Statistica Sinica

Semiparametric Inference for Logitudinal Data with Informative Observation Times and Terminal Event

Shirong Deng, Kin-yat Liu, Wen Su and Xingqiu Zhao

Wuhan University, The Hong Kong Polytechnic University

and University of Hong Kong

Supplementary Materials

Appendix: Proofs

For simplicity of notation, we adopt the empirical process notation of van der Vaart and Wellner (1996) throughout the Appendix. Suppose $X_1, \cdots, X_n$ are independently and identically distributed random variable that follow the distribution P on a measurable space $(\mathcal{X}, \mathcal{A})$. For a measurable function $f : \mathcal{X} \to \mathcal{R}$, we denote $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(X_i)$, $Pf = \int f dP$, and $\mathbb{G}_n f = n^{-\frac{1}{2}}\sum_{i=1}^n \{f(X_i) - Pf\} = \sqrt{n}(\mathbb{P}_n - P)f$. We use the symbol $\lesssim$ to denote the left-hand side is bounded above by a constant times the right-hand side. For any functions μ_1, $\mu_2 \in \mathcal{F}_r$, define $d_1(\mu_1, \mu_2) = \|\mu_1 - \mu_2\|_2$. For the nuisance functional parameter F, it is assumed to be belonged to the following infinite-dimensional parameter space:

$$\mathcal{F}_F = \{F : F(\cdot|\boldsymbol{x}) \text{ is a conditional distribution function on } [0, \tau] \text{ for each } \boldsymbol{x} \in \mathcal{X}\},$$

where $\mathcal{X}$ is the support of the distribution of $\boldsymbol{X}$, and $\|F - F_0\|_\infty = \sup_{u,\boldsymbol{x}} |F(u|\boldsymbol{x}) - F_0(u|\boldsymbol{x})|$ for $F \in \mathcal{F}_F$.

In the Appendix, first we give some lemmas and theorems that will be used in the proof of the theorems in Section 4.

Theorem A.1. *Let $x \mapsto m_{\theta,\eta}(x)$ be measurable functions indexed by parameters (θ, η), and consider* *estimators* $\hat{\theta}_n$ *contained in a set* Θ_n *that, for a given* $\hat{\eta}_n$ *contained in a set* $\mathcal{H}_n$*, maximize the map* $\theta \mapsto \mathbb{P}_n m_{\theta,\hat{\eta}_n}$*. The sets* Θ_n *and* $\mathcal{H}_n$ *need not be metric spaces, but instead we measure the discrepancies between* $\hat{\theta}_n$ *and* θ_0*, and* $\hat{\eta}_n$ *and a limit value* η_0*, by nonnegative functions* $\theta \mapsto d_\eta(\theta,\theta_0)$ *and* $\eta \mapsto d(\eta,\eta_0)$*, which may be arbitrary.*

Assume that for $\phi_n : (0,\infty) \mapsto \mathbb{R}$ *such that* $\delta \mapsto \phi_n(\delta)/\delta^\beta$ *is decreasing to some* $\beta < 2$*, every* $(\theta,\eta) \in \Theta_n \times \mathcal{H}_n$ *with* θ *in a sufficient small neighborhood of* θ_0*, every sufficiently small* $\delta > 0$*,*

$$P(m_{\theta,\eta} - m_{\theta_0,\eta}) \lesssim -d_\eta^2(\theta,\theta_0) + d(\eta,\eta_0)d_\eta(\theta,\theta_0), \tag{A.1}$$

$$E^* \sup_{d_\eta(\theta,\theta_0)<\delta,(\theta,\eta)\in\Theta_n\times\mathcal{H}_n} |\mathbb{G}_n(m_{\theta,\eta} - m_{\theta_0,\eta})| \lesssim \phi_n(\delta). \tag{A.2}$$

Let $\delta_n > 0$ *satisfy* $\phi_n(\delta_n) \leq \sqrt{n}\delta_n^2$ *for every* n*. If* $\mathbb{P}_n m_{\hat{\theta}_n,\hat{\eta}_n} \geq \mathbb{P}_n m_{\theta_0,\hat{\eta}_n} - O_p(\delta_n^2)$ *and* $\hat{\theta}_n$ *converges in outer probability to* θ_0 $(d_{\hat{\eta}_n}(\hat{\theta}_n,\theta_0) = O_p^*(1))$*, then* $d_{\hat{\eta}_n}(\hat{\theta}_n,\theta_0) = O_p^*(\delta_n + d(\hat{\eta}_n,\eta_0))$*. Here,* P^* *and* E^* *are outer probability and expectation as defined in page 258 of van der Vaart (1998).*

Proof: This theorem is similar to Theorem 5.55 of van der Vaart (1998) and Theorem 8.4 of van der Vaart (2002) and thus the proof is omitted.

Lemma A.1. *Under the conditions (C2)-(C4), the class of functions defined as*

$$\{L(\boldsymbol{\theta},\mu,F;\boldsymbol{O}) : \boldsymbol{\theta} \in \Theta, \mu \in \mathcal{F}_r, F \in \mathcal{F}_F\}$$

is a Donsker class.

Proof: Note that

$$\begin{aligned} L(\boldsymbol{\theta}, \mu, F; \boldsymbol{O}) =& \Delta \int_0^\tau \left[Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(\tilde{U} - t) \right]^2 dN(t) \\ &+ (1-\Delta) \int_0^\tau \frac{\int_{\tilde{U}}^\tau \left[Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(u - t) \right]^2 dF(u|\boldsymbol{X})}{1 - F(\tilde{U}|\boldsymbol{X})} dN(t). \end{aligned}$$

It is easy to see that $\mathcal{F}_r \subset C^r[0, \tau]$ is a Donsker Class according to the arguments in P157 of van der Vaart and Wellner (1996). Then by Theorem 2.10.6 of van der Vaart and Wellner (1996), $\{[Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(\tilde{U} - t)]^2 : \boldsymbol{\theta} \in \Theta, \mu \in \mathcal{F}_r\}$ is Donsker since $[Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(\tilde{U} - t)]^2$ is Lipschitz for $\boldsymbol{\theta}$ and μ. Thus $\{\int_0^\tau [Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(\tilde{U} - t)]^2 dN(t) : \boldsymbol{\theta} \in \Theta, \mu \in \mathcal{F}_r\}$ is Donsker by the permanence of the Donsker property for the closure of the convex hull according to Theorem 2.10.3 of van der Vaart and Wellner (1996). And $\mathcal{F}_F$ is a Donsker class.

By integration by parts,

$$\begin{aligned} &\int_{\tilde{U}}^\tau \{Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(u - t)\}^2 dF(u|\boldsymbol{X}) \\ =& \{Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(\tau - t)\}^2 F(\tau|\boldsymbol{X}) - \{Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(\tilde{U} - t)\}^2 F(\tilde{U}|\boldsymbol{X}) \\ &+ 2 \int_{\tilde{U}}^\tau \{Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(u - t)\} \mu'(u - t) F(u|\boldsymbol{X}) du, \end{aligned}$$

where μ' is the first derivative for the function μ. By the permanence of the Donsker property for the closure of the convex hull according to Theorem 2.10.3 of van der Vaart and Wellner (1996), the permanence of the Donsker property for the addition and multiplication of Donsker classes according to Theorem 2.10.7 and Theorem 2.10.10 of van der Vaart and Wellner (1996), and the boundedness of $F(\tau|\boldsymbol{X})$ and $F(\tilde{U}|\boldsymbol{X})$, we have $\{\int_{\tilde{U}}^\tau [Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(u - t)]^2 dF(u|\boldsymbol{X}) : \boldsymbol{\theta} \in \Theta, \mu \in \mathcal{F}_\mu, F \in \mathcal{F}_F\}$ is a Donsker class. Thus, $\{\int_{\tilde{U}}^\tau [Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(u - t)]^2 dF(u|\boldsymbol{X}) / \{1 - F(\tilde{U}|\boldsymbol{X})\} : \boldsymbol{\theta} \in \Theta, \mu \in \mathcal{F}_\mu, F \in \mathcal{F}_F\}$

is a Donsker class by the boundedness of $1 - F(\tilde{U}|\boldsymbol{X})$ from condition C4. Finally, $\{L(\boldsymbol{\theta}, \mu, F; \boldsymbol{O}) : \boldsymbol{\theta} \in \Theta, \mu \in \mathcal{F}_\mu, F \in \mathcal{F}_F\}$ is a Donsker class according to the fact that the classes of the indicator functions $\{\Delta = I(U_i \leq C_i)\}$ and $\{1 - \Delta\}$ are both Donsker and Theorem 2.10.7 of van der Vaart and Wellner (1996). □

Lemma A.2. *Suppose the conditions (C2)-(C4) and (C8) hold.*

(i) For any differentiable function h, we have

$$\begin{aligned}&\left|P\left\{\int_0^\tau (1-\Delta)\int_{\tilde{U}}^\tau h(u-t)\left[\frac{dF(u|\boldsymbol{X})}{1-F(\tilde{U}|\boldsymbol{X})} - \frac{dF(u|\boldsymbol{X})}{1-F(\tilde{U}|\boldsymbol{X})}\right]dN(t)\right\}\right| \\ \lesssim & E\left\{\sum_{j=1}^K \xi(T_{K,j})\left[|h(U-T_{K,j})| + |h'(U-T_{K,j})|\right]\right\}\|F-F_0\|_\infty\end{aligned}$$

for all $F \in \mathcal{F}_F$.

(ii)

$$|P\{L(\boldsymbol{\theta}, \mu, F; \boldsymbol{O}) - L(\boldsymbol{\theta}, \mu, F_0; \boldsymbol{O})\}| \lesssim \|F - F_0\|_\infty$$

for all $\boldsymbol{\theta} \in \Theta$, $\mu \in \mathcal{F}_r$, $F \in \mathcal{F}_F$.

(iii)

$$\begin{aligned}&P\{L(\boldsymbol{\theta}, \mu, F; \boldsymbol{O}) - L(\boldsymbol{\theta}_0, \mu_0, F; \boldsymbol{O})\} \\ =&P\int_0^\tau \Big[\Delta\{(\boldsymbol{\theta}_0-\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{X},t) + (\mu_0-\mu)(\tilde{U}-t)\}^2 \\ &+ \frac{(1-\Delta)}{1-F(\tilde{U}|\boldsymbol{X})}\int_{\tilde{U}}^\tau \{(\boldsymbol{\theta}_0-\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{X},t) + (\mu_0-\mu)(u-t)\}^2 dF(u|\boldsymbol{X})\Big]dN(t)\end{aligned}$$

for all $\boldsymbol{\theta} \in \Theta$, $\mu \in \mathcal{F}_r$, $F \in \mathcal{F}_F$.

Proof: (i) Note that

$$\begin{aligned}
&\left|P\left\{\int_0^\tau (1-\Delta)\int_{\tilde{U}}^\tau h(u-t)\left[\frac{dF(u|\boldsymbol{X})}{1-F(\tilde{U}|\boldsymbol{X})}-\frac{dF_0(u|\boldsymbol{X})}{1-F_0(\tilde{U}|\boldsymbol{X})}\right]dN(t)\right\}\right| \\
\leq&\left|P\left\{\sum_{j=1}^K(1-\Delta)\xi(T_{K,j})\int_{\tilde{U}}^\tau h(u-T_{K,j})\frac{d\{F(u|\boldsymbol{X})-F_0(u|\boldsymbol{X})\}}{1-F_0(\tilde{U}|\boldsymbol{X})}\right\}\right| \\
&+\left|P\left\{\sum_{j=1}^K(1-\Delta)\xi(T_{K,j})\int_{\tilde{U}}^\tau h(u-T_{K,j})\frac{\{F(u|\boldsymbol{X})-F_0(u|\boldsymbol{X})\}dF_0(u|\boldsymbol{X})}{\{1-F_0(\tilde{U}|\boldsymbol{X})\}\{1-F(\tilde{U}|\boldsymbol{X})\}}\right\}\right| \\
=&I_1+I_2.
\end{aligned}$$

We can show that

$$I_1 \leq E\left\{\sum_{j=1}^K \xi(T_{K,j})\left|h'(U-T_{K,j})\right|\right\}\|F-F_0\|_\infty,$$

and

$$I_2 \leq E\left\{\sum_{j=1}^K \xi(T_{K,j})\left|h(U-T_{K,j})\right|\right\}\|F-F_0\|_\infty.$$

Thus, (i) holds. The proofs of (ii) and (iii) are straightforward and thus are omitted.

By Corollary 6.21 of Schumaker(2007), we have the following lemma with the proof omitted.

Lemma A.3. *For $\mu_0 \in \mathcal{F}_r$, there exists a function $\mu_n \in \Psi_{l,\mathcal{I}}$ with $l \geq r$, such that*

$$\|\mu_n-\mu_0\|_\infty = O(n^{-\nu r}), \|\mu_n'-\mu_0'\|_\infty = O(n^{-\nu(r-1)}),$$

where $\|\cdot\|_\infty$ is the sup-norm, μ_n' and μ_0' are the first derivatives for μ_n and μ_0, respectively.

Lemma A.4. *Under condition (C5), $PL(\boldsymbol{\theta},\mu,F_0;\boldsymbol{O})$ has a unique minimizer $(\boldsymbol{\theta}_0,\mu_0)$.*

Proof: Note that under model (2.1),

$$
\begin{aligned}
& PL(\boldsymbol{\theta}, \mu, F_0; \boldsymbol{O}) - PL(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) \\
&= E\left[\sum_{j=1}^{K} \xi(T_{K,j}) \left\{ Y(T_{K,j}) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - \mu(U - T_{K,j}) \right\}^2 \right] \\
&\quad - E\left[\sum_{j=1}^{K} \xi(T_{K,j}) \left\{ Y(T_{K,j}) - \boldsymbol{\theta}_0' \boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - \mu_0(U - T_{K,j}) \right\}^2 \right] \\
&= E\left[\sum_{j=1}^{K} \xi(T_{K,j}) \left\{ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - (\mu - \mu_0)(U - T_{K,j}) \right\}^2 \right] \\
&\geq 0.
\end{aligned}
$$

Thus the lemma follows under condition (C5).

Proof of Theorem 1

By Lemma A.4,

$$
\inf_{(\boldsymbol{\theta}, \mu): d((\boldsymbol{\theta}, \mu), (\boldsymbol{\theta}_0, \mu_0)) > \delta} PL(\boldsymbol{\theta}, \mu, F_0; \boldsymbol{O}) > PL(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) \tag{A.3}
$$

hold for every $\delta > 0$. By the definition of $(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n)$,

$$
\begin{aligned}
& \mathbb{P}_n L(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n; \boldsymbol{O}) \\
\leq & \mathbb{P}_n L(\boldsymbol{\theta}_0, \mu_n, \hat{F}_n; \boldsymbol{O}) \\
= & PL(\boldsymbol{\theta}_0, \mu_n, \hat{F}_n; \boldsymbol{O}) + o_p(1) \\
\lesssim & PL(\boldsymbol{\theta}_0, \mu_n, F_0; \boldsymbol{O}) + \|\hat{F}_n - F_0\|_\infty + o_p(1) \\
\lesssim & PL(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) + d_1^2(\mu_n, \mu_0) + o_p(1) \\
= & \mathbb{P}_n L(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) + o_p(1),
\end{aligned} \tag{A.4}
$$

where μ_n is defined in Lemma A.3, the second equality is obtained by Lemma A.1 that the class of functions $\{L(\boldsymbol{\theta}, \mu, F; O) : \boldsymbol{\theta} \in \Theta, \mu \in \mathcal{F}_\mu, F \in \mathcal{F}_F\}$ is a Donsker class, hence it is Glivenko-Cantelli. The third and forth inequalities are obtained by Lemma A.2 and the consistency for $\hat{F}_n(\cdot)$, and the last equality is obtained by Lemma A.1 and the fact that $d_1^2(\mu_n, \mu_0) = \|\mu_n - \mu_0\|_2^2 \lesssim \|\mu_n - \mu_0\|_\infty^2 = O(n^{-2\nu r}) = o(1)$. Then, we have

$$\begin{aligned}
0 \geq & PL(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) - PL(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, F_0; \boldsymbol{O}) \\
= & \mathbb{P}_n L(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) - \mathbb{P}_n L(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, F_0; \boldsymbol{O}) + o_p(1) \\
\geq & \mathbb{P}_n L(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n; \boldsymbol{O}) - \mathbb{P}_n L(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, F_0; \boldsymbol{O}) + o_p(1) \\
= & PL(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n; \boldsymbol{O}) - PL(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, F_0; \boldsymbol{O}) + o_p(1) \\
= & o_p(1),
\end{aligned} \tag{A.5}$$

where the second and the last but one equalities are obtained by Lemma A.1, the third inequality is obtained by (A.4), the last equality is obtained by Lemma A.2 and the consistency for $\hat{F}_n(\cdot)$.

By the inequality (A.3), for every $\delta > 0$, the event $\{d((\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n), (\boldsymbol{\theta}_0, \mu_0)) > \delta\}$ is contained in the event $\{PL(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, F_0; \boldsymbol{O}) > PL(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O})\}$. The latter sequence of the events going to a null event in view of inequality (A.5), which yields the almost surely convergence of $(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n)$, thus the convergence in probability holds.

Proof of Theorem 2

Our proof is based on Theorem A.1. We'll prove that the conditions in Theorem A.1 are satisfied for our estimators. Denote $g(\boldsymbol{Z}, t, u) = \boldsymbol{\theta}'\boldsymbol{Z}(\boldsymbol{X}, t) + \mu(u - t)$, $g_0(\boldsymbol{Z}, t, u) = \boldsymbol{\theta}_0'\boldsymbol{Z}(\boldsymbol{X}, t) + \mu_0(u - t)$,

$\hat{g}_n(\boldsymbol{Z},t,u)=\hat{\boldsymbol{\theta}}_n'\boldsymbol{Z}(\boldsymbol{X},t)+\hat{\mu}_n(u-t)$, $g_n(\boldsymbol{Z},t,u)=\boldsymbol{\theta}_0'\boldsymbol{Z}(\boldsymbol{X},t)+\mu_n(u-t)$, and

$$\begin{aligned}m(g,F;\boldsymbol{O})=&-\frac{1}{2}L(\boldsymbol{\theta},\mu,F;\boldsymbol{O})\\=&-\frac{1}{2}\int_0^\tau\left[\Delta\{Y(t)-g(\boldsymbol{Z},t,\tilde{U})\}^2+\frac{1-\Delta}{1-F(\tilde{U}|\boldsymbol{X})}\int_{\tilde{U}}^\tau\{Y(t)-g(\boldsymbol{Z},t,u)\}^2dF(u|\boldsymbol{X})\right]dN(t).\end{aligned}$$

In our proof, g and F are the "θ" and "η" in Theorem A.1, respectively,

$$\begin{aligned}d(g,g_0)=\|g-g_0\|_2=&\|(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{Z}(\boldsymbol{X},t)+(\mu-\mu_0)(u-t)\|_2\\=&\left\{E\left[\sum_{j=1}^K\xi(T_{K,j})\{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{Z}(\boldsymbol{X},T_{K,j})+(\mu-\mu_0)(U-T_{K,j})\}^2\right]\right\}^{\frac{1}{2}},\end{aligned}$$

and $d(F,F_0)=\|F-F_0\|_\infty$.

For any $\eta>0$, let

$$\mathcal{F}_\eta=\{g(\boldsymbol{Z},t,u)=\boldsymbol{\theta}'\boldsymbol{Z}(\boldsymbol{X},t)+\mu(u-t):\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\eta,d_1(\mu,\mu_0)\le\eta,\mu\in\Psi_n,\boldsymbol{\theta}\in\Theta\}.$$

By Lemma A.2 (iii), we have $P\{m(g,F_0,;\boldsymbol{O})-m(g_0,F_0;\boldsymbol{O})\}=-\frac{1}{2}\|g-g_0\|_2^2$. And by Lemma A.2 (i) and (iii), for $g\in\mathcal{F}_\eta$ with sufficiently small η,

$$\begin{aligned}&P\{m(g,F;\boldsymbol{O})-m(g_0,F;\boldsymbol{O})\}-P\{m(g,F_0;\boldsymbol{O})-m(g_0,F_0;\boldsymbol{O})\}]\\=&-\frac{1}{2}P\int_0^\tau(1-\Delta)\int_{\tilde{U}}^\tau\{(\boldsymbol{\theta}_0-\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{X},t)+(\mu_0-\mu)(u-t)\}^2\left[\frac{dF(u|\boldsymbol{X})}{1-F(\tilde{U}|\boldsymbol{X})}-\frac{dF_0(u|\boldsymbol{X})}{1-F_0(\tilde{U}|\boldsymbol{X})}\right]dN(t)\\\lesssim&E\left\{\sum_{j=1}^K\xi(T_{K,j})\left[(g-g_0)^2(\boldsymbol{Z},T_{K,j},U)+\left|2(g-g_0)(\boldsymbol{Z},T_{K,j},U)(\mu'-\mu_0')(U-T_{K,j})\right|\right]\right\}\|F-F_0\|_\infty\\\lesssim&\|F-F_0\|_\infty\|g-g_0\|_2.\end{aligned}$$

Therefore,

$$
\begin{aligned}
&P\left\{m(g,F;\boldsymbol{O})-m(g_0,F;\boldsymbol{O})\right\}\\
=&P\left\{m(g,F_0;\boldsymbol{O})-m(g_0,F_0;\boldsymbol{O})\right\}+\left[P\left\{m(g,F;\boldsymbol{O})-m(g_0,F;\boldsymbol{O})\right\}-P\left\{m(g,F_0;\boldsymbol{O})-m(g_0,F_0;\boldsymbol{O})\right\}\right]\\
\lesssim&-\|g-g_0\|_2^2+\|F-F_0\|_\infty\|g-g_0\|_2.
\end{aligned}
$$

By the calculation of Shen and Wong (1994, P597), for $\eta>0$ and any $\varepsilon<\eta$, $\log N_{[]}(\varepsilon,\Psi_n,\|\cdot\|_2)\le C_1q_n\log(\eta/\varepsilon)$, where $q_n=m_n+l$ is the number of spline basis functions and C_1 is a constant. Similar to Lemma A.2 in Huang (1999, P1557), for any $\varepsilon\le\eta$, $\log N_{[]}(\varepsilon,\mathcal{F}_\eta,\|\cdot\|_2)\le C_2q_n\log(\eta/\varepsilon)$ for a constant C_2. Thus, for $\varepsilon>0$, there exists a set of brackets $\{[g_i^l,g_i^r],i=1,\cdots,(\eta/\varepsilon)^{C_2q_n}\}$ such that, for each $g\in\mathcal{F}_\eta$, there is a $[g_s^l,g_s^r]$, s.t. $g_s^l(\boldsymbol{Z},t,u)\le g(\boldsymbol{Z},t,u)\le g_s^r(\boldsymbol{Z},t,u)$ for all $\boldsymbol{Z}$ and $\tau_0\le t\le u\le\tau$ and $\|g_s^r-g_s^l\|_2\le\varepsilon$.

For fixed F, define

$$
\mathcal{M}_\eta(F)=\{m(g,F;\boldsymbol{O})-m(g_0,F;\boldsymbol{O}):g\in\mathcal{F}_\eta\}.
$$

For $i=1,\cdots,(\eta/\varepsilon)^{C_2q_n}$, define

$$
\begin{aligned}
m_i^l(u;\boldsymbol{O})=&\sum_{j=1}^{K}\xi(T_{K,j})\big[2Y(T_{K,j})g_0(\boldsymbol{Z},T_{K,j},u)-g_0^2(\boldsymbol{Z},T_{K,j},u)+\{|g_i^l|\wedge|g_i^r|\}^2(\boldsymbol{Z},T_{K,j},u)\\
&-2Y(T_{K,j})\{g_i^r(\boldsymbol{Z},T_{K,j},u)I(Y(T_{K,j})\ge0)+g_i^l(\boldsymbol{Z},T_{K,j},u)I(Y(T_{K,j})<0)\}\big],
\end{aligned}
$$

and

$$m_i^r(u;\boldsymbol{O}) = \sum_{j=1}^{K} \xi(T_{K,j})\big[2Y(T_{K,j})g_0(\boldsymbol{Z},T_{K,j},u) - g_0^2(\boldsymbol{Z},T_{K,j},u) + \{|g_i^l| \vee |g_i^r|\}^2(\boldsymbol{Z},T_{K,j},u)$$
$$- 2Y(T_{K,j})\{g_i^l(\boldsymbol{Z},T_{K,j},u)I(Y(T_{K,j}) \geq 0) + g_i^r(\boldsymbol{Z},T_{K,j},u)I(Y(T_{K,j}) < 0)\}\big],$$

where $a \vee b = \max\{a, b\}$, and $a \wedge b = \min\{a, b\}$. Moreover, let

$$M_i^r(F,\boldsymbol{O}) = -\frac{1}{2}\left[\Delta m_i^l(\tilde{U};\boldsymbol{O}) + \frac{1-\Delta}{1-F(\tilde{U}|\boldsymbol{X},)}\int_{\tilde{U}}^{\tau} m_i^l(u;\boldsymbol{O})dF(u|\boldsymbol{X})\right],$$

and

$$M_i^l(F,\boldsymbol{O}) = -\frac{1}{2}\left[\Delta m_i^r(\tilde{U};\boldsymbol{O}) + \frac{1-\Delta}{1-F(\tilde{U}|\boldsymbol{X})}\int_{\tilde{U}}^{\tau} m_i^r(u;\boldsymbol{O})dF(u|\boldsymbol{X})\right].$$

Then $M_i^l(F,\boldsymbol{O}) \leq M_i^r(F,\boldsymbol{O})$ and by some calculations, we can show that $\|M_i^r(F,\boldsymbol{O}) - M_i^l(F,\boldsymbol{O})\|_{P,B}^2 \leq C_3\varepsilon^2$ with a positive constant C_3, where $\|\cdot\|_{P,B}$ is the "Bernstein norm" defined as $\|f\|_{P,B} = \{2P(e^{|f|}-1-|f|)\}^{1/2}$ (see van der Varrt and Wellner (1996, p.324)). Thus $\{(M_i^l(F,\boldsymbol{O}), M_i^r(F,\boldsymbol{O})), i = 1, \cdots, (\eta/\varepsilon)^{C_2 q_n}\}$ is the set of brackets for $\mathcal{M}_\eta(F)$, which implies that $\log N_{[]}(\varepsilon, \mathcal{M}_\eta(F), \|\cdot\|_{P,B}) \leq C_2 q_n \log(\eta/\varepsilon)$ for a constant C_2. Therefore, for

$$\mathcal{L}_\eta(F) = \{m(g,F;\boldsymbol{O}) - m(g_0,F;\boldsymbol{O}) : d((\boldsymbol{\theta},\mu),(\boldsymbol{\theta}_0,\mu_0)) \leq \eta, \mu \in \Psi_n, \boldsymbol{\theta} \in \Theta\},$$

we have $\log N_{[]}(\varepsilon, \mathcal{L}_\eta(F), \|\cdot\|_{P,B}) \leq C_4 q_n \log(\eta/\varepsilon)$ for a constant C_4. By Lemma 3.4.3 of van der Varrt and Wellner (1996), we obtain

$$E\|\sqrt{n}(\mathbb{P}_n - P)\|_{\mathcal{L}_\eta(F)} \leq C_5 \tilde{J}_{[]}(\eta, \mathcal{L}_\eta(F), \|\cdot\|_{P,B})\left\{1 + \frac{\tilde{J}_{[]}(\eta, \mathcal{L}_\eta(F), \|\cdot\|_{P,B})}{\sqrt{n}\eta^2}M\right\}, \qquad \text{(A.6)}$$

where C_5 and M are constants and $\|\sqrt{n}(\mathbb{P}_n - P)\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\sqrt{n}(\mathbb{P}_n - P)f|$, and

$$\tilde{J}_{[]}(\eta, \mathcal{L}_\eta(F), \|\cdot\|_{P,B}) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{L}_\eta(F), \|\cdot\|_{P,B})\}^{1/2} d\varepsilon \lesssim q_n^{1/2}\eta.$$

The right hand side of equation (A.6) yields $\phi_n(\eta) = c(q_n^{1/2}\eta + q_n/\sqrt{n})$. It is easy to see that $\phi_n(\eta)/\eta$ is decreasing in η, and $\phi_n(\delta_n) = \delta_n q_n^{1/2} + q_n/\sqrt{n} \le \sqrt{n}\delta_n^2$ for $\delta_n = n^{-\frac{1-\nu}{2}}$ and $0 < \nu < 1/2$.

Next, we show that $\mathbb{P}_n[m(\hat{g}_n, \hat{F}_n; \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})] \ge -O_p(\delta_n^2)$.

Actually

$$\begin{aligned}
&\mathbb{P}_n[m(\hat{g}_n, \hat{F}_n; \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})] \\
=&(\mathbb{P}_n - P)\{m(g_n, \hat{F}_n; \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})\} + P\{m(g_n, \hat{F}_n; \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})\} \\
&+ \mathbb{P}_n\{m(\hat{g}_n, \hat{F}_n; \boldsymbol{O}) - m(g_n, \hat{F}_n; \boldsymbol{O})\} \\
\ge&(\mathbb{P}_n - P)\{m(g_n, \hat{F}_n; \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})\} + P\{m(g_n, \hat{F}_n; \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})\} \\
=&I_1 + I_2.
\end{aligned}$$

For I_1, it can be written as

$$I_1 = n^{-\nu r+\varepsilon}(\mathbb{P}_n - P)\left\{\frac{m(g_n, \hat{F}_n, \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})}{n^{-\nu r+\varepsilon}}\right\}$$

for any $0 < \varepsilon < 1/2 - \nu r$. Now define the class

$$\mathcal{C}(\hat{F}_n) = \left\{\frac{m(\tilde{g}_0, \hat{F}_n; \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})}{n^{-\nu r+\varepsilon}} : \mu \in \Psi_n, \|\mu - \mu_0\|_\infty = O(n^{-\nu r})\right\},$$

where $\tilde{g}_0(\boldsymbol{Z}, t, u) = \boldsymbol{\theta}_0' \boldsymbol{Z}(\boldsymbol{X}, t) + \mu(u - t)$. Then it can be easily prove that $\mathcal{C}(\hat{F}_n)$ is Donsker by the boundedness conditons (C3) and (C4). And for any $f \in \mathcal{C}(\hat{F}_n)$, it has $Pf^2 \to 0$ as $n \to \infty$. Then using corollary 2.3.12 of van der Varrt and Wellner (1996), we obtain that

$$\sqrt{n}(\mathbb{P}_n - P)\frac{m(g_n, \hat{F}_n; \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})}{n^{-\nu r + \varepsilon}} = o_p(1).$$

Thus $I_1 = o_p(n^{-\nu r + \varepsilon})n^{-1/2} = o_p(n^{-2\nu r})$.

For I_2,

$$\begin{aligned} I_2 =& P\{m(g_n, F_0; \boldsymbol{O}) - m(g_0, F_0; \boldsymbol{O})\} \\ &+ P\{m(g_0, F_0; \boldsymbol{O}) - m(g_n, F_0; \boldsymbol{O})\} - P\{m(g_0, \hat{F}_n; \boldsymbol{O}) - m(g_n, \hat{F}_n; \boldsymbol{O})\} \\ \gtrsim& - d_1^2(\mu_n, \mu_0) - \|\hat{F}_n - F_0\|_\infty \|\mu_n - \mu_0\|_\infty \\ \gtrsim& - \|\mu_n - \mu_0\|_\infty^2 - \|\hat{F}_n - F_0\|_\infty \|\mu_n - \mu_0\|_\infty \\ =& - O_p(n^{-2\nu r} + n^{-(\frac{r}{1+2r} + \nu r)}), \end{aligned}$$

where the second inequality holds since by Lemma A.2(i) and (iii), and

$$\begin{aligned} & P\{m(g_0, \hat{F}_n; \boldsymbol{O}) - m(g_n, \hat{F}_n; \boldsymbol{O})\} - P\{m(g_0, F_0; \boldsymbol{O}) - m(g_n, F_0; \boldsymbol{O})\} \\ =& \frac{1}{2} P \int_0^\tau (1 - \Delta) \int_{\tilde{U}}^\tau (\mu_n - \mu_0)^2 (u - t) \left[\frac{d\hat{F}_n(u|\boldsymbol{X})}{1 - \hat{F}_n(\tilde{U}|\boldsymbol{X})} - \frac{dF_0(u|\boldsymbol{X})}{1 - F_0(\tilde{U}|\boldsymbol{X})} \right] dN(t) \\ \lesssim& \|\hat{F}_n - F_0\|_\infty \left[E \left\{ \sum_{j=1}^K \xi(T_{K,j}) \left[|(\mu_n - \mu_0)^2 (U - T_{K,j})| + |2(\mu_n - \mu_0)(U - T_{K,j})(\mu_n' - \mu_0')(U - T_{K,j})| \right] \right\} \right] \\ \lesssim& \|\hat{F}_n - F_0\|_\infty \|\mu_n - \mu_0\|_\infty. \end{aligned}$$

Therefore,

$$\mathbb{P}_n[m(\hat{g}_n, \hat{F}_n; \boldsymbol{O}) - m(g_0, \hat{F}_n; \boldsymbol{O})] \geq -O_p(n^{-2\nu r} + n^{-(\frac{r}{1+2r}+\nu r)}) \geq -O_p(n^{-2\min\{\nu r, \frac{r}{1+2r}\}}).$$

Finally, according to Theorem A.1, we have

$$\|\hat{g}_n - g_0\|_2 = O_p(n^{-\min\{\nu r, \frac{r}{1+2r}, \frac{1-\nu}{2}\}} + \|\hat{F}_n - F_0\|_\infty\|) = O_p(n^{-\min\{\nu r, \frac{1-\nu}{2}, \frac{r}{1+2r}\}}).$$

Since

$$\begin{aligned}
\|\hat{g}_n - g_0\|_2^2 =& E\left\{\sum_{j=1}^{K} \xi(T_{K,j})[(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)'\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) + (\hat{\mu}_n - \mu_0)(U - T_{K,j})]^2\right\} \\
=& E\left\{\sum_{j=1}^{K} \xi(T_{K,j})[(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' E\{\boldsymbol{Z}(\boldsymbol{X}, T_{K,j})|K, \tilde{T}_K, U, C\} + (\hat{\mu}_n - \mu_0)(U - T_{K,j})]^2\right\} \\
&+ E\left\{\sum_{j=1}^{K} \xi(T_{K,j})[(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)'\{\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - E\{\boldsymbol{Z}(\boldsymbol{X}, T_{K,j})|K, \tilde{T}_K, U, C\}\}]^2\right\},
\end{aligned}$$

which yields that $d((\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n), (\boldsymbol{\theta}_0, \mu_0)) = O_p(n^{-\min\{\nu r, \frac{1-\nu}{2}\}})$ under Condition C6.

Theorem A.2. *Given n i.i.d. observations $O_1, \cdots, O_n$ of O, suppose that $(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n)$ of $(\boldsymbol{\theta}, \mu)$ are chosen to maximize the objective function*

$$L_n(\boldsymbol{\theta}, \mu, \hat{F}_n; \{O_i\}_{i=1}^n) = n^{-1}\sum_{i=1}^{n} m(\boldsymbol{\theta}, \mu, \hat{F}_n; O_i)$$

with respect to $(\boldsymbol{\theta}, \mu) \in \Theta \times \mathcal{F}_\mu$, where $\hat{F}_n$ is an estimator of the true parameter $F_0 \in \mathcal{F}_F$. Let $\{\mu_t\}$ be a parameter submodel in $\mathcal{F}_\mu$ passing through μ, that is, $\mu_t \in \mathcal{F}_\mu$ and $\mu_{t=0} = \mu$, and define $\mathcal{H}_\mu = \{h : h =$

$\frac{\partial \mu_t}{\partial t}|_{t=0}\}$ *as the collection of all directions to approach* μ*, and* $\mathcal{H} = \{(\boldsymbol{h}_1, h_2) : \boldsymbol{h}_1 \in \Theta, h_2 \in \mathcal{H}_\mu, \|\boldsymbol{h}_1\| \leq 1, \|h_2\|_\infty \leq 1\}$. *For any* $(\boldsymbol{h}_1, h_2) \in \mathcal{H}$*, define*

$$S_n(\boldsymbol{\theta}, \mu, F)[\boldsymbol{h}_1, h_2] = n^{-1} \sum_{i=1}^{n} \frac{\partial m(\boldsymbol{\theta} + \varepsilon \boldsymbol{h}_1, \mu_\varepsilon, F; O_i)}{\partial \varepsilon}\Big|_{\varepsilon=0} = \mathbb{P}_n \psi(\boldsymbol{\theta}, \mu, F; O)[\boldsymbol{h}_1, h_2],$$

and $S(\boldsymbol{\theta}, \mu, F)[\boldsymbol{h}_1, h_2] = P\psi(\boldsymbol{\theta}, \mu, F; O)[\boldsymbol{h}_1, h_2]$. *Consider the following conditions:*

B1. $\sqrt{n}(S_n - S)(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] - \sqrt{n}(S_n - S)(\boldsymbol{\theta}_0, \mu_0, F_0)[\boldsymbol{h}_1, h_2] = o_p(1)$*;*

B2. $S(\boldsymbol{\theta}_0, \mu_0, F_0)[\boldsymbol{h}_1, h_2] = 0$*,* $S_n(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] = o_p(n^{-\frac{1}{2}})$*;*

B3. $S(\boldsymbol{\theta}, \mu, F_0)$ *is Fréchet differentiable with respect to* $(\boldsymbol{\theta}, \mu)$ *at* $(\boldsymbol{\theta}_0, \mu_0)$ *with a continuous derivative* $\dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}$*, and is Fréchet differentiable with respect to* F *at* $(\boldsymbol{\theta}_0, \mu_0, F_0)$ *with continuous derivatives* $\dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}$*, respectively;*

B4. $S(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] - S(\boldsymbol{\theta}_0, \mu_0, F_0)[\boldsymbol{h}_1, h_2] - \dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \hat{\mu}_n - \mu_0)[\boldsymbol{h}_1, h_2] - \dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{F}_n - F_0)[\boldsymbol{h}_1, h_2] = o_p(n^{-\frac{1}{2}})$.

B5. $\sqrt{n}\dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{F}_n - F_0)[\boldsymbol{h}_1, h_2] + \sqrt{n}S_n(\boldsymbol{\theta}_0, \mu_0, F_0)[\boldsymbol{h}_1, h_2]$ *converges in distribution to a tight Gaussian process on* $l^\infty(\mathcal{H})$.

Then, under conditions (B1)-(B4),

$$\begin{aligned} & - \dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \hat{\mu}_n - \mu_0)[\boldsymbol{h}_1, h_2] \\ =& \dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{F}_n - F_0)[\boldsymbol{h}_1, h_2] + S_n(\boldsymbol{\theta}_0, \mu_0, F_0)[\boldsymbol{h}_1, h_2] + o_p(n^{-\frac{1}{2}}), \end{aligned} \tag{A.7}$$

and $-\sqrt{n}\dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \hat{\mu}_n - \mu_0)[\boldsymbol{h}_1, h_2]$ *converges in distribution to a tight Gaussian process on* $l^\infty(\mathcal{H})$.

Proof: By the conditions B2, B3 and B4,

$$S(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] - \dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \hat{\mu}_n - \mu_0)[\boldsymbol{h}_1, h_2] - \dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{F}_n - F_0)[\boldsymbol{h}_1, h_2] = o_p(n^{-\frac{1}{2}}). \tag{A.8}$$

And by B1 and B2,

$$-\sqrt{n}S(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] - \sqrt{n}S_n(\boldsymbol{\theta}_0, \mu_0, F_0)[\boldsymbol{h}_1, h_2] = o_p(1). \tag{A.9}$$

Combing the equations (A.8) and (A.9), equation (A.7) can be obtained. And the proposition is concluded by Assumption B5. □

Proof of Theorem 3

We show the theorem using Theorem A.2.

For our proposed methods, $L_n(\boldsymbol{\theta}, \mu, \hat{F}_n; \{\boldsymbol{O}_i\}_{i=1}^n) = n^{-1}\sum_{i=1}^n m(\boldsymbol{\theta}, \mu, \hat{F}_n; \boldsymbol{O}_i)$ with $m(\boldsymbol{\theta}, \mu, F; \boldsymbol{O}) = -\frac{1}{2}L(\boldsymbol{\theta}, \mu, F; \boldsymbol{O})$. Let

$$\mathcal{H} = \{(\boldsymbol{h}_1, h_2) : \boldsymbol{h}_1 \in \Theta, h_2 \in \mathcal{F}_r, \|\boldsymbol{h}_1\| \le 1, \|h_2\|_\infty \le 1\},$$

and denote $l^\infty(\mathcal{H})$ as the space of bounded functionals on $\mathcal{H}$ under the supermum norm $\|f\|_\infty = \sup_{h\in\mathcal{H}} |f(h)|$. For fixed F, we define a sequence of maps S_n mapping a neighborhood of $(\boldsymbol{\theta}_0, \mu_0)$, denoted by $\mathcal{U}$, in the parameter space for $(\boldsymbol{\theta}, \mu)$ into $l^\infty(\mathcal{H})$ as

$$S_n(\boldsymbol{\theta}, \mu, F)[\boldsymbol{h}_1, h_2] = n^{-1}\sum_{i=1}^n \frac{\partial m(\boldsymbol{\theta} + \varepsilon\boldsymbol{h}_1, \mu_\varepsilon, F; \boldsymbol{O}_i)}{\partial\varepsilon}\Big|_{\varepsilon=0} = \mathbb{P}_n\psi(\boldsymbol{\theta}, \mu, F; \boldsymbol{O})[\boldsymbol{h}_1, h_2],$$

where

$$
\begin{aligned}
&\psi(\boldsymbol{\theta}, \mu, F; \boldsymbol{O})[\boldsymbol{h}_1, h_2] \\
=&\int_0^{\tau} \Bigg[\Delta \left\{Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(\tilde{U} - t)\right\} \{\boldsymbol{h}_1' \boldsymbol{Z}(\boldsymbol{X}, t) + h_2(\tilde{U} - t)\} \\
&\quad + \frac{(1-\Delta)}{1 - F(\tilde{U}|\boldsymbol{X})} \int_{\tilde{U}}^{\tau} \left\{Y(t) - \boldsymbol{\theta}' \boldsymbol{Z}(\boldsymbol{X}, t) - \mu(u - t)\right\} \{\boldsymbol{h}_1' \boldsymbol{Z}(\boldsymbol{X}, t) + h_2(u - t)\} dF(u|\boldsymbol{X})\Bigg] dN(t).
\end{aligned}
$$

Correspondingly, we define the limit map S: $\mathcal{U} \longrightarrow l^{\infty}(\mathcal{H})$ as

$$
S(\boldsymbol{\theta}, \mu, F)[\boldsymbol{h}_1, h_2] = P\psi(\boldsymbol{\theta}, \mu, F; \boldsymbol{O})[\boldsymbol{h}_1, h_2].
$$

Next we verify the conditions in Theorem A.2.

Note that

$$
\begin{aligned}
&\sqrt{n}(S_n - S)(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] - \sqrt{n}(S_n - S)(\boldsymbol{\theta}_0, \mu_0, F_0)[\boldsymbol{h}_1, h_2] \\
=&\sqrt{n}(\mathbb{P}_n - P)\left(\psi(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n; \boldsymbol{O})[\boldsymbol{h}_1, h_2] - \psi(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_2]\right).
\end{aligned}
$$

For some $\delta > 0$, define

$$
\mathcal{F}_{\delta} = \{\psi(\boldsymbol{\theta}, \mu, F; \boldsymbol{O})[\boldsymbol{h}_1, h_2] - \psi(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_2] : \rho^*((\boldsymbol{\theta}, \mu, F), (\boldsymbol{\theta}_0, \mu_0, F_0)) < \delta, (\boldsymbol{h}_1, h_2) \in \mathcal{H}\},
$$

where $\rho^*((\boldsymbol{\theta}_1, \mu_1, F_1), (\boldsymbol{\theta}_2, \mu_2, F_2)) = \{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^2 + d_1^2(\mu_1, \mu_2) + \|F_1 - F_2\|_{\infty}^2\}^{1/2}$. It is easy to see that $\mathcal{F}_{\delta}$ is a Donsker class by the fact that $\mathcal{H}$ is a Donsker class based on the arguments in Page 154 - 157 of van der Vaart and Wellner (1996) and the similar arguments in the proof of Lemma A.1. And

$$
\begin{aligned}
& P|\psi(\boldsymbol{\theta},\mu,F;\boldsymbol{O})[\boldsymbol{h}_1,h_2]-\psi(\boldsymbol{\theta}_0,\mu_0,F_0;\boldsymbol{O})[\boldsymbol{h}_1,h_2]|^2\\
=& P\bigg| -\int_0^\tau \Delta\{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{Z}(\boldsymbol{X},t)+(\mu-\mu_0)(\tilde{U}-t)\}\{\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X},t)+h_2(\tilde{U}-t)\}dN(t)\\
& -\int_0^\tau (1-\Delta)\int_{\tilde{U}}^\tau \left\{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{Z}(\boldsymbol{X},t)+(\mu-\mu_0)(u-t)\right\}\{\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X},t)+h_2(u-t)\}\frac{dF_0(u|\boldsymbol{X})}{1-F_0(\tilde{U}|\boldsymbol{X})}dN(t)\\
& +\int_0^\tau (1-\Delta)\int_{\tilde{U}}^\tau \left\{Y(t)-\boldsymbol{\theta}'\boldsymbol{Z}(\boldsymbol{X},t)-\mu(u-t)\right\}\{\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X},t)+h_2(u-t)\}\frac{d\{F(u|\boldsymbol{X})-F_0(u|\boldsymbol{X})\}}{1-F(\tilde{U}|\boldsymbol{X})}dN(t)\\
& +\int_0^\tau (1-\Delta)\int_{\tilde{U}}^\tau \left\{Y(t)-\boldsymbol{\theta}'\boldsymbol{Z}(\boldsymbol{X},t)-\mu(u-t)\right\}\{\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X},t)+h_2(u-t)\}\\
& \qquad\qquad \times \frac{\{F(u|\boldsymbol{X})-F_0(u|\boldsymbol{X})\}dF_0(u|\boldsymbol{X})}{\{1-F(\tilde{U}|\boldsymbol{X})\}\{1-F_0(\tilde{U}|\boldsymbol{X})\}}dN(t)\bigg|^2\\
\lesssim & \rho^{*2}((\boldsymbol{\theta},\mu,F),(\boldsymbol{\theta}_0,\mu_0,F_0)),
\end{aligned}
$$

where the last inequality holds by using the similar arguments in the proof of Lemma A.2 under the boundedness condition (C3) and the boundedness of $1-F(\tilde{U}|\boldsymbol{X})$ and $1-F_0(\tilde{U}|\boldsymbol{X})$ by condition (C4). Thus condition B1 of Theorem A.2 holds by Lemma13.3 of Kosorok (2008).

For B2, $S(\boldsymbol{\theta}_0,\mu_0,F_0)[\boldsymbol{h}_1,h_2]=0$. Let h_{2n} be the B-spline approximation of h_2 with $\|h_2-h_{2n}\|_\infty =$

$O(n^{-\nu r})$ by Lemma A.3, then we have $S_n(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_{2n}] = 0$. Thus, for $(\boldsymbol{h}_1, h_2) \in \mathcal{H}$,

$$
\begin{aligned}
&\sqrt{n} S_n(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] \\
=&\sqrt{n}\{S_n(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] - S_n(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_{2n}]\} \\
=&\sqrt{n}(\mathbb{P}_n - P)\left\{\psi(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n; \boldsymbol{O})[\boldsymbol{h}_1, h_2] - \psi(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_2]\right\} \\
&- \sqrt{n}(\mathbb{P}_n - P)\left\{\psi(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n; \boldsymbol{O})[\boldsymbol{h}_1, h_{2n}] - \psi(\boldsymbol{\theta}_0, \mu_0, F_0, ; \boldsymbol{O})[\boldsymbol{h}_1, h_{2n}]\right\} \\
&+ \sqrt{n}\mathbb{P}_n\left\{\psi(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_2] - \psi(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_{2n}]\right\} \\
&+ \sqrt{n}P\left\{\psi(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_2] - \psi(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_{2n}]\right\} \\
&+ \sqrt{n}P\left\{\psi(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n; \boldsymbol{O})[\boldsymbol{h}_1, h_2] - \psi(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_2]\right\} \\
&- \sqrt{n}P\left\{\psi(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n; \boldsymbol{O})[\boldsymbol{h}_1, h_{2n}] - \psi(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_{2n}]\right\} \\
\triangleq& Q_{1n} - Q_{2n} + Q_{3n} + Q_{4n} + Q_{5n} - Q_{6n}.
\end{aligned}
$$

It follows from (B1) that both Q_{1n} and Q_{2n} are $o_p(1)$. Q_{3n} is $o_p(1)$ since

$$
\begin{aligned}
&P|\psi(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_2] - \psi(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O})[\boldsymbol{h}_1, h_{2n}|^2 \\
=&P\bigg|\int_0^\tau \bigg[\Delta\left\{Y(t) - \boldsymbol{\theta}_0'\boldsymbol{Z}(\boldsymbol{X}, t) - \mu_0(\tilde{U} - t)\right\}(h_2 - h_{2n})(\tilde{U} - t) + \frac{(1-\Delta)}{1 - F_0(\tilde{U}|\boldsymbol{X})} \\
&\times \int_{\tilde{U}}^\tau \left\{Y(t) - \boldsymbol{\theta}_0'\boldsymbol{Z}(\boldsymbol{X}, t) - \mu_0(u - t)\right\}(h_2 - h_{2n})(u - t)dF_0(u|\boldsymbol{X})\bigg]dN(t)\bigg|^2 \\
\lesssim&\|h_2 - h_{2n}\|_\infty^2 \to 0 \text{ as } n \to \infty.
\end{aligned}
$$

Furthermore, we can show that $|Q_{4n}| = o_p(1)$ and $|Q_{5n} - Q_{6n}| = o_p(1)$ when $\frac{1}{4r} \le \nu < \frac{1}{2}$. Thus $S_n(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] = o_p(n^{-1/2})$.

For B3, by the smoothness of $S(\boldsymbol{\theta}, \mu, F_0)[\boldsymbol{h}_1, h_2]$ with respect to $(\boldsymbol{\theta}, \mu)$, we have the Fréchet derivative of $S(\boldsymbol{\theta}, \mu, F_0)$ at $(\boldsymbol{\theta}_0, \mu_0)$, denoted by $\dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}$, is a map from the space $\{(\boldsymbol{\theta}, \mu) - (\boldsymbol{\theta}_0, \mu_0) : (\boldsymbol{\theta}, \mu) \in \mathcal{U}\}$ to $l^\infty(\mathcal{H})$ and

$$
\begin{aligned}
&\dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\boldsymbol{\theta} - \boldsymbol{\theta}_0, \mu - \mu_0)[\boldsymbol{h}_1, h_2] \\
=&\left.\frac{dS(\boldsymbol{\theta}_0 + \varepsilon(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \mu_0 + \varepsilon(\mu - \mu_0), F_0)[\boldsymbol{h}_1, h_2]}{d\varepsilon}\right|_{\varepsilon=0} \\
=&-P\int_0^\tau \Bigg[\Delta\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\boldsymbol{Z}(\boldsymbol{X}, t) + (\mu - \mu_0)(\tilde{U} - t)\right\}\{\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X}, t) + h_2(\tilde{U} - t)\} \\
&\quad + \frac{(1-\Delta)}{1 - F_0(\tilde{U}|\boldsymbol{X})}\int_{\tilde{U}}^\tau \left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\boldsymbol{Z}(\boldsymbol{X}, t) + (\mu - \mu_0)(u - t)\right\}\{\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X}, t) + h_2(u - t)\}dF_0(u|\boldsymbol{X})\Bigg]dN(t).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}(F - F_0)[\boldsymbol{h}_1, h_2] \\
=&\left.\frac{dS(\boldsymbol{\theta}_0, \mu_0, F_0 + \varepsilon(F - F_0))[\boldsymbol{h}_1, h_2]}{d\varepsilon}\right|_{\varepsilon=0} \\
=&P\int_{\tilde{U}}^\tau \psi_1(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2]d(F - F_0)(u|\boldsymbol{X}) + \frac{(F - F_0)(\tilde{U}|\boldsymbol{X})}{1 - F_0(\tilde{U}|\boldsymbol{X})}P\int_{\tilde{U}}^\tau \psi_1(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2]dF_0(u|\boldsymbol{X}) \\
=&P\int_{\tilde{U}}^\tau \psi_2(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2]d(F - F_0)(u|\boldsymbol{X}),
\end{aligned}
$$

where

$$
\begin{aligned}
\psi_1(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2] =& \frac{(1-\Delta)}{1 - F_0(\tilde{U}|\boldsymbol{X})}\int_0^\tau \left\{Y(t) - \boldsymbol{\theta}_0'\boldsymbol{Z}(\boldsymbol{X}, t) - \mu_0(u - t)\right\}\{\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X}, t) + h_2(u - t)\}dN(t), \\
\psi_2(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2] =& \psi_1(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2] - \frac{\int_{\tilde{U}}^\tau \psi_1(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2]dF_0(u|\boldsymbol{X})}{1 - F_0(\tilde{U}|\boldsymbol{X})}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\sqrt{n}\{S(\hat{\boldsymbol{\theta}}_n,\hat{\mu}_n,\hat{F}_n,)[\boldsymbol{h}_1,h_2]-S(\boldsymbol{\theta}_0,\mu_0,F_0)[\boldsymbol{h}_1,h_2]\\
&-\dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0,\hat{\mu}_n-\mu_0)[\boldsymbol{h}_1,h_2]-\dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{F}_n-F_0)[\boldsymbol{h}_1,h_2]\\
=&\sqrt{n}P\int_0^\tau(1-\Delta)\int_{\tilde{U}}^\tau\left\{Y(t)-\hat{\boldsymbol{\theta}}_n'\boldsymbol{Z}(\boldsymbol{X},t)-\hat{\mu}_n(u-t)\right\}\{\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X},t)+h_2(u-t)\}\left[\frac{d\hat{F}_n(u|\boldsymbol{X})}{1-\hat{F}_n(\tilde{U}|\boldsymbol{X})}-\frac{dF_0(u|\boldsymbol{X})}{1-F_0(\tilde{U}|\boldsymbol{X})}\right]dN(t)\\
&-\sqrt{n}P\int_{\tilde{U}}^\tau\psi_2(u;\boldsymbol{O})[\boldsymbol{h}_1,h_2]d(\hat{F}_n-F_0)(u|\boldsymbol{X})\\
\lesssim&\sqrt{n}(\|\hat{F}_n-F_0\|_\infty\left[\|\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}_n\|+E\left\{\sum_{j=1}^K\xi(T_{K,j})\left[|(\hat{\mu}_n-\mu_0)(U-T_{K,j})|+|(\hat{\mu}_n'-\mu_0')(U-T_{K,j})|\right]\right\}\right]\\
=&o_p(1),
\end{aligned}
$$

where the last equality holds by Theorem 1 and Lemma A.3. Thus condition B4 is satisfied.

Since

$$
\dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{F}_n-F_0)[\boldsymbol{h}_1,h_2]+S_n(\boldsymbol{\theta}_0,\mu_0,F_0)[\boldsymbol{h}_1,h_2]=\mathbb{P}_n\{\psi(\boldsymbol{\theta}_0,\mu_0,F_0;\boldsymbol{O})[\boldsymbol{h}_1,h_2]+m^*(\boldsymbol{\theta}_0,\mu_0,F_0;\tilde{\boldsymbol{O}})[\boldsymbol{h}_1,h_2]\},
$$

where $m^*(\boldsymbol{\theta}_0,\mu_0,F_0;\tilde{\boldsymbol{O}})[\boldsymbol{h}_1,h_2]=P\int_{\tilde{U}}^\tau\tilde{\psi}_2(u;\boldsymbol{O})[\boldsymbol{h}_1,h_2]d\mathcal{O}(u;\boldsymbol{O};\tilde{\boldsymbol{O}})$ with $\tilde{\psi}_2=\tilde{g}\circ\psi_2$. It can be seen that $\psi(\boldsymbol{\theta}_0,\mu_0,F_0;\boldsymbol{O})[\boldsymbol{h}_1,h_2]+m^*(\boldsymbol{\theta}_0,\mu_0,F_0;\tilde{\boldsymbol{O}})[\boldsymbol{h}_1,h_2]$ is a bounded Lipschitz function with respect to $\mathcal{H}$. Therefore, condition B5 holds since $\mathcal{H}$ is a Donsker Class.

Finally, by the Theorem A.2, we have

$$
\begin{aligned}
&-\sqrt{n}\dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0,\hat{\mu}_n-\mu_0)[\boldsymbol{h}_1,h_2]\\
=&\sqrt{n}\dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{F}_n-F_0)[\boldsymbol{h}_1,h_2]+\sqrt{n}S_n(\boldsymbol{\theta}_0,\mu_0,F_0)[\boldsymbol{h}_1,h_2]+o_p(1)
\end{aligned}
\tag{A.10}
$$

is asymptotically normal distributed.

Especially, if we take $h_2(u - T_{K,j}) = -\boldsymbol{h}_1' E\{\boldsymbol{Z}(\boldsymbol{X}, T_{K,j})|K, T_{K,j}, U = u, C\}$, then

$$E\left\{\int_0^\tau \{\boldsymbol{h}_1'\boldsymbol{Z}(\boldsymbol{X}, t) + h_2(U - t)\}(\hat{\mu}_n - \mu_0)(U - t)dN(t)\right\} = 0,$$

and for this h_2, we have

$$\begin{aligned}
&-\sqrt{n}\dot{S}_{1,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \hat{\mu}_n - \mu_0)[\boldsymbol{h}_1, h_2] \\
=&\boldsymbol{h}_1' E\left\{\sum_{j=1}^K \xi(T_{K,j})\left[\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - E\{\boldsymbol{Z}(\boldsymbol{X}, T_{K,j})|K, T_{K,j}, U, C\}\right]^{\otimes 2}\right\}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
\triangleq&\boldsymbol{h}_1'\boldsymbol{J}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),
\end{aligned}$$

and

$$\dot{S}_{2,(\boldsymbol{\theta}_0,\mu_0,F_0)}(\hat{F}_n - F_0)[\boldsymbol{h}_1, h_2] + S_n(\boldsymbol{\theta}_0, \mu_0, F_0)[\boldsymbol{h}_1, h_2] = \boldsymbol{h}_1'\mathbb{P}_n\{\psi^*(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) + m^{**}(\boldsymbol{\theta}_0, \mu_0, F_0; \tilde{\boldsymbol{O}})\},$$

where

$$\begin{aligned}
&\psi^*(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) \\
=&\sum_{j=1}^K \xi(T_{K,j})\Bigg[\Delta\left\{Y(T_{K,j}) - \boldsymbol{\theta}_0'\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - \mu_0(\tilde{U} - T_{K,j})\right\}[\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - E\{\boldsymbol{Z}(\boldsymbol{X}, T_{K,j})|K, T_{K,j}, U, C\}] \\
&+ \frac{(1-\Delta)}{1 - F_0(\tilde{U}|\boldsymbol{X})}\int_{\tilde{U}}^\tau \left\{Y(T_{K,j}) - \boldsymbol{\theta}_0'\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - \mu_0(u - T_{K,j})\right\} \\
&\qquad\qquad \times [\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - E\{\boldsymbol{Z}(\boldsymbol{X}, T_{K,j})|K, T_{K,j}, U = u, C\}]\, dF_0(u|\boldsymbol{X})\Bigg],
\end{aligned}$$

and $m^{**}(\boldsymbol{\theta}_0, \mu_0, F_0; \tilde{\boldsymbol{O}})\} = P \int_{\tilde{U}}^{\tau} \tilde{\psi}_2^*(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2] d\mathcal{O}(u; \boldsymbol{O}; \tilde{\boldsymbol{O}})$ with $\tilde{\psi}_2^* = \tilde{g} \circ \psi_2^*$ and

$$
\begin{aligned}
\psi_2^*(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2] =& \psi_1^*(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2] - \frac{\int_{\tilde{U}}^{\tau} \psi_1^*(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2] dF_0(u|\boldsymbol{X})}{1 - F_0(\tilde{U}|\boldsymbol{X})}, \\
\psi_1^*(u; \boldsymbol{O})[\boldsymbol{h}_1, h_2] =& \frac{(1-\Delta)}{1 - F_0(\tilde{U}|\boldsymbol{X})} \sum_{j=1}^{K} \left\{ Y(T_{K,j}) - \boldsymbol{\theta}_0' \boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - \mu_0(u - T_{K,j}) \right\} \\
& \times \left[\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) - E\{\boldsymbol{Z}(\boldsymbol{X}, T_{K,j}) | K, T_{K,j}, U = u, C\} \right].
\end{aligned}
$$

Finally, according to (A.10), by using the multivariate central limit theorem, $\boldsymbol{h}_1' \boldsymbol{J} \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converges in distribution to a mean zero normal random vector with variance matrix $\boldsymbol{h}_1' \boldsymbol{Q} \boldsymbol{h}_1$ as n tends to infinity, where $Q = E[\{\psi^*(\boldsymbol{\theta}_0, \mu_0, F_0; \boldsymbol{O}) + m^{**}(\boldsymbol{\theta}_0, \mu_0, F_0; \tilde{\boldsymbol{O}})\}^{\otimes 2}]$. Therefore, by the delta method, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converges in distribution to a mean zero normal random vector with variance matrix $\boldsymbol{J}^{-1} \boldsymbol{Q} \boldsymbol{J}^{-1}$ as n tends to infinity. This completes the proof of the theorem. □

REFERENCES

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *The Annals of Statistics*, **27**, 1536–1563.

Kalbfleisch, J. D., and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time data*, 2nd ed. Hoboken: Wiley.

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.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer series in statistics. Springer.

Schumaker, L. (2007). *Spline Functions: Basic Theory (Third Edition)*, Cambridge University Press.

Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics*, **22**, 580–615.

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

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

van der Vaart, A. W. and Weller, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer-Verlag.