Semiparametric Reversed Mean Model for Recurrent Event Process with Informative Terminal Event

Wen Su[1*], Li Liu[2*], Guosheng Yin[1], Xingqiu Zhao[3] and Ying Zhang[4]

[1]*Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong*

[2]*School of Mathematics and Statistics, Wuhan University, Wuhan, China*

[3]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong* [4]*Department of Biostatistics, University of Nebraska Medical Center, Omaha, NE, USA*

Supplementary Material

In this Appendix, we provide lemmas, the asymptotic distributional theory of semi-parametric M-estimators with a nuisance functional parameter, and proofs of Theorems 3.1–3.3 and Corollary 3.1.

S1 Notation

When θ takes a value of $\widehat{\theta}_n$, we denote A_j and S as $\hat{A}_{nj}$ and $\hat{S}_n$, and represent them as A_{0j} and S_0 respectively when $\theta = \theta_0$. Define $\boldsymbol{H}_1 = \Big\{\boldsymbol{h} = (h_1, h_2) : h_1 \in \mathcal{R}, h_2 \in$

*Co-first author

$\mathcal{H}_r, \|h_1\| \leq 1, \|h_2\|_\infty \leq 1\Big\}$, and $l^\infty(\boldsymbol{H}_1)$ be the space of bounded functionals on $\boldsymbol{H}_1$ under the supermum norm $\|f\|_\infty = \sup_{\mathbf{h}\in\boldsymbol{H}_1} |f(\mathbf{h})|$. Suppose that F_η are the parameter paths in $\mathcal{F}$ through F, that is, $F_\eta \in \mathcal{F}$ and $F_\eta\big|_{\eta=0} = F$. Let $\boldsymbol{H}_2 = \Big\{h : h = \frac{\partial F_\eta}{\partial \eta}\Big|_{\eta=0}\Big\}$, and let P and $\mathbb{P}_n$ denote probability measure and empirical measure.

For any $\boldsymbol{h} = (h_1, h_2), \boldsymbol{h}_1 = (h_{11}, h_{12}), \boldsymbol{h}_2 = (h_{21}, h_{22}) \in \boldsymbol{H}_1$, $h_3 \in \boldsymbol{H}_2$, we denote

$$
\begin{aligned}
m_{11}(\theta, F; D)[\boldsymbol{h}_1, \boldsymbol{h}_2] &= \nabla^2_\beta m(\theta, F; D)\\
&= -\Delta h_{11}^T Z^{\otimes 2} h_{21} \sum_{j=1}^K \triangle\Lambda_j(Y) e^{\beta^T Z} - \frac{1-\Delta}{\overline{F}(Y|Z)} h_{11}^T Z^{\otimes 2} h_{21} \int_Y^\infty \sum_{j=1}^K \triangle\Lambda_j(u) e^{\beta^T Z} dF(u|Z),\\
m_{12}(\theta, F; D)[\boldsymbol{h}_1, \boldsymbol{h}_2] &= \frac{\partial m_1(\beta, \Lambda_\eta, F; D)[h_{11}]}{\partial \eta}\Big|_{\eta=0}\\
&= -\Delta h_{11}^T Z \sum_{j=1}^K e^{\beta^T Z} \triangle h_{22j}(Y) - \frac{1-\Delta}{\overline{F}(Y|Z)} h_{11}^T Z \int_Y^\infty \sum_{j=1}^K e^{\beta^T Z} \triangle h_{22j}(u) dF(u|Z),
\end{aligned}
$$

$$
\begin{aligned}
m_{21}(\theta, F; D)[\boldsymbol{h}_1, \boldsymbol{h}_2] &= m_{12}(\theta, F; D)[\boldsymbol{h}_2, \boldsymbol{h}_1],\\
m_{22}(\theta, F; D)[\boldsymbol{h}_1, \boldsymbol{h}_2] &= \frac{\partial^2 m(\beta, \Lambda_{\eta_j}, F; D)}{\partial \eta_1 \partial \eta_2}\Big|_{\eta_1=\eta_2=0} = -\Delta \sum_{j=1}^K \frac{\triangle N_j}{(\triangle\Lambda_j(Y))^2} \triangle h_{12j}(Y) \triangle h_{22j}(Y)\\
&\quad - \frac{1-\Delta}{\overline{F}(Y|Z)} \int_Y^\infty \sum_{j=1}^K \frac{\triangle N_j}{(\triangle\Lambda_j(u))^2} \triangle h_{12j}(u) \triangle h_{22j}(u) dF(u|Z),\\
m_{13}(\theta, F; D)[\boldsymbol{h}, h_3] &= \frac{\partial m_1(\theta, F_\eta; D)[h_1]}{\partial \eta}\Big|_{\eta=0} = \frac{1-\Delta}{\overline{F}(Y|Z)} h_1^T Z \int_Y^\infty \sum_{j=1}^K A_j(u) \triangle\Lambda_j(u) dh_3(u|Z)\\
&\quad - \frac{1-\Delta}{\overline{F}^2(Y|Z)} h_3(Y|Z) h_1^T Z \int_Y^\infty \sum_{j=1}^K A_j(u) \triangle\Lambda_j(u) dF(u|Z),
\end{aligned}
$$

$$m_{23}(\theta, F; D)[\boldsymbol{h}, h_3] = \frac{\partial m_2(\theta, F_\eta; D)[h_2]}{\partial \eta}\Big|_{\eta=0} = \frac{1-\Delta}{\overline{F}(Y|Z)} \int_Y^\infty \sum_{j=1}^K A_j(u) \triangle h_{2j}(u) dh_3(u|Z)$$

$$- \frac{1-\Delta}{\overline{F}^2(Y|Z)} h_3(Y|Z) \int_Y^\infty \sum_{j=1}^K A_j(u) \triangle h_{2j}(u) dF(u|Z),$$

$$m^{(11)}(\theta, F; D)[\boldsymbol{h}_1, \boldsymbol{h}_2] = m_{11}(\theta, F; D)[\boldsymbol{h}_1, \boldsymbol{h}_2] + m_{12}(\theta, F; D)[\boldsymbol{h}_1, \boldsymbol{h}_2] + m_{21}(\theta, F; D)[\boldsymbol{h}_1, \boldsymbol{h}_2]$$

$$+ m_{22}(\theta, F; D)[\boldsymbol{h}_1, \boldsymbol{h}_2],$$

$$m^{(12)}(\theta, F; D)[\boldsymbol{h}, h_3] = m_{13}(\theta, F; D)[\boldsymbol{h}, h_3] + m_{23}(\theta, F; D)[\boldsymbol{h}, h_3]$$

with $\triangle h_j(u) = h(u - T_{K,j-1}) - h(u - T_{K,j})$.

S2 Lemmas

Lemma S2.1. Under Conditions (C1)–(C4), the class of functions $\{m(\theta, F; D) : \theta \in \Theta, F \in \mathcal{F}, \Lambda$ is uniformly bounded $\}$ is Donsker, where $m(\theta, F; D)$ is defined in Section 3.

Proof. First, the class $\{\Lambda : \Lambda \in \Psi, \Lambda$ is uniformly bounded $\}$ is Donsker, since Ψ is a monotone and uniformly bounded functional class. By Conditions (C1)–(C3), $\{S(u), \theta \in \Theta\}$ is Donsker by Theorem 2.10.6 of van der Vaart and Wellner (1996). By Theorem 2.10.3 of van der Vaart and Wellner (1996) and the performance of the Donsker property for the closure of the convex hull, we have $\{\int_Y^\tau S(u) dF(u|Z), \theta \in \Theta, F \in \mathcal{F}\}$ is Donsker. Condition (C4) ensures that $\int_Y^\tau S(u) dF(u|Z)/\overline{F}(Y|Z) < \infty$, which shows that

$\{\int_Y^\tau S(u)dF(u|Z)/\overline{F}(Y|Z), \theta \in \Theta, F \in \mathcal{F}\}$ is Donsker from Theorem 2.10.6 of van der Vaart and Wellner (1996). This completes the proof of Lemma S2.1. □

Lemma S2.2. (i). Under Condition (C5), for any $F \in \mathcal{F}_\delta$ and any differentiable function g,

$$P\left|\frac{1-\Delta}{\overline{F}(Y|Z)}\sum_{j=1}^{K}\int_Y^\tau \triangle g_j(u)dF(u|Z) - \frac{1-\Delta}{\overline{F}_0(Y|Z)}\sum_{j=1}^{K}\int_Y^\tau \triangle g_j(u)dF_0(u|Z)\right|$$
$$\lesssim\left(E\left[\sum_{j=1}^{K}(\triangle \dot{g}_j(U))^2 + E\left[\sum_{j=1}^{K}(\triangle g_j(U))^2\right]\right]\right)^{1/2}\|F-F_0\|_\infty,$$

where $\triangle g_j(u) = g(u - T_{K,j-1}) - g(u - T_{K,j})$.

(ii). Under Conditions (C1)–(C5), we have for small enough δ

$$P(m(\theta_0, F; D) - m(\theta_0, F_0; D)) \lesssim \|F - F_0\|_\infty, \tag{S2.1}$$

$$\left|P(m(\theta, F; D) - m(\theta, F_0; D)) - P(m(\theta_0, F; D) - m(\theta_0, F_0; D))\right|$$
$$\lesssim d_1(\theta, \theta_0)\|F - F_0\|_\infty. \tag{S2.2}$$

Proof. Using integration by parts and Condition (C5), we can show part (i). Part (ii) can be derived using Taylor expansion and part (i). □

Lemma S2.3. Assume that for given $F \in \mathcal{F}_\eta$, and for arbitrary function $\phi_n : (0, \infty) \to R$

such that $\delta \to \phi_n(\delta)/\delta^\iota$ is decreasing for some $\iota < 2$, every $\delta > 0$ and $\theta \in \Theta_n$,

$$P(m(\theta, F; D) - m(\theta_0, F; D)) \lesssim -d_1^2(\theta, \theta_0) + d_1(\theta, \theta_0)\|F - F_0\|_\infty,$$

$$E \sup_{\theta \in \Theta_{n\delta}} |\sqrt{n}(\mathbb{P}_n - P)(m(\theta, F; D) - m(\theta_0, F; D))| \leq \phi_n(\delta).$$

Let $r_n > 0$ satisfy $\phi_n(r_n) \leq \sqrt{n} r_n^2$ for every $n \in \mathbb{N}$. If $\widehat{\theta}_n \in \Theta_n$ is a consistent estimator for θ satisfying $\mathbb{P}_n m(\widehat{\theta}_n, F; D) \geq \mathbb{P}_n m(\theta_0, F; D) - O_p(r_n^2)$, then $d_1(\widehat{\theta}_n, \theta_0) = O_p(r_n + \|F - F_0\|_\infty)$.

Proof. The lemma can be proved by following the same techniques in proving Theorem 5.55 in van der Vaart (1998). □

Lemma S2.4. Let $\mathcal{G} = \{h : [0, \tau] \to [0, M]\}$, where M is a finite positive constant. Define

$$\begin{aligned}
\mathcal{G}_\delta(F) =& \{m(\theta, F; D) - m(\theta_0, F; D) : \theta \in \Theta_{n\delta}\}, \\
\mathcal{G}_{1,\delta}(F)[\boldsymbol{h}] =& \{m^{(1)}(\theta, F; D)[\boldsymbol{h}] - m^{(1)}(\theta_0, F; D)[\boldsymbol{h}] : \theta \in \Theta_{n\delta}\}, \\
\mathcal{G}_{2,\delta}(F)[h] =& \Big\{ V_n\Big(\theta, \theta_0, h, F\Big) : \theta \in \Theta_{n\delta} \Big\}
\end{aligned}$$

for $F \in \mathcal{F}_\eta$, $\boldsymbol{h} \in \boldsymbol{H}_1$ and $h \in \mathcal{G}$. Then under Conditions (C1)–(C4) and (C8)–(C11), we have for any $0 < \epsilon < \delta$,

$$\log N_{[]}(\epsilon, \mathcal{G}_{\delta}(F), \|\cdot\|_{P,B}) \le c q_n \log(\delta/\epsilon), \tag{S2.3}$$

$$\log N_{[]}(\epsilon, \mathcal{G}_{1,\delta}(F)[\boldsymbol{h}], \|\cdot\|_{P,B}) \le c q_n \log(\delta/\epsilon), \tag{S2.4}$$

$$\log N_{[]}(\epsilon, \mathcal{G}_{2,\delta}(F)[h], \|\cdot\|_{P,B}) \le c q_n \log(\delta/\epsilon), \tag{S2.5}$$

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

Proof. The lemma can be verified by using Lemma 7.1 of Wellner and Zhang (2007) and the arguments similar to those used in proving Lemma 1 of Zhao and Zhang (2017, Supplementary Material). □

S3 Asymptotic distributional theory of semiparametric M-estimators with a nuisance functional parameter

Suppose that $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample taken from the distribution of X, and $l_n(\theta, F; \mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} m(\theta, F; X_i)$ is an objective function based on $\mathbf{X}$, where $\theta = (\beta, \Lambda)$ with β being an unknown d-dimensional parameter vector in $\mathcal{R}$ and Λ being an unknown function in the class Ψ, and F is a nuisance functional parameter. Let Ψ_n be the sieve parameter space satisfying $\Psi_n \subseteq \Psi_{n+1} \subseteq \cdots \subseteq \Psi, \quad$ for $n \ge 1$. Assume that $\widehat{F}_n$ is a consistent estimator of F and $\widehat{\theta}_n$ is the estimator of Λ_0 by maximizing $l_n(\theta, \widehat{F}_n; \mathbf{X})$ in the sieve parameter space $\mathcal{R} \times \Psi_n$.

For $\boldsymbol{h} \in \boldsymbol{H}_1$, define a sequence of maps G_n of $\mathcal{U}$, a neighborhood of (θ_0, F_0), into $l^\infty(\boldsymbol{H}_1)$ by

$$G_n(\theta, F)[\boldsymbol{h}] = \frac{\partial}{\partial \delta} l_n(\theta_\delta, F; \mathbf{X})\big|_{\delta=0} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \delta} m(\theta_\delta, F; X_i)\big|_{\delta=0} = \mathbb{P}_n m^{(1)}(\theta, F; X)[\boldsymbol{h}],$$

and let $G(\theta, F)[\boldsymbol{h}] = Pm^{(1)}(\theta, F; X)[\boldsymbol{h}]$, where P and $\mathbb{P}_n$ denote 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 of $\widehat{\theta}_n$, we propose the following working conditions that can be verified for a specific problem using empirical process theorems.

(B1). $\sqrt{n}(G_n - G)(\widehat{\theta}_n, \widehat{F}_n)[\boldsymbol{h}] - \sqrt{n}(G_n - G)(\theta_0, F_0)[\boldsymbol{h}] = o_p(1)$.

(B2). $G(\theta_0, F_0)[\boldsymbol{h}] = 0$ and $G_n(\widehat{\theta}_n, \widehat{F}_n)[\boldsymbol{h}] = o_p(n^{-1/2})$.

(B3). $G(\theta, F)[\boldsymbol{h}]$ is Fréchet-differentiable with respect to θ and F with the continuous derivative $\dot{G}_{1,\theta,F}[\boldsymbol{h}]$ and $\dot{G}_{2,\theta,F}[\boldsymbol{h}]$, respectively.

(B4). $G(\widehat{\theta}_n, \widehat{F}_n)[\boldsymbol{h}] - G(\theta_0, F_0)[\boldsymbol{h}] - \dot{G}_{1,\theta_0,F_0}(\widehat{\theta}_n - \theta_0)[\boldsymbol{h}] - \dot{G}_{2,\theta_0,F_0}(\widehat{F}_n - F_0)[\boldsymbol{h}] = o_p(n^{-1/2})$.

(B5). $\sqrt{n}G_n(\theta_0, F_0)[\boldsymbol{h}] + \sqrt{n}\dot{G}_{2,\theta_0,F_0}(\widehat{F}_n - F_0)[\boldsymbol{h}]$ converges in distribution to a tight Gaussian process on $l^\infty(\boldsymbol{H}_1)$.

Conditions (B2), (B3) and (B5) are the similar to the analytical conditions in Theorem 3.3.1 of van der Vaart and Wellner (1996); (B1) and (B4) require the remainder of the Taylor expansions are negligible, which are weaker than those in van der Vaart and Wellner

(1996).

Theorem S3.1. Under Assumptions (B1)–(B5), we have for any $\boldsymbol{h} \in \boldsymbol{H}_1$,

$$-\sqrt{n}\dot{G}_{1,\theta_0,F_0}(\widehat{\theta}_n - \theta_0)[\boldsymbol{h}] = \sqrt{n}G_n(\theta_0, F_0)[\boldsymbol{h}] + \sqrt{n}\dot{G}_{2,\theta_0,F_0}(\widehat{F}_n - F_0)[\boldsymbol{h}] + o_p(1),$$

and $-\sqrt{n}\dot{G}_{1,\theta_0,F_0}(\widehat{\theta}_n - \theta_0)[\boldsymbol{h}]$ converges in distribution to a tight Gaussian process on $l^\infty(\boldsymbol{H}_1)$.

Proof. From Assumptions (B2)–(B4), we have

$$-\sqrt{n}\dot{G}_{1,\theta_0,F_0}(\widehat{\theta}_n - \theta_0)[\boldsymbol{h}] - \sqrt{n}\dot{G}_{2,\theta_0,F_0}(\widehat{F}_n - F_0)[\boldsymbol{h}] = -\sqrt{n}G(\widehat{\theta}_n, \widehat{F}_n)[\boldsymbol{h}] + o_p(1).$$

We can then get by Assumptions (B1) and (B2)

$$-\sqrt{n}G(\widehat{\theta}_n, \widehat{F}_n)[\boldsymbol{h}] = \sqrt{n}(G_n - G)(\theta_0, F_0)[\boldsymbol{h}] + o_p(1). \tag{S3.1}$$

Thus, it follows from (S3.1)

$$-\sqrt{n}\dot{G}_{1,\theta_0,F_0}(\widehat{\theta}_n - \theta_0)[\boldsymbol{h}] = \sqrt{n}G_n(\theta_0, F_0)[\boldsymbol{h}] + \sqrt{n}\dot{G}_{2,\theta_0,F_0}(\widehat{F}_n - F_0)[\boldsymbol{h}] + o_p(1).$$

At last, the proposition is concluded by Assumption (B5). □

Theorem S3.1 concludes the general functional asymptotic normality for semiparametric $M-$estimators with nuisance parameters. This conclusion extends both Theorem

3.3 in Wellner and Zhang (2007) and Theorem 2.1 in Zhao and Zhang (2017) to a more general case. Applying this conclusion, we obtain the asymptotic normality of the regression coefficient estimator and the two-stage functional estimator for panel count data with terminal event.

S4 Proofs of Main Results

Proof of Theorem 3.1

Proof. First we note that

$$
\begin{aligned}
&Pm(\theta_0, F_0; D) - Pm(\theta, F_0; D)\\
=&E\left(\sum_{j=1}^{K}\left\{\triangle N_j \log\left(\frac{\triangle\Lambda_{0j}(U)e^{\beta_0^T Z}}{\triangle\Lambda_j(U)e^{\beta^T Z}}\right) - \left[\frac{\triangle\Lambda_{0j}(U)e^{\beta_0^T Z}}{\triangle\Lambda_j(U)e^{\beta^T Z}} + 1\right]\triangle\Lambda_j(U)e^{\beta^T Z}\right\}\right)\\
=&\int \exp(\beta^T z)\triangle\Lambda(u,t)\phi\left[\frac{\triangle\Lambda_0(u,t)\exp(\beta_0^T z)}{\triangle\Lambda(u,t)\exp(\beta^T z)}\right]d\nu_1(u,t,z),
\end{aligned}
$$

where $\triangle\Lambda(u,t) = \Lambda(u) - \Lambda(t)$ and $\phi(x) = x\log(x) - x + 1$. Since $\phi(x) \geq 0$ for $x > 0$ with equality holding only at $x = 1$, we have $Pm(\theta_0, F_0; D) \geq Pm(\theta, F_0; D)$ and $Pm(\theta_0, F_0; D) = Pm(\theta, F_0; D)$ if and only if

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

Then following the arguments in the proof of Theorem 3.1 in Wellner and Zhang (2007), under Conditions (C6) and (C7), we can get that $\theta = \theta_0$ a.e. with respect to μ_1. In addition, since the metrics d_1 and d_2 are equivalent, it follows $\theta = \theta_0$ a.e. with respect to μ_2.

Noting that $\Lambda_0 \in \mathcal{H}_r$, there exists a $\Lambda_n \in \Psi_n$ such that $\|\Lambda_n - \Lambda_0\|_\infty = O(n^{-r\gamma})$ according to (C8) and Lemma A1 of Lu, Zhang and Huang (2007). For any given $\epsilon > 0$, let $\tilde{\theta}_n = (\hat{\beta}_n, (1-\epsilon)\hat{\Lambda}_n + \epsilon\Lambda_n)) = (\hat{\beta}_n, \hat{\Lambda}_n) + \epsilon(0, \Lambda_n - \hat{\Lambda}_n)$. By the definition of $\widehat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n)$, it follows that

$$
\begin{aligned}
0 \geq & \lim_{\epsilon \to 0} \frac{\mathbb{P}_n m(\widehat{\theta}_n + \epsilon(0, \Lambda_n - \hat{\Lambda}_n), \widehat{F}_n; D) - \mathbb{P}_n m(\widehat{\theta}_n, \widehat{F}_n; D)}{\epsilon} \\
= & \mathbb{P}_n \Bigg(\Delta \sum_{j=1}^{K} \hat{A}_{nj}(Y)(\triangle\Lambda_{nj}(Y) - \triangle\hat{\Lambda}_{nj}(Y)) \\
& + \frac{1-\Delta}{1-\widehat{F}_n(Y|Z)} \int_Y^\tau \sum_{j=1}^{K} \hat{A}_{nj}(u)(\triangle\Lambda_{nj}(u) - \triangle\hat{\Lambda}_{nj}(u)) d\widehat{F}_n(u|Z) \Bigg).
\end{aligned}
$$

By (C1)–(C4) and the strong law of large numbers, we obtain that

$$
\begin{aligned}
&\mathbb{P}_n\Bigg(\Delta\sum_{j=1}^{K}\Big(\triangle N_j\frac{\triangle\Lambda_{nj}(Y)}{\triangle\hat{\Lambda}_{nj}(Y)}+\triangle\hat{\Lambda}_{nj}(Y)e^{\hat{\beta}_n^T Z}\Big)\\
&\quad+\frac{1-\Delta}{1-\widehat{F}_n(Y|Z)}\int_Y^{\tau}\sum_{j=1}^{K}\Big(\triangle N_j\frac{\triangle\Lambda_{nj}(u)}{\triangle\hat{\Lambda}_{nj}(u)}+\triangle\hat{\Lambda}_{nj}(u)e^{\hat{\beta}_n^T Z}\Big)d\widehat{F}_n(u|Z)\Bigg)\\
\leq&\mathbb{P}_n\Bigg(\Delta\sum_{j=1}^{K}(\triangle N_j+\triangle\Lambda_{nj}(Y)e^{\hat{\beta}_n^T Z})\\
&\quad+\frac{1-\Delta}{1-\widehat{F}_n(Y|Z)}\int_Y^{\tau}\sum_{j=1}^{K}(\triangle N_j+\triangle\Lambda_{nj}(u)e^{\hat{\beta}_n^T Z})d\widehat{F}_n(u|Z)\Bigg)\\
\leq&c\mathbb{P}_n\Bigg(\Delta\sum_{j=1}^{K}(\triangle N_j+\triangle\Lambda_{nj}(Y))+\frac{1-\Delta}{1-\widehat{F}_n(Y|Z)}\int_Y^{\tau}\sum_{j=1}^{K}(\triangle N_j+\triangle\Lambda_{nj}(u))d\widehat{F}_n(u|Z)\\
\xrightarrow{a.s.}&cP\Bigg(\Delta\sum_{j=1}^{K}(\triangle N_j+\triangle\Lambda_{0j}(Y))+\frac{1-\Delta}{\overline{F}_0(Y|Z)}\int_Y^{\tau}\sum_{j=1}^{K}(\triangle N_j+\triangle\Lambda_{0j}(u))dF_0(u|Z)\Bigg).
\end{aligned}
$$

On the other hand, we have

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{P}_n\bigg(\Delta\sum_{j=1}^{K}\Big(\triangle N_j\frac{\triangle\Lambda_{nj}(Y)}{\triangle\hat{\Lambda}_{nj}(Y)}+\triangle\hat{\Lambda}_{nj}(Y)e^{\hat{\beta}_n^TZ}\Big)\\
&+\frac{1-\Delta}{1-\widehat{F}_n(Y|Z)}\int_Y^\tau\sum_{j=1}^{K}\Big(\triangle N_j\frac{\triangle\Lambda_{nj}(u)}{\triangle\hat{\Lambda}_{nj}(u)}+\triangle\hat{\Lambda}_{nj}(u)e^{\hat{\beta}_n^TZ}\Big)d\widehat{F}_n(u|Z)\bigg)\\
\geq& c\limsup_{n\to\infty}\big(\hat{\Lambda}_n(b_2)-\hat{\Lambda}_n(b_1)\big)\mathbb{P}_n\bigg(\Delta 1_{\{Y-T_{K,K}\in[0,b_1],Y\in[b_2,\tau]\}}\\
&+\frac{1-\Delta}{1-\widehat{F}_n(Y|Z)}\int_Y^\tau 1_{\{u-T_{K,K}\in[0,b_1],u\in[b_2,\tau]\}}d\widehat{F}_n(u|Z)\bigg)\\
\geq& c\limsup_{n\to\infty}\big(\hat{\Lambda}_n(b_2)-\hat{\Lambda}_n(b_1)\big)E\big[1_{\{U-T_{K,K}\in[0,b_1],U\in[b_2,\tau]\}}\big]\\
=& c\limsup_{n\to\infty}\big(\hat{\Lambda}_n(b_2)-\hat{\Lambda}_n(b_1)\big)\mu_3([0,b_1]\times[b_2,\tau])
\end{aligned}$$

for all $0\leq b_1\leq b_2\leq\tau$. Thus, if $\mu_3([0,b_1]\times[b_2,\tau])>0$ and $\mu_3(\{0\}\times\{\tau\})>0$ specially, $\hat{\Lambda}_n$ is uniformly bounded almost surely on $[0,\tau]$.

Recalling that $\|\Lambda_n-\Lambda_0\|_\infty=O(n^{-r\gamma})$, we have

$$\begin{aligned}
&\mathbb{P}_n m(\Lambda_n,\beta_0,\widehat{F}_n;D)-\mathbb{P}_n m(\theta_0,\widehat{F}_n;D)\\
=&(\mathbb{P}_n-P)\big(m(\Lambda_n,\beta_0,\widehat{F}_n;D)-m(\theta_0,\widehat{F}_n;D)\big)+P\big(m(\Lambda_n,\beta_0,\widehat{F}_n;D)-m(\theta_0,\widehat{F}_n;D)\big)\\
:=&I_1+I_2.
\end{aligned}$$

Using Lemma S2.1 and the fact that a Donsker class is also a Glivanko-Cantilli class, we get that $|I_1|=o_p(1)$. In addition, we note $|I_2|\leq cE(N(T_{K,K})+1)\|\Lambda_n-\Lambda_0\|_\infty=$

$O_p(n^{-r\gamma}) = o_p(1)$. Thus, we have

$$\mathbb{P}_n m(\Lambda_n, \beta_0, \widehat{F}_n; D) = \mathbb{P}_n m(\theta_0, \widehat{F}_n; D) + o_p(1).$$

It then follows by the definition of $\hat{\theta}_n$ that

$$\begin{aligned}
\mathbb{P}_n m(\hat{\theta}_n, \hat{F}_n; D) \geq \sup_{\Lambda_n \in \Psi_n} \mathbb{P}_n m(\Lambda_n, \beta_0, \hat{F}_n; D) &\geq \mathbb{P}_n m(\theta_0, \hat{F}_n; D) + o_p(1) \\
&= \mathbb{P}_n m(\theta_0, F_0; D) + \mathbb{P}_n(m(\theta_0, \hat{F}_n; D) - m(\theta_0, F_0; D)) + o_p(1) \\
&= \mathbb{P}_n m(\theta_0, F_0; D) + (\mathbb{P}_n - P)\big(m(\theta_0, \hat{F}_n; D) - m(\theta_0, F_0; D)\big) \\
&\quad + P(m(\theta_0, \hat{F}_n; D) - m(\theta_0, F_0; D)) + o_p(1) \\
&= \mathbb{P}_n m(\theta_0, F_0; D) + o_p(1), \qquad \text{(S4.1)}
\end{aligned}$$

where the last equality is from Lemmas S2.1 and S2.2. Using Lemma S2.1 again, we get that

$$\begin{aligned}
0 &\leq P(m(\theta_0, F_0; D) - m(\hat{\theta}_n, F_0; D)) \\
&= \mathbb{P}_n m(\theta_0, F_0; D) - \mathbb{P}_n m(\hat{\theta}_n, F_0; D) + o_p(1) \\
&\leq \mathbb{P}_n m(\hat{\theta}_n, \hat{F}_n; D) - \mathbb{P}_n m(\hat{\theta}_n, F_0; D) + o_p(1) \\
&= P(m(\hat{\theta}_n, \hat{F}_n; D) - m(\hat{\theta}_n, F_0; D)) + o_p(1) \\
&= o_p(1), \qquad \text{(S4.2)}
\end{aligned}$$

where the second inequality follows from (S4.1) and the last equality is from (S2.1) and (S2.2) in Lemma S2.2. Noting that for any $\delta > 0$

$$\sup_{d_1(\theta,\theta_0)>\delta} Pm(\theta, F_0; D) < Pm(\theta_0, F_0; D)$$

and that θ_0 is the unique maximizer of $Pm(\theta, F_0)$, we have

$$\{d_1(\hat{\theta}_n, \theta_0) > \delta\} \subset \{Pm(\hat{\theta}_n, F_0; D) < Pm(\theta_0, F_0; D)\},$$

with the sequence of the events on the right going to a null event in view of inequality (S4.2). Hence $d_1(\hat{\theta}_n, \theta_0)$ converges to 0 almost surely by Theorem 5.8 in van der Vaart (2002). So $d_2(\hat{\theta}_n, \theta_0)$ converges to 0 almost surely as well since the metrics d_1 and d_2 are equivalent. This completes the proof of Theorem 3.1. □

Proof of Theorem 3.2

Proof. We conclude the theorem through checking the conditions in Lemma S2.3. Using the triangle inequality, we have

$$
\begin{aligned}
& Pm(\theta,\widehat{F}_n;D)-Pm(\theta_0,\widehat{F}_n;D) \\
=& P\Big(\big(m(\theta,\widehat{F}_n;D)-m(\theta,F_0;D)\big)-\big(m(\theta_0,\widehat{F}_n;D)-m(\theta_0,F_0;D)\big)\Big) \\
& -P(m(\theta_0,F_0;D)-m(\theta,F_0;D)) \\
\leq& \Big|P\Big(\big(m(\theta,\widehat{F}_n;D)-m(\theta,F_0;D)\big)-\big(m(\theta_0,\widehat{F}_n;D)-m(\theta_0,F_0;D)\big)\Big)\Big| \\
& -P(m(\theta_0,F_0;D)-m(\theta,F_0;D)) \\
:=& I_1-I_2.
\end{aligned}
$$

From Lemma S2.2, it follows that for $\theta\in\Theta_\delta$, $I_1\lesssim d_1(\theta,\theta_0)\|\widehat{F}_n-F_0\|_\infty$. We now consider the term I_2. Since

$$
\begin{aligned}
I_2 =& P(m(\theta_0,F_0;D)-m(\theta,F_0;D)) \\
=& E\left[\sum_{j=1}^{K}\triangle\Lambda_j(U)e^{\beta^TZ}\phi\left(\frac{\triangle\Lambda_{0j}(U)e^{\beta_0^TZ}}{\triangle\Lambda_j(U)e^{\beta^TZ}}\right)\right],
\end{aligned}
$$

and $\phi(x) \geq (x-1)^2/4$ for $0 \leq x \leq 5$, we have for any θ in a sufficiently small neighborhood of θ_0,

$$\begin{aligned} I_2 \geq & \frac{1}{4} E\left[\sum_{j=1}^{K} \triangle\Lambda_j(U) e^{\beta^T Z} \left(\frac{\triangle\Lambda_{0j}(U) e^{\beta_0^T Z}}{\triangle\Lambda_j(U) e^{\beta^T Z}} - 1\right)^2\right] \\ \geq & c \int \left(\triangle\Lambda(u_1, u_2) e^{\beta^T z} - \triangle\Lambda_0(u_1, u_2) e^{\beta_0^T z}\right)^2 d\nu_1(u_1, u_2, z) \end{aligned}$$

by (C1) and (C2), where $\triangle\Lambda(u_1, u_2) = \Lambda(u_1) - \Lambda(u_2)$. Let $g(t) = \triangle\Lambda_t(U_1, U_2)\exp(\beta_t^T Z)$, where $\triangle\Lambda_t = \triangle\Lambda_0 + t\triangle h_2$ and $\beta_t = \beta_0 + th_1$ for $0 \leq t \leq 1$ with $\triangle h_2 = \triangle\Lambda - \triangle\Lambda_0$, $h_1 = \beta - \beta_0$ and $(U_1, U_2, Z) \sim \nu_1$. Then $\triangle\Lambda(U_1, U_2) e^{\beta^T Z} - \triangle\Lambda_0(U_1, U_2) e^{\beta_0^T Z} = g(1) - g(0) = \dot{g}(\xi)$ for some $\xi \in (0, 1)$ by the mean value theorem. Since

$$\dot{g}(\xi) = \exp({\beta^{\xi}}^T Z)[\triangle h_2(U_1, U_2)(1 + \xi h_1^T Z) + h_1^T Z \triangle\Lambda_0(U_1, U_2)]$$

with $\beta^{\xi} = \beta_0 + \xi h_1$, we then get that

$$I_2 \geq c \int \left(\triangle h_2(u_1, u_2) + h_1^T z[\triangle\Lambda_0(u_1, u_2) + \xi\triangle h_2(u_1, u_2)]\right)^2 d\nu_1(u_1, u_2, z) = \nu_1(g_1 w + g_2)^2,$$

where $g_1(U_1, U_2, Z) = h_1^T Z \triangle\Lambda_0(U_1, U_2)$, $g_2(U_1, U_2) = \triangle h_2(U_1, U_2)$ and $w(U_1, U_2, Z) = 1 + \xi\triangle h_2(U_1, U_2)/\triangle\Lambda_0(U_1, U_2)$. According to Lemma 8.8 of van der Vaart (2002), we need to bound $[\nu_1(g_1 g_2)]^2$ by a constant less than one times $\nu_1(g_1^2)\nu_1(g_2^2)$. To the end, writing expectations under ν_1 as E_1, we have by the Cauchy-Schwartz inequality

$$\begin{aligned}
[E_1(g_1g_2)]^2 =&(E_1[E_1(g_1g_2|U_1,U_2)])^2 \leq E_1(g_2^2)E_1([E_1(g_1|U_1,U_2)]^2)\\
=&E_1(g_2^2)E_1(\triangle\Lambda_0^2(U_1,U_2)[E_1(h_1^TZ|U_1,U_2)]^2)\\
=&E_1(g_2^2)E_1(\triangle\Lambda_0^2(U_1,U_2)[E_1(h_1^T[Z-(Z-E_1(Z|U_1,U_2))]^{\otimes 2}h_1|U_1,U_2)])\\
\leq&(1-\varpi)E_1(g_2^2)E_1(\triangle\Lambda_0^2(U_1,U_2)h_1^TE_1[ZZ^T|U_1,U_2]h_1)\\
=&(1-\varpi)E_1(g_2^2)E_1(g_1^2),
\end{aligned}$$

where the last inequality follows from (C12). By van der Vaart's lemma,

$$\nu_1(g_1w+g_2)^2 \geq c(\nu_1(g_1^2)+\nu_1(g_2^2)) = cd_1^2(\theta,\theta_0).$$

Thus, $I_2 \gtrsim d_1^2(\theta,\theta_0)$. It follows then

$$P(m(\theta,\widehat{F}_n;D)-m(\theta_0,\widehat{F}_n;D)) \lesssim -d_1^2(\theta,\theta_0)+d_1(\theta,\theta_0)\|\widehat{F}_n-F_0\|_\infty. \qquad \text{(S4.3)}$$

Similar to the proof of Lemma S2.4, we can show that

$$\|m(\theta,\widehat{F}_n;D)-m(\theta_0,\widehat{F}_n;D)\|_{P,B}^2 \lesssim d_1(\theta,\theta_0)\|\widehat{F}_n-F_0\|_\infty + d_1^2(\theta,\theta_0) \lesssim \delta^2$$

for $\|\widehat{F}_n - F_0\|_\infty \lesssim O_p(\delta)$ and $d_1(\theta, \theta_0) = O_p(\delta^2)$. Therefore, Lemma S2.4 gives that

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

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

$$\begin{aligned}
& E \sup_{\theta \in \Theta_{n\delta}} \sqrt{n} |(\mathbb{P}_n - P)(m(\theta, \widehat{F}_n; D) - m(\theta_0, \widehat{F}_n; D))| \\
\leq & J_{[]}(\delta, \mathcal{G}_\delta(\widehat{F}_n), \|\cdot\|_{P,B})(1 + J_{[]}(\delta, \mathcal{G}_\delta(\widehat{F}_n), \|\cdot\|_{P,B})/(\delta^2 n^{1/2})) \\
\lesssim & q_n^{1/2} \delta + q_n n^{-1/2}.
\end{aligned}$$

Setting $\phi_n(\delta) = q_n^{1/2}\delta + q_n n^{-1/2}$ and $r_n = n^{-(1-\gamma)/2}$, it can be obtained that $\phi_n(\delta)/\delta$ is a decreasing function of δ and $\phi_n(r_n)/r_n^2 = O(n^{1/2})$.

Since there exists $\Lambda_n \in \Psi_n$ of order $l \geq r + 2$ such that $\|\Lambda_n - \Lambda_0\|_\infty = O(n^{-r\gamma})$ for $0 < \gamma < 1/2$ by (C8) and Lemma A1 in Lu, Zhang and Huang (2007), we have

$$\begin{aligned}
& \mathbb{P}_n m(\widehat{\theta}_n, \widehat{F}_n; D) - \mathbb{P}_n m(\theta_0, \widehat{F}_n; D) \\
= & \mathbb{P}_n \big(m(\widehat{\theta}_n, \widehat{F}_n; D) - m(\Lambda_n, \beta_0, \widehat{F}_n; D) \big) + (\mathbb{P}_n - P) \big(m(\Lambda_n, \beta_0, \widehat{F}_n; D) - m(\theta_0, \widehat{F}_n; D) \big) \\
& + P \big(m(\Lambda_n, \beta_0, \widehat{F}_n; D) - m(\theta_0, \widehat{F}_n; D) \big) \\
:= & I_1 + I_2 + I_3,
\end{aligned}$$

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

$$\tilde{\mathcal{G}}(F) = \Big\{ \frac{m(\theta, F; D) - m(\theta_0, F; D)}{n^{-r\gamma+\epsilon}} : \theta \in \Theta_n, \|\theta - \theta_0\|_\infty = O(n^{-r\gamma}) \Big\}$$

for any $0 < \epsilon < 1/2 - r\gamma$ and $F \in \mathcal{F}$. It can be shown that $\tilde{\mathcal{G}}(\widehat{F}_n)$ is P-Donsker by (C1)–(C4). In addition, we note that $Pf^2 \to 0$ as $n \to \infty$ for any $f \in \tilde{\mathcal{G}}(\widehat{F}_n)$. By Corollary 2.3.12 of van der Vaart and Wellner (1996), we have

$$I_2 = n^{-r\gamma+\epsilon}(\mathbb{P}_n - P)\Big(\frac{m(\theta, F; D) - m(\theta_0, F; D)}{n^{-r\gamma+\epsilon}}\Big) = O_p(n^{-r\gamma+\epsilon}n^{-1/2}) = o(n^{-2r\gamma}).$$

Following the deduction for (S4.3), we have $I_3 \geq -O_p(n^{-2r\gamma})$. Hence,

$$\mathbb{P}_n m(\widehat{\theta}_n, \widehat{F}_n; D) - \mathbb{P}_n m(\theta_0, \widehat{F}_n; D) \geq -O_p(n^{-2r\gamma}).$$

The conditions in Lemma S2.3 imply that $n^{-2r\gamma} = O(n^{-(1-\gamma)})$, which gives the rate of convergence of $n^{r/(1+2r)}$ by choosing $\gamma = 1/(1+2r)$. This completes the proof of Theorem 3.2. □

Proof of Theorem 3.3

Proof. To conclude the theorem, it suffices to verify Conditions (B1)–(B5) by Theorem S3.1. Note that $G(\theta_0, F_0)[\boldsymbol{h}] = Pm^{(1)}(\theta_0, F_0; D)[\boldsymbol{h}] = 0$ for any $\boldsymbol{h} \in \boldsymbol{H}_1$, and the Fréchet derivatives are

$$\dot{G}_{1,\theta,F}(\theta-\theta_0)[\boldsymbol{h}] = P(m^{(11)}(\theta,F;D)[\boldsymbol{h},\theta-\theta_0]),$$

$$\dot{G}_{2,\theta,F}(F-F_0)[\boldsymbol{h}] = P(m^{(12)}(\theta,F;D)[\boldsymbol{h},F-F_0]).$$

We only show (B5) since (B1)–(B4) can be verified by using the techniques similar to those used in the proofs of Theorem 3.1 in Zhao and Zhang (2017).

Let $\mathcal{S}_n(\theta,F) = G_n(\theta,F) + \dot{G}_{2,\theta,F}(\widehat{F}_n - F)$. Then $\mathcal{S}_n$ is a map from $\mathcal{U}$ to $l^\infty(\boldsymbol{H}_1)$. Recall that $G_n(\theta_0,F_0)[\boldsymbol{h}] = n^{-1}\sum_{i=1}^n m^{(1)}(\theta_0,F_0;D_i)[\boldsymbol{h}]$ and $\dot{G}_{2,\theta_0,F_0}(\widehat{F}_n - F_0)[\boldsymbol{h}] = P\Big(\int_Y^\tau \psi(u;D)[\boldsymbol{h}]\, d[\widehat{F}_n(u|Z) - F_0(u|Z)]\Big) = n^{-1}\sum_{i=1}^n P\Big(\int_Y^\tau \tilde{\psi}(u;D)[\boldsymbol{h}]dO(u;D;\tilde{D}_i)\Big) := \mathbb{P}_n m^{**}(\theta_0,F_0;D)[\boldsymbol{h}]$ with $\psi_1(u;D)[\boldsymbol{h}] = \frac{1-\Delta}{\overline{F}_0(Y|Z)}\sum_{j=1}^K A_j(u)(h_1^T Z \triangle\Lambda_j(u) + \triangle h_j(u))$, $\psi_2(u;D)[\boldsymbol{h}] = \int_Y^\tau \psi_1(u;D)[\boldsymbol{h}]dF_0(u|Z$ $\psi(u;D) = \psi_1(u;D) - \psi_2(u;D)$ and $\tilde{\psi} = g\circ\psi$. It can be seen that $\mathcal{S}_n$ is a bounded Lipschitz function with respect to $\boldsymbol{H}_1$. Therefore, (B5) holds since $\boldsymbol{H}_1$ is a Donsker class.

Thus, we get that

$$-\sqrt{n}\dot{G}_{1,\theta_0,F_0}(\widehat{\theta}_n - \theta_0)[\boldsymbol{h}] = \sqrt{n}G_n(\theta_0,F_0)[\boldsymbol{h}] + \sqrt{n}\dot{G}_{2,\theta_0,F_0}(\widehat{F}_n - F_0)[\boldsymbol{h}] + o_p(1) \tag{S4.4}$$

is asymptotically normal distributed.

Especially, if we take $\boldsymbol{h} = (h_1, h)$ with $\|h_1\| = 1, h_1 \in \mathcal{R}$ and

$$\begin{aligned}\triangle h_j(u) =& - \triangle\Lambda_{0j} h_1^T E(Ze^{\beta_0^T Z}|U = u, K, T_{K,j}, T_{K,j-1})/E(e^{\beta_0^T Z}|U = u, K, T_{K,j}, T_{K,j-1}) \\ :=& - \triangle\Lambda_{0j} h_1^T R(u, K, T_{K,j}, T_{K,j-1}),\end{aligned}$$

then $P\big(m_{12}(\theta_0, F_0; D)[\boldsymbol{h}, \widehat{\theta}_n - \theta_0] + m_{22}(\theta_0, F_0; D)[\boldsymbol{h}, \widehat{\theta}_n - \theta_0]\big) = 0$ and

$$\begin{aligned}-\dot{G}_{1,\theta_0,F_0}(\widehat{\theta}_n - \theta_0)[\boldsymbol{h}] =& - P\big(m_{11}(\theta_0, F_0; D)[\boldsymbol{h}, \widehat{\theta}_n - \theta_0] + m_{21}(\theta_0, F_0; D)[\boldsymbol{h}, \widehat{\theta}_n - \theta_0]\big) \\ =& h_1^T E\left(\sum_{j=1}^{K} \triangle\Lambda_{0j}(U) e^{\beta_0^T Z}\big[Z - R(U, K, T_{K,j}, T_{K,j-1})\big] Z^T\right)(\hat{\beta}_n - \beta_0) \\ =& h_1^T E\left(\sum_{j=1}^{K} \triangle\Lambda_{0j}(U) e^{\beta_0^T Z}[Z - R(U, K, T_{K,j}, T_{K,j-1})]^{\otimes 2}\right)(\hat{\beta}_n - \beta_0).\end{aligned}$$

Note that

$$\begin{aligned}& G_n(\theta_0, F_0)[\boldsymbol{h}] + \dot{G}_{2,\theta_0,F_0}(\widehat{F}_n - F_0)[\boldsymbol{h}] \\ =& \mathbb{P}_n m^{(1)}(\theta_0, F_0; D)[\boldsymbol{h}] + P m^{(12)}(\theta_0, F_0; D)[\boldsymbol{h}, \widehat{F}_n - F_0] \\ =& h_1^T \mathbb{P}_n\left(\Delta \sum_{j=1}^{K} (\triangle N_j - e^{\beta_0^T Z} \triangle\Lambda_{0j}(Y))[Z - R(Y, K, T_{K,j}, T_{K,j-1})]\right) \\ & + h_1^T \mathbb{P}_n\left(\frac{1 - \Delta}{\overline{F}_0(Y|Z)} \int_Y^\tau \sum_{j=1}^{K} (\triangle N_j - e^{\beta_0^T Z} \triangle\Lambda_{0j}(u))[Z - R(u, K, T_{K,j}, T_{K,j-1})] dF_0(u|Z)\right) \\ & + h_1^T \frac{1}{n} \sum_{i=1}^{n} P\left(\int_Y^\tau \tilde{\rho}(u; D) dO(u; D; \tilde{D}_i)\right), \\ :=& h_1^T \mathbb{P}_n m^*(\theta_0, F_0; D),\end{aligned}$$

where

$$\rho_1(u;D) = \sum_{j=1}^{K} A_j(u)\triangle\Lambda_{0j}(u)(Z - R(u,K,T_{K,j},T_{K,j-1})),$$
$$\rho(u;D) = \frac{1-\Delta}{\overline{F}_0(Y|Z)}\Bigg(\rho_1(u;D) - \int_Y^{\tau} \frac{\rho_1(u;D)}{\overline{F}_0(Y|Z)} dF_0(u|Z)\Bigg),$$

and $\tilde{\rho} = g \circ \rho$. Then (S4.4) reduces to

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N_d(0, \Sigma_1^{-1}\Sigma_2(\Sigma_1^{-1})^T),$$

where

$$\Sigma_1 = E\Bigg(\sum_{j=1}^{K} \triangle\Lambda_{0j}(U)e^{\beta_0^T Z}[Z - R(U,K,T_{K,j},T_{K,j-1})]^{\otimes 2}\Bigg),$$
$$\Sigma_2 = E\Big(m^*(\theta_0, F_0; D)^{\otimes 2}\Big).$$

Besides, we take $\boldsymbol{h} = (0, h)$. Define

$$\tilde{m}^*(\theta_0, F_0; D)[h] = E\Big(Z^T \sum_{j=1}^{K} e^{\beta_0^T Z}\triangle h_j(U)\Big)\Sigma^{-1} m^*(\theta_0, F_0; D)[\boldsymbol{h}_1, \boldsymbol{h}],$$
$$\mathbb{P}_n m^{**}(\theta_0, F_0; D)[h] = \frac{1}{n}\sum_{i=1}^{n} P\Big(\int_Y^{\tau} \tilde{\phi}(u;D)[h] dO(u;D;\tilde{D}_i)\Big),$$

where

$$\begin{aligned}\phi_1(u;D)[h] =& \sum_{j=1}^{K} A_j(u)\triangle h_j(u),\\ \phi(u;D)[h] =& \frac{1-\Delta}{\overline{F}_0(Y|Z)}\left(\phi_1(u;D)[h] - \int_Y^\tau \frac{\phi_1(u;D)[h]}{\overline{F}_0(Y|Z)} dF_0(u|Z)\right),\end{aligned}$$

and $\tilde{\phi} = g \circ \phi$. Let

$$m^\dagger(\theta_0, F_0; D)[h] = m_2(\theta_0, F_0; D)[h] - \tilde{m}^*(\theta_0, F_0; D)[h] + m^{**}(\theta_0, F_0; D)[h].$$

Then $P\big(m_{11}(\theta_0, F_0; D)[\boldsymbol{h}, \widehat{\theta}_n - \theta_0] + m_{12}(\theta_0, F_0; D)[\boldsymbol{h}, \widehat{\theta}_n - \theta_0]\big) = 0$, and

$$\begin{aligned}-\dot{G}_{1,\theta_0,F_0}(\widehat{\theta}_n - \theta_0)[\boldsymbol{h}] &= -P\big(m_{21}(\theta_0, F_0; D)[\boldsymbol{h}, \widehat{\theta}_n - \theta_0] + m_{22}(\theta_0, F_0; D)[\boldsymbol{h}, \widehat{\theta}_n - \theta_0]\big)\\ &= \mathbb{P}_n(\tilde{m}^*(\theta_0, F_0; D)[h]) + P\left(\sum_{j=1}^{K} \frac{\triangle N_j}{\big(\triangle\Lambda_{0j}(U)\big)^2} \triangle h_j(U)\big(\triangle\hat{\Lambda}_{nj}(U) - \triangle\Lambda_{0j}(U)\big)\right),\end{aligned}$$

$$G_n(\theta_0, F_0)[\boldsymbol{h}] + \dot{G}_{2,\theta_0,F_0}(\widehat{F}_n - F_0)[\boldsymbol{h}] = \mathbb{P}_n\Big(m_2(\theta_0, F_0; D)[h] + m^{**}(\theta_0, F_0; D)[h]\Big).$$

Applying (S4.4) again, we get

$$\sqrt{n}\int \frac{\triangle\hat{\Lambda}_n(t,s) - \triangle\Lambda_0(t,s)}{\triangle\Lambda_0(t,s)} e^{\beta_0 z} \triangle h(t,s) d\nu_1(s,t,z) \xrightarrow{d} N(0, \sigma_1^2[h]),$$

where $\triangle\Lambda(t,s) = \Lambda(t) - \Lambda(s)$, $\triangle h(t,s) = h(t) - h(s)$ and $\sigma_1^2[h] = E(m^\dagger(\theta_0, F_0; D)[h])^2$.

This completes the proof of Theorem 3.3. □

Proof of Corollary 3.1

Proof. Under Conditions (C4), (C5) and (C13), Lemma A.3 of Kong et. al. (2018) gives that $\widehat{F}_n$ is $n^{1/2}-$consistent in a finite interval. Moreover,

$$\int_0^\tau \psi(u;D)d\big[\widehat{F}_n(u|Z) - F_0(u|Z)\big] = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \tilde{\psi}(u;D)dO(u;D;\tilde{D}_i),$$

where $\tilde{\psi}(u;D) = g\circ\psi = \psi(u;D)e^{\zeta_0^T Z}$ and $O(u;D;\tilde{D}_i) = \overline{F}_0(u|Z)A_1(F_0;u,Z;\tilde{Z}_i,\tilde{\Delta}_i,\tilde{Y}_i)$ with A_1 is defined as in Lemma A.3 of Kong et al. (2018). Therefore, the asymptotic normality in Theorem 3.3 holds. This completes the proof of Corollary 3.1. □

References

Kong, S., Nan, B., Kalbfleisch, J. D., Saran, R. and Hirth, R. (2018). Conditional modeling of longitudinal data with terminal event. *Journal of the American Statistical Association* **113**, 357–368.

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.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge.

van der Vaart, A. W. (2002). *Semiparametric Statistics.* in Lectures on Probability Theory and Statistics, Ecole dÉte de Probabilites de Saint-Flour XXIX99, ed. P. Bernard, Berlin: Springer-Verlag, 330–457.

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

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

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