A spline-based nonparametric analysis for interval-censored bivariate survival data

Yuan Wu

Duke University Medical Center, Durham, NC 27705

Ying Zhang

University of Nebraska Medical Center, Omaha, NE 68198

Junyi Zhou

Indiana University Fairbanks School of Public Health, Indianapolis, IN 46202

Supplementary Material

THIS SUPPLEMENTARY MATERIAL CONTAINS TECHNICAL DETAILS INCLUDING LEMMAS AND THEIR PROOFS NECESSARY FOR THE MAIN PAPER.

S1. Derivation of the log likelihood function

The log likelihood of the model parameter $\boldsymbol{\theta}$ based on the n observations can be written as

$$
\begin{aligned}
&l_n(\boldsymbol{\theta};\text{data}) \\
&\quad= \sum_{k=1}^{n} \{\delta_{1,k}^{(1)}\delta_{2,k}^{(1)} \log \Pr(T_1 \leq u_{1,k}, T_2 \leq u_{2,k}) \\
&\qquad + \delta_{1,k}^{(1)}\delta_{2,k}^{(2)} \log \Pr(T_1 \leq u_{1,k}, u_{2,k} < T_2 \leq v_{2,k}) \\
&\qquad + \delta_{1,k}^{(1)}\delta_{2,k}^{(3)} \log \Pr(T_1 \leq u_{1,k}, T_2 > v_{2,k}) \\
&\qquad + \delta_{1,k}^{(2)}\delta_{2,k}^{(1)} \log \Pr(u_{1,k} < T_1 \leq v_{1,k}, T_2 \leq u_{2,k}) \\
&\qquad + \delta_{1,k}^{(2)}\delta_{2,k}^{(2)} \log \Pr(u_{1,k} < T_1 \leq v_{1,k}, u_{2,k} < T_2 \leq v_{2,k}) \\
&\qquad + \delta_{1,k}^{(2)}\delta_{2,k}^{(3)} \log \Pr(u_{1,k} < T_1 \leq v_{1,k}, T_2 > v_{2,k}) \\
&\qquad + \delta_{1,k}^{(3)}\delta_{2,k}^{(1)} \log \Pr(T_1 > v_{1,k}, T_2 \leq u_{2,k}) \\
&\qquad + \delta_{1,k}^{(3)}\delta_{2,k}^{(2)} \log \Pr(T_1 > v_{1,k}, u_{2,k} < T_2 \leq v_{2,k}) \\
&\qquad + \delta_{1,k}^{(3)}\delta_{2,k}^{(3)} \log \Pr(T_1 > v_{1,k}, T_2 > v_{2,k})\}.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&l_n(\boldsymbol{\theta};\text{data})\\
&\quad=\sum_{k=1}^{n}\{\delta_{1,k}^{(1)}\delta_{2,k}^{(1)}\log F_0(u_{1,k},u_{2,k})\\
&\qquad+\delta_{1,k}^{(1)}\delta_{2,k}^{(2)}\log[F_0(u_{1,k},v_{2,k})-F_0(u_{1,k},u_{2,k})]\\
&\qquad+\delta_{1,k}^{(1)}\delta_{2,k}^{(3)}\log[F_1(u_{1,k})-F_0(u_{1,k},v_{2,k})]\\
&\qquad+\delta_{1,k}^{(2)}\delta_{2,k}^{(1)}\log[F_0(v_{1,k},u_{2,k})-F_0(u_{1,k},u_{2,k})]\\
&\qquad+\delta_{1,k}^{(2)}\delta_{2,k}^{(2)}\log[F_0(v_{1,k},v_{2,k})-F_0(u_{1,k},v_{2,k})-F_0(v_{1,k},u_{2,k})+F_0(u_{1,k},u_{2,k})]\\
&\qquad+\delta_{1,k}^{(2)}\delta_{2,k}^{(3)}\log[F_1(v_{1,k})-F_0(v_{1,k},v_{2,k})-F_1(u_{1,k})+F_0(u_{1,k},v_{2,k})]\\
&\qquad+\delta_{1,k}^{(3)}\delta_{2,k}^{(1)}\log[F_2(u_{2,k})-F_0(v_{1,k},u_{2,k})]\\
&\qquad+\delta_{1,k}^{(3)}\delta_{2,k}^{(2)}\log[F_2(v_{2,k})-F_2(u_{2,k})-F_0(v_{1,k},v_{2,k})+F_0(v_{1,k},u_{2,k})]\\
&\qquad+\delta_{1,k}^{(3)}\delta_{2,k}^{(3)}\log[1-F_1(v_{1,k})-F_2(v_{2,k})+F_0(v_{1,k},v_{2,k})]\}.\text{”}
\end{aligned}
$$

S2. Theorem Proofs

Proof of the consistency for Theorem 1

The key point for the proof is to find Θ_n as described in Theorem 1. First, for any element $\boldsymbol{\theta}_n \in \Theta_n$, with a small positive number c_{knot} two knot sequences are selected as

$$\Big\{(\xi_i)_{i=1}^{p_n+l} : (\xi_i)_{i=1}^{p_n+l} \in \boldsymbol{\xi},\ \text{for } \boldsymbol{\xi} \text{ defined by (2.2) in the main paper },\quad \frac{\min_{i:l\le i\le p_n}(\xi_{i+1}-\xi_i)}{\max_{i:l\le i\le p_n}(\xi_{i+1}-\xi_i)} > c_{knot}\Big\} \tag{S2.1}$$

and

$$\Big\{(\eta_j)_{j=1}^{q_n+l} : (\eta_j)_{j=1}^{q_n+l} \in \boldsymbol{\eta},\ \text{for } \boldsymbol{\eta} \text{ defined by (2.3) in the main paper },\quad \frac{\min_{j:l\le j\le q_n}(\eta_{j+1}-\eta_j)}{\max_{j:l\le j\le q_n}(\eta_{j+1}-\eta_j)} > c_{knot}\Big\}. \tag{S2.2}$$

Second, for any $\boldsymbol{\theta}_n \in \Theta_n$ we make conditions in (2.8) in the main paper a little stronger by introducing a small $c^* > 0$ and updating those conditions to the following

$$\begin{aligned}
& c_* \le F_{n,0}(u_1, u_2), \\
& F_{n,0}(u_1, u_2) + c_* \le F_{n,0}(v_1, u_2), \\
& F_{n,0}(u_1, u_2) + c_* \le F_{n,0}(u_1, v_2), \\
& \{F_{n,0}(v_1, v_2) - F_{n,0}(u_1, v_2)\} - \{F_{n,0}(v_1, u_2) - F_{n,0}(u_1, u_2)\} \ge c_*, \\
& F_{n,1}(u_1) - F_{n,0}(u_1, v_2) \ge c_*, \\
& F_{n,2}(u_2) - F_{n,0}(v_1, u_2) \ge c_*, \\
& \{F_{n,1}(v_1) - F_{n,1}(u_1)\} - \{F_{n,0}(v_1, v_2) - F_{n,0}(u_1, v_2)\} \ge c_*, \\
& \{F_{n,2}(v_2) - F_{n,2}(u_2)\} - \{F_{n,0}(v_1, v_2) - F_{n,0}(v_1, u_2)\} \ge c_*, \\
& \{1 - F_{n,1}(v_1)\} - \{F_{n,2}(v_2) - F_{n,0}(v_1, v_2)\} \ge c_*.
\end{aligned} \tag{S2.3}$$

That is, for

$$\mathcal{D} = \{(u_1, v_1, u_2, v_2) : u_1 \in [\tau_{1,l}, \tau_{1,h}], v_1 \in [\tau_{1,l}, \tau_{1,h}], \\ u_2 \in [\tau_{2,l}, \tau_{2,h}], v_2 \in [\tau_{2,l}, \tau_{2,h}], u_1 + \tau_d \leq v_1, u_2 + \tau_d \leq v_2\} \tag{S2.4}$$

we define

$$\Theta_n = \{\boldsymbol{\theta}_n : \text{(S2.3) holds for } (u_1, v_1, u_2, v_2) \in \mathcal{D} \text{ for } \mathcal{D} \text{ defined by (S2.4)}, \\ \text{knot sequences for } \boldsymbol{\theta}_n \text{ by (S2.1) and (S2.2)}\}, \tag{S2.5}$$

where Then it is clear that $\Theta_n \subset \Psi_n$. In what follows, we will show that the above defined Θ_n guarantees the consistency in Theorem 1.

We apply Theorem 5.7 in van der Vaart (1998) to show the consistency. By the proof of Theorem 5.7 in van der Vaart (1998), we need to find a class containing both $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n$. For this purpose, Θ is constructed as

$$\Theta = \{\theta = (F_0, F_1, F_2) : \text{(S2.3) holds for } (u_1, v_1, u_2, v_2) \in \mathcal{D} \\ \text{with } c_* \text{ replaced by } \tilde{c}_* \text{ for } 0 < \tilde{c}_* < c_* \text{ in (S2.3)}\}. \tag{S2.6}$$

Then it is easy to see $\Theta_n \subset \Theta$. If $\tilde{c}_*$ is a sufficiently small positive number, Condition C3 implies $\boldsymbol{\theta}_0 \in \Theta$. Then by adjusting $\tilde{c}_*$, we have Θ contains both Θ_n and $\boldsymbol{\theta}_0$ and hence Θ contains both $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$.

For variable $\boldsymbol{\theta} \in \Theta$, (S2.3) guarantees the log likelihood function $l(\boldsymbol{\theta}; \boldsymbol{X})$ is uniformly bounded. We denote $\mathbb{M}(\boldsymbol{\theta}) = Pl(\boldsymbol{\theta}; \boldsymbol{X})$ and $\mathbb{M}_n(\boldsymbol{\theta}) = \mathbb{P}_n l(\boldsymbol{\theta}; \boldsymbol{X})$

with $\mathbb{P}_n f(X) = \frac{1}{n}\sum_{i=1}^n f(x_i)$. First, we verify $\sup_{\boldsymbol{\theta}\in\Theta} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}(\boldsymbol{\theta})| \to_p 0$. Denote $\mathcal{L} = \{l(\boldsymbol{\theta};\boldsymbol{x}) : \boldsymbol{\theta} \in \Theta\}$, then

$$\sup_{\boldsymbol{\theta}\in\Theta} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}(\boldsymbol{\theta})| = \sup_{l(\boldsymbol{\theta};\boldsymbol{x})\in\mathcal{L}} |(\mathbb{P}_n - P)\, l(\boldsymbol{\theta};\boldsymbol{X})| \,.$$

Hence, it suffices to show $\mathcal{L}$ is a P-Glivenko-Cantelli. Let

$$A = \left\{ \frac{\log\left[F_0(v_1,v_2) - F_0(v_1,u_2) - F_0(u_1,v_2) + F_0(u_1,u_2)\right]}{\log \tilde{c}^*} : \right.$$
$$\left. \boldsymbol{\theta} = (F_0, F_1, F_2) \in \Theta,\ \text{for } (u_1,v_1,u_2,v_2) \in \mathcal{D} \text{ with } \mathcal{D} \text{ defined by (S2.4)} \right\},$$

where $F_0(v_1,v_2) - F_0(v_1,u_2) - F_0(u_1,v_2) + F_0(u_1,u_2) \geq \tilde{c}^*$ in $\mathcal{D}$ with $\tilde{c}^* < c^*$ for c^* given in (S2.3), as discussed in the construction of Θ. For $(u_1,u_2,v_1,v_2) \in \mathcal{D}$ with $\mathcal{D}$ given by (S2.4), define two classes of indicator functions

$$\mathcal{G}_1 = \{1_{[\tau_{1,l},u_1]\times[\tau_{1,l},v_1]\times[\tau_{2,l},u_2]\times[\tau_{2,l},v_2]}\}$$

and

$$\mathcal{G}_2 = \{1_{[\tau_{1,l},u_1]\times[\tau_{1,l},\tau_{1,h}]\times[\tau_{2,l},u_2]\times[\tau_{2,l},\tau_{2,h}]}\}.$$

(S2.3) implies that each element in A is increasing in v_1 and v_2 but decreasing in u_1 and u_2, then we have

$$A \subseteq \overline{\text{sconv}}\left(\mathcal{G}_2 - \mathcal{G}_1\right), \tag{S2.7}$$

where $\overline{\text{sconv}}\left(\cdot\right)$ is the closure of symmetric convex hull (van der Vaart and Wellner, 1996). For $\mathcal{G}_1$ and $\mathcal{G}_2$, it is easily shown that they have VC-index values (van der Vaart and Wellner, 1996) $V(\mathcal{G}_1) = 5$ and $V(\mathcal{G}_2) = 3$. Therefore, Theorem 2.6.7 in van der Vaart and Wellner (1996) implies

$$N\left(\epsilon, \mathcal{G}_1, L_2\left(Q_{U_1,U_2,V_1,V_2}\right)\right) \le c\left(\frac{1}{\epsilon}\right)^8 \tag{S2.8}$$

and

$$N\left(\epsilon, \mathcal{G}_2, L_2\left(Q_{U_1,U_2,V_1,V_2}\right)\right) \le c\left(\frac{1}{\epsilon}\right)^4, \tag{S2.9}$$

for any probability measure Q_{U_1,U_2,V_1,V_2} of (U_1, U_2, V_1, V_2). Since the envelop functions of $\mathcal{G}_1$ and $\mathcal{G}_2$ are both 1, then by (S2.8), (S2.9) and Theorem 2.6.9 in van der Vaart and Wellner (1996) it follows that

$$\log N\left(\epsilon, \overline{\mathrm{sconv}}\left(\mathcal{G}_1\right), L_2\left(\left(Q_{U_1,U_2,V_1,V_2}\right)\right)\right) \le c\left(\frac{1}{\epsilon}\right)^{8/5}$$

and

$$\log N\left(\epsilon, \overline{\mathrm{sconv}}\left(\mathcal{G}_2\right), L_2\left(\left(Q_{U_1,U_2,V_1,V_2}\right)\right)\right) \le c\left(\frac{1}{\epsilon}\right)^{4/3}$$

for any Q_{U_1,U_2,V_1,V_2}. Hence for any Q_{U_1,U_2,V_1,V_2},

$$\begin{aligned}\log N\left(\epsilon, \overline{\mathrm{sconv}}\left(\mathcal{G}_2\right) - \overline{\mathrm{sconv}}\left(\mathcal{G}_1\right), L_2\left(\left(Q_{U_1,U_2,V_1,V_2}\right)\right)\right) \le & c\left\{\left(\frac{1}{\epsilon}\right)^{8/5} + \left(\frac{1}{\epsilon}\right)^{4/3}\right\} \\ = & c\left(\frac{1}{\epsilon}\right)^{8/5}.\end{aligned}$$

By (S2.7), we have $A \subseteq \overline{\mathrm{sconv}}\left(\mathcal{G}_2\right) - \overline{\mathrm{sconv}}\left(\mathcal{G}_1\right)$, then

$$\log N\left(\epsilon, A, L_2\left(\left(Q_{U_1,U_2,V_1,V_2}\right)\right)\right) = c\left(\frac{1}{\epsilon}\right)^{8/5}. \tag{S2.10}$$

Now let

$$A' = \Big\{ \delta_1^{(2)} \delta_2^{(2)} \log \left[F_0(v_1, v_2) - F_0(v_1, u_2) - F_0(u_1, v_2) + F_0(u_1, u_2) \right]$$
$$: \boldsymbol{\theta} = (F_0, F_1, F_2) \in \Theta, (u_1, v_1, u_2, v_2) \in \mathcal{D} \text{ for } \mathcal{D} \text{ defined by (S2.4)} \Big\} .$$

We can find a positive number c_0 as the envelop function for A'. Based on (S2.10), using the same arguments as those given on on page 1626–1627 of Wu and Zhang (2012) leads to

$$\int_0^1 \sup_Q \sqrt{\log N \left(\epsilon c_0, A', L_2(Q) \right)} d\epsilon \leq c \int_0^1 \left(\frac{1}{\epsilon} \right)^{4/5} d\epsilon < \infty,$$

where Q can be any probability measure for $\boldsymbol{X}$. Hence, by Theorem 2.5.2 in van der Vaart and Wellner (1996), A' is a P-Donsker. Similarly, we can show all other items in $\mathcal{L}$ are P-Donskers. Therefore $\mathcal{L}$ is a P-Donsker as well by the fact that a finite sum of Donskers is a Donkser. Since P-Donsker is also P-Glivenko-Cantelli, then $\sup_{l(\boldsymbol{\theta};\boldsymbol{x}) \in \mathcal{L}} \| (\mathbb{P}_n - P) l(\boldsymbol{\theta}; \boldsymbol{X}) \| \to_p 0$.

Second, we verify $\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}) \geq c d^2(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \Theta$, which is completed by Lemma 1.

Finally, we verify $\mathbb{M}_n \left(\hat{\boldsymbol{\theta}}_n \right) - \mathbb{M}_n \left(\boldsymbol{\theta}_0 \right) \geq o_P(1)$.

It follows by C1 and C2, Lemma 0.2 in Wu and Zhang (2012) and Jackson's Theorem on page 149 in De Boor (2001) that there exists $\boldsymbol{\theta}_n = (F_{n,0}, F_{n,1}, F_{n,2})$ such that $\|F_{n,0} - F_{0,0}\|_\infty \leq c \left(n^{-(p+r)\kappa} \right)$, $\|F_{n,1} - F_{0,1}\|_\infty \leq c \left(n^{-(p+r)\kappa} \right)$ and $\|F_{n,2} - F_{0,2}\|_\infty \leq c \left(n^{-(p+r)\kappa} \right)$. Then we know that for large

n, this $\boldsymbol{\theta}_n$ can be chosen from Θ_n define by (S2.5). Since $\hat{\boldsymbol{\theta}}_n$ maximizes $\mathbb{M}_n(\boldsymbol{\theta}_n)$ over Θ_n, $\mathbb{M}_n\left(\hat{\boldsymbol{\theta}}_n\right) - \mathbb{M}_n(\boldsymbol{\theta}_n) > 0$. Hence,

$$\begin{aligned}\mathbb{M}_n\left(\hat{\boldsymbol{\theta}}_n\right) - \mathbb{M}_n(\boldsymbol{\theta}_0) &= \mathbb{M}_n\left(\hat{\boldsymbol{\theta}}_n\right) - \mathbb{M}_n(\boldsymbol{\theta}_n) + \mathbb{M}_n\left(\boldsymbol{\theta}_n\right) - \mathbb{M}_n(\boldsymbol{\theta}_0) \\ &\geq \mathbb{M}_n\left(\boldsymbol{\theta}_n\right) - \mathbb{M}_n(\boldsymbol{\theta}_0) = \mathbb{P}_n\{l(\boldsymbol{\theta}_n;\boldsymbol{X})\} - \mathbb{P}_n\{l(\boldsymbol{\theta}_0;\boldsymbol{X})\} \\ &= (\mathbb{P}_n - P)\{l(\boldsymbol{\theta}_n;\boldsymbol{X}) - l(\boldsymbol{\theta}_0;\boldsymbol{X})\} \\ &\quad + P\{l(\boldsymbol{\theta}_n;\boldsymbol{X}) - l(\boldsymbol{\theta}_0;\boldsymbol{X})\}.\end{aligned}$$

Define

$$\mathcal{L}_n = \Big\{l(\boldsymbol{\theta}_n;\boldsymbol{x}) : \boldsymbol{\theta}_n = (F_{n,0}, F_{n,1}, F_{n,2}) \in \Theta_n, \|F_{n,0} - F_{0,0}\|_\infty \leq c\left(n^{-(p+r)\kappa}\right), \\ \|F_{n,1} - F_{0,1}\|_\infty \leq c\left(n^{-(p+r)\kappa}\right), \|F_{n,1} - F_{0,1}\|_\infty \leq c\left(n^{-(p+r)\kappa}\right)\Big\}.$$

For $\Delta_1^{(1)}\Delta_2^{(1)} \log \frac{F_{n,0}(U_1,U_2)}{F_{0,0}(U_1,U_2)}$ (the first term in $l(\boldsymbol{\theta}_n;\boldsymbol{X}) - l(\boldsymbol{\theta}_0;\boldsymbol{X})$), we have $1/2 \leq \frac{F_{n,0}(U_1,U_2)}{F_{0,0}(U_1,U_2)} < 2$ for large n by $\|F_{n,0} - F_{0,0}\|_\infty \leq c\left(n^{-(p+r)\kappa}\right)$. Then $\left|\log \frac{F_{n,0}(U_1,U_2)}{F_{0,0}(U_1,U_2)}\right| \leq c\left|\frac{F_{n,0}(U_1,U_2)}{F_{0,0}(U_1,U_2)} - 1\right|$. Hence,

$$\begin{aligned}P\left\{\Delta_1^{(1)}\Delta_2^{(1)} \log \frac{F_{n,0}(U_1,U_2)}{F_{0,0}(U_1,U_2)}\right\}^2 &\leq P_{U_1,U_2}\left\{\log \frac{F_{n,0}(U_1,U_2)}{F_{0,0}(U_1,U_2)}\right\}^2 \\ &\leq cP_{U_1,U_2}\left\{\frac{F_{n,0}(U_1,U_2)}{F_{0,0}(U_1,U_2)} - 1\right\}^2 \\ &\leq cP_{U_1,U_2}\left\{F_{n,0}(U_1,U_2) - F_{0,0}(U_1,U_2)\right\}^2 \to 0.\end{aligned}$$

And we can show the similar results for other terms in $l(\boldsymbol{\theta}_n;\boldsymbol{X}) - l(\boldsymbol{\theta}_0;\boldsymbol{X})$. Then, by $\left(\sum_{i=1}^9 a_i\right)^2 \leq 9\sum_{i=1}^9 a_i^2$, we have $P\{l(\boldsymbol{\theta}_n;\boldsymbol{X}) - l(\boldsymbol{\theta}_0;\boldsymbol{X})\}^2 \to 0$ as

$n \to \infty$. Therefore,

$$[\text{var}\{l(\boldsymbol{\theta}_n; \boldsymbol{X}) - l(\boldsymbol{\theta}_0; \boldsymbol{X})\}]^{1/2} \leq \left[P\{l(\boldsymbol{\theta}_n; \boldsymbol{X}) - l(\boldsymbol{\theta}_0; \boldsymbol{X})\}^2\right]^{1/2} \to 0.$$

Since we already showed $\mathcal{L}$ is a P-Donsker, by the fact that both $l(\boldsymbol{\theta}_n; \boldsymbol{x})$ and $l(\boldsymbol{\theta}_0; \boldsymbol{x})$ are in $\mathcal{L}$, Corollary 2.3.12 of van der Vaart and Wellner (1996) results in that

$$(\mathbb{P}_n - P)\{l(\boldsymbol{\theta}_n; \boldsymbol{X}) - l(\boldsymbol{\theta}_0; \boldsymbol{X})\} = o_P\left(n^{-1/2}\right). \tag{S2.11}$$

In addition, By Cauchy-Schwartz inequality,

$$\begin{aligned}|P\{l(\boldsymbol{\theta}_n; \boldsymbol{X}) - l(\boldsymbol{\theta}_0; \boldsymbol{X})\}| &\leq P|l(\boldsymbol{\theta}_n; \boldsymbol{X}) - l(\boldsymbol{\theta}_0; \boldsymbol{X})| \\ &\leq c\left[P\{l(\boldsymbol{\theta}_n; \boldsymbol{X}) - l(\boldsymbol{\theta}_0; \boldsymbol{X})\}^2\right]^{1/2} \to 0.\end{aligned}$$

Then $P\{l(\boldsymbol{\theta}_n; \boldsymbol{X}) - l(\boldsymbol{\theta}_0; \boldsymbol{X})\} \geq -o(1)$. Hence,

$$\mathbb{M}_n\left(\hat{\boldsymbol{\theta}}_n\right) - \mathbb{M}_n\left(\boldsymbol{\theta}_0\right) \geq o_P\left(n^{-1/2}\right) - o(1) \geq -o_P(1).$$

The consistency is proved. □

Proof of the rate of convergence for Theorem 1

We derive the rate of convergence by verifying the conditions of Theorem 3.4.1 of van der Vaart and Wellner (1996).

Let $\boldsymbol{\theta}_n \in \Theta_n$ with $\boldsymbol{\theta}_n$ satisfying $d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) \leq c\left(n^{-(p+r)\kappa}\right)$. We verify that for every n and any $\delta > \delta_n = n^{-(p+r)\kappa}$,

$$\sup_{\delta/2 < d(\boldsymbol{\theta}, \boldsymbol{\theta}_n) < \delta, \boldsymbol{\theta} \in \Theta_n} \{\mathbb{M}(\boldsymbol{\theta}) - \mathbb{M}(\boldsymbol{\theta}_n)\} \leq -c\delta^2.$$

For $x \geq M_l > 0$, we can show that there exists $c_{M_l} > 0$ such that $x\log(x) - x + 1 \leq c_{M_l}(x-1)^2$. With the similar arguments as in the proof of Lemma 1, we can show that

$$\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}_n) \leq cd^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) \leq cn^{-2(p+r)\kappa}.$$

Then by the result of Lemma 1, it follows that

$$\mathbb{M}(\boldsymbol{\theta}) - \mathbb{M}(\boldsymbol{\theta}_n) = \mathbb{M}(\boldsymbol{\theta}) - \mathbb{M}(\boldsymbol{\theta}_0) + \mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}_n) \leq -c\delta^2 + cn^{-2(p+r)\kappa} = -c\delta^2.$$

Next, we need to find a function $\psi(\cdot)$, for $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ such that

$$E_P \sup_{\delta/2 < d(\boldsymbol{\theta}, \boldsymbol{\theta}_n) \leq \delta, \boldsymbol{\theta} \in \Theta_n} [\mathbb{G}_n\{l(\boldsymbol{\theta}; \boldsymbol{X}) - l(\boldsymbol{\theta}_n; \boldsymbol{X})\}]_+ \leq c\psi(\delta)$$

and $\psi(\delta)/\delta^\alpha$ is decreasing in δ, for some $\alpha < 2$, and for $\gamma_n \leq \delta_n^{-1}$, it satisfies

$$\gamma_n^2 \psi(1/\gamma_n) \leq c\sqrt{n} \text{ for every } n.$$

Let $\mathcal{L}_{n,\delta} = \{l(\boldsymbol{\theta}; \boldsymbol{x}) - l(\boldsymbol{\theta}_n; \boldsymbol{x}) : \boldsymbol{\theta} = (F_0, F_1, F_2) \in \Theta_n \text{ and } \delta/2 < d(\boldsymbol{\theta}, \boldsymbol{\theta}_n) \leq \delta\}$. C4 implies the density of the probability measure P has a positive lower bound. Then $d(\boldsymbol{\theta}, \boldsymbol{\theta}_n) \leq \delta$ implies

$$\left\{\int_{[\tau_{1,l}, \tau_{1,h}] \times [\tau_{2,l}, \tau_{2,h}]} (F_0(t_1, t_2) - F_{n,0}(t_1, t_2))^2 dt_2 dt_1\right\}^{1/2} \leq c\delta,$$

$$\left\{\int_{[\tau_{1,l}, \tau_{1,h}]} (F_1(t_1) - F_{n,1}(t_1))^2 dt_1\right\}^{1/2} \leq c\delta$$

and

$$\left\{\int_{[\tau_{2,l}, \tau_{2,h}]} (F_2(t_2) - F_{n,2}(t_2))^2 dt_2\right\}^{1/2} \leq c\delta.$$

Hence, we can use Lemma 2 and Lemma 3 with some algebra to show that

$$N_{[\ ]}(\epsilon, \mathcal{L}_{n,\delta}, \|\cdot\|_\infty) \le (\delta/\epsilon)^{cp_nq_n}.$$

Then obviously,

$$N_{[\ ]}\{\epsilon, \mathcal{L}_{n,\delta}, L_2(P)\} \le (\delta/\epsilon)^{cp_nq_n}. \tag{S2.12}$$

Next, for any $l(\boldsymbol{\theta};\boldsymbol{x}) - l(\boldsymbol{\theta}_n;\boldsymbol{x}) \in \mathcal{L}_{n,\delta}$, since $d(\boldsymbol{\theta}, \boldsymbol{\theta}_n) \le \delta$. Lemma 4 and Lemma 7.1 in Wellner and Zhang (2007) with Condition C1–C4, imply that for small δ and sufficiently large n, $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_n$ are both very close to $\boldsymbol{\theta}_0$ in terms of $\|\cdot\|_\infty$ in the domain of censoring times, $\mathcal{D}$ as defined by (S2.4). Hence, for any $l(\boldsymbol{\theta};\boldsymbol{x}) - l(\boldsymbol{\theta}_n;\boldsymbol{x}) \in \mathcal{L}_{n,\delta}$, it can be shown that $P\{l(\boldsymbol{\theta};\boldsymbol{x}) - l(\boldsymbol{\theta}_n;\boldsymbol{x})\}^2 \le c\delta^2$. Also by $\mathcal{L}_{n,\delta}$ being a uniformly bounded class, Lemma 3.4.2 of van der Vaart and Wellner (1996) indicates that

$$E_P\|\mathbb{G}_n\|_{\mathcal{L}_{n,\delta}} \le c\tilde{J}_{[\ ]}\{\delta, \mathcal{L}_{n,\delta}, L_2(P)\}\left[1 + \frac{\tilde{J}_{[\ ]}\{\delta, \mathcal{L}_{n,\delta}, L_2(P)\}}{\delta^2\sqrt{n}}\right],$$

where $\tilde{J}_{[\ ]}\{\delta, \mathcal{L}_{n,\delta}, L_2(P)\} = \int_0^\delta \sqrt{1 + \log N_{[\ ]}\{\epsilon, \mathcal{L}_{n,\delta}, L_2(P)\}}d\epsilon \le c(p_nq_n)^{1/2}\delta$, by (S2.12). Then $E_P \sup_{\delta/2 < d(\boldsymbol{\theta},\boldsymbol{\theta}_n) \le \delta, \boldsymbol{\theta}\in\Theta_n}[\mathbb{G}_n\{l(\boldsymbol{\theta};\boldsymbol{X}) - l(\boldsymbol{\theta}_n;\boldsymbol{X})\}]_+ \le E_P\|\mathbb{G}_n\|_{\mathcal{L}_{n,\delta}}$ indicates that

$$\psi(\delta) = (p_nq_n)^{1/2}\delta + (p_nq_n)/n^{1/2}.$$

It is easy to see that $\psi(\delta)/\delta$ is a decreasing function of δ. Then $p_n = q_n = n^\kappa$ implies that if $r_n = n^{\min\{(p+r)\kappa,(1-2\kappa)/2\}}$, $r_n \le \delta_n^{-1}$ and $r_n^2\psi(1/r_n) \le cn^{1/2}$.

Since $\mathbb{M}_n\left(\hat{\boldsymbol{\theta}}_n\right)-\mathbb{M}_n(\boldsymbol{\theta}_n)\geq 0$ and $d\left(\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_n\right)\leq d\left(\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0\right)+d(\boldsymbol{\theta}_0,\boldsymbol{\theta}_n)\rightarrow$ 0 in probability. Therefore, it follows by Theorem 3.4.1 in van der Vaart and Wellner (1996) that $r_n d\left(\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_n\right)=O_P(1)$. Hence, by $d(\boldsymbol{\theta}_n,\boldsymbol{\theta}_0)\leq cn^{-(p+r)\kappa}$

$$r_n d\left(\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0\right)\leq r_n d\left(\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_n\right)+r_n d(\boldsymbol{\theta}_n,\boldsymbol{\theta}_0)\leq O_P(1)+r_n cn^{-(p+r)\kappa}=O_P(1)$$

This establishes the convergence rate. □

Proof of Theorem 2

First, we use Riesz representation theorem for Hilbert space (Halmos, 1982) to show some intermediate results for the proof.

By the regularity condition C4 and Cauchy-Schwarz inequality, it can be shown that $\left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right|\leq cd(0,\boldsymbol{w})$, where $d(\cdot,\cdot)$ is defined by (3.1) in the main paper. By C3 and C4 we can also show that $d(0,\boldsymbol{w})\leq c\|\boldsymbol{w}\|$ with $\|\cdot\|$ being the Fisher information norm defined by (3.4) in the main paper. Hence we have $\left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right|\leq c\|\boldsymbol{w}\|$, which results in

$$\left\|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}\right\|_{*,\infty}=\sup_{w\in\mathfrak{W},\|\boldsymbol{w}\|>0}\frac{\left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right|}{\|\boldsymbol{w}\|}<\infty. \tag{S2.13}$$

By (3.6) in the main paper we know that $\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]$ defined by (3.5) in the main paper is linear in $\boldsymbol{w}$, by Riesz representation theorem there exists $\boldsymbol{w}^*=(w_0^*,w_1^*,w_2^*)'\in\overline{\mathfrak{W}}$ with $\overline{\mathfrak{W}}$ being the completion of $\mathfrak{W}$, such that for any $\boldsymbol{w} \in \overline{\mathfrak{W}}$

$$\langle \boldsymbol{w}^*, \boldsymbol{w} \rangle = \frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]. \tag{S2.14}$$

and

$$\|\boldsymbol{w}^*\| = \sup_{w \in \overline{\mathfrak{W}}, \|\boldsymbol{w}\|>0} \frac{\left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right|}{\|\boldsymbol{w}\|} = \left\|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}\right\|_{*,\infty} \tag{S2.15}$$

due to the fact that

$$\sup_{w \in \overline{\mathfrak{W}}, \|\boldsymbol{w}\|>0} \frac{\left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right|}{\|\boldsymbol{w}\|} = \sup_{w \in \mathfrak{W}, \|\boldsymbol{w}\|>0} \frac{\left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right|}{\|\boldsymbol{w}\|}.$$

Therefore, $\|\boldsymbol{w}^*\|$ is bounded by (S2.13).

In what follows we establish the asymptotic normality using (S2.14) and (S2.15).

We define $r[\boldsymbol{\theta}, \boldsymbol{\theta}_0; \boldsymbol{x}] \equiv l(\boldsymbol{\theta}; \boldsymbol{x}) - l(\boldsymbol{\theta}_0; \boldsymbol{x}) - \frac{l(\boldsymbol{\theta}_0; \boldsymbol{x})}{d\boldsymbol{\theta}}[\boldsymbol{\theta} - \boldsymbol{\theta}_0]$. Lemma 5 shows that we can find small ϵ_n and the spline function vector $\boldsymbol{w}_n^*$ (the approximation for $\boldsymbol{w}^*$), such that

$$\begin{aligned} &P\left(r\left[\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0; \boldsymbol{X}\right] - r\left[\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*, \boldsymbol{\theta}_0; \boldsymbol{X}\right]\right) \\ &\quad = \pm\epsilon_n \left\langle \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \boldsymbol{w}^* \right\rangle + \epsilon_n o_P\left(n^{-1/2}\right), \end{aligned} \tag{S2.16}$$

$$\frac{1}{n}\sum_{i=1}^n \frac{dl(\boldsymbol{\theta}_0; \boldsymbol{x}_i)}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^*] = (\mathbb{P}_n - P)\left\{\frac{dl(\boldsymbol{\theta}_0; \boldsymbol{x}_i)}{d\boldsymbol{\theta}}[\boldsymbol{w}^*]\right\} + o_P\left(n^{-1/2}\right), \tag{S2.17}$$

and

$$(\mathbb{P}_n - P)\left(r\left[\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0; \boldsymbol{X}\right] - r\left[\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*, \boldsymbol{\theta}_0; \boldsymbol{X}\right]\right) = \pm\epsilon_n o_P\left(n^{-1/2}\right). \tag{S2.18}$$

For $\hat{\boldsymbol{\theta}}_n$ being the vector of sieve NPMLEs, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left\{l\left(\hat{\boldsymbol{\theta}}_n;\boldsymbol{x}_i\right)-l\left(\hat{\boldsymbol{\theta}}_n\pm\epsilon_n\boldsymbol{w}_n^*;\boldsymbol{x}_i\right)\right\}\geq 0.$$

It immediately follows that

$$\begin{aligned}
&\mp\epsilon_n\frac{1}{n}\sum_{i=1}^{n}\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{x}_i)}{d\boldsymbol{\theta}}\left[\boldsymbol{w}_n^*\right]\\
&+(\mathbb{P}_n-P)\left(r\left[\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0;\boldsymbol{X}\right]-r\left[\hat{\boldsymbol{\theta}}_n\pm\epsilon_n\boldsymbol{w}_n^*,\boldsymbol{\theta}_0;\boldsymbol{X}\right]\right)\\
&+P\left(r\left[\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0;\boldsymbol{X}\right]-r\left[\hat{\boldsymbol{\theta}}_n\pm\epsilon_n\boldsymbol{w}_n^*,\boldsymbol{\theta}_0;\boldsymbol{X}\right]\right)\geq 0.
\end{aligned}\tag{S2.19}$$

Hence by (S2.17), (S2.18), (S2.16) and (S2.19),

$$\pm\epsilon_n(\mathbb{P}_n-P)\left\{\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}\left[\boldsymbol{w}^*\right]\right\}\pm\epsilon_n\left\langle\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0,\boldsymbol{w}^*\right\rangle+\epsilon_n o_P\left(n^{-1/2}\right)\geq 0.$$

This leads to the conclusion that

$$\left|\sqrt{n}\left\langle\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0,\boldsymbol{w}^*\right\rangle-\sqrt{n}(\mathbb{P}_n-P)\left\{\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}\left[\boldsymbol{w}^*\right]\right\}\right|\leq o_P(1)$$

and hence

$$\sqrt{n}\left\langle\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0,\boldsymbol{w}^*\right\rangle=\sqrt{n}(\mathbb{P}_n-P)\left\{\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}\left[\boldsymbol{w}^*\right]\right\}+o_P(1).$$

Then by $P\left\{\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}\left[\boldsymbol{w}^*\right]\right\}=0$ and central limit theorem, we have

$$\sqrt{n}\left\langle\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0,\boldsymbol{w}^*\right\rangle\rightarrow_d N\left(0,\|\boldsymbol{w}^*\|^2\right)\tag{S2.20}$$

By (3.6) in the main paper, we can easily get

$$\begin{aligned}
&\left|\rho\left(\hat{\boldsymbol{\theta}}_n\right) - \rho(\boldsymbol{\theta}_0) - \frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right]\right| \\
&\qquad = \left|\int_{\tau_{1,l}}^{\tau_{1,h}} \left\{\hat{F}_{n,1}(t_1) - F_{0,1}(t_1)\right\} dt_1 \int_{\tau_{2,l}}^{\tau_{2,h}} \left\{\hat{F}_{n,2}(t_2) - F_{0,2}(t_2)\right\} dt_2\right| \\
&\qquad \le c\left[\int_{\tau_{1,l}}^{\tau_{1,h}} \left\{\hat{F}_{n,1}(t_1) - F_{0,1}(t_1)\right\}^2 dt_1\right]^{1/2} \\
&\qquad\qquad\qquad \cdot \left[\int_{\tau_{2,l}}^{\tau_{2,h}} \left\{\hat{F}_{n,2}(t_2) - F_{0,2}(t_2)\right\}^2 dt_2\right]^{1/2} \\
&\qquad \le c d^2\left(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0\right),
\end{aligned}$$

where the last inequality holds by C2 and C4. Hence by Theorem 1 and $p + r > 3$ we have

$$\left|\rho\left(\hat{\boldsymbol{\theta}}_n\right) - \rho(\boldsymbol{\theta}_0) - \frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right]\right| \le O_P\left(n^{-\frac{2(p+r)}{2(p+r)+2}}\right) = o_P\left(n^{-1/2}\right).$$

It is easy to see that $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \in \mathfrak{W}$. Then by (S2.14), we have

$$\left|\rho\left(\hat{\boldsymbol{\theta}}_n\right) - \rho(\boldsymbol{\theta}_0) - \left\langle\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \boldsymbol{w}^*\right\rangle\right| = o_P\left(n^{-1/2}\right). \tag{S2.21}$$

Finally, by (S2.15), (S2.20) and (S2.21), we obtain

$$\sqrt{n}\left\{\rho\left(\hat{\boldsymbol{\theta}}_n\right) - \rho\left(\boldsymbol{\theta}_0\right)\right\} \to_d N\left(0, \left\|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}\right\|_{*,\infty}^2\right). \quad \square$$

Proof of Theorem 3

For the first part, we prove that the proposed plug-in estimator $\rho\left(\hat{\boldsymbol{\theta}}_n\right)$ is a path-wise regular estimator for $\rho(\boldsymbol{\theta}_0)$.

By (S2.21) in the proof of Theorem 2, we have

$$\begin{aligned}\sqrt{n}\left\{\rho\left(\hat{\boldsymbol{\theta}}_n\right)-\rho(\boldsymbol{\theta}_{n,h})\right\} &= \sqrt{n}\left\{\rho\left(\hat{\boldsymbol{\theta}}_n\right)-\rho(\boldsymbol{\theta}_0)\right\}-\sqrt{n}\left\{\rho\left(\boldsymbol{\theta}_{n,h}\right)-\rho(\boldsymbol{\theta}_0)\right\}\\ &= \sqrt{n}\left\langle\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0,\boldsymbol{w}^*\right\rangle-\sqrt{n}\left\langle\frac{s_nh}{\sqrt{n}}\boldsymbol{w},\boldsymbol{w}^*\right\rangle+o_P(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{x}_i)}{d\boldsymbol{\theta}}\left[\boldsymbol{w}^*\right]-s_nh\left\langle\boldsymbol{w},\boldsymbol{w}^*\right\rangle+o_P(1)\end{aligned}$$

On the other hand, for the directional derivatives $\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}]$ and $\frac{d^2l(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\boldsymbol{w}][\tilde{\boldsymbol{w}}]$ defined, respectively, by (3.2) and (3.3) in the main paper, we can easily derive the following local asymptotic normality (LAN):

$$\begin{aligned}\log\prod_{i=1}^{n}\frac{dP_{\boldsymbol{\theta}_{n,h}}}{dP_{\boldsymbol{\theta}_0}}(\boldsymbol{X}) &= \frac{s_nh}{\sqrt{n}}\sum_{i=1}^{n}\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{x}_i)}{d\boldsymbol{\theta}}\left[\boldsymbol{w}\right]+\frac{s_n^2h^2}{2n}\sum_{i=1}^{n}\frac{d^2l(\boldsymbol{\theta}_0;\boldsymbol{x}_i)}{d\boldsymbol{\theta}^2}[\boldsymbol{w}][\boldsymbol{w}]\\ &\quad+\mathrm{Rem}_n\\ &= \frac{s_nh}{\sqrt{n}}\sum_{i=1}^{n}\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{x}_i)}{d\boldsymbol{\theta}}\left[\boldsymbol{w}\right]-\frac{s_n^2h^2}{2}\|\boldsymbol{w}\|^2+o_P(1).\end{aligned} \tag{S2.22}$$

Then by multivariate central limit theorem, Slutsky's Theorem and the fact that $s_n\to 1$, we have

$$\begin{bmatrix}\sqrt{n}\left\{\rho\left(\hat{\boldsymbol{\theta}}_n\right)-\rho(\boldsymbol{\theta}_{n,h})\right\}\\ \log\prod_{i=1}^{n}\frac{dP_{\boldsymbol{\theta}_{n,h}}}{dP_{\boldsymbol{\theta}_0}}(\boldsymbol{X})\end{bmatrix}\to_d N_2\left\{\begin{pmatrix}-h\langle\boldsymbol{w},\boldsymbol{w}^*\rangle\\ -\frac{h^2}{2}\|\boldsymbol{w}\|^2\end{pmatrix},\begin{pmatrix}\|\boldsymbol{w}^*\|^2 & h\langle\boldsymbol{w},\boldsymbol{w}^*\rangle\\ h\langle\boldsymbol{w},\boldsymbol{w}^*\rangle & h^2\|\boldsymbol{w}\|^2\end{pmatrix}\right\}.$$

Now Example 6.7 (Le Cam's third lemma) in van der Vaart (1998) implies that

$$\sqrt{n}\left\{\rho\left(\hat{\boldsymbol{\theta}}_n\right)-\rho(\boldsymbol{\theta}_{n,h})\right\}\xrightarrow{P_{\boldsymbol{\theta}_{n,h}}}_d N(0,\|\boldsymbol{w}^*\|^2),$$

where $\xrightarrow{P_{\boldsymbol{\theta}_{n,h}}}_d$ means converging in distribution under the measure $P_{\boldsymbol{\theta}_{n,h}}$. By the same argument, we also have

$$\sqrt{n}\left\{\rho\left(\hat{\boldsymbol{\theta}}_n\right)-\rho(\boldsymbol{\theta}_{n,-h})\right\}\xrightarrow{P_{\boldsymbol{\theta}_{n,-h}}}_d N(0,\|\boldsymbol{w}^*\|^2).$$

Hence,

$$\limsup \Pr_{\boldsymbol{\theta}_{n,h}}\left\{\rho\left(\hat{\boldsymbol{\theta}}_n\right)<\rho\left(\boldsymbol{\theta}_{n,h}\right)\right\}\le \liminf \Pr_{\boldsymbol{\theta}_{n,-h}}\left\{\rho\left(\hat{\boldsymbol{\theta}}_n\right)<\rho\left(\boldsymbol{\theta}_{n,-h}\right)\right\},$$

which means that $\rho\left(\hat{\boldsymbol{\theta}}_n\right)$ is a path-wise regular estimator for $\rho(\boldsymbol{\theta}_0)$.

For the second part, we prove that the lower bound of the asymptotic variances for all path-wise regular estimators for $\rho(\boldsymbol{\theta}_0)$ equals to $\sup_{\boldsymbol{w}\in\mathfrak{W},\|\boldsymbol{w}\|>0}\left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right|^2/\|\boldsymbol{w}\|^2$.

It is equivalent to show the following result about concentration probabilities as

$$\limsup \Pr\left\{\sqrt{n}\left|T_n-\rho(\boldsymbol{\theta}_0)\right|<h\right\}\le \Pr\left[\left|N\left\{0,\sup_{\boldsymbol{w}\in\mathfrak{W},\|\boldsymbol{w}\|>0}\frac{\left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right|^2}{\|\boldsymbol{w}\|^2}\right\}\right|<h\right],$$

for any $h>0$. It is also equivalent to show that for any $\boldsymbol{w}\in\mathfrak{W}$ with $\|\boldsymbol{w}\|>0$ and $h>0$

$$\limsup \Pr\left\{\sqrt{n}\left|T_n-\rho(\boldsymbol{\theta}_0)\right|<h\right\}\le \Pr\left[\left|N\left\{0,\frac{\left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right|^2}{\|\boldsymbol{w}\|^2}\right\}\right|<h\right]. \tag{S2.23}$$

If $\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}] = 0$ then it is obvious that (S2.23) holds, since the right hand side equals to 1. In what follows we show that for any $\boldsymbol{w} \in \mathfrak{W}$ with $\|\boldsymbol{w}\| > 0$ and $\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}] \neq 0$ (S2.23) also holds.

(a) For any fixed $\boldsymbol{w} \in \mathfrak{W}$ with $\|\boldsymbol{w}\| > 0$ and $\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}] \neq 0$, it is true that $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + s\boldsymbol{w}$ and s are one-to-one locally. Then we re-parameterize $P_{\boldsymbol{\theta}}$ by s and denote $P_s = P_{\boldsymbol{\theta}}$. Let $s_{n,h} = s_0 + \frac{s_n h}{\sqrt{n}}$ for $s_0 = 0$, $s_n \to 1$ and any $h > 0$. Then $P_{s_0} = P_{\boldsymbol{\theta}_0}$ and $P_{s_{n,h}} = P_{\boldsymbol{\theta}_{n,h}}$. Hence by LAN (S2.22), we have LAN

$$\log \prod_{i=1}^{n} \frac{dP_{s_{n,h}}}{dP_{s_0}}(\boldsymbol{X}) = \frac{s_n h}{\sqrt{n}} \sum_{i=1}^{n} \frac{dl(\boldsymbol{\theta}_0; \boldsymbol{x}_i)}{d\boldsymbol{\theta}}[\boldsymbol{w}] - \frac{s_n^2 h^2}{2}\|\boldsymbol{w}\|^2 + o_P(1).$$

(b) By the regularity conditions C1–C3 and the construction of $\mathfrak{W}$, for each $\boldsymbol{w}$ there exists a small neighborhood of s_0 ($s_0 = 0$), denoted as δ_{s_0}, such that for each $s \in \delta_{s_0}$, $\boldsymbol{\theta}_0 + s\boldsymbol{w}$ corresponds to a joint distribution and $l(\boldsymbol{\theta}_0 + s\boldsymbol{w}; \boldsymbol{X})$ is bounded. We denote $\lambda_{\boldsymbol{w}}(s) \equiv \rho(\boldsymbol{\theta}_0 + s\boldsymbol{w})$. It is easy to see that $\lambda'_{\boldsymbol{w}}(s) = \frac{d\rho(\boldsymbol{\theta}_0 + s\boldsymbol{w})}{d\boldsymbol{\theta}}[\boldsymbol{w}]$ is continuous function of s by (3.6) in the main paper, for each $s \in \delta_{s_0}$. In addition, $\lambda'_{\boldsymbol{w}}(s_0) = \frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}] \neq 0$.

(c) Since T_n is the path-wise regular estimator for $\rho(\boldsymbol{\theta}_0)$. Then for any fixed $\boldsymbol{w} \in \mathfrak{W}$ with $\|\boldsymbol{w}\| > 0$ and $\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}] \neq 0$, T_n is also the regular estimator of $\lambda_{\boldsymbol{w}}(s_0)$ (Wong, 1992). That is,

$$\limsup \Pr_{\boldsymbol{\theta}_{n,h}} \{T_n < \lambda_{\boldsymbol{w}}(s_{n,h})\} \leq \liminf \Pr_{\boldsymbol{\theta}_{n,-h}} \{T_n < \lambda_{\boldsymbol{w}}(s_{n,-h})\}.$$

The preceding arguments (a), (b) and (c) justify the conditions for Proposition 14 in Wong (1992). Hence, by Proposition 14 in Wong (1992) we have that

$$\limsup \Pr\left\{\sqrt{n}\,|T_n - \lambda_{\boldsymbol{w}}(s_0)| < h\right\} \le \Pr\left[\left|N\left\{0, \frac{\lambda'_{\boldsymbol{w}}(s_0)^2}{\|\boldsymbol{w}\|^2}\right\}\right| < h\right].$$

It implies that for any fixed $\boldsymbol{w} \in \mathfrak{W}$ with $\|\boldsymbol{w}\| > 0$ and $\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}] \neq 0$ and all $h > 0$, (S2.23) holds. This completes the second part of the proof.

By the first and the second parts of the proved result. We conclude that the asymptotic variance for $\rho\left(\hat{\boldsymbol{\theta}}_n\right)$ reaches the lower bound for all path-wise regular estimators for $\rho(\boldsymbol{\theta}_0)$. $\square$

S3. Technical lemmas and proofs

Lemma 1. *Let $\mathbb{M}(\theta) = Pl(\boldsymbol{\theta}; \boldsymbol{X})$ and Θ defined by (S2.6) contains $\boldsymbol{\theta}_0$. Then we have*

$$\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}) \ge cd^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}),$$

for any $\boldsymbol{\theta} \in \Theta$,

Proof of Lemma 1

Note that

$$
\begin{aligned}
&\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}) \\
&\quad = P\{l(\boldsymbol{\theta}_0; \boldsymbol{X}) - l(\boldsymbol{\theta}; \boldsymbol{X})\} \\
&\quad = P\left\{\Delta_1^{(1)}\Delta_2^{(1)} \log \frac{F_{0,0}(U_1, U_2)}{F_0(U_1, U_2)} + \Delta_1^{(1)}\Delta_2^{(2)} \log \frac{F_{0,0}(U_1, V_2) - F_{0,0}(U_1, U_2)}{F_0(U_1, V_2) - F_0(U_1, U_2)}\right. \\
&\qquad + \Delta_1^{(1)}\Delta_2^{(3)} \log \frac{F_{0,1}(U_1) - F_{0,1}(U_1, V_2)}{F_1(U_1) - F_0(U_1, V_2)} \\
&\qquad + \Delta_1^{(2)}\Delta_2^{(1)} \log \frac{F_{0,0}(V_1, U_2) - F_{0,0}(U_1, U_2)}{F_0(V_1, U_2) - F_0(U_1, U_2)} \\
&\qquad + \Delta_1^{(2)}\Delta_2^{(2)} \log \frac{F_{0,0}(V_1, V_2) - F_{0,0}(V_1, U_2) - F_{0,0}(U_1, V_2) + F_{0,0}(U_1, U_2)}{F_0(V_1, V_2) - F_0(V_1, U_2) - F_0(U_1, V_2) + F_0(U_1, U_2)} \\
&\qquad + \Delta_1^{(2)}\Delta_2^{(3)} \log \frac{F_{0,1}(V_1) - F_{0,0}(V_1, V_2) - F_{0,1}(U_1) + F_{0,0}(U_1, V_2)}{F_1(V_1) - F_0(V_1, V_2) - F_1(U_1) + F_0(U_1, V_2)} \\
&\qquad + \Delta_1^{(3)}\Delta_2^{(1)} \log \frac{F_{0,2}(U_2) - F_{0,0}(V_1, U_2)}{F_2(U_2) - F_0(V_1, U_2)} \\
&\qquad + \Delta_1^{(3)}\Delta_2^{(2)} \log \frac{F_{0,2}(V_2) - F_{0,2}(U_2) - F_{0,0}(V_1, V_2) + F_{0,0}(V_1, U_2)}{F_2(V_2) - F_2(U_2) - F_0(V_1, V_2) + F_0(V_1, U_2)} \\
&\qquad \left. + \Delta_1^{(3)}\Delta_2^{(3)} \log \frac{1 - F_{0,2}(V_2) - F_{0,1}(V_1) + F_{0,0}(V_1, V_2)}{1 - F_2(V_2) - F_1(V_1) + F_0(V_1, V_2)}\right\}.
\end{aligned}
$$

Then by the independence between (U_1, U_2, V_1, V_2) and (T_1, T_2), we have

$$
\begin{aligned}
&\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}) \\
&= P_{U_1,U_2,V_1,V_2}\left[F_0(U_1,U_2) m\left\{\frac{F_{0,0}(U_1,U_2)}{F_0(U_1,U_2)}\right\}\right. \\
&\quad + \{F_0(U_1,V_2) - F_0(U_1,U_2)\} m\left\{\frac{F_{0,0}(U_1,V_2) - F_{0,0}(U_1,U_2)}{F_0(U_1,V_2) - F_0(U_1,U_2)}\right\} \\
&\quad + \{F_1(U_1) - F_0(U_1,V_2)\} m\left\{\frac{F_{0,1}(U_1) - F_{0,1}(U_1,V_2)}{F_1(U_1) - F_0(U_1,V_2)}\right\} \\
&\quad + \{F_0(V_1,U_2) - F_0(U_1,U_2)\} m\left\{\frac{F_{0,0}(V_1,U_2) - F_{0,0}(U_1,U_2)}{F_0(V_1,U_2) - F_0(U_1,U_2)}\right\} \\
&\quad + \{F_0(V_1,V_2) - F_0(V_1,U_2) - F_0(U_1,V_2) + F_0(U_1,U_2)\} \\
&\qquad \cdot m\left\{\frac{F_{0,0}(V_1,V_2) - F_{0,0}(V_1,U_2) - F_{0,0}(U_1,V_2) + F_{0,0}(U_1,U_2)}{F_0(V_1,V_2) - F_0(V_1,U_2) - F_0(U_1,V_2) + F_0(U_1,U_2)}\right\} \\
&\quad + \{F_1(V_1) - F_0(V_1,V_2) - F_1(U_1) + F_0(U_1,V_2)\} \\
&\qquad \cdot m\left\{\frac{F_{0,1}(V_1) - F_{0,0}(V_1,V_2) - F_{0,1}(U_1) + F_{0,0}(U_1,V_2)}{F_1(V_1) - F_0(V_1,V_2) - F_1(U_1) + F_0(U_1,V_2)}\right\} \\
&\quad + \{F_2(U_2) - F_0(V_1,U_2)\} m\left\{\frac{F_{0,2}(U_2) - F_{0,0}(V_1,U_2)}{F_2(U_2) - F_0(V_1,U_2)}\right\} \\
&\quad + \{F_2(V_2) - F_2(U_2) - F_0(V_1,V_2) + F_0(V_1,U_2)\} \\
&\qquad \cdot m\left\{\frac{F_{0,2}(V_2) - F_{0,2}(U_2) - F_{0,0}(V_1,V_2) + F_{0,0}(V_1,U_2)}{F_2(V_2) - F_2(U_2) - F_0(V_1,V_2) + F_0(V_1,U_2)}\right\} \\
&\quad + \{1 - F_2(V_2) - F_1(V_1) + F_0(V_1,V_2)\} \\
&\qquad \left.\cdot m\left\{\frac{1 - F_{0,2}(V_2) - F_{0,1}(V_1) + F_{0,0}(V_1,V_2)}{1 - F_2(V_2) - F_1(V_1) + F_0(V_1,V_2)}\right\}\right],
\end{aligned}
\tag{S3.24}
$$

where $m(x) = x\log(x) - x + 1$. For $0 < x \leq M_h$, we can show that there exists $c_{M_h} > 0$ such that $m(x) \geq c_{M_h}(x-1)^2$. Then by the fact that the distribution functions are bounded and (S2.3), we have

$$\begin{aligned} &P_{U_1,U_2,V_1,V_2}\left[F_0(U_1,U_2)m\left\{\frac{F_{0,0}(U_1,U_2)}{F_0(U_1,U_2)}\right\}\right] \\ &\quad\geq cP_{U_1,U_2,V_1,V_2}\left[F_0(U_1,U_2)\left\{\frac{F_{0,0}(U_1,U_2)}{F_0(U_1,U_2)}-1\right\}^2\right] \\ &\quad\geq cP_{U_1,U_2,V_1,V_2}\left\{F_{0,0}(U_1,U_2)-F_0(U_1,U_2)\right\}^2. \end{aligned}$$

We can show the similar results as above for other terms of the right hand side in (S3.24). Hence,

$$
\begin{aligned}
&\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}) \\
&\quad \geq c P_{U_1,U_2,V_1,V_2} \left[F_{0,0}(U_1,U_2) - F_0(U_1,U_2)\right]^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2} \left[\{F_{0,0}(U_1,V_2) - F_0(U_1,V_2)\} - \{F_{0,0}(U_1,U_2) - F_0(U_1,U_2)\}\right]^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2} \left[\{F_{0,1}(U_1) - F_1(U_1)\} - \{F_{0,0}(U_1,V_2) - F_0(U_1,V_2)\}\right]^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2} \left[\{F_{0,0}(V_1,U_2) - F_0(V_1,U_2)\} - \{F_{0,0}(U_1,U_2) - F_0(U_1,U_2)\}\right]^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2} \left[\{F_{0,0}(V_1,V_2) - F_0(V_1,V_2)\} - \{F_{0,0}(V_1,U_2) - F_0(V_1,U_2)\}\right. \\
&\qquad\qquad \left. - \{F_{0,0}(U_1,V_2) - F_0(U_1,V_2)\} + \{F_{0,0}(U_1,U_2) - F_0(U_1,V_2)\}\right]^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2} \left[\{F_{0,1}(V_1) - F_1(V_1)\} - \{F_{0,0}(V_1,V_2) - F_0(V_1,V_2)\}\right. \\
&\qquad\qquad \left. - \{F_{0,1}(U_1) - F_1(U_1)\} + \{F_{0,0}(U_1,V_2) - F_0(U_1,V_2)\}\right]^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2} \left[\{F_{0,2}(U_2) - F_2(U_2)\} - \{F_{0,0}(V_1,U_2) - F_0(V_1,U_2)\}\right]^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2} \left[\{F_{0,2}(V_2) - F_2(V_2)\} - \{F_{0,2}(U_2) - F_2(U_2)\}\right. \\
&\qquad\qquad \left. - \{F_{0,0}(V_1,V_2) - F_0(V_1,V_2)\} + \{F_{0,0}(V_1,U_2) - F_0(V_1,U_2)\}\right]^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2} \left[-\{F_{0,2}(V_2) - F_2(V_2)\} - \{F_{0,1}(V_1) - F_1(V_1)\}\right. \\
&\qquad\qquad \left. + \{F_{0,0}(V_1,V_2) - F_0(V_1,V_2)\}\right]^2
\end{aligned}
\tag{S3.25}
$$

By $a^2 + b^2 \geq (a+b)^2/2$ we have

$$
\begin{aligned}
&\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}) \\
&\quad \geq c P_{U_1,U_2,V_1,V_2}\{F_{0,0}(U_1,U_2) - F_0(U_1,U_2)\}^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2}\left[\{F_{0,0}(U_1,V_2) - F_0(U_1,V_2)\} - \{F_{0,0}(U_1,U_2) - F_0(U_1,U_2)\}\right]^2 \\
&\quad \geq c P_{U_1,U_2,V_1,V_2}\{F_{0,0}(U_1,V_2) - F_0(U_1,V_2)\}^2 = c\|F_{0,0} - F_0\|^2_{L_2(P_{U_1,V_2})}.
\end{aligned}
$$

By $a^2 + b^2 + c^2 \geq (a+b+c)^2/3$ we have

$$
\begin{aligned}
&\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}) \\
&\quad \geq c P_{U_1,U_2,V_1,V_2}\{F_{0,0}(U_1,U_2) - F_0(U_1,U_2)\}^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2}\left[\{F_{0,0}(U_1,V_2) - F_0(U_1,V_2)\} - \{F_{0,0}(U_1,U_2) - F_0(U_1,U_2)\}\right]^2 \\
&\qquad + c P_{U_1,U_2,V_1,V_2}\left[\{F_{0,1}(U_1) - F_1(U_1)\} - \{F_{0,0}(U_1,V_2) - F_0(U_1,V_2)\}\right]^2 \\
&\quad \geq c P_{U_1,U_2,V_1,V_2}\{F_{0,1}(U_1) - F_1(U_1)\}^2 = c\|F_{0,1} - F_1\|^2_{L_2(P_{U_1})}.
\end{aligned}
$$

By general relationship $\sum_{i=1}^{j} a_i^2 \geq \left(\sum_{i=1}^{j} a_i\right)^2/j$ and using similar arguments as above for (S3.25), we can show that $\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta})$ is greater than the product of a positive constant and each of the terms in $d^2(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ defined by (3.1) in the main paper. This results in

$$
\begin{aligned}
&\mathbb{M}(\boldsymbol{\theta}_0) - \mathbb{M}(\boldsymbol{\theta}) \\
&\quad \geq c\|F_{0,0} - F_0\|^2_{L_2(P_{U_1,U_2})} + c\|F_{0,0} - F_0\|^2_{L_2(P_{U_1,V_2})} \\
&\qquad + c\|F_{0,0} - F_0\|^2_{L_2(P_{V_1,U_2})} + c\|F_{0,0} - F_0\|^2_{L_2(P_{V_1,V_2})} \\
&\qquad + c\|F_{0,1} - F_1\|^2_{L_2(P_{U_1})} + c\|F_{0,1} - F_1\|^2_{L_2(P_{V_1})} \\
&\qquad + c\|F_{0,2} - F_2\|^2_{L_2(P_{U_2})} + c\|F_{0,2} - F_2\|^2_{L_2(P_{V_2})} \\
&\quad \geq c d^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}). \quad \square
\end{aligned}
$$

Lemma 2. *Let* $\left\{B_i^{(1),l}\right\}_{i=1}^p$ *and* $\left\{B_i^{(2),l}\right\}_{j=1}^q$ *be the two sets of B-spline basis functions with the knot sequences* $\boldsymbol{\xi} = (\xi_i)_{i=1}^{p+1}$ *and* $\boldsymbol{\eta} = (\eta_j)_{j=1}^{q+1}$ *satisfying* $0 = \xi_1 = \cdots = \xi_l < \xi_{l+1} < \cdots < \xi_p < \xi_{p+1} = \xi_{p+l} = \tau_1$ *with* $\frac{\min_{i:l\leq i\leq p}(\xi_{i+1}-\xi_i)}{\max_{i:l\leq i\leq p}(\xi_{i+1}-\xi_i)} > c_{knot}$ *and* $0 = \eta_1 = \cdots = \eta_l < \eta_{l+1} < \cdots < \eta_q < \eta_{q+1} = \eta_{q+l} = \tau_2$ *with* $\frac{\min_{j:l\leq j\leq q}(\eta_{j+1}-\eta_j)}{\max_{j:l\leq j\leq q}(\eta_{j+1}-\eta_j)} > c_{knot}$*, respectively, for a small positive* c_{knot}*. Define*

$$
\begin{aligned}
\Phi_\delta = \Bigg\{ \phi : \phi(s,t) = &\sum_{i=1}^p \sum_{j=1}^q \alpha_{i,j} B_i^{(1),l}(s) B_j^{(2),l}(t), \\
&0 < \tau_{1,l} \leq s \leq \tau_{1,h} < \tau_1, 0 < \tau_{2,l} \leq t \leq \tau_{2,h} < \tau_2, \\
&\int_{[\tau_{1,l},\tau_{1,h}]\times[\tau_{2,l},\tau_{2,h}]} \phi^2(s,t)\,dtds \leq \delta^2 \Bigg\}.
\end{aligned}
$$

Then for $\epsilon < \delta$*, we have*

$$
\log N_{[\,]}(\epsilon, \Phi_\delta, \|\cdot\|_\infty) \leq cpq \log(\delta/\epsilon).
$$

Proof of Lemma 2.

Denote $\langle f,g\rangle = \int_{[\tau_{1,l},\tau_{1,h}]\times[\tau_{2,l},\tau_{2,h}]} f(s,t)g(s,t)dtds$. The Gram-Schmidt process based on $\langle f,g\rangle$ leads to the set of orthogonal basis functions $\{O_k(\cdot,\cdot)\}_{k=1}^K$ for Φ_δ, with $K \le pq$ and equal to the number of elements of $\left\{B_i^{(1),l}B_j^{(2),l}\right\}_{(i,j)\in\mathcal{I}}$, where for each member $(i,j)\in\mathcal{I}$, $B_i^{(1),l}B_j^{(2),l}$ has a support on $[\tau_{1,l},\tau_{1,h}]\times[\tau_{2,l},\tau_{2,h}]$. So any $\phi\in\Phi_\delta$ can be written as $\phi(s,t)=\sum_{k=1}^K \omega_k O_k(s,t)$, where $\langle O_k, O_{k'}\rangle \ge c\left(\frac{1}{K}\right)$ for $k=k'$ because the construction of knot sequences $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ implies that the support for every basis $B_i^{(1),l}B_j^{(2),l}$ has an area greater than $c\left(\frac{1}{K}\right)$ and $\langle O_k, O_{k'}\rangle = 0$ for $k\ne k'$. Then

$$\sum_{k=1}^K \omega_k^2 c\left(\frac{1}{K}\right) \le \sum_{k=1}^K \omega_k^2 \langle O_k, O_k\rangle = \langle \phi,\phi\rangle \le \delta^2.$$

Hence,

$$\sum_{k=1}^K \omega_k^2 \le cK\delta^2 \le cpq\delta^2. \tag{S3.26}$$

Let

$$S = \{\boldsymbol{\omega} = (\omega_1,\cdots,\omega_K)' : \sum_{k=1}^K \omega_k^2 \le cpq\delta^2\}.$$

Lemma 0.4 of Wu and Zhang (2012) indicates that there exist ϵ-balls $B_1, B_2, \cdots, B_{[(\delta/\epsilon)^{cpq}]}$ centered at $\boldsymbol{\omega}^{(1)} = \left(\omega_1^{(1)},\cdots,\omega_K^{(1)}\right)'$, $\boldsymbol{\omega}^{(2)} = \left(\omega_1^{(2)},\cdots,\omega_K^{(2)}\right)'$,

$\cdots$,

$\boldsymbol{\omega}^{([(\delta/\epsilon)^{cpq}])} = \left(\omega_1^{([(\delta/\epsilon)^{cpq}])},\cdots,\omega_K^{([(\delta/\epsilon)^{cpq}])}\right)'$, respectively, which cover S. For $m=1,\cdots,[(\delta/\epsilon)^{cpq}]$, define

$$\psi^{(m)}(s,t) = \sum_{k=1}^K \omega_k^{(m)} O_k(s,t).$$

On the other hand, we consider any $\phi(s,t) = \sum_{k=1}^{K} \omega_k O_k(s,t) \in \Phi_\delta$ with its coefficients $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_K)' \in S$ defined by (S3.26). Since ϵ-balls $B_1, B_2, \cdots, B_{[(\delta/\epsilon)^{cpq}]}$ cover S, there exists a d with $1 \le d \le [(\delta/\epsilon)^{cpq}]$, such that

$$\|\boldsymbol{\omega} - \boldsymbol{\omega}^{(d)}\|_\infty = \max_{k:1\le k\le K} \left|\omega_k - \omega_k^{(d)}\right| \le \epsilon.$$

Then, for any $(s,t) \in [\tau_{1,l}, \tau_{1,h}] \times [\tau_{2,l}, \tau_{2,h}]$,

$$\begin{aligned}
\left|\phi(s,t) - \psi^{(d)}(s,t)\right| &= \left|\sum_{k=1}^{K} \left(\omega_k - \omega_k^{(d)}\right) O_k(s,t)\right| \\
&\le \max_{k:1\le k\le K} \left|\omega_k - \omega_k^{(d)}\right| \sum_{k=1}^{K} |O_k(s,t)| \\
&\le c \max_{k:1\le k\le K} \left|\omega_k - \omega_k^{(d)}\right| \left\{ c \sum_{i=1}^{p} \sum_{j=1}^{q} B_i^{(1),l}(s) B_j^{(2),l}(t) \right\} \\
&\le c \max_{k:1\le k\le K} \left|\omega_k - \omega_k^{(d)}\right| \le c\epsilon,
\end{aligned}$$

where we use the fact that $\sum_{k=1}^{K} |O_k(s,t)| \le c \sum_{i=1}^{p} \sum_{j=1}^{q} B_i^{(1),l}(s) B_j^{(2),l}(t)$ with c not related to p and q due to the structure of O_k by the Gram-Schmidt process and the fact that each B-spline basis function only has support on l sub-intervals. It implies that

$$\left\|\phi - \psi^{(d)}\right\|_\infty \le c\epsilon.$$

For $d = 1, \cdots, [(\delta/\epsilon)^{cpq}]$, let $\Psi_\epsilon^{(d)} = \left\{\psi : \|\psi - \psi^{(d)}\|_\infty \le c\epsilon, \psi \in \Psi\right\}$, where

$$\Psi = \Bigg\{\psi : \psi(s,t) = \sum_{k=1}^{K} \omega_k O_k(s,t),$$
$$0 < \tau_{1,l} \le s \le \tau_{1,h} < \tau_1, 0 < \tau_{2,l} \le t \le \tau_{2,h} < \tau_2, \Bigg\}.$$

Then, $\phi \in \Psi_\epsilon^{(d)}$ for some $1 \le d \le [(\delta/\epsilon)^{cpq}]$. Hence, $\left\{\Psi_\epsilon^{(d)}\right\}_{d=1}^{[(\delta/\epsilon)^{cpq}]}$ cover Φ_δ. Therefore, ϵ-covering number of Φ_δ is bounded by $[(\delta/\epsilon)^{cpq}]$. By the fact $N_{[\ ]}(2\epsilon, \Phi_\delta, \|\cdot\|_\infty) \le N(\epsilon, \Phi_\delta, \|\cdot\|_\infty)$, it is true that

$$\log N_{[\ ]}(\epsilon, \Phi_\delta, \|\cdot\|_\infty) \le cpq \log(\delta/\epsilon). \quad \square$$

Lemma 3. *Let $\left\{B_i^{(1),l}\right\}_{i=1}^{p}$ be a set of B-spline basis functions with the knot sequence $\boldsymbol{\xi}$ satisfying $0 = \xi_1 = \cdots = \xi_l < \xi_{l+1} < \cdots < \xi_p < \xi_{p+1} = \xi_{p+l} = \tau_1$ with $\frac{\min_{i:l \le i \le p}(\xi_{i+1} - \xi_i)}{\max_{i:l \le i \le p}(\xi_{i+1} - \xi_i)} > c_{knot}$ for a small positive number c_{knot}. Define*

$$\Phi_\delta = \Bigg\{\phi : \phi(s) = \sum_{i=1}^{p} \beta_i B_i^{(1),l}(s),$$
$$0 < \tau_{1,l} \le s \le \tau_{1,h} < \tau_1, \int_{[\tau_{1,l}, \tau_{1,h}]} \phi^2(s) ds \le \delta^2 \Bigg\}.$$

Then for $\epsilon < \delta$, we have

$$\log N_{[\ ]}(\epsilon, \Phi_\delta, \|\cdot\|_\infty) \le cp \log(\delta/\epsilon).$$

Proof of Lemma 3.

The proof follows exactly the same lines as those for Lemma 2, so it is not presented here. □

Lemma 4. *Suppose* $\Lambda_0(s,t)$ *and* $\Lambda(s,t)$ *are both nondecreasing in* s *and in* t *in the domain* $[\tau_{1,l}, \tau_{1,h}] \times [\tau_{2,l}, \tau_{2,h}]$ *satisfying* $\|\Lambda - \Lambda_0\|_{L^2(\mu)} \leq \eta$. *Then*

$$\sup_{(s,t)\in[\tau_{1,l},\tau_{1,h}]\times[\tau_{2,l},\tau_{2,h}]} |\Lambda(s,t) - \Lambda_0(s,t)| \leq c\eta^{2/3}$$

if

(1) $\Lambda_0(s,t)$ *has mixed derivative* $\frac{\partial^2 \Lambda_0(s,t)}{\partial s \partial t}$ *and there exits a constant* $0 < f_0 < \infty$ *such that* $1/f_0 \leq \frac{\partial^2 \Lambda_0(s,t)}{\partial s \partial t} \leq f_0$ *on* $[\tau_{1,l}, \tau_{1,h}] \times [\tau_{2,l}, \tau_{2,h}]$.

(2) The probability measure μ *is absolute continuous with respect to Lebesgue measure with mixed derivative* $\frac{\partial^2 \mu(s,t)}{\partial s \partial t}$ *satisfying* $\frac{\partial^2 \mu(s,t)}{\partial s \partial t} \geq c_0$, *for some positive* c_0.

Proof of Lemma 4

Suppose that $(s', t') \in [\tau_{1,l}, \tau_{1,h}] \times [\tau_{2,l}, \tau_{2,h}]$ satisfies

$$|\Lambda(s',t') - \Lambda_0(s',t')| \geq (1/2) \sup_{(s,t)\in[\tau_{1,l},\tau_{1,h}]\times[\tau_{2,l},\tau_{2,h}]} |\Lambda(s,t) - \Lambda_0(s,t)| \equiv \xi/2.$$

Then either $\Lambda(s',t') \geq \Lambda_0(s',t') + \xi/2$ or $\Lambda_0(s',t') \geq \Lambda(s',t') + \xi/2$. In the following, we only show the inequality for the first case, $\Lambda(s',t') \geq \Lambda_0(s',t') + \xi/2$, as the arguments are parallel for the second case.

There exists (s'', t'') with $s'' \geq s', t'' \geq t'$, such that $\Lambda_0(s'', t'') = \Lambda_0(s', t') + \xi/2$ by Condition (1).

Then

$$\begin{aligned}
\eta^2 &\geq \int \{\Lambda(s,t) - \Lambda_0(s,t)\}^2 d\mu(s,t) \\
&\geq \int_{s'}^{s''} \int_{t'}^{t''} \{\Lambda(s,t) - \Lambda_0(s,t)\}^2 \frac{\partial^2 \mu(s,t)}{\partial s \partial t} ds dt \\
&\geq \int_{s'}^{s''} \int_{t'}^{t''} \{\Lambda_0(s'',t'') - \Lambda_0(s,t)\}^2 \frac{\partial^2 \mu(s,t)}{\partial s \partial t} ds dt \\
&\geq c_0 \int_{s'}^{s''} \int_{t'}^{t''} \{\Lambda_0(s'',t'') - \Lambda_0(s,t)\}^2 ds dt,
\end{aligned}$$

where

$$\Lambda_0(s'', t'') - \Lambda_0(s,t) = \int_{s}^{s''} \int_{t}^{t''} \frac{\partial^2 \Lambda_0(x,y)}{\partial x \partial y} dx dy \geq (1/f_0)(s'' - s)(t'' - t).$$

Hence

$$\begin{aligned}
\eta^2 &\geq c \int_{s'}^{s''} \int_{t'}^{t''} (s'' - s)^2 (t'' - t)^2 ds dt \geq c(s'' - s')^3 (t'' - t')^3 \\
&= c \left\{ \int_{s'}^{s''} \int_{t'}^{t''} ds dt \right\}^3 \geq (c/f_0^3) \left\{ \int_{s'}^{s''} \int_{t'}^{t''} \frac{\partial^2 \Lambda_0(s,t)}{\partial s \partial t} ds dt \right\}^3 \\
&\geq c\{\Lambda_0(s'', t'') - \Lambda_0(s', t')\}^3 = c\xi^3. \quad \square
\end{aligned}$$

Lemma 5. *Given that C1–C4 hold and $p + r > 3$ in C1 and C2. There exist ϵ_n and spline function vector $\boldsymbol{w}_n^*$, such that (S2.16), (S2.17), and (S2.18) given in the proof of Theorem 2 hold.*

Proof of Lemma 5

Before proving the three main results given in the proof of Theorem 2, we show some intermediate results that are imperative to the proof of the lemma.

Since for any $\boldsymbol{w} \in \mathfrak{W}$, $\langle \boldsymbol{w}^*, \boldsymbol{w} \rangle = \frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]$, with $\langle \boldsymbol{w}^*, \boldsymbol{w} \rangle$ given by the Fisher information inner product and $\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]$ given by (3.6) in the main paper. Using the regularity conditions C1-C4, it can be shown that $\boldsymbol{w}^* = (w_0^*, w_1^*, w_2^*)'$ is a vector of piecewise continuous functions with bounded derivatives $\partial w_0^*(t_1, t_2)/\partial t_1$, $\partial w_0^*(t_1, t_2)/\partial t_2$, $dw_1^*(t_1)/dt_1$ and $dw_2^*(t_2)/dt_2$. Then by Jackson's Theorem in De Boor (2001) with Lemma 0.2 in Wu and Zhang (2012), there exist spline functions $w_{n,0}^*(\cdot, \cdot)$, $w_{n,1}^*(\cdot)$, $w_{n,2}^*(\cdot)$, such that for $\boldsymbol{w}_n^* = (w_{n,0}^*, w_{n,1}^*, w_{n,2}^*)'$, we have

$$\|\boldsymbol{w}_n^* - \boldsymbol{w}^*\| \le cn^{-\kappa_n} = o\left(n^{-\frac{1}{2(p+r)+2}}\right), \tag{S3.27}$$

by choosing $\kappa_n > \frac{1}{2(p+r)+2}$, where n^{κ_n} is the number of uniformly distributed interior knots and we use the fact that Fisher information norm (3.4) in the main paper is bounded by the infinity norm $\|\cdot\|_\infty$.

In what follows, we establish that for $\tilde{\boldsymbol{\theta}} = \{\tilde{F}_{n,0}(\cdot, \cdot), \tilde{F}_{n,1}(\cdot), \tilde{F}_{n,2}(\cdot)\}'$

between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n$,

$$\left| P\left\{ \frac{d^2 l(\tilde{\boldsymbol{\theta}};\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0] - \frac{d^2 l(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0] \right\} \right| = o_P\left(n^{-1-\kappa^*}\right), \tag{S3.28}$$

for some small $\kappa^* > 0$. Taking the absolute value of the first term in $\frac{d^2 l(\tilde{\boldsymbol{\theta}};\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0] - \frac{d^2 l(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0]$, we have

$$T_{n,1} \equiv \Delta_1^{(1)}\Delta_2^{(1)} \left| -\frac{1}{\tilde{F}^2_{n,0}(U_1,U_2)} + \frac{1}{F^2_{0,0}(U_1,U_2)} \right| \left\{ \hat{F}_{n,0}(U_1,U_2) - F_{0,0}(U_1,U_2) \right\}^2.$$

By Theorem 1, we have

$$\left\{ P_{U_1,U_2}\left(\hat{F}_{n,0}-F_{0,0}\right)^2 \right\}^{1/2} = O_P\left(n^{-\frac{p+r}{2(p+r)+2}}\right).$$

Then for any small $\epsilon > 0$, there exists an $M > 0$ such that for all positive integer n,

$$\Pr\left[\left\{ P_{U_1,U_2}\left(\hat{F}_{n,0}-F_{0,0}\right)^2 \right\}^{1/2} \le M n^{-\frac{p+r}{2(p+r)+2}} \right] > 1-\epsilon.$$

Given $\left\{ P_{U_1,U_2}\left(\hat{F}_{n,0}-F_{0,0}\right)^2 \right\}^{1/2} \le M n^{-\frac{p+r}{2(p+r)+2}}$, using Lemma 4 and Conditions C3 and C4 together with the fact that $\tilde{\boldsymbol{\theta}}$ is between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n$ leads to

$$T_{n,1} \le c_M \left| \hat{F}_{n,0}(U_1,U_2) - F_{0,0}(U_1,U_2) \right|^3$$

and by $p+r>3$,

$$\begin{aligned} P_{U_1,U_2}T_{n,1} &\le c_M P_{U_1,U_2}\left|\hat{F}_{n,0}-F_{0,0}\right|^3 \\ &\le c_M P_{U_1,U_2}\left\{\hat{F}_{n,0}-F_{0,0}\right\}^2\left\|\hat{F}_{n,0}-F_{0,0}\right\|_\infty \\ &\le c'_M n^{-\frac{2(p+r)}{2(p+r)+2}} n^{-\frac{(2/3)(p+r)}{2(p+r)+2}} = c'_M n^{-\frac{2(p+r)+(2/3)(p+r)}{2(p+r)+2}} \\ &= c'_M n^{-\kappa_0^*} n^{-1-\kappa^*}, \end{aligned}$$

for some $\kappa_0^*>0$ and $\kappa^*>0$. We also know for any small $\epsilon'>0$, there exists an integer $N>0$ such that for $n>N$

$$c'_M n^{-\kappa_0^*}<\epsilon'.$$

In summary, for any $\epsilon>0$ and $\epsilon'>0$, there exists an integer N, such that for $n>N$

$$\begin{aligned} \Pr\left\{P_{U_1,U_2}\left(\frac{T_{n,1}}{n^{-1-\kappa^*}}\right)<\epsilon'\right\} &\ge \Pr\left\{P_{U_1,U_2}\left(\frac{T_{n,1}}{n^{-1-\kappa^*}}\right)\le c'_M n^{-\kappa_0^*}\right\} \\ &\ge \Pr\left[\left\{P_{U_1,U_2}\left(\hat{F}_{n,0}-F_{0,0}\right)^2\right\}^{1/2}\le Mn^{-\frac{p+r}{2(p+r)+2}}\right]\ge 1-\epsilon, \end{aligned}$$

by the fact that the event $\left[\left\{P_{U_1,U_2}\left(\hat{F}_{n,0}-F_{0,0}\right)^2\right\}^{1/2}\le Mn^{-\frac{p+r}{2(p+r)+2}}\right]$ is contained in the event $\left\{P_{U_1,U_2}\left(\frac{T_{n,1}}{n^{-1-\kappa^*}}\right)\le c'_M n^{-\kappa_0^*}\right\}$. Hence $P_{U_1,U_2}T_{n,1}=o_P\left(n^{-1-\kappa^*}\right)$.

Proceeding in the same manner for all other terms in

$$\frac{d^2 l(\tilde{\boldsymbol{\theta}};\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0]-\frac{d^2 l(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0]$$

results (S3.28).

Since $\boldsymbol{w}^*$ is piecewise continuous, it is bounded in a finite interval. Then for large n, $\boldsymbol{w}_n^*$ is uniformly bounded. Hence, by similar arguments we can show that for $\tilde{\boldsymbol{\theta}}$ between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n \pm \epsilon_{n,0}\boldsymbol{w}_n^*$ for any $\epsilon_{n,0}$ with $\epsilon_{n,0} = o\left(n^{-1/2}\right)$

$$\left| P\left\{ \frac{d^2 l(\tilde{\boldsymbol{\theta}};\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n \pm \epsilon_{n,0}\boldsymbol{w}_n^* - \boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n \pm \epsilon_{n,0}\boldsymbol{w}_n^* - \boldsymbol{\theta}_0] \right.\right.$$
$$\left.\left. - \frac{d^2 l(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n \pm \epsilon_{n,0}\boldsymbol{w}_n^* - \boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n \pm \epsilon_{n,0}\boldsymbol{w}_n^* - \boldsymbol{\theta}_0] \right\}\right|$$
$$= o_P\left(n^{-1-\kappa^*}\right).$$

Now if we let $\epsilon_n = n^{-1/2-\kappa^*}$, then $\epsilon_n = o\left(n^{-1/2}\right)$ and $\epsilon_n o_P\left(n^{-1/2}\right) = o_P\left(\epsilon_n n^{-1/2}\right) = o_P\left(n^{-1-\kappa^*}\right)$. Hence, we have

$$\left| P\left\{ \frac{d^2 l(\tilde{\boldsymbol{\theta}};\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0] - \frac{d^2 l(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0] \right\}\right|$$
$$= \epsilon_n o_P\left(n^{-1/2}\right) \tag{S3.29}$$

and

$$\left| P\left\{ \frac{d^2 l(\tilde{\boldsymbol{\theta}};\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n \pm \epsilon_n\boldsymbol{w}_n^* - \boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n \pm \epsilon_n\boldsymbol{w}_n^* - \boldsymbol{\theta}_0] \right.\right.$$
$$\left.\left. - \frac{d^2 l(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n \pm \epsilon_n\boldsymbol{w}_n^* - \boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n \pm \epsilon_n\boldsymbol{w}_n^* - \boldsymbol{\theta}_0] \right\}\right|$$
$$= \epsilon_n o_P\left(n^{-1/2}\right). \tag{S3.30}$$

Next, we use Corollary 19.35 in van der Vaart (1998) to establish

$$(\mathbb{P}_n - P)\left\{\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^* - \boldsymbol{w}^*]\right\} = o_P(n^{-1/2}). \tag{S3.31}$$

The first term in $\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^* - \boldsymbol{w}^*]$ is $\Delta_1^{(1)}\Delta_2^{(1)}\left(w_{n,0}^* - w_0^*\right)(U_1,U_2)/F_{0,0}(U_1,U_2)$.

Let

$$\mathcal{F}^{(1)} = \left\{\frac{\Delta_1^{(1)}\Delta_2^{(1)}\left(w_{n,0}^* - w_0^*\right)(U_1,U_2)}{F_{0,0}(U_1,U_2)}\right\}$$

be a single element set. Since there exists a positive constant c_w, such that

$$\left|\frac{\Delta_1^{(1)}\Delta_1^{(1)}\left(w_{n,0}^* - w_0^*\right)(U_1,U_2)}{F_{0,0}(U_1,U_2)}\right| \le c_w\left|w_{n,0}^* - w_0^*\right|$$

by Conditions C3 and C4. Then $c_w\left|w_{n,0}^* - w_0^*\right|$ is an envelope function for $\mathcal{F}^{(1)}$. As we mentioned previously, it can be shown by Conditions C3 and C4 that Fisher information distance $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$ defined by (3.4) in the main paper can be bounded by the distance $d(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ defined by (3.1) in the main paper. Therefore, (S3.27) implies $\left\{P_{U_1,U_2}\left(c_w\left|w_{n,0}^* - w_0^*\right|\right)^2\right\}^{1/2} \le cn^{-\frac{1}{2(p+r)+2}}$. Then by Corollary 19.35 in van der Vaart (1998), we have

$$\begin{aligned}
E_P\|\mathbb{G}_n\|_{\mathcal{F}^{(1)}} &\le cJ_{[\,]}\left\{cn^{-\frac{1}{2(p+r)+2}}, \mathcal{F}^{(1)}, L_2(P)\right\} \\
&= \int_0^{cn^{-\frac{1}{2(p+r)+2}}} \sqrt{1+\log N_{[\,]}\{\epsilon,\mathcal{F}^{(1)},L_2(P)\}}d\epsilon \\
&= \int_0^{cn^{-\frac{1}{2(p+r)+2}}} \sqrt{1+\log 1}d\epsilon = cn^{-\frac{1}{2(p+r)+2}}
\end{aligned}$$

using the fact that $\mathcal{F}^{(1)}$ is a single element set. Then it follows from

Markov's inequality that

$$(\mathbb{P}_n - P)\left\{\frac{\Delta_1^{(1)}\Delta_2^{(1)}\left(w_{n,0}^* - w_0^*\right)(U_1, U_2)}{F_{0,0}(U_1, U_2)}\right\} = o_P(n^{-1/2}).$$

Proceeding in the same manner for all other terms in $\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^* - \boldsymbol{w}^*]$ results in (S3.31).

Similarly, in what follows we use Corollary 19.35 in van der Vaart (1998) to establish

$$(\mathbb{P}_n - P)\left\{\frac{dl(\tilde{\boldsymbol{\theta}};\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^*] - \frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^*]\right\} = o_P(n^{-1/2}), \qquad \text{(S3.32)}$$

where $\tilde{\boldsymbol{\theta}} = (\tilde{F}_{n,0}(\cdot,\cdot), \tilde{F}_{n,1}(\cdot), \tilde{F}_{n,2}(\cdot))'$ is a spline function vector between $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*$.

The first term in $\frac{dl(\tilde{\boldsymbol{\theta}};\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^*] - \frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^*]$ is $\Delta_1^{(1)}\Delta_2^{(1)}w_{n,0}^*(U_1,U_2)\left\{\frac{1}{\tilde{F}_{n,0}(U_1,U_2)} - \frac{1}{F_{0,0}(U_1,U_2)}\right\}$. So we define

$$f^*(F_{n,0}) \equiv \Delta_1^{(1)}\Delta_2^{(1)}w_{n,0}^*\left\{\frac{1}{F_{n,0}} - \frac{1}{F_{0,0}}\right\}$$

and let

$$\mathcal{F}^{(2)} = \left[f^*(F_{n,0}) : \left\{P_{U_1,U_2}(F_{n,0} - F_{0,0})^2\right\}^{1/2} \le cn^{-\frac{p+r}{2(p+r)+2}}\right],$$

where $F_{n,0}$ is between $F_{n,0}^{(0)}$ and $F_{n,0}^{(0)} \pm \epsilon_n w_{n,0}^*$ with $\boldsymbol{\theta}_n = \left(F_{n,0}^{(0)}, \cdot, \cdot\right) \in \Theta_n$ for Θ_n defined by (S2.5). Lemma 4 and the fact that $\left\|\epsilon_n w_{n,0}^*\right\|_\infty \le c\epsilon_n$ ($\boldsymbol{w}_n^*$ is uniformly bounded) imply $\|F_{n,0} - F_{0,0}\|_\infty = cn^{-\frac{2(p+r)}{6(p+r)+6}}$. Hence, we have

$$F_{n,0} \ge F_{0,0} - cn^{-\frac{2(p+r)}{6(p+r)+6}}.$$

It follows that

$$\left|\Delta_1^{(1)}\Delta_2^{(1)}w_{n,0}^*(U_1,U_2)\left\{\frac{1}{F_{n,0}(U_1,U_2)}-\frac{1}{F_{0,0}(U_1,U_2)}\right\}\right|$$
$$\leq \left|w_{n,0}^*\right|\left\{\frac{1}{F_{0,0}-cn^{-\frac{2(p+r)}{6(p+r)+6}}}-\frac{1}{F_{0,0}}\right\}$$
$$\leq cn^{-\frac{2(p+r)}{6(p+r)+6}}\left|w_{n,0}^*\right|,$$

by Conditions C3 and C4. Hence, $cn^{-\frac{2(p+r)}{6(p+r)+6}}\left|w_{n,0}^*\right|$ is an envelope function for $\mathcal{F}^{(2)}$, with

$$\left[P_{U_1,U_2}\left\{cn^{-\frac{2(p+r)}{6(p+r)+6}}\left|w_{n,0}^*\right|\right\}^2\right]^{1/2}\leq cn^{-\frac{2(p+r)}{6(p+r)+6}}.$$

On the other hand, Lemma 2 implies that by choosing $p_n=q_n=n^\kappa$ and $\kappa=\frac{1}{2(p+r)+2}$, ϵ-bracketing number with $\|\cdot\|_\infty$-norm for set

$$\left[F_{n,0}^{(0)}:\left\{P_{U_1,U_2}\left(F_{n,0}^{(0)}-F_{0,0}\right)^2\right\}^{1/2}\leq cn^{-\frac{p+r}{2(p+r)+2}}\right]$$

is equal to $\left\{cn^{-\frac{p+r}{2(p+r)+2}}/\epsilon\right\}^{cn^{\frac{2}{2(p+r)+2}}}$, where $F_{n,0}^{(0)}$ satisfies $\boldsymbol{\theta}_n=\left(F_{n,0}^{(0)},\cdot,\cdot\right)\in\Theta_n$ for Θ_n defined by (S2.5). Then by Conditions C3 and C4 with some algebra we can show that

$$\log N_{[\ ]}\left\{\epsilon,\mathcal{F}^{(2)},L_2(P)\right\}=cn^{\frac{2}{2(p+r)+2}}\log\left\{\frac{cn^{-\frac{p+r}{2(p+r)+2}}}{\epsilon}\right\}.$$

Now by Corollary 19.35 in van der Vaart (1998), we have

$$\begin{aligned} E_P\|\mathbb{G}_n\|_{\mathcal{F}^{(2)}} &\le cJ_{[\,]}\left\{cn^{-\frac{2(p+r)}{6(p+r)+6}}, \mathcal{F}^{(2)}, L_2(P)\right\} \\ &= \int_0^{cn^{-\frac{2(p+r)}{6(p+r)+6}}} \sqrt{1+\log N_{[\,]}\{\epsilon, \mathcal{F}^{(2)}, L_2(P)\}}d\epsilon \\ &\le \int_0^{cn^{-\frac{2(p+r)}{6(p+r)+6}}} cn^{\frac{3}{6(p+r)+6}} n^{-\frac{3/2(p+r)}{6(p+r)+6}} \epsilon^{-1/2} d\epsilon \\ &= cn^{\frac{3-5/2(p+r)}{6(p+r)+6}}. \end{aligned} \tag{S3.33}$$

Since $\left\|\epsilon_n w^*_{n,0}\right\|_\infty \le c\epsilon_n$ and $\epsilon_n = o\left(n^{-1/2}\right)$, Theorem 1 implies

$$\left\{P_{U_1,U_2}\left(\tilde{F}_{n,0} - F_{0,0}\right)^2\right\}^{1/2} = O_P\left(n^{-\frac{p+r}{2(p+r)+2}}\right).$$

Then for any small $\epsilon_0 > 0$, there exists an $M > 0$ such that for all positive integer n

$$\Pr\left[\left\{P_{U_1,U_2}\left(\tilde{F}_{n,0} - F_{0,0}\right)^2\right\}^{1/2} \le Mn^{-\frac{p+r}{2(p+r)+2}}\right] > 1-\epsilon_0. \tag{S3.34}$$

If $\left\{P_{U_1,U_2}\left(\tilde{F}_{n,0} - F_{0,0}\right)^2\right\}^{1/2} \le Mn^{-\frac{p+r}{2(p+r)+2}}$, then $f^*\left(\tilde{F}_{n,0}\right) \in \mathcal{F}^{(2)}$. So we know that

$$E_P\left[\left|\mathbb{G}_n\left\{f^*\left(\tilde{F}_{n,0}\right)\right\}\right| \middle| \left\{P_{U_1,U_2}\left(\tilde{F}_{n,0} - F_{0,0}\right)^2\right\}^{1/2} \le Mn^{-\frac{p+r}{2(p+r)+2}}\right] \le E_P\|\mathbb{G}_n\|_{\mathcal{F}^{(2)}}.$$

Hence, by (S3.33) we have

$$E_P\left[\left|\mathbb{G}_n\left\{f^*\left(\tilde{F}_{n,0}\right)\right\}\right|\left|\left\{P_{U_1,U_2}\left(\tilde{F}_{n,0}-F_{0,0}\right)^2\right\}^{1/2}\le Mn^{-\frac{p+r}{2(p+r)+2}}\right]$$
$$\le c_M n^{\frac{3-5/2(p+r)}{6(p+r)+6}}=o(1),$$

since $p+r>3$. Then conditional Markov's inequality implies that, for any small ϵ_1 and ϵ_2 there exists an integer $N>0$ such that for $n>N$ we have

$$\Pr\left[\left|(\mathbb{P}_n-P)\left\{\frac{f^*\left(\tilde{F}_{n,0}\right)}{n^{-1/2}}\right\}\right|<\epsilon_1\left|\left\{P_{U_1,U_2}\left(\tilde{F}_{n,0}-F_{0,0}\right)^2\right\}^{1/2}\le Mn^{-\frac{p+r}{2(p+r)+2}}\right]$$
$$>1-\epsilon_2.$$

Now by (S3.34) and the definition of conditional probability we have for $n>N$

$$\Pr\left[\left|(\mathbb{P}_n-P)\left\{\frac{f^*\left(\tilde{F}_{n,0}\right)}{n^{-1/2}}\right\}\right|<\epsilon_1\right]>(1-\epsilon_2)(1-\epsilon_0)>1-\epsilon_2-\epsilon_0.$$

Finally, for any small ϵ_1, ϵ, if we let $\epsilon_0=\epsilon_2=\epsilon/2$, by the preceding display there exists an integer $N>0$, such that for $n>N$

$$\Pr\left[\left|(\mathbb{P}_n-P)\left\{\frac{f^*\left(\tilde{F}_{n,0}\right)}{n^{-1/2}}\right\}\right|<\epsilon_1\right]>1-\epsilon.$$

That is,

$$(\mathbb{P}_n-P)\left\{f^*\left(\tilde{F}_{n,0}\right)\right\}=o_P\left(n^{-1/2}\right).$$

Proceeding in the same manner for all other terms in $\frac{dl(\tilde{\boldsymbol{\theta}};\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^*]-\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^*]$ results in (S3.32).

In what follows we establish the three main results for this lemma.

First, we verify (S2.16) given in the proof of Theorem 2 holds.

$$\begin{aligned}
P\left(r\left[\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0; \boldsymbol{X}\right]\right) &= P\left\{l\left(\hat{\boldsymbol{\theta}}_n; \boldsymbol{X}\right) - l(\boldsymbol{\theta}_0; \boldsymbol{X}) - \frac{dl(\boldsymbol{\theta}_0; \boldsymbol{X})}{d\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right]\right\} \\
&= \frac{1}{2}P\left\{\frac{d^2 l(\boldsymbol{\theta}_0; \boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0]\right\} \\
&+ \frac{1}{2}P\left\{\frac{d^2 l(\tilde{\boldsymbol{\theta}}; \boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0]\right. \\
&\qquad \left. - \frac{d^2 l(\boldsymbol{\theta}_0; \boldsymbol{X})}{d\boldsymbol{\theta}^2}[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0]\right\},
\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ is between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n$. Then by the definition of Fisher information norm $\|\cdot\|$ and (S3.29)

$$P\left(r\left[\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0; \boldsymbol{X}\right]\right) = -\frac{\left\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right\|^2}{2} + \epsilon_n o_P\left(n^{-1/2}\right). \tag{S3.35}$$

Similarly, by (S3.30)

$$P\left(r\left[\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*, \boldsymbol{\theta}_0; \boldsymbol{X}\right]\right) = -\frac{\left\|\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^* - \boldsymbol{\theta}_0\right\|^2}{2} + \epsilon_n o_P\left(n^{-1/2}\right). \tag{S3.36}$$

By (S3.35) and (S3.36), we have

$$
\begin{aligned}
&P\left(r\left[\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0; \boldsymbol{X}\right] - r\left[\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*, \boldsymbol{\theta}_0; \boldsymbol{X}\right]\right) \\
&= \frac{\left\|\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^* - \boldsymbol{\theta}_0\right\|^2 - \left\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right\|^2}{2} + \epsilon_n o_P\left(n^{-1/2}\right) \\
&= \pm\epsilon_n \left\langle \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \boldsymbol{w}_n^* \right\rangle + \frac{\|\epsilon_n \boldsymbol{w}_n^*\|^2}{2} + \epsilon_n o_P\left(n^{-1/2}\right) \\
&= \pm\epsilon_n \left\langle \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \boldsymbol{w}^* \right\rangle \pm \epsilon_n \left\langle \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \boldsymbol{w}_n^* - \boldsymbol{w}^* \right\rangle + \frac{\|\epsilon_n \boldsymbol{w}_n^*\|^2}{2} + \epsilon_n o_P\left(n^{-1/2}\right).
\end{aligned}
\tag{S3.37}
$$

By Theorem 1 and (S3.27) we have

$$
\begin{aligned}
\left|\left\langle \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \boldsymbol{w}_n^* - \boldsymbol{w}^* \right\rangle\right| &\leq \left\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right\| \|\boldsymbol{w}_n^* - \boldsymbol{w}^*\| \\
&\leq O_P\left(n^{-\frac{p+r}{2(p+r)+2}}\right) o\left(n^{-\frac{1}{2(p+r)+2}}\right) = o_P\left(n^{-1/2}\right).
\end{aligned}
\tag{S3.38}
$$

In addition, since $\boldsymbol{w}_n^*$ is uniformly bounded, we have

$$
\frac{\|\epsilon_n \boldsymbol{w}_n^*\|^2}{2} = \epsilon_n o\left(n^{-1/2}\right). \tag{S3.39}
$$

Then, (S3.37), (S3.38) and (S3.39) imply (S2.16),

$$
\begin{aligned}
P\left(r\left[\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0; \boldsymbol{X}\right] - r\left[\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*, \boldsymbol{\theta}_0; \boldsymbol{X}\right]\right) \\
= \pm\epsilon_n \left\langle \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \boldsymbol{w}^* \right\rangle + \epsilon_n o_P\left(n^{-1/2}\right),
\end{aligned}
$$

Second, we verify (S2.17) given in the proof of Theorem 2.

By (S3.31) and $P\left\{\frac{dl(\boldsymbol{\theta}_0;\boldsymbol{X})}{d\boldsymbol{\theta}}\left[\boldsymbol{w}_n^*\right]\right\} = 0$, it is clear that (S2.17),

$$
\frac{1}{n}\sum_{i=1}^{n} \frac{dl(\boldsymbol{\theta}_0; \boldsymbol{x}_i)}{d\boldsymbol{\theta}}\left[\boldsymbol{w}_n^*\right] = (\mathbb{P}_n - P)\left\{\frac{dl(\boldsymbol{\theta}_0; \boldsymbol{X})}{d\boldsymbol{\theta}}\left[\boldsymbol{w}^*\right]\right\} + o_P\left(n^{-1/2}\right),
$$

holds.

Finally, we verify (S2.18) given in the proof of Theorem 2. Note that

$$\begin{aligned}
&(\mathbb{P}_n - P)\left(r\left[\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0; \boldsymbol{X}\right] - r\left[\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*, \boldsymbol{\theta}_0; \boldsymbol{X}\right]\right) \\
&\quad = (\mathbb{P}_n - P)\left\{l\left(\hat{\boldsymbol{\theta}}_n; X\right) - l\left(\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*; X\right) - \frac{dl(\boldsymbol{\theta}_0; \boldsymbol{X})}{d\boldsymbol{\theta}}\left[\mp \epsilon_n \boldsymbol{w}_n^*\right]\right\} \\
&\quad = \mp \epsilon_n (\mathbb{P}_n - P)\left\{\frac{dl(\tilde{\boldsymbol{\theta}}; \boldsymbol{X})}{d\boldsymbol{\theta}}[\boldsymbol{w}_n^*] - \frac{dl(\boldsymbol{\theta}_0; \boldsymbol{X})}{d\boldsymbol{\theta}}\left[\boldsymbol{w}_n^*\right]\right\}
\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ is between $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*$, then by (S3.32) it immediately follows that (S2.18),

$$(\mathbb{P}_n - P)\left(r\left[\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0; \boldsymbol{X}\right] - r\left[\hat{\boldsymbol{\theta}}_n \pm \epsilon_n \boldsymbol{w}_n^*, \boldsymbol{\theta}_0; \boldsymbol{X}\right]\right) = \epsilon_n o_P\left(n^{-1/2}\right),$$

holds. The proof is complete. □

Bibliography

De Boor, C. (2001). *A Practical Guide to Splines, Revised Ed.* New York: Springer.

Halmos, P. (1982). *A Hilbert Space Problem Book*. New York: Springer.

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

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

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

Wong, W. H. (1992). On asymptotic efficiency in estimation theory. *Statistical Science 2*, 47–68.

Wu, Y. and Y. Zhang (2012). Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data. *Annals of Statistics 40*, 1609–1636.