\Delta)\frac{\int_Y^\infty S_0(u)\Big[\sum_{j=1}^K A_{\xi,j}(u)h_{1j}(u)\Big]\Big[\sum_{j=1}^K A_{\xi,j}(u)h_j(u)\Big] d\widehat{F}_n(u)}{\int_Y^\infty S_0(u) d\widehat{F}_n(u)} \\
&- (1-\Delta)\frac{\int_Y^\infty S_0(u)\Big[\sum_{j=1}^K A_{0j}(u)h_{1j}(u)\Big]\Big[\sum_{j=1}^K A_{0j}(u)h_j(u)\Big] d\widehat{F}_n(u)}{\int_Y^\infty S_0(u) d\widehat{F}_n(u)} + o_p(1)\Bigg)\Bigg| \\
\lesssim & d_1^2(\widehat{\Lambda}_n, \Lambda_0) + o_p(\|\widehat{F}_n - F_0\|_\infty), \\
|I_{12}| =& \left| P\left((1-\Delta)\frac{\int_Y^\infty S_0(u)\Big[\sum_{j=1}^K \Big(\frac{\triangle N_j}{(\triangle\Lambda_j^\xi(u))^2} - \frac{\triangle N_j}{\triangle\Lambda_{0j}^2(u)}\Big) h_{1j}(u) h_j(u)\Big] d\widehat{F}_n(u)}{\int_Y^\infty S_0(u) d\widehat{F}_n(u)}\right) + o_p(1)\right| \\
\lesssim & d_1^2(\widehat{\Lambda}_n, \Lambda_0) + o_p(\|\widehat{F}_n - F_0\|_\infty),
\end{aligned}$$

$$
\begin{aligned}
|I_{13}| = & \Bigg| P\Bigg( (1-\Delta)\Big(\int_Y^\infty S_0(u) d\widehat{F}_n(u)\Big)^{-2} \Bigg( \int_Y^\infty S_0(u) \sum_{j=1}^K A_{\xi,j}(u) h_{1j}(u) d\widehat{F}_n(u) \\
& \times \int_Y^\infty S_0(u) \sum_{j=1}^K A_{\xi,j}(u) h_j(u) d\widehat{F}_n(u) \\
& - \int_Y^\infty S_0(u) \sum_{j=1}^K A_{0j}(u) h_{1j}(u) d\widehat{F}_n(u) \int_Y^\infty S_0(u) \sum_{j=1}^K A_{0j}(u) h_j(u) d\widehat{F}_n(u) \Bigg)\Bigg) + o_p(1) \Bigg| \\
\lesssim & d_1^2(\widehat{\Lambda}_n, \Lambda_0) + o_p(\|\widehat{F}_n - F_0\|_\infty).
\end{aligned}
$$

We then split $I_2$ into $I_{21} - I_{22}$, where

$$
\begin{aligned}
|I_{21}| = & \Bigg| P\Bigg( (1-\Delta) \int_Y^\infty S_0(u) \sum_{j=1}^K A_{0j}(u) h_{1j}(u) d[\widehat{F}_n(u) - F_0(u)] \\
& \times \left[ \frac{1}{\int_Y^\infty S_0(u) dF^\xi(u)} - \frac{1}{\int_Y^\infty S_0(u) dF_0(u)} \right] \Bigg| = o_p(\|\widehat{F}_n - F_0\|_\infty), \\
|I_{22}| = & \Bigg| P\Bigg( (1-\Delta)\Big(\int_Y^\infty S_0(u) dF_0(u)\Big)^{-2} \int_Y^\infty S_0(u) \sum_{j=1}^K A_{0j}(u) h_j(u) d[F^\xi(u) - F_0(u)] \\
& \times \int_Y^\infty S_0(u) d[\widehat{F}_n(u) - F_0(u)] \Bigg) \Bigg| = o_p(\|\widehat{F}_n - F_0\|_\infty).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
& G(\widehat{\Lambda}_n, \widehat{F}_n)[h] - G(\Lambda_0, F_0)[h] - \dot{G}_{1,\Lambda_0,\widehat{F}_n}(\widehat{\Lambda}_n - \Lambda_0)[h] - \dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h] \\
\lesssim & d_1^2(\widehat{\Lambda}_n, \Lambda_0) + o_p(\|\widehat{F}_n - F_0\|_\infty) = o_p(n^{-1/2}).
\end{aligned}
$$

(e) We show (B5). Let $\mathcal{S}_n(\Lambda, F) = G_n(\Lambda, F) + \dot{G}_{2,\Lambda,F}(\widehat{F}_n - F)$. Then $\mathcal{S}_n$ is a map from $\mathcal{U}$ to $l^\infty(\mathcal{H}_r)$. Write

$$
\psi(u; D) = \frac{S_0(u)}{\int_Y^\infty S_0(u) dF_0(u)} \quad \text{and} \quad \zeta(u; D)[h] = \sum_{j=1}^K A_j(u) h_j(u).
$$

Then

$$
\begin{aligned}
m_{12}(\Lambda_0, F_0; D)[F - F_0, h] = & (1-\Delta) \int_Y^\infty \psi(u; D) \\
& \times \left( \zeta(u; D)[h] - \int_Y^\infty \psi(s; D) \zeta(s; D)[h] dF_0(s) \right) d[F(u) - F_0(u)].
\end{aligned}
$$

Define $\overline{\zeta}(u;D)[h]=\zeta(u;D)[h]-\int_Y^\infty \psi(s;D)\zeta(s;D)[h]dF_0(s)$ and $\phi(u;D)[h]=\psi(u;D)\overline{\zeta}(u;D)[h]$. By Lemma 7, we have

$$\begin{aligned}\dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n-F_0)[h]=&P\left((1-\Delta)\int_Y^\infty \phi(u;D)[h]d[\widehat{F}_n(u)-F_0(u)]\right)\\ =&\frac{1}{n}\sum_{i=1}^n P\left((1-\Delta_i)\int_{Y_i}^\tau \frac{\tilde{\phi}(u;D)[h]}{1-H_0(u-)}dM(u;\tilde{D}_i)\right)\\ :=&\mathbb{P}_n\big(\kappa(\Lambda_0,F_0;D)[h]\big),\end{aligned}$$

where $\{\tilde{D}_i, i=1,\dots,n\}$ represents the i.i.d. sample estimating $F_0$ in Stage 1.

Recalling that $G_n(\Lambda_0,F_0)[h]=\mathbb{P}_n\big(m_1(\Lambda_0,F_0;D)[h]\big)$, it can be seen that $\mathcal{S}_n$ is a bounded Lipschitz function with respect to $\mathcal{H}_r$. Therefore, (B5) holds since $\mathcal{H}_r$ is a Donsker class. So

$$\begin{aligned}&-\sqrt{n}\dot{G}_{1,\Lambda_0,\widehat{F}_n}(\widehat{\Lambda}_n-\Lambda_0)[h]\\ &\quad=\sqrt{n}(G_n-G)(\Lambda_0,F_0)[h]+\sqrt{n}\dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n-F_0)[h]+o_p(1)\xrightarrow{d}N(0,\sigma_1^2[h]),\end{aligned}$$

where $\sigma_1^2[h]=E\big(m_1(\Lambda_0,F_0;D)[h]+\kappa(\Lambda_0,F_0;D)[h]\big)^2$. This completes the proof of Theorem 4.

## 8.6. Proof of Theorem 5

(i) Note that

$$\begin{aligned}Q_n=&\sqrt{n}\mathbb{P}_n\Big(V_n\big(\widehat{\Lambda}_1,\widehat{\Lambda}_2,h_n,\widehat{F}_n\big)\Big)\\ =&\sqrt{n}\mathbb{P}_n\Big(V_n\big(\widehat{\Lambda}_1,\Lambda_0,h_n,\widehat{F}_n\big)\Big)-\sqrt{n}\mathbb{P}_n\Big(V_n\big(\widehat{\Lambda}_2,\Lambda_0,h_n,\widehat{F}_n\big)\Big):=I_1-I_2.\end{aligned}$$

We split $I_1$ into three parts $I_1=I_{11}+I_{12}+I_{13}$, where $I_{11}=\sqrt{n}(\mathbb{P}_n-P)\Big(V_n\big(\widehat{\Lambda}_1,\Lambda_0,h_n,\widehat{F}_n\big)\Big)$, $I_{12}=\sqrt{n}P\Big(V_n\big(\widehat{\Lambda}_1,\Lambda_0,h_n-h_0,\widehat{F}_n\big)\Big)$, and $I_{13}=\sqrt{n}P\Big(V_n\big(\widehat{\Lambda}_1,\Lambda_0,h_0,\widehat{F}_n\big)\Big)$. For $I_{11}$, it can be shown that $|V_n\big(\widehat{\Lambda}_1,\Lambda_0,h,\widehat{F}_n\big)|_{P,B}\lesssim\delta$ for $h\in\mathcal{G}$ and $n^{-1/2}\lesssim O(\delta)$. From (4), we have

$$J_{[]}(\delta,\mathcal{G}_{2,\delta}(\widehat{F}_n)[h],\|\cdot\|_{P,B})=\int_0^\delta\big(1+\log N_{[]}(\epsilon,\mathcal{G}_{2,\delta}(\widehat{F}_n)[h],\|\cdot\|_{P,B})\big)^{1/2}d\epsilon\lesssim q_n^{1/2}\delta.$$

Hence, from Lemma 3.4.3 of van de Vaart and Wellner (1996),

$$\begin{aligned}&E\left\{\sup_{V_n\in\mathcal{G}_{2,\delta}(\widehat{F}_n)[h]}\big|\sqrt{n}(\mathbb{P}_n-P)V_n\Big(\widehat{\Lambda}_1,\Lambda_0,h,\widehat{F}_n\Big)\big|\right\}\\ &\lesssim J_{[]}(\delta,\mathcal{G}_{2,\delta}(\widehat{F}_n)[h],\|\cdot\|_{P,B})\big(1+J_{[]}(\delta,\mathcal{G}_{2,\delta}(\widehat{F}_n)[h],\|\cdot\|_{P,B})/(\delta^2n^{1/2})\big).\end{aligned}$$

It follows from $d_1(\widehat{\Lambda}_1, \Lambda_0) = O_p(n^{-r/(1+2r)})$ and $\|\widehat{F}_n - F_0\|_\infty = O_p(n^{-1/2})$ that

$$E\left\{\sup_{V_n \in \mathcal{G}_{2,\delta}(\widehat{F}_n)[h_n]} \left|\sqrt{n}(\mathbb{P}_n - P)V_n\left(\widehat{\Lambda}_1, \Lambda_0, h_n, \widehat{F}_n\right)\right|\right\} = o(1),$$

which yields that $I_{11} = o_p(1)$.

For $I_{12}$, by the Cauchy–Schwartz inequality and Lemma 4, we can show that

$$I_{12}^2 \lesssim n\big(d_1^2(h_n, h_0) + \|\widehat{F}_n - F_0\|_\infty d_1(h_n, h_0)\big)\big(d_1^2(\widehat{\Lambda}_1, \Lambda_0) + \|\widehat{F}_n - F_0\|_\infty d_1(\widehat{\Lambda}_1, \Lambda_0)\big) = o_p(1).$$

At last, from the proof of Theorem 4, we have

$$I_{13} = \sqrt{n}\dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h_0] + \sqrt{n}G_{n_1}(\Lambda_0, F_0)[h_0] + o_p(1).$$

Thus, $I_1 = \sqrt{n}\dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h_0] + \sqrt{n}G_{n_1}(\Lambda_0, F_0)[h_0] + o_p(1)$. Similarly, we have

$$I_2 = \sqrt{n}\dot{G}_{2,\Lambda_0,F_0}(\widehat{F}_n - F_0)[h_0] + \sqrt{n}G_{n_2}(\Lambda_0, F_0)[h_0] + o_p(1).$$

Hence, we obtain that

$$Q_n = \sqrt{\frac{n}{n_1}}\sqrt{n_1}(\mathbb{P}_{n_1} - P)m_1(\Lambda_0, F_0; D)[h_0] - \sqrt{\frac{n}{n_2}}\sqrt{n_2}(\mathbb{P}_{n_2} - P)m_1(\Lambda_0, F_0; D)[h_0] + o_p(1),$$

where $\mathbb{P}_{n_r}$ is the empirical measure based on the sample from group $r, r = 1, 2$. Since $\mathbb{P}_{n_1}$ and $\mathbb{P}_{n_2}$ are independent, it follows that $Q_n$ converges in distribution to $N(0, \sigma_2^2)$.

(ii) To show that $\widehat{\sigma}_0^2 - \sigma_0^2 = o_p(1)$, we note that $\sigma_0^2 = P(m_1^2(\Lambda_0, F_0; D)[h_0])$ and $\widehat{\sigma}_0^2 = \mathbb{P}_n(m_1^2(\widehat{\Lambda}_0, \widehat{F}_n; D)[h_n])$. Then

$$\begin{aligned}\widehat{\sigma}_0^2 - \sigma_0^2 =& \mathbb{P}_n\big(m_1^2(\widehat{\Lambda}_0, \widehat{F}_n; D)[h_n] - m_1^2(\Lambda_0, F_0; D)[h_n]\big) \\ &+ \mathbb{P}_n\big(m_1^2(\Lambda_0, F_0; D)[h_n] - m_1^2(\Lambda_0, F_0; D)[h_0]\big) \\ &+ (\mathbb{P}_n - P)m_1^2(\Lambda_0, F_0; D)[h_0] := I_1 + I_2 + I_3.\end{aligned}$$

It can be easily shown that $I_1 = o_p(1)$ and $I_3 = o_p(1)$. We now consider $I_2$. By Conditions (C1), (C3), (C7) and (C8), we have

$$\begin{aligned}&\big|m_1(\Lambda_0, F_0; D)[h_n] - m_1(\Lambda_0, F_0; D)[h_0]\big| = \big|m_1(\Lambda_0, F_0; D)[h_n - h_0]\big| \\ &\le c_1(N(T_{K,K}) + 1)\left(\Delta\sum_{j=1}^K |h_{n,j}(Y) - h_{0,j}(Y)| + (1 - \Delta)\frac{\sum_{j=1}^K \int_Y^\infty S_0(u)|h_{n,j}(u) - h_{0,j}(u)|dF_0(u)}{\int_Y^\infty S_0(u)dF_0(u)}\right)\end{aligned}$$

with probability 1 for some constant $c_1$, and

$$\big|m_1(\Lambda_0, F_0; D)[h_n] + m_1(\Lambda_0, F_0; D)[h_0]\big| = \big|m_1(\Lambda_0, F_0; D)[h_n + h_0]\big| \le c_2 K(N(T_{K,K}) + 1)$$

with probability 1 for some constant $c_2$. Thus by Conditions (C2) and (C3),

$$E\big|m_1^2(\Lambda_0, F_0; D)[h_n]) - m_1^2(\Lambda_0, F_0; D)[h_0]\big|$$

$$\leq c\Big(E(N(T_{K,K})+1)^4\Big)^{1/2}\left(E\sum_{j=1}^{K}\big(h_{n,j}(U)-h_{0,j}(U)\big)^2\right)^{1/2}=o(n^{-1/2(1+2r)})\to 0.$$

This completes the proof of Theorem 5.

## Acknowledgments

The authors would like to thank the two referees, the Associate Editor and Editor-in-Chief (Professor Mark Podolskij) for their constructive and insightful comments and suggestions that greatly improved the paper. This research was supported in part by the Research Grants Council of Hong Kong (15301218, 17308420), the Natural Science Foundation of China (11771366, 12171374, 11971362), and the CAS AMSS-PolyU Joint Laboratory of Applied Mathematics.

## Supplementary Materials

The Supplement Materials contain proofs of Lemmas 1–5.

## References

[1] Akritas, M. (2000). The central limit theorem under censoring. *Bernoulli* **6**, 1109–1120.
[2] Andersen, P. K., Borgan, Ø, Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, Berlin.
[3] 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.
[4] Balakrishnan, N. and Zhao, X. (2009). New multi-sample nonparametric tests for panel count data. *Ann. Statist.* **37**, 1112–1149.
[5] Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Appl. Statist.* **33**, 38–44.
[6] Chan, K-C. and Wang, M-C. (2010). Backward estimation of stochastic processes with failure events as time origins. *Ann. Appl. Statist.* **4**, 1602–1620.
[7] Chan, K-C. and Wang, M-C. (2017). Semiparametric modeling and estimation of the terminal behavior of recurrent marker processes before failure events. *J. Am. Statist. Assoc.* **112**, 351–362.
[8] Cheng, S. C. and Wei, L. J. (2000). Inferences for a semiparametric model with panel data. *Biometrika* **87**, 89–97.
[9] Diao, G., Zeng, D., Hu, K. and Ibrahimb, J.G. (2017). Modeling event count data in the presence of informative dropout with application to bleeding and transfusion events in myelodysplastic syndrome. *Statist. Med.* **36**, 3475–3494.
[10] Hinde. J. (1982). Compound Poisson regression models. In *GLIM 82: Proceedings of the International Conference in Generalized Linear Models*, R. Gilchrist (ed), pp. 109–121, Berlin: Springer-Verlag.
[11] Hu X. J., Lagakos, S. W. and Lockhart, R. A. (2009). Generalized least squares estimation of the mean function of a counting process based on panel counts. *Statist. Sinica* **19**, 561–580.
[12] Hu, X. J., Sun, J. and Wei, L. J. (2003). Regression parameter estimation from panel counts. *Scand. J. Statist.* **30**, 25–43.
[13] Huang, C.-Y. and Wang, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Am. Statist. Assoc.* **99**, 1153–1165

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

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

[16] Kong, S. C. and Nan, B. (2016). Semiparametric approach to regression with a covariate subject to a detection limit. *Biometrika* **103**, 161–174.

[17] Kong, S. C., Nan, B., Kalbfleisch, J. D., Sarah, R., and Hirth, R. (2018). Conditional modeling of longitudinal data with terminal event. *J. Am. Statist. Assoc.* **113**, 357-368.

[18] Li, Z., Tosteson, TD., and Bakitas, MA. (2013). Joint modeling quality of life and survival using a terminal decline model in palliative care studies. *Statist. Med.* **32**, 1394-1406.

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

[20] Lu, M., Zhang, Y. and Huang, J. (2009). Semiparametric estimation methods for panel count cata using monotone B-splines. *J. Am. Statist. Assoc.* **113**, 294–305.

[21] Ma, L. and Sundaram, R. (2018). Analysis of Gap Times Based on Panel Count Data With Informative Observation Times and Unknown Start Time. *J. Am. Statist. Assoc.* **104**, 1060–1070.

[22] Schumaker, L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press.

[23] Sun, J. and Fang, H. B. (2003). A nonparametric test for panel count data. *Biometrika* **90**, 199–208.

[24] Sun, J. and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel data. *Statist. Sinica* **5**, 279–289.

[25] Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. R. Statist. Soc. Ser. B* **62**, 293–302.

[26] Sun, L., Song, X., Zhou, J. and Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *J. Am. Statist. Assoc.* **107**, 688–700.

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

[28] Thall, P. F. and Lachin, J. M. (1988). Analysis of recurrent events: nonparametric methods for random-interval count data. *J. Am. Statist. Assoc.* **83**, 339–347.

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

[30] 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.

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

[32] Zeng, D. and Cai, J. (2010). A semiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika* **97**, 699–712.

[33] Zeng, Y., Vaupel, J., Xiao, Z., Liu, Y. and Zhang, C. (2017). Chinese Longitudinal Healthy Longevity Survey (CLHLS), 1998-2014. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2017-04-11. https://doi.org/10.3886/ICPSR36692.v1

[34] Zhang, Y. (2006). Nonparametric k-sample tests for panel count data. *Biometrika* **93**, 777–790.

[35] Zhang, Y. and Jamshidian, M. (2003). The gamma-frailty Poisson model for the nonparametric estimation of panel count data. *Biometrics* **59**, 1099–1106.

[36] Zhao, H., Li, Y. and Sun, J. (2013). Analyzing panel count data with a dependent observation process and a terminal event. *Can. J. Statist.* **41**, 174–191.

[37] Zhao, H., Zhang, Y., Zhao, X. and Yu, Z. (2019). A nonparametric regression model for panel count data analysis. *Statist. Sinica 29*, 809–826.

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

[39] Zhao, X., Zhou, J. and Sun, L. (2011). Semiparametric transformation models with time-varying coefficients for recurrent and terminal events. *Biometrics* **67**, 404–414.

[40] Zhou, J., Zhang, H., Sun, L. and Sun, J. (2017). Joint analysis of panel count data with an informative observation process and a dependent terminal event. *Lifetime Data Anal.* **23**, 560–584.

*Submitted to Bernoulli*

# Nonparametric Inference for Reversed Mean Models with Panel Count Data

LI LIU[1] WEN SU[2] GUOSHENG YIN[2] XINGQIU ZHAO[3] and YING ZHANG[4]

[1]*School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, 430072, China*
*E-mail:* `lliu.math@whu.edu.cn`
[2]*Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong*
*E-mail:* `jenna.wen.su@connect.hku.hk`
*E-mail:* `gyin@hku.hk; Corresponding author`
[3]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong*
*E-mail:* `xingqiu.zhao@polyu.edu.hk`
[4]*Department of Biostatistics, University of Nebraska Medical Center, Omaha, NE, USA*
*E-mail:* `ying.zhang@unmc.edu`

## Supplementary Materials

## Appendix A: Proofs of Lemmas 1–5

*Proof of Lemma 1*

By Jensen's inequality, we have

$$\begin{aligned} Pm(\Lambda_0, F_0; D) - Pm(\Lambda, F_0; D) =& P(-\log(L(\Lambda, F_0; D)/L(\Lambda_0, F_0; D))) \\ \geq& -\log P(L(\Lambda, F_0; D)/L(\Lambda_0, F_0; D)) \\ =& 0. \end{aligned} \tag{A.1}$$

This equality holds if and only if $L(\Lambda, F_0; D) = L(\Lambda_0, F_0; D)$ a.e. $P$. By Jensen's inequality, we have

$$\begin{aligned} & P\big(L(\Lambda, F_0; D)/L(\Lambda_0, F_0; D)\big) \\ =& P\big(E_{\underline{\triangle N}|\Delta, Y, \underline{T}, K}\big[L(\Lambda, F_0; D)/L(\Lambda_0, F_0; D)\big]\big) \\ =& P\Bigg(E_{\underline{\triangle N}|\Delta, Y, \underline{T}, K}\Bigg[\Big(\exp\Big\{\sum_{j=1}^{K}\Big[\triangle N_j \log\Big(\frac{\triangle\Lambda_j(Y)}{\triangle\Lambda_{0j}(Y)}\Big) - \big(\triangle\Lambda_j(Y) - \triangle\Lambda_{0j}(Y)\big)\Big]\Big\}\Big)^{\Delta} \\ & \times \Big(\int_Y^\infty S(u)dF_0(u)\Big/\int_Y^\infty S_0(u)dF_0(u)\Big)^{1-\Delta}\Bigg]\Bigg) \\ \geq& P\Bigg(\Bigg(\exp\Big\{\sum_{j=1}^{K}\Big[\triangle\Lambda_{0j}(Y)\log\Big(\frac{\triangle\Lambda_j(Y)}{\triangle\Lambda_{0j}(Y)}\Big) - \big(\triangle\Lambda_j(Y) - \triangle\Lambda_{0j}(Y)\big)\Big]\Big\}\Bigg)^{\Delta} \end{aligned}$$

*The first two authors contribute equally to this work.

$$\times \left( \frac{1}{\overline{F}_0(Y)} \int_Y^\infty \exp\Big\{ \sum_{j=1}^K \Big[ \triangle\Lambda_{0j}(Y) \log\Big( \frac{\triangle\Lambda_j(Y)}{\triangle\Lambda_{0j}(Y)} \Big) - \big( \triangle\Lambda_j(Y) - \triangle\Lambda_{0j}(Y) \big) \Big] \Big\} dF_0(u) \right)^{1-\Delta} \Bigg)$$

$$= P\Bigg( \Delta \exp\Big\{ - \sum_{j=1}^K \Big[ \triangle\Lambda_j(Y) \phi\Big( \frac{\triangle\Lambda_{0j}(Y)}{\triangle\Lambda_j(Y)} \Big) \Big] \Big\} + \frac{1-\Delta}{\overline{F}_0(Y)} \int_Y^\infty \exp\Big\{ - \sum_{j=1}^K \Big[ \triangle\Lambda_j(u) \phi\Big( \frac{\triangle\Lambda_{0j}(u)}{\triangle\Lambda_j(u)} \Big) \Big] \Big\} dF_0(u) \Bigg) \Bigg\}$$

$$= \int \exp\Big\{ - \Lambda(u_1, u_2) \phi\Big( \frac{\Lambda_0(u_1, u_2)}{\Lambda(u_1, u_2)} \Big) \Big\} d\mu_1(u_1, u_2), \tag{A.2}$$

where $\Lambda(u_1, u_2) = \Lambda(u_1) - \Lambda(u_2)$, and $\phi(x) = x\log(x) - x + 1 \geq 0$ with equality if and only if $x = 1$. Since the equality in (A.2) holds if and only if $L(\Lambda, F_0; D) = L(\Lambda_0, F_0; D)$ a.e. $P$, which is equivalent to $\int \exp\Big\{ - \Lambda(u_1, u_2)\phi\Big(\frac{\Lambda_0(u_1,u_2)}{\Lambda(u_1,u_2)}\Big)\Big\} d\mu_1(u_1, u_2) = 1$, i.e., $\Lambda = \Lambda_0$ a.e. with respect to $\mu_1$. Combining this with (A.1) yields that $Pm(\Lambda, F_0; D)$ has the unique maximizer at $\Lambda = \Lambda_0$ a.e. with respect to $\mu_1$ or $d_1(\Lambda, \Lambda_0) = 0$. In addition, noting the relation that $d_1(\Lambda_1, \Lambda_2)/2 \leq d_2(\Lambda_1, \Lambda_2) \leq M_1 d_1(\Lambda_1, \Lambda_2)$ by (4.5) in Wellner and Zhang (2000) and $P(K \leq M) = 1$ for $M = \tau/s_0$ from Condition (C7), the metrics $d_1$ and $d_2$ are equivalent. Therefore, $Pm(\Lambda_0, F_0; D) - Pm(\Lambda, F_0; D) \geq 0$ with equality if and only if $\Lambda = \Lambda_0$ a.e. with respect to $\mu_2$. This completes the proof of Lemma 1.

*Proof of Lemma 2*

First, we claim that the class $\{\Lambda : \Lambda \in \Psi, \Lambda \text{ is uniformly bounded}\}$ is Donsker by noting that $\Psi$ is the monotone and uniformly bounded functional class. As $\omega_j(u) = \triangle N_j \log \triangle\Lambda_j(u) - \triangle\Lambda_j(u)$ is Lipschitz and square integrable on $[0, \tau)$ from (C2), (C7) and (C8), it belongs to the Donsker class by Theorem 2.10.6 of van der Vaart and Wellner (1996). By $P(K \leq M) = 1$ for $M = \tau/s_0$ from Condition (C7), both $\sum_{j=1}^K \omega_j(u)$ and $S(u)$ belong to the Donsker class by Theorem 2.10.6 of van der Vaart and Wellner (1996). By Theorem 2.10.3 of van der Vaart and Wellner (1996), $\{\int_Y^\infty S(u)dF(u), \Lambda \in \Psi, F \in \mathcal{F}\}$ is Donsker. Conditions (C1)-(C3) and (C7)-(C8) ensure that $\log \int_Y^\infty S(u)dF(u) < \infty$, which yields that $\{\log \int_Y^\infty S(u)dF(u), \Lambda \in \Psi, F \in \mathcal{F}\}$ is Donsker from Theorem 2.10.6 of van der Vaart and Wellner (1996). This completes the proof of Lemma 2.

*Proof of Lemma 3*

We only show that (2) of Lemma 3 holds by two steps as (3) and (4) of Lemma 3 can be obtained similarly.

Step (i) We construct the bracketing set of the class $\mathcal{G}_\delta(F)$.

Under Condition (C6), it follows from Shen and Wong (1994) that there exists a set of brackets $\{[\Lambda_v^l, \Lambda_v^r] : d_1(\Lambda_v^r, \Lambda_v^l) \leq \epsilon, v = 1, \ldots, (\delta/\epsilon)^{c_1 q_n}\}$ to cover $\Phi_n$ for any $\epsilon < \delta$.

Next, we show that for sufficiently small $\epsilon > 0$ and $\delta > 0$, there exist positive constants $\gamma_1$ and $\gamma_2$ such that $\triangle\Lambda_{vj}^r(u) - \triangle\Lambda_{vj}^l(u) \leq \gamma_1$ and $\triangle\Lambda_{vj}^l(u) \geq \gamma_2$ for all $u - T_{k,j} \in [0, \tau]$ and $v = 1, \ldots, (\delta/\epsilon)^{c_1 q_n}$.

In fact, for any $v = 1, \ldots, (\delta/\epsilon)^{c_1 q_n}$, there is $\Lambda \in \Psi_{n\delta}$ such that $d_1(\Lambda_v^r, \Lambda) \leq \epsilon$ and $d_1(\Lambda, \Lambda_v^l) \leq \epsilon$, which implies that $d_1(\Lambda_v^r, \Lambda_0) \leq \epsilon_1$ ($\epsilon_1 = \sqrt{\epsilon^2 + \delta^2}$) and $d_1(\Lambda_v^l, \Lambda_0) \leq \epsilon_1$. By $d_1(\Lambda_1, \Lambda_2)/2 \leq d_2(\Lambda_1, \Lambda_2) \leq M_1 d_1(\Lambda_1, \Lambda_2)$ and the boundness of $K$, the metrics $d_1$ and $d_2$ are equivalent. Thus, for any $\Lambda \in \Psi_{n\delta}$, we have $d_2(\Lambda_v^r, \Lambda_0) \leq M_1\epsilon_1$ and $d_2(\Lambda_v^l, \Lambda_0) \leq M_1\epsilon_1$. Then Lemma 7.1 of Wellner and Zhang (2007) indicates that $\sup_{u\in[0,\tau]} |\Lambda_v^r(u) - \Lambda_0(u)| \leq \epsilon_2$ and $\sup_{u\in[0,\tau]} |\Lambda_v^l(u) - \Lambda_0(u)| \leq \epsilon_2$ for a sufficiently small $\epsilon_2 > 0$ ($\epsilon_2$ can be taken as $(M_1\epsilon_1/c)^{2/3}$ in view of Lemma 7.1 in Wellner and Zhang (2007)). It follows that $\triangle\Lambda_{vj}^r(u) \leq \triangle\Lambda_{0j}(u) + 2\epsilon_2$ and $\triangle\Lambda_{vj}^l(u) \geq \triangle\Lambda_{0j}(u) - 2\epsilon_2$. Therefore, taking $\gamma_1 = 4\epsilon_2$ and $\gamma_2 = s_0/M_0 - 2\epsilon_2$, where $s_0$ and $M_0$ are defined as in (C7) and (C8), we have $\triangle\Lambda_{vj}^r(u) - \triangle\Lambda_{vj}^l(u) \leq \gamma_1$ and $\triangle\Lambda_{vj}^l(u) \geq \gamma_2$.

Define

$$A_{1,v}^l = \sum_{j=1}^{K} \left[\triangle N_j\{\log\triangle\Lambda_{vj}^l(Y) - \log\triangle\Lambda_{0j}(Y)\} - \{\triangle\Lambda_{vj}^r(Y) - \triangle\Lambda_{0j}(Y)\}\right],$$

$$A_{1,v}^r = \sum_{j=1}^{K} \left[\triangle N_j\{\log\triangle\Lambda_{vj}^r(Y) - \log\triangle\Lambda_{0j}(Y)\} - \{\triangle\Lambda_{vj}^l(Y) - \triangle\Lambda_{0j}(Y)\}\right],$$

and write $\omega_{0,j}(u) = \triangle N_j \log\triangle\Lambda_{0j}(u) - \triangle\Lambda_{0j}(u)$.

It can be seen that $\sum_{j=1}^{K}\{\omega_j(Y) - \omega_{0j}(Y)\} \in [A_{1,v}^l, A_{1,v}^r]$ for some $v = 1, 2, \ldots, (\delta/\epsilon)^{c_1 q_n}$. Taking

$$\tilde{A}_v^l(u) = \sum_{j=1}^{K} \left\{\triangle N_j \log\triangle\Lambda_{vj}^l(u) - \triangle\Lambda_{vj}^r(u) - \log\triangle N_j!\right\},$$

$$\tilde{A}_v^r(u) = \sum_{j=1}^{K} \left\{\triangle N_j \log\triangle\Lambda_{vj}^r(u) - \triangle\Lambda_{vj}^l(u) - \log\triangle N_j!\right\}),$$

$$A_{2,v}^l = \log\left(\int_Y^\infty \exp\{\tilde{A}_v^l(u)\}dF(u)\right) - \log\left(\int_Y^\infty S_0(u)dF(u)\right),$$

$$A_{2,v}^r = \log\left(\int_Y^\infty \exp(\tilde{A}_v^r(u))dF(u)\right) - \log\left(\int_Y^\infty S_0(u)dF(u)\right),$$

we have $\log(\int_Y^\infty S(u)dF(u)) - \log(\int_Y^\infty S_0(u)dF(u)) \in [A_{2,v}^l, A_{2,v}^r]$ for some $v = 1, 2, \ldots, (\delta/\epsilon)^{c_1 q_n}$. Therefore, $m(\Lambda, F; D) - m(\Lambda_0, F; D) \in [L_v^l, L_v^r]$ for some $v = 1, 2, \ldots, (\delta/\epsilon)^{c_1 q_n}$, where

$$L_v^l = \Delta A_{1,v}^l + (1-\Delta)A_{2,v}^l, \quad L_v^r = \Delta A_{1,v}^r + (1-\Delta)A_{2,v}^r. \tag{A.3}$$

(ii) We show that $\|L_v^r - L_v^l\|_{P,B}^2 \lesssim \epsilon^2$, where $L_v^l$ and $L_v^r$ are defined as in (A.3).

Note that

$$|L_v^r - L_v^l| \leq \Delta|A_{1,v}^r - A_{1,v}^l| + (1-\Delta)|A_{2,v}^r - A_{2,v}^l| := I_1 + I_2,$$

where

$$I_1 = \Delta\Big|\sum_{j=1}^{K} \left(\triangle N_j[\log\triangle\Lambda_{vj}^r(Y) - \log\triangle\Lambda_{vj}^l(Y)] + [\triangle\Lambda_{vj}^r(Y) - \triangle\Lambda_{vj}^l(Y)]\right)\Big|.$$

Since $\log y = \log x + (x + \xi(y-x))^{-1}(y-x)$ for $0 < x \le y$ and $\xi \in (0,1)$, we have

$$\log \triangle\Lambda_{vj}^r \le \log \triangle\Lambda_{vj}^l + \gamma_2^{-1}(\triangle\Lambda_{vj}^r - \triangle\Lambda_{vj}^l).$$

Taking $\epsilon_2 \le s_0/(4M_0)$, it yields that

$$\begin{aligned} I_1 \le & \Delta \sum_{j=1}^K \Big( \triangle N_j |\log \triangle\Lambda_{vj}^r(Y) - \log \triangle\Lambda_{vj}^l(Y)| + |\triangle\Lambda_{vj}^r(Y) - \triangle\Lambda_{vj}^l(Y)| \Big) \\ \le & c(N(T_{K,K})+1)\Delta \sum_{j=1}^K |\triangle\Lambda_{vj}^r(Y) - \triangle\Lambda_{vj}^l(Y)|. \end{aligned}$$

Similarly, using Condition (C3), we can show that

$$\begin{aligned} I_2 \le & (1-\Delta)|A_{2,v}^r - A_{2,v}^l| \\ \le & c(N(T_{K,K})+1)\frac{1-\Delta}{\overline{F}(Y)} \int_Y^\infty \sum_{j=1}^K |\triangle\Lambda_{vj}^r(u) - \triangle\Lambda_{vj}^l(u)| dF(u). \end{aligned}$$

Therefore, by Condition (C3), we obtain that

$$\begin{aligned} \|L_v^r - L_v^l\|_{P,B}^2 \le & P\Big( |L_v^r - L_v^l|^2 \cdot e^{|L_v^r - L_v^l|} \Big) \\ \lesssim & P\Bigg( e^{cN(T_{K,K})}(N(T_{K,K})^2+1) \Bigg[ \Delta \sum_{j=1}^K |\triangle\Lambda_{vj}^r(Y) - \triangle\Lambda_{vj}^l(Y)|^2 \\ & + \frac{1-\Delta}{\overline{F}(Y)} \int_Y^\infty \sum_{j=1}^K |\triangle\Lambda_{vj}^r(u) - \triangle\Lambda_{vj}^l(u)|^2 dF(u) \Bigg] \Bigg) \\ \lesssim & \epsilon^2. \end{aligned}$$

Thus, it follows from (i) and (ii) that

$$\log N_{[]}(\epsilon, \mathcal{L}_{\eta,\delta}, \|\cdot\|_{P,B}) \le c q_n \log(\delta/\epsilon).$$

This completes the proof of Lemma 3.

*Proof of Lemma 4*

(i) Using integration by parts, we have

$$\begin{aligned} & \left( \int_y^\infty \phi(x) d(F(x) - F_0(x)) \right)^2 \\ = & \left( \Big[ \phi(x)(F(x) - F_0(x)) \Big]_y^\infty - \int_y^\infty \dot\phi(x)(F(x) - F_0(x)) dx \right)^2 \end{aligned}$$

$$
\begin{aligned}
&\leq\bigg(|\phi(y)(F(y)-F_0(y))|+\int_y^\infty|\dot\phi(x)|dx\cdot\|F(x)-F_0(x)\|_\infty\bigg)^2\\
&=\bigg(|\phi(y)(F(y)-F_0(y))|+\int_y^\infty\frac{|\dot\phi(x)|}{f_0(x)}dF_0(x)\cdot\|F(x)-F_0(x)\|_\infty\bigg)^2\\
&\lesssim\int_y^\infty\dot\phi^2(x)dF_0(x)\cdot\|F-F_0\|_\infty^2+\phi^2(y)\|F-F_0\|_\infty^2
\end{aligned}
$$

by Condition (C4).

(ii) Using the fact that

$$\log(x)-\log(x_0)=\frac{h}{x_0}-\frac{h^2}{(x^\xi)^2}$$

for $h=x-x_0$ and $x^\xi=x_0+\xi h$ with some $\xi\in(0,1)$, we have

$$
\begin{aligned}
&P\big(m(\Lambda,F;D)-m(\Lambda,F_0;D)\big)\\
=&P\bigg((1-\Delta)\frac{\int_Y^\infty S(u)d[F(u)-F_0(u)]}{\int_Y^\infty S(u)dF_0(u)}\bigg)-P\bigg((1-\Delta)\bigg(\frac{\int_Y^\infty S(u)d[F(u)-F_0(u)]}{\int_Y^\infty S(u)dF^\xi(u)}\bigg)^2\bigg)\\
:=&I_1-I_2,
\end{aligned}
$$

where $F^\xi=F_0+\xi(F-F_0)$ for some $\xi\in(0,1)$. For $I_1$, we set

$$g(t)=\frac{\int_Y^\infty\exp\Big(\sum_{j=1}^K\Big[\triangle N_j\log(\triangle\Lambda_{0j}+th_j)-(\triangle\Lambda_{0j}+th_j)-\log\triangle N_j!\Big]\Big)d[F(u)-F_0(u)]}{\int_Y^\infty\exp\Big(\sum_{j=1}^K\Big[\triangle N_j\log(\triangle\Lambda_{0j}+th_j)-(\triangle\Lambda_{0j}+th_j)-\log\triangle N_j!\Big]\Big)dF_0(u)}$$

with $h_j=\triangle\Lambda_j-\triangle\Lambda_{0j}$. Then $I_1=P((1-\Delta)g(0))+P((1-\Delta)\dot g(\xi))$ for some $\xi\in(0,1)$. We first note that $P((1-\Delta)g(0))=P\Big((1-\Delta)[F_0(Y)-F(Y)]/\overline{F}_0(Y)\Big)$. In fact, using $\sum_{\underline{\triangle N}}$ to express the summation regarding to $\underline{\triangle N}$ running through its all possible values, we have

$$
\begin{aligned}
P\Big((1-\Delta)g(0)\Big)=&P\Big((1-\Delta)E_{\underline{\triangle N}|\Delta=0,Y,\underline{T},K}(g(0))\Big)\\
=&P\left((1-\Delta)\frac{\int_Y^\infty\sum_{\underline{\triangle N}}S_0(u)d[F(u)-F_0(u)]}{\overline{F}_0(Y)}\right)\\
=&P\bigg((1-\Delta)\frac{\int_Y^\infty d[F(u)-F_0(u)]}{\overline{F}_0(Y)}\bigg)\\
=&P\bigg((1-\Delta)\frac{F_0(Y)-F(Y)}{\overline{F}_0(Y)}\bigg).
\end{aligned}
$$

Moreover,

$$\begin{aligned}
\dot{g}(\xi) =& \frac{\int_Y^\infty S_\xi(u) \sum_{j=1}^K A_{\xi,j}(u) h_j(u) d[F(u) - F_0(u)]}{\int_Y^\infty S_\xi(u) dF_0(u)} \\
&- \frac{\int_Y^\infty S_\xi(u) d[F(u) - F_0(u)] \cdot \int_Y^\infty S_\xi(u) \sum_{j=1}^K A_{\xi,j}(u) h_j(u) dF_0(u)}{\Big(\int_Y^\infty S_\xi(u) dF_0(u)\Big)^2} \\
:=& I_{11} - I_{12},
\end{aligned}$$

where $\triangle\Lambda_j^\xi = \triangle\Lambda_{0j} + \xi(\triangle\Lambda_j - \triangle\Lambda_{0j})$, and $A_{\xi,j}(u)$ and $S_\xi(u)$ represent the values of $A_j(u)$ and $S(u)$ at $\triangle\Lambda_j(u) = \triangle\Lambda_{0j}(u) + \xi h_j$ for $\xi \in (0,1)$, respectively. We use these notations in the sequel proofs as well. Then, using the conclusion of part (i) and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
|P((1-\Delta)I_{11})| = & \Bigg| P\Bigg( (1-\Delta) \int_Y^\infty [(S_\xi(u) - S_0(u)) + S_0(u)] \sum_{j=1}^K A_{\xi,j}(u) h_j(u) \\
& \times d[F(u) - F_0(u)] \frac{1}{\int_Y^\infty S_0(u) dF_0(u)} \cdot \frac{\int_Y^\infty S_0(u) dF_0(u)}{\int_Y^\infty S_\xi(u) dF_0(u)} \Bigg) \Bigg| \\
= & \Bigg| P\Bigg( \Bigg\{ (1-\Delta) \int_Y^\infty \xi S_{\xi_*}(u) \Big[ \sum_{j=1}^K A_{\xi_*,j}(u) h_j(u) \Big] \\
& \times \Big[ \sum_{j=1}^K A_{\xi,j}(u) h_j(u) \Big] d[F(u) - F_0(u)] \cdot \frac{1}{\int_Y^\infty S_0(u) dF_0(u)} \\
& + (1-\Delta) \int_Y^\infty S_0(u) \sum_{j=1}^K A_{\xi,j}(u) h_j(u) d[F(u) - F_0(u)] \frac{1}{\int_Y^\infty S_0(u) dF_0(u)} \Bigg\} \Bigg) (1 + o_p(1)) \Bigg| \\
= & 2\xi \Bigg| P\Bigg( \frac{1-\Delta}{\overline{F}_0(Y)} \int_Y^\infty \sum_{j=1}^K \frac{h_j^2(u)}{\triangle\Lambda_{0j}(u)} d[F(u) - F_0(u)] \Bigg) (1 + o_p(1)) \Bigg| \\
\lesssim & d_1(\Lambda, \Lambda_0) \|F - F_0\|_\infty,
\end{aligned}$$

where $\triangle\Lambda_j^{\xi_*}(u)$, $A_{\xi_*,j}(u)$ and $S_{\xi_*}(u)$ are defined similarly as $\triangle\Lambda_j^\xi(u)$, $A_{\xi,j}(u)$ and $S_\xi(u)$ for some $\xi_* \in (0,\xi)$. Similar to the proof of $I_{11}$, it can be seen that $|P((1-\Delta)I_{12})| \lesssim d_1(\Lambda,\Lambda_0)\|F - F_0\|_\infty$ and $|P((1-\Delta)I_2)| \lesssim \|F - F_0\|_\infty^2$. Thus,

$$\begin{aligned}
& \sup_{\Lambda \in \Psi_\delta^0} \Big| P\Big( m(\Lambda, F; D) - m(\Lambda, F_0; D) - (1-\Delta) \frac{F_0(Y) - F(Y)}{\overline{F}_0(Y)} \Big) \Big| \\
\leq & |P((1-\Delta)I_{11})| + |P((1-\Delta)I_{12})| + |P((1-\Delta)I_2)| \\
\lesssim & d_1(\Lambda, \Lambda_0) \|F - F_0\|_\infty + \|F - F_0\|_\infty^2.
\end{aligned}$$

This completes the proof of Lemma 4.

*Proof of Lemma 5*

The lemma follows from Theorem 5.55 of van der Vaart (1998). For each $n \in \mathbb{N}$, $j \in \mathbb{Z}$ and $M > 0$, let $S_{n,j,M}$ be the set

$$\{\Lambda \in \Psi_n : 2^{j-1} r_n < d_1(\Lambda, \Lambda_0) \leq 2^j r_n, d(F, F_0) \leq 2^{-M} d_1(\Lambda, \Lambda_0)\}.$$

Then the intersection of the events $\hat{\Lambda}_n \in \Psi_n$ and $d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^M (r_n + d(F, F_0))$ is contained in the union of the events $\{\hat{\Lambda}_n \in S_{n,j,M}\}$ over $j \geq M$. By the definition of $\hat{\Lambda}_n$, the supremum of $\mathbb{P}_n(m(\Lambda, F; D) - m(\Lambda_0, F; \underline{D}))$ over the set of parameters $\Lambda \in S_{n,j,M}$ is not less than $-R_n = -O_p(r_n^2)$ on the event $\{\hat{\Lambda}_n \in S_{n,j,M}\}$. So we conclude that for some constant $c$,

$$\begin{aligned}
&P(\{d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^M (r_n + d(F, F_0)), \hat{\Lambda}_n \in \Psi_n\}) \\
\leq &\sum_{j \geq M, 2^j \leq \delta / r_n} P\left( \sup_{\Lambda \in S_{n,j,M}} \mathbb{P}_n(m(\Lambda, F; D) - m(\Lambda_0, F; D)) \geq -c r_n^2 \right) \\
&+ P(d_1(\hat{\Lambda}_n, \Lambda_0) \geq \delta) + P(R_n \geq c r_n^2). \qquad \text{(A.4)}
\end{aligned}$$

Since $\hat{\Lambda}_n$ is consistent for $\Lambda_0$ and $R_n = O_p(r_n^2)$, the last two terms on the right side converge to 0 as $n \to \infty$ for every $\delta > 0$. Then for every $j$ involved in the sum, we have for every $\Lambda \in S_{n,j,M}$ and sufficiently large $M$,

$$\begin{aligned}
P(m(\Lambda, F; D) - m(\Lambda_0, F; D)) &\lesssim -d_1^2(\Lambda, \Lambda_0) + d^2(F, F_0) + d_1(\Lambda, \Lambda_0) d(F, F_0) \\
&\lesssim -(1 - 2^{-M} - 2^{-2M}) d_1^2(\Lambda, \Lambda_0) \leq -c_1 2^{2j} r_n^2
\end{aligned}$$

for some constant $c_1$. Taking large enough $M$ such that $c \leq c_1 2^{2M-1}$ and using the Markov's inequality, we have for $j \geq M$,

$$\begin{aligned}
&P\Big( \sup_{\Lambda \in S_{n,j,M}} \mathbb{P}_n(m(\Lambda, F; D) - m(\Lambda_0, F; D)) \geq -c r_n^2 \Big) \\
\leq &P\Big( \sup_{\Lambda \in S_{n,j,M}} (\mathbb{P}_n - P)(m(\Lambda, F; D) - m(\Lambda_0, F; D)) \geq c_1 2^{2j-1} r_n^2 \Big) \\
\leq &\frac{1}{c_1 2^{2j-1} r_n^2} E \sup_{\Lambda \in S_{n,j,M}} |(\mathbb{P}_n - P)(m(\Lambda, F; D) - m(\Lambda_0, F; D))| \\
\leq &\frac{\phi_n(2^j r_n)}{c_1 2^{2j-1} r_n^2 \sqrt{n}}. \qquad \text{(A.5)}
\end{aligned}$$

Since $\phi_n(\delta)/\delta^\beta$ is decreasing for some $\beta < 2$, then $\phi_n(c_2 r_n) \leq c_2^\beta \phi_n(r_n) \leq c_2^\beta \sqrt{n} r_n^2$ for $c_2 > 1$. Therefore,

$$\frac{\phi_n(2^j r_n)}{c_1 2^{2j-1} r_n^2 \sqrt{n}} \leq \frac{c_3 2^{\beta j} \sqrt{n} r_n^2}{2^{2j} r_n^2 \sqrt{n}} \leq c_3 2^{-(2-\beta) j}. \qquad \text{(A.6)}$$

Equations (A.4)–(A.6) yield that

$$P(d_1(\widehat{\Lambda}_n, \Lambda_0) \geq 2^M(r_n + d(F, F_0))) \leq \sum_{j \geq M} c_3 2^{-(2-\beta)j}$$

as $M \to \infty$, which means that

$$d_1(\widehat{\Lambda}_n, \Lambda_0) = O_p(r_n + d(F, F_0)).$$

This completes the proof of Lemma 5.

# References

[1] Shen, X. and Wong, W.H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **14**, 590–606.
[2] van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge.
[3] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.
[4] 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.
[5] Wellner, J. A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.* **35**, 2106–2142.