NONPARAMETRIC ESTIMATION AND TESTING FOR PANEL COUNT DATA WITH INFORMATIVE TERMINAL EVENT

Xiangbin Hu[1], Li Liu[2], Ying Zhang[3], and Xingqiu Zhao[1]

[1] *Hong Kong Polytechnic University,* [2] *Wuhan University and* [3] *University of Nebraska Medical Center*

Supplementary Material

This Supplementary Material contains calculations of the loss function, some preliminary lemmas, proofs of Theorems 1–5, and some additional works on simulation studies and the real data analysis.

S1 Calculation of Loss Function

For the first part of $\ell_n(\Lambda, \hat{F}_n; X)$, replacing $\Lambda(s)$ by $\mathbf{I}^T(s)\boldsymbol{\alpha}$, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\left\{\triangle N_{i,j}-\triangle\mathbf{I}_j(Y_i)^T\boldsymbol{\alpha}\right\}^2\\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\left\{\boldsymbol{\alpha}^T\triangle\mathbf{I}_j(Y_i)\triangle\mathbf{I}_j(Y_i)^T\boldsymbol{\alpha}-2\triangle N_{i,j}\triangle\mathbf{I}_j(Y_i)^T\boldsymbol{\alpha}+\triangle N_{i,j}^2\right\}\\
=&\boldsymbol{\alpha}^T\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\triangle\mathbf{I}_j(Y_i)\triangle\mathbf{I}_j(Y_i)^T\right\}\boldsymbol{\alpha}-\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\triangle N_{i,j}\triangle\mathbf{I}_j(Y_i)\right\}^T\boldsymbol{\alpha}\\
+&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\triangle N_{i,j}^2\\
=&:\boldsymbol{\alpha}^T\mathbf{A}_1\boldsymbol{\alpha}-2\mathbf{B}_1^T\boldsymbol{\alpha}+C_1.
\end{aligned}
$$

For the second part, the KM estimator is a nondecreasing step function

$$
\hat{F}_n(u)=\sum_{l=1}^{L}f_l 1_{[t_l,t_{l+1})}(u),
$$

where $t_1=0$, $t_{L+1}=\tau$, and $\{[t_l,t_{l+1}):l=1,\cdots,L\}$ is a partition of $[0,\tau)$. Furthermore, for each subject, we assume Y_i to be in the interval $[t_{l_i},t_{l_i+1})$,

and take $f_{L+1} = f_L$. Then we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}(1-\Delta_i)\frac{\int_{Y_i}^{\infty}\{\triangle N_{i,j} - \triangle \mathbf{I}_j(u)^T\boldsymbol{\alpha}\}^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y_i)} \\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}(1-\Delta_i)\frac{\sum_{l=l_i}^{L}\left[(f_{l+1}-f_l)\{\triangle N_{i,j} - \triangle \mathbf{I}_j(t_{l+1}^{(1)})^T\boldsymbol{\alpha}\}^2\right]}{1-\hat{F}_n(Y_i)} \\
=&\sum_{i=1}^{n}\frac{1-\Delta_i}{n(1-\hat{F}_n(Y_i))}\sum_{j=1}^{K_i}\sum_{l=l_i}^{L}\Big[\big(f_{l+1}-f_l\big)\Big\{\boldsymbol{\alpha}^T\triangle\mathbf{I}_j(t_{l+1}^{(1)})\triangle\mathbf{I}_j(t_{l+1}^{(1)})^T\boldsymbol{\alpha} \\
&-2\triangle N_{i,j}\triangle\mathbf{I}_j(t_{l+1}^{(1)})^T\boldsymbol{\alpha} + \triangle N_{i,j}^2\Big\}\Big] \\
=&\boldsymbol{\alpha}^T\left[\sum_{i=1}^{n}\frac{1-\Delta_i}{n(1-\hat{F}_n(Y_i))}\sum_{j=1}^{K_i}\sum_{l=l_i}^{L}\Big\{(f_{l+1}-f_l)\triangle\mathbf{I}_j(t_{l+1}^{(1)})\triangle\mathbf{I}_j(t_{l+1}^{(1)})^T\Big\}\right]\boldsymbol{\alpha} \\
&-2\left[\sum_{i=1}^{n}\frac{1-\Delta_i}{n(1-\hat{F}_n(Y_i))}\sum_{j=1}^{K_i}\left\{\triangle N_{i,j}\sum_{l=l_i}^{L}(f_{l+1}-f_l)\triangle\mathbf{I}_j(t_{l+1}^{(1)})^T\right\}\right]^T\boldsymbol{\alpha} \\
&+\sum_{i=1}^{n}\frac{(1-\Delta_i)(f_{L+1}-f_{l_i})}{n(1-\hat{F}_n(Y_i))}\sum_{j=1}^{K_i}\triangle N_{i,j}^2 \\
&=:\boldsymbol{\alpha}^T\mathbf{A}_2\boldsymbol{\alpha} - 2\mathbf{B}_2^T\boldsymbol{\alpha} + C_2.
\end{aligned}
$$

Thus, to obtain the estimator $\hat{\Lambda}_n$, we should minimize $\boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha} - 2\mathbf{B}^T\boldsymbol{\alpha} + C$ under the constraints that $\alpha_l \geq 0$ for $l = 1, \cdots, q_n$, where $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$, $\mathbf{B} = \mathbf{B}_1 + \mathbf{B}_2$ and $C = C_1 + C_2$.

S2 Lemmas

Lemma 1. (i) *Suppose that Condition* (C2) *holds. For sufficiently small* δ*, any* $F \in \mathcal{F}_\delta$ *and any differentiable function* g*, we have*

$$\begin{aligned}&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\left|\frac{\int_Y^\infty g(u-T_j)dF(u)}{1-F(Y)}-\frac{\int_Y^\infty g(u-T_j)dF_0(u)}{1-F_0(Y)}\right|\right]\\ \lesssim&\left(E\left[\sum_{j=1}^{K}|g'(U-T_j)|\right]+E\left[\sum_{j=1}^{K}|g(U-T_j)|\right]\right)||F-F_0||_\infty.\end{aligned} \tag{S2.1}$$

(ii) *In addition, suppose that Conditions* (C1) *and* (C3) *hold. It follows that for all* $\Lambda \in \Phi$*,*

$$|\mathcal{P}m(\Lambda, F; X) - \mathcal{P}m(\Lambda, F_0; X)| \lesssim d_2(F, F_0).$$

Thus, we have $\mathcal{P}m(\Lambda, \hat{F}_n; X) - \mathcal{P}m(\Lambda, F_0; X) = o_p(1)$ *for all* $\Lambda \in \Phi$*.*

Proof. (i) By direct calculations, we have

$$\begin{aligned}&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\left|\frac{\int_Y^\infty g(u-T_j)dF_0(u)}{1-F_0(Y)}-\frac{\int_Y^\infty g(u-T_j)dF(u)}{1-F(Y)}\right|\right]\\ =&\mathcal{P}\left[\frac{1-\Delta}{(1-F_0(Y))(1-F(Y))}\sum_{j=1}^{K}\left|(1-F_0(Y))\int_Y^\infty g(u-T_j)d(F_0(u)-F(u))\right.\right.\\ &\left.\left.+\ (F_0(Y)-F(Y))\int_Y^\infty g(u-T_j)dF_0(u)\right|\right]\\ \leq&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{|\int_Y^\infty g(u-T_j)d(F_0(u)-F(u))|}{1-F(Y)}\right]\\ +&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty |g(u-T_j)|dF_0(u)}{(1-F_0(Y))(1-F(Y))}\right]||F-F_0||_\infty.\end{aligned}$$

Using integration by parts, we have

$$\begin{aligned}
&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{|\int_Y^\infty g(u-T_j)d(F_0(u)-F(u))|}{1-F(Y)}\right]\\
=&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\left|-\big(F_0(Y)-F(Y)\big)g(Y-T_j)-\int_Y^\infty g'(u-T_j)(F_0(u)-F(u))du\right|}{1-F(Y)}\right]\\
\leq&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{|g(Y-T_j)|}{1-F(Y)}\right]||F-F_0||_\infty+\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g'(u-T_j)|du}{1-F(Y)}\right]||F-F_0||_\infty.
\end{aligned}$$

Condition (C2) implies that $1-F_0(Y)$ is larger than a positive constant. Hence, for sufficiently small δ and any $F\in\mathcal{F}_\delta$, $1-F(Y)$ is also larger than a positive constant. It follows that

$$\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{|g(Y-T_j)|}{1-F(Y)}\right]\lesssim\mathcal{P}\left[\Delta\sum_{j=1}^{K}|g(Y-T_j)|\right]\leq\mathcal{P}\left[\sum_{j=1}^{K}|g(U-T_j)|\right]$$

and

$$\begin{aligned}
&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g(u-T_j)|dF_0(u)}{(1-F_0(Y))(1-F(Y))}\right]\\
\lesssim&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g(u-T_j)|dF_0(u)}{1-F_0(Y)}\right]\leq\mathcal{P}\left[\sum_{j=1}^{K}|g(U-T_j)|\right].
\end{aligned}$$

Moreover, by the second part of Condition (C2), we obtain

$$\begin{aligned}
&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g'(u-T_j)|du}{1-F(Y)}\right]\lesssim\mathcal{P}\left[\frac{1-\Delta}{1-F_0(Y)}\sum_{j=1}^{K}\int_Y^\infty\frac{|g'(u-T_j)|}{f_0(u)}dF_0(u)\right]\\
\lesssim&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g'(u-T_j)|dF_0(u)}{1-F_0(Y)}\right]\leq\mathcal{P}\left[\sum_{j=1}^{K}|g'(U-T_j)|\right].
\end{aligned}$$

Thus, (S2.1) holds.

(ii) By the first part of the Lemma, Conditions (C1) and (C3), we have

$$
\begin{aligned}
&|\mathcal{P}[m(\Lambda, F; X) - m(\Lambda, F_0; X)]| \\
\leq &\mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left|\frac{\int_Y^\infty \left(\triangle N_j - \triangle\Lambda_j(u)\right)^2 dF(u)}{1-F(Y)} - \frac{\int_Y^\infty \left(\triangle N_j - \triangle\Lambda_j(u)\right)^2 dF_0(u)}{1-F_0(Y)}\right|\right] \\
\lesssim &\left(E\left[\sum_{j=1}^{K}\left\{\left|\Lambda_j'(U)\left(\triangle N_j - \triangle\Lambda_j(U)\right)\right| + \left(\triangle N_j - \triangle\Lambda_j(U)\right)^2\right\}\right]\right)||F - F_0||_\infty \\
\lesssim &d_2(F, F_0).
\end{aligned}
$$

□

Lemma 2. *Suppose that Conditions* (C2), (C4) *and* (C6) *hold. Then for sufficiently small* δ, $\{m(\Lambda, F; X) : \Lambda \in \Phi,\ F \in \mathcal{F}_\delta,\ \Lambda$ *is uniformly bounded*$\}$ *is Donsker, where*

$$m(\Lambda, F; X) = \sum_{j=1}^{K}\Delta\{\triangle N_j - \triangle\Lambda_j(Y)\}^2 + \sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty \{\triangle N_j - \triangle\Lambda_j(u)\}^2 dF(u)}{1-F(Y)}.$$

Proof. Note that functions in $\mathcal{F}_\delta$ and Φ are monotone and uniformly bounded. It follows from Section 3 of van der Vaart (1996) that $\{\Lambda \in \Phi : \Lambda$ is uniformly bounded$\}$ and $\mathcal{F}_\delta$ are Donsker. Since $\{\triangle N_j - \triangle\Lambda_j(Y)\}^2$ is Lipschitz for Λ, by Theorem 2.10.6 of van der Vaart and Wellner (1996),

$$\left\{[\triangle N_j - \triangle\Lambda_j(Y)]^2 : \Lambda \in \Phi \text{ is uniformly bounded}\right\}$$

is Donsker. Note that

$$\left\{\int_Y^\infty \left[\triangle N_j - \triangle\Lambda_j(u)\right]^2 dF(u) : \Lambda \in \Phi \text{ is uniformly bounded}\right\}$$

is a subset of the convex combinations of functions in

$$\big\{[\triangle N_j - \triangle\Lambda_j(Y)]^2 : \Lambda \in \Phi \text{ is uniformly bounded}\big\}.$$

By Theorem 2.10.1 and Theorem 2.10.3 of van der Vaart and Wellner (1996),

$$\left\{\int_Y^\infty [\triangle N_j - \triangle\Lambda_j(u)]^2 dF(u) : \Lambda \in \Phi \text{ is uniformly bounded}\right\}$$

is Donsker. Since $1-F(Y)$ and K are bounded from Conditions (C2), (C4) and (C6),

$$\{m(\Lambda, F; X) : \Lambda \in \Phi,\ F \in \mathcal{F}_\delta,\ \Lambda \text{ is uniformly bounded}\}$$

is Donsker by Theorem 2.10.6 of van der Vaart and Wellner (1996). □

Lemma 3 (Rate of Convergence of M-estimator with Nuisance Parameter)**.** *Suppose that for every $\Lambda \in \Phi_n$, sufficiently large n and sufficiently small η,*

$$\mathcal{P}(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \lesssim -d_1(\Lambda, \Lambda_0)^2 + d_1(\Lambda, \Lambda_0)d_2(\hat{F}_n, F_0) + d_2(\hat{F}_n, F_0)^2$$

and

$$E \sup_{\{\Lambda\in\Phi_n : d_1(\Lambda,\Lambda_0)<\eta\}} |(\mathbb{P}_n - \mathcal{P})(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X))| \lesssim \frac{\phi_n(\eta)}{\sqrt{n}}$$

hold, where $\phi_n(\eta)$ satisfies that $\eta \mapsto \phi_n(\eta)/\eta^\alpha$ is decreasing for some $\alpha < 2$. Let $r_n > 0$ satisfy $\phi_n(r_n) \lesssim \sqrt{n} r_n^2$. If the sequence $\hat{\Lambda}_n$ satisfies

$$\mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \geq \mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - O_p(r_n^2)$$

and converges in probability to Λ_0, then $d_1(\hat{\Lambda}_n, \Lambda_0) = O_p(r_n + d_2(\hat{F}_n, F_0))$.

Proof. This Lemma is similar to Theorem 5.55 of van der Vaart (1998). In order to verify $d_1(\hat{\Lambda}_n, \Lambda_0) = O_p(r_n + d_2(\hat{F}_n, F_0))$, we need to prove that for sufficiently large n,

$$\lim_{M\to\infty} P(\hat{\Lambda}_n \in \Phi_n : d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^M(d_2(\hat{F}_n, F_0) + r_n)) = 0.$$

Then we divide Φ_n into shells

$$S_{n,j,M} = \{\Lambda \in \Phi_n : 2^j r_n \leq d_1(\Lambda, \Lambda_0) < 2^{j+1} r_n, 2^M d_2(\hat{F}_n, F_0) \leq d_1(\Lambda, \Lambda_0)\}.$$

For any $\Lambda \in S_{n,j,M}$, we have $2d_1(\Lambda, \Lambda_0) \geq 2^M d_2(\hat{F}_n, F_0) + 2^j r_n$. Hence,

$$\left\{\hat{\Lambda}_n \in \Phi_n : 2d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^M(d_2(\hat{F}_n, F_0) + r_n)\right\} \subseteq \bigcup_{j\geq M} \{\hat{\Lambda}_n \in S_{n,j,M}\}.$$

It follows that

$$P\left(\hat{\Lambda}_n \in \Phi_n : d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^{M-1}(d_2(\hat{F}_n, F_0) + r_n)\right) \leq P\left(\hat{\Lambda}_n \in \bigcup_{j\geq M} S_{n,j,M}\right). \tag{S2.2}$$

Furthermore, since $\hat{\Lambda}_n$ satisfies that $\mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \geq \mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - O_p(r_n^2)$, for $\hat{\Lambda}_n \in S_{n,j,M}$ we can find a variable $R_n = O_p(r_n^2)$ such that

$$\sup_{\Lambda\in S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\hat{\Lambda}_n, \hat{F}_n; X)) \geq -R_n.$$

Then for any constant κ, we have

$$\begin{aligned} &P\left(\hat{\Lambda}_n \in \bigcup_{j\geq M} S_{n,j,M}\right) \\ \leq &P\left(\sup_{\Lambda\in\bigcup_{j\geq M} S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq -\kappa r_n^2\right) + P(R_n \geq \kappa r_n^2). \end{aligned} \tag{S2.3}$$

By (S2.2) and (S2.3), we obtain

$$
\begin{aligned}
&P\left(\hat{\Lambda}_n \in \Phi_n : d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^M (d_2(\hat{F}_n, F_0) + r_n)\right) \\
\leq & P(d_1(\hat{\Lambda}_n, \Lambda_0) \geq \eta) + P(R_n \geq \kappa r_n^2) \\
+ & \sum_{j \geq M+1, 2^{j+1} \leq \eta / r_n} P\left(\sup_{\Lambda \in S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq -\kappa r_n^2\right).
\end{aligned}
$$

Since $\hat{\Lambda}_n$ is consistent and $R_n = O_p(r_n^2)$, $P(d_1(\hat{\Lambda}_n, \Lambda_0) \geq \eta)$ and $P(R_n \geq \kappa r_n^2)$ can be arbitrarily small for sufficiently large n by the choice of η and κ. Thus, we need to prove the limitation of the summation on the right hand side is 0 as M goes to infinity.

Note that for any positive integer M, $1/4 \leq 1 - 2^{-M} - 2^{-2M} < 1$. Then for all $\Lambda \in S_{n,j,M}$,

$$
\begin{aligned}
&\mathcal{P}(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \\
\lesssim & - d_1(\Lambda, \Lambda_0)^2 + d_1(\Lambda, \Lambda_0) d_2(\hat{F}_n, F_0) + d_2(\hat{F}_n, F_0)^2 \\
\leq & - (1 - 2^{-M} - 2^{-2M}) d_1(\Lambda, \Lambda_0)^2 \leq -2^{2j} r_n^2.
\end{aligned}
$$

Hence, $\mathcal{P}(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \leq -c_1 2^{2j} r_n^2$ for some constant c_1. Taking M with $M \geq \frac{1}{2} \log_2(2\kappa / c_1)$, then by the Markov's inequality, for

$j \geq M+1$ and sufficiently large n with $r_n \leq 2^{-(j+1)}\eta$, we have

$$
\begin{aligned}
&P\left(\sup_{\Lambda \in S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq -\kappa r_n^2\right) \\
\leq & P\left(\sup_{\Lambda \in S_{n,j,M}} (\mathbb{P}_n - \mathcal{P})(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq \frac{c_1}{2} 2^{2j} r_n^2\right) \\
\leq & \frac{2}{c_1 2^{2j} r_n^2} E \sup_{\Lambda \in S_{n,j,M}} \left|(\mathbb{P}_n - \mathcal{P})(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X))\right| \lesssim \frac{\phi_n(2^{(j+1)} r_n)}{2^{2j} r_n^2 \sqrt{n}}.
\end{aligned}
$$

Since $\phi_n(\eta)/\eta^\alpha$ is decreasing for some $\alpha < 2$, we have $\phi_n(c\eta) \leq c^\alpha \phi_n(\eta)$ for any $c > 1$. Then $\phi_n(r_n) \lesssim \sqrt{n} r_n^2$ ensures that

$$
\frac{\phi_n(2^{(j+1)} r_n)}{2^{2j} r_n^2 \sqrt{n}} \lesssim \frac{2^{\alpha(j+1)} \sqrt{n} r_n^2}{2^{2j} r_n^2 \sqrt{n}} = \frac{1}{2^{(2-\alpha)j-\alpha}}.
$$

Thus,

$$
\sum_{j \geq M+1, 2^{j+1} \leq \eta/r_n} P\left(\sup_{\Lambda \in S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq 0\right) \lesssim \sum_{j \geq M+1} \frac{1}{2^{(2-\alpha)j-\alpha}}.
$$

Noting that $\sum_{j \geq M+1} \frac{1}{2^{(2-\alpha)j-\alpha}}$ tends to 0 as M approaches to infinity, this lemma is concluded. □

Lemma 4. *Suppose that Conditions* (C1), (C2), (C4)–(C6), (C8), (C9) *and* (C11) *hold. Define the class*

$$
\mathcal{M}_\eta(F) = \{m(\Lambda, F; X) - m(\Lambda_0, F; X) : \Lambda \in \Phi_n, d_1(\Lambda, \Lambda_0) \leq \eta\}
$$

for each $F \in \mathcal{F}_\delta$. For any $\varepsilon < \eta$ and sufficiently small δ, we have

$$
\log N_{[]}(\varepsilon, \mathcal{M}_\eta(F), ||\cdot||_{P,B}) \lesssim q_n \log(\eta/\varepsilon),
$$

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

Proof. By the calculation in Shen and Wong (1994, Page 597), under Condition (C5), for any $\varepsilon < \eta$, there is a set of brackets

$$\left\{[\Lambda_i^L, \Lambda_i^U] : ||\triangle\Lambda_i^U(s_1, s_2) - \triangle\Lambda_i^L(s_1, s_2)||_{L_2(\mu_1)} \leq \varepsilon, i = 1, \cdots, (\eta/\varepsilon)^{c_0 q_n}\right\}$$

such that for all $\Lambda \in \{\Lambda \in \Phi_n : d_1(\Lambda, \Lambda_0) \leq \eta\}$, we can find an interval $[\triangle\Lambda_i^L, \triangle\Lambda_i^U]$ satisfying $\triangle\Lambda(s_1, s_2) \in [\triangle\Lambda_i^L(s_1, s_2), \triangle\Lambda_i^U(s_1, s_2)]$. This implies that

$$||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_i^L(s_1, s_2)||_{L_2(\mu_1)} \leq \varepsilon,$$

$$||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_i^U(s_1, s_2)||_{L_2(\mu_1)} \leq \varepsilon.$$

Noting that $||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_{L_2(\mu_1)} \leq \eta$, we have

$$||\triangle\Lambda_i^L(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_{L_2(\mu_1)} \leq (\varepsilon^2 + \eta^2)^{1/2},$$

$$||\triangle\Lambda_i^U(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_{L_2(\mu_1)} \leq (\varepsilon^2 + \eta^2)^{1/2}.$$

By Lemma 7.1 of Wellner and Zhang (2007), under Conditions (C8) and (C9), for any $\triangle\Lambda(s_1, s_2)$ satisfying $||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_{L_2(\mu_1)} \leq \varepsilon^*$, we have $||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_\infty \leq (\varepsilon^*/c_1)^{2/3}$ for some constant c_1. Hence,

$$0 \vee \left(\triangle\Lambda_0(s_1, s_2) - ((\varepsilon^2 + \eta^2)^{1/2}/c_1)^{2/3}\right) \leq \triangle\Lambda_i^L(s_1, s_2)$$

$$\leq \triangle\Lambda_i^U(s_1, s_2) \leq \triangle\Lambda_0(s_1, s_2) + ((\varepsilon^2 + \eta^2)^{1/2}/c_1)^{2/3},$$

which implies that $\triangle\Lambda_i^U(s_1, s_2) - \triangle\Lambda_i^L(s_1, s_2)$ are uniformly bounded by $2((\varepsilon^2 + \eta^2)^{1/2}/c_1)^{2/3}$.

We turn to consider the ε-bracket of $m(\Lambda, F; X) - m(\Lambda_0, F; X)$. Note that

$$
\begin{aligned}
& m(\Lambda, F; X) - m(\Lambda_0, F; X) \\
= & \sum_{j=1}^{K} \Delta \Big[(\triangle N_j - \triangle \Lambda_j(Y))^2 - (\triangle N_j - \triangle \Lambda_{0,j}(Y))^2 \Big] \\
+ & \sum_{j=1}^{K} (1-\Delta) \left[\frac{\int_Y^\infty (\triangle N_j - \triangle \Lambda_j(u))^2 - (\triangle N_j - \triangle \Lambda_{0,j}(u))^2 \, dF(u)}{1 - F(Y)} \right] \\
= & \sum_{j=1}^{K} \Delta \Big[\{\triangle \Lambda_j(Y)\}^2 - 2\triangle N_j \triangle \Lambda_j(Y) - \{\triangle \Lambda_{0,j}(Y)\}^2 + 2\triangle N_j \triangle \Lambda_{0,j}(Y) \Big] \\
+ & \sum_{j=1}^{K} \frac{1-\Delta}{1-F(Y)} \int_Y^\infty \Big[\{\triangle \Lambda_j(u)\}^2 - 2\triangle N_j \triangle \Lambda_j(u) - \{\triangle \Lambda_{0,j}(u)\}^2 + 2\triangle N_j \triangle \Lambda_{0,j}(u) \Big] dF(u).
\end{aligned}
$$

By Conditions (C2) and (C4), $1 - F(Y)$ is positive and has uniform upper and lower bounds. Hence,

$$
m_i^L(\Lambda_i^L, \Lambda_i^U, F; X) \le m(\Lambda, F; X) - m(\Lambda_0, F; X) \le m_i^U(\Lambda_i^L, \Lambda_i^U, F; X),
$$

where

$$
\begin{aligned}
& m_i^L(\Lambda_i^L, \Lambda_i^U, F; X) \\
= & \sum_{j=1}^{K} \Delta \Big[\{\triangle \Lambda_{i,j}^L(Y)\}^2 - 2\triangle N_j \triangle \Lambda_{i,j}^U(Y) - \{\triangle \Lambda_{0,j}(Y)\}^2 + 2\triangle N_j \triangle \Lambda_{0,j}(Y) \Big] \\
+ & \sum_{j=1}^{K} \frac{1-\Delta}{1-F(Y)} \int_Y^\infty \Big[\{\triangle \Lambda_{i,j}^L(u)\}^2 - 2\triangle N_j \triangle \Lambda_{i,j}^U(u) - \{\triangle \Lambda_{0,j}(u)\}^2 + 2\triangle N_j \triangle \Lambda_{0,j}(u) \Big] dF(u)
\end{aligned}
$$

and

$$
\begin{aligned}
&m_i^U(\Lambda_i^L, \Lambda_i^U, F; X) \\
=&\sum_{j=1}^{K} \Delta\Big[\{\triangle\Lambda_{i,j}^U(Y)\}^2 - 2\triangle N_j \triangle\Lambda_{i,j}^L(Y) - \{\triangle\Lambda_{0,j}(Y)\}^2 + 2\triangle N_j \triangle\Lambda_{0,j}(Y)\Big] \\
+&\sum_{j=1}^{K} \frac{1-\Delta}{1-F(Y)} \int_Y^\infty \Big[\{\triangle\Lambda_{i,j}^U(u)\}^2 - 2\triangle N_j \triangle\Lambda_{i,j}^L(u) - \{\triangle\Lambda_{0,j}(u)\}^2 + 2\triangle N_j \triangle\Lambda_{0,j}(u)\Big] dF(u)
\end{aligned}
$$

for $i = 1, \cdots, (\eta/\varepsilon)^{c_0 q_n}$.

We also need to verify $||m_i^L(\Lambda_i^L, \Lambda_i^U, F; X) - m_i^U(\Lambda_i^L, \Lambda_i^U, F; X)||_{P,B}^2 \lesssim \varepsilon^2$. According to

$$
||f||_{P,B}^2 = 2P(e^{|f|} - 1 - |f|) = 2P\left(\sum_{n=2}^{\infty} \frac{|f|^n}{n!}\right) \leq 2P\left(|f|^2 e^{|f|}\right),
$$

$||f||_{P,B}^2 \lesssim \varepsilon^2$ is followed from $P(|f|^2 e^{|f|}) \lesssim \varepsilon^2$. Note that

$$
\begin{aligned}
&\{\triangle\Lambda_{i,j}^U(u)\}^2 - \{\triangle\Lambda_{i,j}^L(u)\}^2 + 2\triangle N_j(\triangle\Lambda_{i,j}^U(u) - \triangle\Lambda_{i,j}^L(u)) \\
=&\left(\triangle\Lambda_{i,j}^U(u) + \triangle\Lambda_{i,j}^L(u) + 2\triangle N_j\right)\left(\triangle\Lambda_{i,j}^U(u) - \triangle\Lambda_{i,j}^L(u)\right) \\
\lesssim&\triangle N_j\left(\triangle\Lambda_{i,j}^U(u) - \triangle\Lambda_{i,j}^L(u)\right) + \left(\triangle\Lambda_{i,j}^U(u) - \triangle\Lambda_{i,j}^L(u)\right).
\end{aligned}
$$

Since Λ_i^L, Λ_i^U and $1 - F(Y)$ are uniformly bounded, we have

$$
e^{|m_i^U(\Lambda_i^L, \Lambda_i^U, F; X) - m_i^L(\Lambda_i^L, \Lambda_i^U, F; X)|} \lesssim e^{cN(T_K)}
$$

with some constant c. By Cauchy-Schwarz inequality and Condition (C11),

we obtain

$$
\begin{aligned}
&\mathcal{P}\left(|m_i^U(\Lambda_i^L,\Lambda_i^U,F;X)-m_i^L(\Lambda_i^L,\Lambda_i^U,F;X)|^2 e^{|m_i^U(\Lambda_i^L,\Lambda_i^U,F;X)-m_i^L(\Lambda_i^L,\Lambda_i^U,F;X)|}\right)\\
\lesssim&\mathcal{P}\left(e^{cN(T_K)}\left|m_i^U(\Lambda_i^L,\Lambda_i^U,F;X)-m_i^L(\Lambda_i^L,\Lambda_i^U,F;X)\right|^2\right)\\
\lesssim&\mathcal{P}\left[\left\{\Delta\sum_{j=1}^K(\triangle N_j+1)\left(\triangle\Lambda_{i,j}^U(Y)-\triangle\Lambda_{i,j}^L(Y)\right)\right.\right.\\
&\left.\left.+\ (1-\Delta)\sum_{j=1}^K\frac{\int_Y^\infty(\triangle N_j+1)\left(\triangle\Lambda_{i,j}^U(u)-\triangle\Lambda_{i,j}^L(u)\right)dF(u)}{1-F(Y)}\right\}^2\right]\\
\lesssim&\mathcal{P}\left[\sum_{j=1}^K\left\{\Delta\left(\triangle\Lambda_{i,j}^U(Y)-\Lambda_{i,j}^L(Y)\right)^2+(1-\Delta)\frac{\int_Y^\infty\left(\triangle\Lambda_{i,j}^U(u)-\triangle\Lambda_{i,j}^L(u)\right)^2dF(u)}{1-F(Y)}\right\}\right]\\
\lesssim&\delta+||\triangle\Lambda_i^U(s_1,s_2)-\triangle\Lambda_i^L(s_1,s_2)||_{L_2(\mu_1)}^2\lesssim\varepsilon^2.
\end{aligned}
\tag{S2.4}
$$

This implies that $||m_i^L-m_i^U||_{P,B}^2\lesssim\varepsilon^2$. That means $N_{[]}(\varepsilon,\mathcal{M}_\eta(F),||\cdot||_{P,B})\leq(\eta/\varepsilon)^{c_0q_n}$ and $\log N_{[]}(\varepsilon,\mathcal{M}_\eta(F),||\cdot||_{P,B})\lesssim q_n\log(\eta/\varepsilon)$. □

S3 Proof of Theorem 1

Proof. First, for any $\Lambda\in\Phi$, we have

$$
\begin{aligned}
&\mathcal{P}m(\Lambda,F_0;X)-\mathcal{P}m(\Lambda_0,F_0;X)\\
=&\mathcal{P}\left[\sum_{j=1}^K\Delta\{(\triangle N_j-\triangle\Lambda_j(Y))^2-(\triangle N_j-\triangle\Lambda_{0,j}(Y))^2\}\right.\\
&\left.+\sum_{j=1}^K(1-\Delta)\frac{\int_Y^\infty\{(\triangle N_j-\triangle\Lambda_j(u))^2-(\triangle N_j-\triangle\Lambda_{0,j}(u))^2\}dF_0(u)}{1-F_0(Y)}\right]
\end{aligned}
$$

$$
\begin{aligned}
=&\mathcal{P}\left[\sum_{j=1}^{K}\Delta\left(2\triangle N_j-\triangle\Lambda_j(Y)-\triangle\Lambda_{0,j}(Y)\right)\left(\triangle\Lambda_{0,j}(Y)-\triangle\Lambda_j(Y)\right)\right.\\
&\left.+\sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty\left(2\triangle N_j-\triangle\Lambda_j(u)-\triangle\Lambda_{0,j}(u)\right)\left(\triangle\Lambda_{0,j}(u)-\triangle\Lambda_j(u)\right)dF_0(u)}{1-F_0(Y)}\right]\\
=&\mathcal{P}\left[\sum_{j=1}^{K}\Delta\big\{\triangle\Lambda_{0,j}(Y)-\triangle\Lambda_j(Y)\big\}^2+\sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty\big\{\triangle\Lambda_{0,j}(u)-\triangle\Lambda_j(u)\big\}^2dF_0(u)}{1-F_0(Y)}\right]\\
=&d_1(\Lambda,\Lambda_0)^2.
\end{aligned}
$$

Thus, to draw the conclusion, we only need to consider $\mathcal{P}m(\Lambda,F_0;X)-\mathcal{P}m(\Lambda_0,F_0;X)$.

Since $\hat{\Lambda}_n$ is the minimizer of $\mathbb{P}_n m(\Lambda,\hat{F}_n;X)$ with respect to $\Lambda\in\Phi_n$, for any direction function $h\in\Phi_n$, we obtain

$$
\begin{aligned}
0&=\lim_{\varepsilon\to 0}\frac{\mathbb{P}_nL(\hat{\Lambda}_n+\varepsilon h,\hat{F}_n;X)-\mathbb{P}_nL(\hat{\Lambda}_n,\hat{F}_n;X)}{\varepsilon}\\
&=-2\mathbb{P}_n\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle N_j-\triangle\hat{\Lambda}_{n,j}(Y)\big)\triangle h_j(Y)\right.\right.\\
&\left.\left.+\ (1-\Delta)\frac{\int_Y^\infty\big(\triangle N_j-\triangle\hat{\Lambda}_{n,j}(u)\big)\triangle h_j(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right].
\end{aligned}
\tag{S3.5}
$$

Taking $h(s)=s$, by Conditions (C1) and (C6), it follows that

$$
\begin{aligned}
&\mathbb{P}_n\left[\sum_{j=1}^{K}\left\{(T_j-T_{j-1})\left(\Delta\triangle\hat{\Lambda}_{n,j}(Y)+(1-\Delta)\frac{\int_Y^\infty\triangle\hat{\Lambda}_{n,j}(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right)\right\}\right]\\
=&\mathbb{P}_n\left[\sum_{j=1}^{K}\left\{\triangle N_j(T_j-T_{j-1})\right\}\right]\xrightarrow{a.s.}E\left[\sum_{j=1}^{K}\triangle\Lambda_{0,j}(U)(T_j-T_{j-1})\right]\le M_3\Lambda_0(\tau)\tau,
\end{aligned}
$$

where the almost surely convergence follows from the strong law of large

number. Moreover, for the left hand side, by Condition (C7), we have

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}_n \left[\sum_{j=1}^{K} \left\{ (T_j - T_{j-1}) \left(\Delta \triangle \hat{\Lambda}_{n,j}(Y) + (1-\Delta) \frac{\int_Y^\infty \triangle \hat{\Lambda}_{n,j}(u) d\hat{F}_n(u)}{1-\hat{F}_n(Y)} \right) \right\} \right] \\
\gtrsim &\limsup_{n\to\infty} \mathbb{P}_n \left[\Delta \{ \hat{\Lambda}_n(Y) - \hat{\Lambda}_n(Y - T_K) \} + (1-\Delta) \frac{\int_Y^\infty \{ \hat{\Lambda}_n(u) - \hat{\Lambda}_n(u - T_K) \} d\hat{F}_n(u)}{1-\hat{F}_n(Y)} \right] \\
\geq &\limsup_{n\to\infty} \triangle \hat{\Lambda}_n(b_1, b_2) \mathbb{P}_n \left[\Delta 1_{\{Y - T_K \in [0,b_1], Y \in [b_2,\tau]\}} + (1-\Delta) \frac{\int_Y^\infty 1_{\{u - T_K \in [0,b_1], u \in [b_2,\tau]\}} d\hat{F}_n(u)}{1-\hat{F}_n(Y)} \right] \\
= &\limsup_{n\to\infty} \triangle \hat{\Lambda}_n(b_1, b_2) E \left[1_{\{U - T_K \in [0,b_1], U \in [b_2,\tau]\}} \right] = \limsup_{n\to\infty} \triangle \hat{\Lambda}_n(b_1, b_2) \mu_2([0, b_1] \times [b_2, \tau]).
\end{aligned}$$

Hence, for every $0 \leq b_1 \leq b_2 \leq \tau$ satisfying $\mu_2([0, b_1] \times [b_2, \tau]) > 0$, we have

$$\triangle \hat{\Lambda}_n(s_1, s_2) 1_{\{(s_1,s_2) \in [b_1,b_2] \times [b_1,b_2]\}}$$

is uniformly bounded. In particular, if $\mu_2(\{0\} \times \{\tau\}) > 0$, then $\hat{\Lambda}_n(s)$ is uniformly bounded.

By Lemma A1 of Lu, Zhang, and Huang (2007), under Condition (C5), there is $\Lambda_n^* \in \Phi_n$ such that $||\Lambda_n^* - \Lambda_0||_\infty = O(n^{-\nu r})$. This implies that

$$\begin{aligned}
&\mathcal{P}m(\Lambda_n^*, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X) \\
= &\mathcal{P} \left[\sum_{j=1}^{K} \Delta \big(\triangle \Lambda_{0,j}(Y) - \triangle \Lambda_{n,j}^*(Y) \big)^2 + \sum_{j=1}^{K} (1-\Delta) \frac{\int_Y^\infty \big(\triangle \Lambda_{0,j}(u) - \triangle \Lambda_{n,j}^*(u) \big)^2 dF_0(u)}{1-F_0(Y)} \right] \\
\lesssim &||\Lambda_n^* - \Lambda_0||_\infty^2 = O(n^{-2\nu r}) = o(1).
\end{aligned}$$

Note that

$$
\begin{aligned}
&0 \le \mathcal{P}m(\hat{\Lambda}_n, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X) \\
=&\mathcal{P}m(\hat{\Lambda}_n, F_0; X) - \mathcal{P}m(\hat{\Lambda}_n, \hat{F}_n; X) + \mathcal{P}m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) \\
+&\mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\Lambda_n^*, \hat{F}_n; X) + \mathbb{P}_n m(\Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\Lambda_n^*, \hat{F}_n; X) \\
+&\mathcal{P}m(\Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\Lambda_n^*, F_0; X) + \mathcal{P}m(\Lambda_n^*, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X).
\end{aligned}
$$

By Lemma 1, $\mathcal{P}m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathcal{P}m(\hat{\Lambda}_n, F_0; X) = o_p(1)$ and $\mathcal{P}m(\Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\Lambda_n^*, F_0; X) = o_p(1)$. By Lemma 2, the class of functions

$$\{m(\Lambda, F; X) : \Lambda \in \Phi_n, F \in \mathcal{F}_\delta\}$$

is Donsker. Hence it is Glivenko-Cantelli, and we have

$$(\mathbb{P}_n - \mathcal{P})m(\Lambda_n^*, \hat{F}_n; X) = o_p(1) \text{ and } (\mathbb{P}_n - \mathcal{P})m(\hat{\Lambda}_n, \hat{F}_n; X) = o_p(1)$$

since $\hat{\Lambda}_n$ is uniformly bounded. According to the definition of $\hat{\Lambda}_n$,

$$\mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\Lambda_n^*, \hat{F}_n; X) \le 0.$$

Hence, $d_1(\hat{\Lambda}_n, \Lambda_0) = \mathcal{P}m(\hat{\Lambda}_n, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X) = o_p(1)$. □

S4 Proof of Theorem 2

Proof. To apply Lemma 3, we need to verify that for every $\Lambda \in \Phi_n$ and sufficiently large n, the inequalities

$$\mathcal{P}\big(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)\big) \lesssim -d_1(\Lambda, \Lambda_0)^2 + d_1(\Lambda, \Lambda_0) d_2(\hat{F}_n, F_0) + d_2(\hat{F}_n, F_0)^2$$

and

$$E \sup_{\{\Lambda\in\Phi_n: d_1(\Lambda,\Lambda_0)<\eta\}} \left| \big(\mathbb{P}_n - \mathcal{P}\big)\big(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)\big) \right| \lesssim \frac{\phi_n(\eta)}{\sqrt{n}}$$

hold. By some calculations,

$$\begin{aligned}
&\mathcal{P}\big(m(\Lambda_0, F; X) - m(\Lambda, F; X)\big) \\
=&\mathcal{P}\left[\sum_{j=1}^{K} \Delta \left\{ \big(\triangle N_j - \triangle\Lambda_{0,j}(Y)\big)^2 - \big(\triangle N_j - \triangle\Lambda_j(Y)\big)^2 \right\}\right. \\
&\left. + \sum_{j=1}^{K} (1-\Delta) \left\{ \frac{\int_Y^\infty \big(\triangle N_j - \triangle\Lambda_{0,j}(u)\big)^2 - \big(\triangle N_j - \triangle\Lambda_j(u)\big)^2 dF(u)}{1-F(Y)} \right\}\right] \\
=&\mathcal{P}\left[\sum_{j=1}^{K} \Delta \big\{2\triangle N_j - \triangle\Lambda_{0,j}(Y) - \triangle\Lambda_j(Y)\big\}\big\{\triangle\Lambda_j(Y) - \triangle\Lambda_{0,j}(Y)\big\}\right. \\
&\left. + \sum_{j=1}^{K} (1-\Delta) \frac{\int_Y^\infty \big\{2\triangle N_j - \triangle\Lambda_{0,j}(u) - \triangle\Lambda_j(u)\big\}\big\{\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u)\big\} dF(u)}{1-F(Y)}\right] \\
=&-\mathcal{P}\left[\sum_{j=1}^{K} \left\{ \Delta(\triangle\Lambda_j(Y) - \triangle\Lambda_{0,j}(Y))^2 + (1-\Delta) \frac{\int_Y^\infty (\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u))^2 dF(u)}{1-F(Y)} \right\}\right] \\
=&\mathcal{P}\left[\sum_{j=1}^{K} (1-\Delta) \left\{ \frac{\int_Y^\infty (\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u))^2 dF_0(u)}{1-F_0(Y)} - \frac{\int_Y^\infty (\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u))^2 dF(u)}{1-F(Y)} \right\}\right] \\
&-d_1(\Lambda, \Lambda_0)^2.
\end{aligned}$$

Moreover, by the first part of Lemma 1, under Condition (C1), we have

$$\mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left\{\frac{\int_Y^\infty(\triangle\Lambda_j(u)-\triangle\Lambda_{0,j}(u))^2dF_0(u)}{1-F_0(Y)}-\frac{\int_Y^\infty(\triangle\Lambda_j(u)-\triangle\Lambda_{0,j}(u))^2dF(u)}{1-F(Y)}\right\}\right]$$

$$\lesssim\mathcal{P}\left[\sum_{j=1}^{K}2\left|\{\triangle\Lambda_j'(U)-\triangle\Lambda_{0,j}'(U)\}\{\triangle\Lambda_j(U)-\triangle\Lambda_{0,j}(U)\}\right|\right]||F-F_0||_\infty$$

$$+\mathcal{P}\left[\sum_{j=1}^{K}\{\triangle\Lambda_j(U)-\triangle\Lambda_{0,j}(U)\}^2\right]||F-F_0||_\infty$$

$$\lesssim d_1(\Lambda,\Lambda_0)d_2(F,F_0)+d_1(\Lambda,\Lambda_0)^2d_2(F,F_0).$$

This implies that

$$\mathcal{P}m(\Lambda_0,\hat{F}_n;X)-\mathcal{P}m(\Lambda,\hat{F}_n;X)\lesssim -d_1(\Lambda,\Lambda_0)^2+d_1(\Lambda,\Lambda_0)d_2(\hat{F}_n,F_0)+d_1(\Lambda,\Lambda_0)^2d_2(\hat{F}_n,F_0).$$

Second, we need to find a $\phi_n(\eta)$ such that

$$E\sup_{\{\Lambda\in\Phi_n:d_1(\Lambda,\Lambda_0)<\eta\}}|(\mathbb{P}_n-\mathcal{P})(m(\Lambda,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X))|\lesssim\frac{\phi_n(\eta)}{\sqrt{n}}.$$

By Lemma 4, we have $\log N_{[]}(\varepsilon,\mathcal{M}_\eta(\hat{F}_n),||\cdot||_{P,B})\lesssim q_n\log(\eta/\varepsilon)$, where

$$\mathcal{M}_\eta(\hat{F}_n)=\{m(\Lambda,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X):\Lambda\in\Phi_n,d_1(\Lambda,\Lambda_0)\le\eta\}.$$

For all Λ satisfying that $\Lambda\in\Phi_n,d_1(\Lambda,\Lambda_0)\le\eta$, note that

$$|m(\Lambda,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X)|$$

$$=\sum_{j=1}^{K}\Delta\left|\big(\triangle\Lambda_j(Y)-\triangle\Lambda_{0,j}(Y)\big)\big(\triangle\Lambda_j(Y)+\triangle\Lambda_{0,j}(Y)-2\triangle N_j\big)\right|$$

$$+\sum_{j=1}^{K}(1-\Delta)\left|\frac{\int_Y^\infty\big(\triangle\Lambda_j(u)-\triangle\Lambda_{0,j}(u)\big)\big(\triangle\Lambda_j(u)+\triangle\Lambda_{0,j}(u)-2\triangle N_j\big)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right|$$

$$\lesssim\sum_{j=1}^{K}\left[(\triangle N_j+1)\left\{\Delta|\triangle\Lambda_j(Y)-\triangle\Lambda_{0,j}(Y)|+(1-\Delta)\frac{\int_Y^\infty|\triangle\Lambda_j(u)-\triangle\Lambda_{0,j}(u)|d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right].$$

Similar to the proof of (S2.4), since $\hat{\Lambda}_n$ is uniformly bounded, it follows that $e^{|m(\hat{\Lambda}_n,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X)|} \lesssim e^{cN(T_K)}$ and

$$
\begin{aligned}
&\mathcal{P}\left[e^{|m(\hat{\Lambda}_n,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X)|}|m(\hat{\Lambda}_n,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X)|^2\right]\\
\lesssim&\mathcal{P}\left[e^{cN(T_K)}\sum_{j=1}^{K}\left\{\Delta(\triangle N_j+1)^2(\triangle\Lambda_j(Y)-\triangle\Lambda_{0,j}(Y))^2\right.\right.\\
&\left.\left.+\ (1-\Delta)(\triangle N_j+1)^2\frac{\int_Y^\infty(\triangle\Lambda_j(u)-\triangle\Lambda_{0,j}(u))^2d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right]\\
\lesssim&\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta(\triangle\Lambda_j(Y)-\triangle\Lambda_{0,j}(Y))^2+(1-\Delta)\frac{\int_Y^\infty(\triangle\Lambda_j(u)-\triangle\Lambda_{0,j}(u))^2d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right]\\
\lesssim&d_1^2(\hat{\Lambda}_n,\Lambda_0)+d_1(\hat{\Lambda}_n,\Lambda_0)d_2(\hat{F}_n,F_0).
\end{aligned}
$$

That means that for sufficiently large n with $d_2(\hat{F}_n,F_0)\lesssim\eta$, we have

$$||m(\hat{\Lambda}_n,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X)||_{P,B}^2\lesssim\eta^2.$$

Then Lemma 3.4.3 of van der Vaart and Wellner (1996) yields that

$$E||n^{1/2}(\mathbb{P}_n-\mathcal{P})||_{\mathcal{M}_\eta(\hat{F}_n)}\lesssim J_{[]}(\eta,\mathcal{M}_\eta(\hat{F}_n),||\cdot||_{P,B})\left\{1+\frac{J_{[]}(\eta,\mathcal{M}_\eta(\hat{F}_n),||\cdot||_{P,B})}{\eta^2n^{1/2}}\right\},$$

where $J_{[]}(\eta,\mathcal{M}_\eta(\hat{F}_n),||\cdot||_{P,B}):=\int_0^\eta\{1+\log N_{[]}(\varepsilon,\mathcal{M}_\eta(\hat{F}_n),||\cdot||_{P,B})\}^{1/2}d\varepsilon\lesssim q_n^{1/2}\eta$. It follows that

$$E\sup_{\{\Lambda\in\Phi_n:d_1(\Lambda,\Lambda_0)<\eta\}}\sqrt{n}|(\mathbb{P}_n-\mathcal{P})(m(\Lambda,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X))|\lesssim q_n^{1/2}\eta+q_nn^{-1/2}.$$

Set $\phi_n(\eta)=q_n^{1/2}\eta+q_nn^{-1/2}$. It is clear that $\phi_n(\eta)/\eta$ is decreasing about η. Moreover, $r_n^2\phi(1/r_n)=q_n^{1/2}r_n+n^{-1/2}q_nr_n^2$, where $r_n=O(n^a)$. Note that

$q_n = O(n^{\nu})$ with $0 < \nu < 1/2$. It follows that

$$r_n^2 \phi(\frac{1}{r_n}) = O(n^{a+\frac{\nu}{2}} + n^{2a+\nu-\frac{1}{2}}).$$

Thus, $a \leq (1-\nu)/2$ ensures $r_n^2 \phi(1/r_n) \lesssim n^{1/2}$. This implies that $r_n = O(n^{(1-\nu)/2})$.

According to the proof of Theorem 1 and the definition of $\hat{\Lambda}_n$, we have

$$\begin{aligned}
&\mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \\
=&\mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\Lambda_n^*, \hat{F}_n; X) + \mathbb{P}_n m(\Lambda_n^*, \hat{F}_n; X) - \mathcal{P} m(\Lambda_n^*, \hat{F}_n; X) \\
&+\mathcal{P} m(\Lambda_n^*, \hat{F}_n; X) - \mathcal{P} m(\Lambda_0, \hat{F}_n; X) + \mathcal{P} m(\Lambda_0, \hat{F}_n; X) - \mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \\
\leq& n^{-\nu r+\varepsilon}(\mathbb{P}_n - \mathcal{P}) \left(\frac{m(\Lambda_n^*, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X)}{n^{-\nu r+\varepsilon}} \right) + O_p(n^{-2\nu r})
\end{aligned}$$

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

$$\widetilde{\mathcal{M}}_n = \left\{ \frac{m(\Lambda, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X)}{n^{-\nu r+\varepsilon}} : \Lambda \in \Phi_n, ||\Lambda - \Lambda_0||_\infty = O(n^{-\nu r}) \right\}.$$

Similar to the proof of Theorem 2 in Lu, Zhang, and Huang (2009), under Conditions (C2) and (C5), we have $\widetilde{\mathcal{M}}_n$ is Donsker, and $\mathcal{P}\tilde{m}^2 \to 0$ as $n \to \infty$ for any $\tilde{m} \in \widetilde{\mathcal{M}}_n$. Hence,

$$n^{-\nu r+\varepsilon}(\mathbb{P}_n - \mathcal{P}) \left(\frac{m(\Lambda_n^*, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X)}{n^{-\nu r+\varepsilon}} \right) = o_p(n^{-\nu r+\varepsilon-1/2}) = o_p(n^{-2\nu r}).$$

This implies that $\mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \leq O_p(n^{-2\nu r})$. Noting that $\hat{\Lambda}_n$ is needed to satisfy $\mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \geq \mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - O_p(r_n^{-2})$, so we should take ν such that $O_p(n^{-2\nu r}) \leq O_p(r_n^{-2})$. Since $r_n = O(n^{(1-\nu)/2})$,

it follows that $\nu \geq 1/(1+2r)$. Taking $\nu = 1/(1+2r)$ and by Lemma 3, we have $d_1(\hat{\Lambda}_n, \Lambda_0) = O_p(n^{-r/(1+2r)} + n^{-1/2}) = O_p(n^{-r/(1+2r)})$. □

S5 Proof of Theorem 3

Proof. (i) To prove this part, following Zhao and Zhang (2017), we need to verify the following conditions.

(B1) $Q(\Lambda_0, F_0)[h] = 0$ and $Q_n(\hat{\Lambda}_n, \hat{F}_n)[h] = o_p(n^{-1/2})$.

(B2) $\sqrt{n}(Q_n - Q)(\hat{\Lambda}_n, \hat{F}_n)[h] - \sqrt{n}(Q_n - Q)(\Lambda_0, F_0)[h] = o_p(1)$.

(B3) $Q(\hat{\Lambda}_n, \hat{F}_n)[h] - Q(\Lambda_0, F_0)[h] - \dot{Q}^{(1)}_{\Lambda_0, \hat{F}_n}(\hat{\Lambda}_n - \Lambda_0)[h] - \dot{Q}^{(2)}_{\Lambda_0, F_0}(\hat{F}_n - F_0)[h] = o_p(n^{-1/2})$.

For (B1), since $E(\tilde{N}(t)|U = u) = \Lambda_0(u - t)$, we have $Q(\Lambda_0, F_0)[h] = 0$. Note that $\hat{\Lambda}_n$ is obtained by minimizing the loss function $\mathbb{P}_n m(\Lambda, \hat{F}_n; X)$. By (S3.5), we have $Q_n(\hat{\Lambda}_n, \hat{F}_n)[h_n] = 0$ for all $h_n \in \Phi_n$. According to Lemma A1 of Lu, Zhang, and Huang (2007) and the properties of spline functions, for any $h \in \mathcal{H}_r$, there is an $h_n \in \Phi_n$ such that $||h_n - h||_\infty = O(n^{-r\nu}) = O(n^{-r/(1+2r)})$ and $||h_n' - h'||_\infty = o(1)$, where h' is the derivative of h. Next, to prove $Q_n(\hat{\Lambda}_n, \hat{F}_n)[h] = o_p(n^{-1/2})$, we need to show that

$Q_n(\hat{\Lambda}_n, \hat{F}_n)[h - h_n] = \mathbb{P}_n \psi(\hat{\Lambda}_n, \hat{F}_n; X)[h - h_n] = o_p(n^{-1/2})$. Note that

$$
\begin{aligned}
&Q_n(\hat{\Lambda}_n, \hat{F}_n)[h - h_n] = \Big[Q_n(\hat{\Lambda}_n, \hat{F}_n)[h - h_n] - Q_n(\hat{\Lambda}_n, F_0)[h - h_n]\Big] \\
&+ \Big[Q_n(\hat{\Lambda}_n, F_0)[h - h_n] - Q_n(\Lambda_0, F_0)[h - h_n]\Big] + Q_n(\Lambda_0, F_0)[h - h_n] \\
&= I_{1n} + I_{2n} + I_{3n}.
\end{aligned}
$$

For the first term, by Lemma 1, we have

$$
\begin{aligned}
&\mathcal{P}|I_{1n}| = \mathcal{P}|Q_n(\hat{\Lambda}_n, \hat{F}_n)[h - h_n] - Q_n(\hat{\Lambda}_n, F_0)[h - h_n]| \\
&\leq \mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left|\frac{\int_Y^\infty \big(\triangle N_j - \triangle\hat{\Lambda}_{n,j}(u)\big)\cdot(\triangle h_j(u) - \triangle h_{n,j}(u))dF_0(u)}{1 - F_0(Y)}\right.\right. \\
&\left.\left. - \frac{\int_Y^\infty \big(\triangle N_j - \triangle\hat{\Lambda}_{n,j}(u)\big)\cdot(\triangle h_j(u) - \triangle h_{n,j}(u))d\hat{F}_n(u)}{1 - \hat{F}_n(Y)}\right|\right] \\
&\lesssim d_2(\hat{F}_n, F_0)(||h - h_n||_\infty + ||h' - h_n'||_\infty) = o_p(n^{-1/2}).
\end{aligned}
$$

For I_{2n}, by the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
&\mathcal{P}|I_{2n}| = \mathcal{P}|Q_n(\hat{\Lambda}_n, F_0)[h - h_n] - Q_n(\Lambda_0, F_0)[h - h_n]| \\
&\leq \mathcal{P}\left[\sum_{k=1}^{K}\left|\Delta\{\triangle\Lambda_{0,j}(Y) - \triangle\hat{\Lambda}_{n,j}(Y)\}\cdot\{\triangle h_j(Y) - \triangle h_{n,j}(Y)\}\right.\right. \\
&\left.\left. + (1-\Delta)\frac{\int_Y^\infty\{\triangle\Lambda_{0,j}(u) - \triangle\hat{\Lambda}_{n,j}(u)\}\cdot\{\triangle h_j(u) - \triangle h_{n,j}(u)\}dF_0(u)}{1 - F_0(Y)}\right|\right] \\
&\lesssim \mathcal{P}\left[\sum_{k=1}^{K}\left\{\Delta|\triangle\Lambda_{0,j}(Y) - \triangle\hat{\Lambda}_{n,j}(Y)|\right.\right. \\
&\left.\left. + (1-\Delta)\frac{\int_Y^\infty\left|\triangle\Lambda_{0,j}(u) - \triangle\hat{\Lambda}_{n,j}(u)\right|dF_0(u)}{1 - F_0(Y)}\right\}\right]||h - h_n||_\infty \\
&\leq d_1(\hat{\Lambda}_n, \Lambda_0)||h - h_n||_\infty = o_p(n^{-1/2}).
\end{aligned}
$$

For the third term, note that $\mathcal{P}\psi(\Lambda_0, F_0; X_i)[h - h_n] = 0$. By the independence between X_i and X_j when $i \neq j$, we obtain

$$\mathcal{P}I_{3n}^2 = \mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}\psi(\Lambda_0, F_0; X_i)[h - h_n]\right)^2 = n^{-1}\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}\psi^2(\Lambda_0, F_0; X_i)[h - h_n]\right)$$
$$\lesssim n^{-1}\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta|\triangle N_j - \triangle\Lambda_{0,j}(Y)| + (1-\Delta)\frac{\int_Y^\infty |\triangle N_j - \triangle\Lambda_{0,j}(u)|\, dF_0(u)}{1 - F_0(Y)}\right\}\right]^2 ||h - h_n||_\infty^2$$
$$\lesssim n^{-1}||h - h_n||_\infty^2.$$

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

For (B2), note that

$$\sqrt{n}(Q_n - Q)(\hat{\Lambda}_n, \hat{F}_n)[h] - \sqrt{n}(Q_n - Q)(\Lambda_0, F_0)[h]$$
$$= \sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\hat{\Lambda}_n, \hat{F}_n; X)[h] - \psi(\Lambda_0, F_0; X)[h]).$$

For a bounded function $h \in \mathcal{H}_r$, define

$$\bar{\Psi}_\eta(h) = \{\psi(\Lambda, F; X)[h] - \psi(\Lambda_0, F_0; X)[h] : \Lambda \in \Phi_n, F \in \mathcal{F}$$
$$d_1(\Lambda, \Lambda_0) < \eta, d_2(F, F_0) < \eta, \Lambda \text{ is uniformly bounded}\}.$$

Similar to the proof of Lemma 2, $\bar{\Psi}_\eta(h)$ is Donsker. According to Condition

(C6) and Lemma 1, we obtain

$$
\begin{aligned}
&\mathcal{P}(\psi(\Lambda, F; X)[h] - \psi(\Lambda_0, F_0; X)[h])^2 \\
=&\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta(\triangle\Lambda_{0,j}(Y) - \triangle\Lambda_j(Y))\triangle h_j(Y) + (1-\Delta)\frac{\int_Y^\infty(\triangle\Lambda_{0,j}(u) - \triangle\Lambda_j(u))\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right.\right. \\
&\left.\left. + (1-\Delta)\left(\frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_j(u)\}\triangle h_j(u)dF(u)}{1-F(Y)} + \frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_j(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right)\right\}\right]^2 \\
\lesssim&\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta(\triangle\Lambda_{0,j}(Y) - \triangle\Lambda_j(Y))^2\triangle h_j^2(Y) + (1-\Delta)\frac{\int_Y^\infty(\triangle\Lambda_{0,j}(u) - \triangle\Lambda_j(u))^2\triangle h_j^2(u)dF_0(u)}{1-F_0(Y)}\right\}\right] \\
&+\mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left|\frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_j(u)\}\triangle h_j(u)dF(u)}{1-F(Y)} - \frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_j(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right|^2\right] \\
\lesssim& d_1(\Lambda, \Lambda_0)^2 + d_2(F, F_0)^2.
\end{aligned}
$$

It follows that for any function $\bar{\psi} \in \bar{\Psi}_\eta(h)$,

$$
\sup_{\bar{\psi}\in\bar{\Psi}_\eta(h)} \rho_{\mathcal{P}}(\bar{\psi}) \leq \sup_{\bar{\psi}\in\bar{\Psi}_\eta(h)} \mathcal{P}(\bar{\psi}^2)^{1/2} \lesssim \eta \to 0
$$

as $\eta \to 0$ for the seminorm $\rho_{\mathcal{P}}(\bar{\psi}) = \{\mathcal{P}(\bar{\psi} - \mathcal{P}\bar{\psi})^2\}^{1/2}$. Then by Corollary 2.3.12 of van der Vaart and Wellner (1996), we have

$$
\sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\hat{\Lambda}_n, \hat{F}_n; X)[h] - \psi(\Lambda_0, F_0; X)[h]) = o_p(1), \tag{S5.6}
$$

and (B2) holds.

Next, we verify (B3). Note that

$$
\begin{aligned}
&\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n - F_0)[h] \\
=&\mathcal{P}\left[(1-\Delta)\frac{1-\hat{F}_n(Y)}{1-F_0(Y)}\sum_{j=1}^{K}\left\{\frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_{0,j}(u)\}\triangle h_j(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right.\right. \\
&\left.\left. - \frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_{0,j}(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right\}\right].
\end{aligned}
$$

By the conclusion of Lemma 1, we have

$$
\begin{aligned}
&|Q(\Lambda_0,\hat{F}_n)[h]-Q(\Lambda_0,F_0)[h]-\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]| \\
=&\left|\mathcal{P}\left[(1-\Delta)\frac{\hat{F}_n(Y)-F_0(Y)}{1-F_0(Y)}\sum_{j=1}^{K}\left\{\frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right.\right.\right. \\
&\left.\left.\left.-\ \frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right\}\right]\right| \\
\lesssim&||\hat{F}_n-F_0||_\infty\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\left|\frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right.\right. \\
&\left.\left.-\ \frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right|\right]\lesssim||\hat{F}_n-F_0||^2_\infty=o_p(n^{-1/2}).
\end{aligned}
$$

Moreover, $Q(\hat{\Lambda}_n,\hat{F}_n)[h]-Q(\Lambda_0,\hat{F}_n)[h]-\dot{Q}^{(1)}_{\Lambda_0,\hat{F}_n}(\hat{\Lambda}_n-\Lambda_0)[h]=0$. It follows that

$$
Q(\hat{\Lambda}_n,\hat{F}_n)[h]-Q(\Lambda_0,F_0)[h]-\dot{Q}^{(1)}_{\Lambda_0,\hat{F}_n}(\hat{\Lambda}_n-\Lambda_0)[h]-\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]=o_p(n^{-1/2}).
$$

Thus, (B1)-(B3) hold. By (B1)and (B3), we have

$$
Q(\hat{\Lambda}_n,\hat{F}_n)[h]-\dot{Q}^{(1)}_{\Lambda_0,\hat{F}_n}(\hat{\Lambda}_n-\Lambda_0)[h]-\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]=o_p(n^{-1/2}).
$$

By (B1) and (B2), we have $-Q(\hat{\Lambda}_n,\hat{F}_n)[h]=Q_n(\Lambda_0,F_0)[h]+o_p(n^{-1/2})$. Combining the above two equations, it follows that

$$
-\sqrt{n}\dot{Q}^{(1)}_{\Lambda_0,\hat{F}_n}(\hat{\Lambda}_n-\Lambda_0)[h]=\sqrt{n}\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]+\sqrt{n}Q_n(\Lambda_0,F_0)[h]+o_p(1).
$$

(ii) To prove the second part, we need to rewrite $\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]+Q_n(\Lambda_0,F_0)[h]$. Let

$$
\tilde{\varphi}_{\Lambda,F}(u,a;X)[h]=(1-F(u))\bar{\varphi}_{\Lambda,F}(u;X)[h]-\int_u^a\bar{\varphi}_{\Lambda,F}(s;X)[h]dF(s).
$$

Note that $\hat{F}_n$ is the KM estimator based on the data $\{(\tilde{Y}_i, \tilde{\Delta}_i) : i = 1, \cdots, n\}$. Setting $G_0(s) = \mathcal{P}1_{\{\tilde{Y} \geq s\}}$, then for any constant a, by Propositions 3 and 4 of Akritas (2000), we have

$$\int_0^a \bar{\varphi}_{\Lambda_0,F_0}(u;X)[h]d(\hat{F}_n - F_0)(u) = \frac{1}{n}\sum_{i=1}^n \int_0^a \frac{\tilde{\varphi}_{\Lambda_0,F_0}(u,a;X)[h]}{G_0(u)} d\tilde{M}_i(u) + o_p(n^{-1/2}),$$

where $\tilde{M}_i(u) = 1_{\{\tilde{Y}_i \leq u, \tilde{\Delta}_i = 1\}} - \int_{-\infty}^u \{1_{\{\tilde{Y}_i \geq s\}}/(1 - F_0(s))\} dF_0(s)$. It follows that

$$\begin{aligned}
&\int_Y^\infty \bar{\varphi}_{\Lambda_0,F_0}(u;X)[h]d(\hat{F}_n - F_0)(u) \\
=&\frac{1}{n}\sum_{i=1}^n \left\{ \int_0^\infty \frac{\tilde{\varphi}_{\Lambda_0,F_0}(u,\infty;X)[h]}{G_0(u)} d\tilde{M}_i(u) - \int_0^Y \frac{\tilde{\varphi}_{\Lambda_0,F_0}(u,Y;X)[h]}{G_0(u)} d\tilde{M}_i(u) \right\} + o_p(n^{-1/2}) \\
=&\frac{1}{n}\sum_{i=1}^n \left\{ 1_{\{\tilde{\Delta}_i = 1\}} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(\tilde{Y}_i,\infty;X)[h]}{G_0(\tilde{Y}_i)} - \int_0^{\tilde{Y}_i} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(u,\infty;X)[h]}{G_0(u)(1-F_0(u))} dF_0(u) \right. \\
&\left. + \int_0^{\tilde{Y}_i \wedge Y} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(u,Y;X)[h]}{G_0(u)(1-F_0(u))} dF_0(u) - 1_{\{\tilde{\Delta}_i = 1, Y \geq \tilde{Y}_i\}} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(\tilde{Y}_i,Y;X)[h]}{G_0(\tilde{Y}_i)} \right\} + o_p(n^{-1/2}).
\end{aligned}$$

Hence, we obtain $\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n - F_0)[h] = \mathbb{P}_n\{\mathcal{P}\varphi(\Lambda_0, F_0; X; \tilde{Y}, \tilde{\Delta})[h]\} + o_p(n^{-1/2})$, where

$$\begin{aligned}
&\varphi(\Lambda, F; X; \tilde{Y}, \tilde{\Delta})[h] \\
=&1_{\{\tilde{\Delta}=1\}} \frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y},\infty;X)[h]}{G_0(\tilde{Y})} - \int_0^{\tilde{Y}} \frac{\tilde{\varphi}_{\Lambda,F}(u,\infty;X)[h]}{G_0(u)(1-F_0(u))} dF(u) \\
&+ \int_0^{\tilde{Y}\wedge Y} \frac{\tilde{\varphi}_{\Lambda,F}(u,Y;X)[h]}{G_0(u)(1-F_0(u))} dF(u) - 1_{\{\tilde{\Delta}=1, Y \geq \tilde{Y}\}} \frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y},Y;X)[h]}{G_0(\tilde{Y})}.
\end{aligned}$$

Noting that $Q_n(\Lambda_0, F_0)[h] = \mathbb{P}_n\psi(\Lambda_0, F_0; X)[h]$, we have

$$-\sqrt{n}\dot{Q}^{(1)}_{\Lambda_0,\hat{F}_n}(\hat{\Lambda}_n - \Lambda_0)[h] \xrightarrow{d} N(0, \sigma_0^2),$$

where $\sigma_0^2 = E[\{\mathcal{P}\varphi(\Lambda_0, F_0; X; \tilde{Y}, \tilde{\Delta})[h] + \psi(\Lambda_0, F_0; X)[h]\}^2]$. □

S6 Proof of Theorem 4

Proof. Setting $U_n^{(l)} = \sqrt{n}\mathbb{P}_n\varsigma(\hat{\Lambda}_l, \hat{F}_n; X)[h_n]$ for $l = 1, 2$, we have $U_n = U_n^{(1)} - U_n^{(2)}$ and $U_n^{(l)} = U_{1n}^{(l)} + U_{2n}^{(l)} + U_{3n}^{(l)}$, where

$$U_{1n}^{(l)} = \sqrt{n}(\mathbb{P}_n - \mathcal{P})\varsigma(\hat{\Lambda}_l, \hat{F}_n; X)[h_n],$$

$$U_{2n}^{(l)} = \sqrt{n}\mathcal{P}\varsigma(\hat{\Lambda}_l, \hat{F}_n; X)[h_n - h],$$

$$U_{3n}^{(l)} = \sqrt{n}\mathcal{P}\varsigma(\hat{\Lambda}_l, \hat{F}_n; X)[h].$$

For $U_{1n}^{(l)}$, similar to the proof of (S5.6), we have

$$\sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\hat{\Lambda}_l, \hat{F}_n; X)[h_n] - \psi(\Lambda_0, F_0; X)[h_n]) = o_p(1)$$

and

$$\sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\Lambda_0, \hat{F}_n; X)[h_n] - \psi(\Lambda_0, F_0; X)[h_n]) = o_p(1).$$

Noting that $\varsigma(\Lambda, F; X)[h_n] = \psi(\Lambda_0, F; X)[h_n] - \psi(\Lambda, F; X)[h_n]$, it follows that $U_{1n}^{(l)} = o_p(1)$.

For $U_{2n}^{(l)}$, by Cauchy-Schwarz inequality, we obtain

$$
\begin{aligned}
U_{2n}^{(l)} \leq \sqrt{n}\mathcal{P}\Bigg[\sum_{j=1}^{K}\Bigg\{&\Delta\big(\triangle\hat{\Lambda}_{l,j}(Y)-\triangle\Lambda_{0,j}(Y)\big)^2 \\
+\ (1-\Delta)&\frac{\int_Y^\infty\big(\triangle\hat{\Lambda}_{l,j}(u)-\triangle\Lambda_{0,j}(u)\big)^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\Bigg\}\Bigg]^{1/2} \\
\times\mathcal{P}\Bigg[\sum_{j=1}^{K}\Bigg\{&\Delta\big(\triangle h_{n,j}(Y)-\triangle h_j(Y)\big)^2 \\
+\ (1-\Delta)&\frac{\int_Y^\infty\big(\triangle h_{n,j}(u)-\triangle h_j(u)\big)^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\Bigg\}\Bigg]^{1/2}.
\end{aligned}
$$

By Lemma 1, similar to the proof of the first inequality in Theorem 2, we have

$$
\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle\hat{\Lambda}_{l,j}(Y)-\triangle\Lambda_{0,j}(Y)\big)^2+(1-\Delta)\frac{\int_Y^\infty\big(\triangle\hat{\Lambda}_{l,j}(u)-\triangle\Lambda_{0,j}(u)\big)^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right]
$$

$$
\lesssim d_1(\hat{\Lambda}_l,\Lambda_0)^2+d_1(\hat{\Lambda}_l,\Lambda_0)d_2(\hat{F}_n,F_0)=O_p(n^{-2r/(1+2r)})
$$

and

$$
\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle h_{n,j}(Y)-\triangle h_j(Y)\big)^2+(1-\Delta)\frac{\int_Y^\infty\big(\triangle h_{n,j}(u)-\triangle h_j(u)\big)^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right]
$$

$$
\lesssim d_1(h_n,h)^2+d_1(h_n,h)d_2(\hat{F}_n,F_0)=o_p(n^{-1/(1+2r)}).
$$

Hence, $U_{2n}^{(l)}=o_p(1)$.

For $U_{3n}^{(l)}$, by Theorem 3, we have

$$
U_{3n}^{(l)}=\sqrt{n}\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]+\sqrt{n}\mathbb{P}_{n_l}\psi(\Lambda_0,F_0;X)[h]+o_p(1),
$$

where $\mathbb{P}_{n_l}$ is the empirical measure based on group l. Thus,

$$\begin{aligned}
&U_n = U_{3n}^{(1)} - U_{3n}^{(2)} + o_p(1) \\
=&\sqrt{\frac{n}{n_1}}\sqrt{n_1}\mathbb{P}_{n_1}\psi(\Lambda_0, F_0; X)[h] - \sqrt{\frac{n}{n_2}}\sqrt{n_2}\mathbb{P}_{n_2}\psi(\Lambda_0, F_0; X)[h] + o_p(1).
\end{aligned}$$

Note that $\mathbb{P}_{n_1}$ and $\mathbb{P}_{n_2}$ are independent, and $\sqrt{n_l}\mathbb{P}_{n_l}\psi(\Lambda_0, F_0; X)[h] \xrightarrow{d} N(0, \sigma_0^2)$. Thus, we have $U_n \xrightarrow{d} N(0, (\frac{1}{p} + \frac{1}{1-p})\sigma_0^2)$.

Finally, we need to prove that $\hat{\sigma}_n^2 - \sigma_0^2 = o_p(1)$. Note that $\sigma_0^2 = \mathcal{P}\psi^2(\Lambda_0, F_0; X)[h]$ and $\hat{\sigma}_n^2 = \mathbb{P}_n\psi^2(\hat{\Lambda}_n, \hat{F}_n; X)[h_n]$. Thus, we have

$$\begin{aligned}
&\hat{\sigma}_n^2 - \sigma_0^2 = \mathbb{P}_n\{\psi^2(\hat{\Lambda}_n, \hat{F}_n; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h_n]\} \\
&+\mathbb{P}_n\{\psi^2(\Lambda_0, F_0; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h]\} + (\mathbb{P}_n - \mathcal{P})\psi^2(\Lambda_0, F_0; X)[h].
\end{aligned}$$

By the consistency of $(\hat{\Lambda}_n, \hat{F}_n)$ and the law of large numbers, we have

$$\mathbb{P}_n\{\psi^2(\hat{\Lambda}_n, \hat{F}_n; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h_n]\} = o_p(1)$$

and

$$(\mathbb{P}_n - \mathcal{P})\psi^2(\Lambda_0, F_0; X)[h] = o_p(1).$$

Then we only need to consider $\mathbb{P}_n\{\psi^2(\Lambda_0, F_0; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h]\}$. Since Λ_0, F_0, h and h_n are bounded functions, we have

$$\begin{aligned}
&|\psi(\Lambda_0, F_0; X)[h_n] + \psi(\Lambda_0, F_0; X)[h]| = |\psi(\Lambda_0, F_0; X)[h_n + h]| \\
\lesssim&\left|\sum_{j=1}^{K}\left[\Delta\{\triangle N_j - \triangle\Lambda_{0,j}(Y)\} + (1-\Delta)\frac{\int_Y^\infty\{(\triangle N_j - \triangle\Lambda_{0,j}(u))\}dF_0(u)}{1 - F_0(Y)}\right]\right| \\
\lesssim&(N(T_K) + \Lambda_0(\tau))
\end{aligned}$$

with probability 1. Thus, by Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
&\mathcal{P}\left|\psi^2(\Lambda_0, F_0; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h]\right| \\
\lesssim &\mathcal{P}\left[\left|\psi(\Lambda_0, F_0; X)[h_n] - \psi(\Lambda_0, F_0; X)[h]\right|(N(T_K) + \Lambda_0(\tau))\right] \\
\lesssim &\mathcal{P}\left[(N(T_K) + \Lambda_0(\tau))\sum_{j=1}^{K}\left\{\Delta|\triangle N_j - \triangle\Lambda_{0,j}(Y)||\triangle h_{n,j}(Y) - \triangle h_j(Y)|\right.\right. \\
&\left.\left. + (1-\Delta)\frac{\int_Y^\infty |\triangle N_j - \triangle\Lambda_{0,j}(u)||\triangle h_{n,j}(u) - \triangle h_j(u)|dF_0(u)}{1 - F_0(Y)}\right\}\right] \\
\lesssim &\mathcal{P}\left[(N(T_K) + \Lambda_0(\tau))^2\sum_{j=1}^{K}\left\{\Delta|\triangle h_{n,j}(Y) - \triangle h_j(Y)|\right.\right. \\
&\left.\left. + (1-\Delta)\frac{\int_Y^\infty |\triangle h_{n,j}(u) - \triangle h_j(u)|dF_0(u)}{1 - F_0(Y)}\right\}\right] \\
\leq &\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle h_{n,j}(Y) - \triangle h_j(Y)\big)^2 + (1-\Delta)\frac{\int_Y^\infty \big(\triangle h_{n,j}(u) - \triangle h_j(u)\big)^2 dF_0(u)}{1 - F_0(Y)}\right\}\right]^{\frac{1}{2}} \\
&\times\mathcal{P}\left[(N(T_K) + \Lambda_0(\tau))^4\right]^{\frac{1}{2}} \\
\lesssim &d_1(h_n, h) = o(1).
\end{aligned}
$$

Therefore, $\mathbb{P}_n\{\psi^2(\Lambda_0, F_0; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h]\} = o_p(1)$ and $\hat{\sigma}_n^2 - \sigma_0^2 = o_p(1)$. □

S7 Proof of Theorem 5

Proof. (i) First we write

$$
\tilde{U}_n = \sum_{l=1}^{2}\frac{n_l}{\sqrt{n}}\mathbb{P}_{n_l}\left(\varsigma(\hat{\Lambda}_1, \hat{F}_l; X^{(l)})[h_n] - \varsigma(\hat{\Lambda}_2, \hat{F}_l; X^{(l)})[h_n]\right)
$$

for $l = 1, 2$. Note that

$$\mathbb{P}_{n_l}(\varsigma(\Lambda, \hat{F}_l; X^{(l)})[h_n]) = (\mathbb{P}_{n_l} - \mathcal{P}_l)\varsigma(\Lambda, \hat{F}_l; X^{(l)})[h_n] + \mathcal{P}_l\varsigma(\Lambda, \hat{F}_l; X^{(l)})[h_n - h]$$

$$+\mathcal{P}_l(\varsigma(\Lambda, \hat{F}_l; X^{(l)})[h] - \varsigma(\Lambda, F_l; X^{(l)})[h]) + \mathcal{P}_l\varsigma(\Lambda, F_l; X^{(l)})[h].$$

We next show that

$$\mathbb{P}_{n_l}(\varsigma(\hat{\Lambda}_1, \hat{F}_l; X^{(l)})[h_n]) = \mathcal{P}_l\varsigma(\hat{\Lambda}_1, F_l; X^{(l)})[h] + o_p(n^{-1/2}) \tag{S7.7}$$

and

$$\mathbb{P}_{n_l}(\varsigma(\hat{\Lambda}_2, \hat{F}_l; X^{(l)})[h_n]) = \mathcal{P}_l\varsigma(\hat{\Lambda}_2, F_l; X^{(l)})[h] + o_p(n^{-1/2}). \tag{S7.8}$$

According to the proof of Theorem 4, we have

$$\sqrt{n_l}(\mathbb{P}_{n_l} - \mathcal{P}_l)\varsigma(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[h_n] = o_p(1) \text{ and } \sqrt{n_l}\mathcal{P}_l\varsigma(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[h_n - h] = o_p(1).$$

Moreover, according to Lemma 1, this implies that

$$\begin{aligned}
&\left|\mathcal{P}_l\left(\varsigma(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[h] - \varsigma(\hat{\Lambda}_l, F_l; X^{(l)})[h]\right)\right| \\
\lesssim &\left(\mathcal{P}_l\left[\sum_{j=1}^{K^{(l)}}\left|\triangle\hat{\Lambda}_{l,j}(U^{(l)}) - \triangle\Lambda_{0,j}(U^{(l)})\right|\right]\right. \\
&\left. + \mathcal{P}_l\left[\sum_{j=1}^{K^{(l)}}\left|\triangle\hat{\Lambda}'_{l,j}(U^{(l)}) - \triangle\Lambda'_{0,j}(U^{(l)})\right|\right]\right)||\hat{F}_l - F_l||_\infty \\
\lesssim &(||\hat{\Lambda}_l - \Lambda_0||_{L_2(\mu_l)} + ||\hat{\Lambda}'_l - \Lambda'_0||_{L_2(\mu_l)})||\hat{F}_l - F_l||_\infty.
\end{aligned} \tag{S7.9}$$

Applying Lemma 3.5 and Corollary 2.1 of He and Shi (1994), we have $||\hat{\Lambda}'_l - \Lambda'_0||_{L_2(\mu_l)} = O_p(n^{-(r-1)/(1+2r)}) = o_p(1)$. This gives $\mathcal{P}_l(\varsigma(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[h] - \varsigma(\hat{\Lambda}_l, F_l; X^{(l)})[h]) = o_p(n^{-1/2})$. Thus, (S7.7) and (S7.8) hold.

Moreover, we have

$$\mathcal{P}_l\left[\sum_{j=1}^{K^{(l)}}\triangle h_j(U^{(l)})\right]=\int E\left[\sum_{j=1}^{K^{(l)}}\triangle h_j(u)\middle|U^{(l)}=u\right]dF_l(u)$$
$$=\int E\left[\sum_{j=1}^{K^{(r)}}\triangle h_j(u)\middle|U^{(r)}=u\right]\frac{f_l(u)}{f_r(u)}dF_r(u)=\mathcal{P}_r\left[\sum_{j=1}^{K^{(r)}}\triangle h_j(U^{(r)})\frac{f_l(U^{(r)})}{f_r(U^{(r)})}\right]. \tag{S7.10}$$

Hence, (S7.7), (S7.8) and (S7.10) yield

$$\begin{aligned}\tilde{U}_n=&\sum_{l=1}^{2}\frac{n_l}{\sqrt{n}}\mathcal{P}_l\left(\varsigma(\hat{\Lambda}_1,F_l;X^{(l)})[h]-\varsigma(\hat{\Lambda}_2,F_l;X^{(l)})[h]\right)+o_p(1)\\
=&\sqrt{n}\sum_{l=1}^{2}\mathcal{P}_l\varsigma(\hat{\Lambda}_1,F_l;X^{(l)})[p_lh]-\sqrt{n}\sum_{l=1}^{2}\mathcal{P}_l\varsigma(\hat{\Lambda}_2,F_l;X^{(l)})[p_lh]+o_p(1)\\
=&\frac{\sqrt{n_1}}{\sqrt{p_1}}\mathcal{P}_1\varsigma(\hat{\Lambda}_1,F_1;X^{(1)})[w_1]-\frac{\sqrt{n_2}}{\sqrt{p_2}}\mathcal{P}_2\varsigma(\hat{\Lambda}_2,F_2;X^{(2)})[w_2]+o_p(1).\end{aligned}$$

Set $G_l(s)=\mathcal{P}_l1_{\{\tilde{Y}^{(l)}\geq s\}}$ and

$$\varphi_l(\Lambda,F;X;\tilde{Y},\tilde{\Delta})[w]=1_{\{\tilde{\Delta}=1\}}\frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y},\infty;X)[w]}{G_l(\tilde{Y})}-\int_0^{\tilde{Y}}\frac{\tilde{\varphi}_{\Lambda,F}(u,\infty;X)[w]}{G_l(u)(1-F(u))}dF(u)$$
$$+\int_0^{\tilde{Y}\wedge Y}\frac{\tilde{\varphi}_{\Lambda,F}(u,Y;X)[w]}{G_l(u)(1-F(u))}dF(u)-1_{\{\tilde{\Delta}=1,Y>\tilde{Y}\}}\frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y},Y;X)[w]}{G_l(\tilde{Y})}.$$

By Theorem 3 and (S7.9), we can get

$$\sqrt{n_l}\mathcal{P}_l\varsigma(\hat{\Lambda}_l,F_l;X^{(l)})[w_l]=\sqrt{n_l}\mathcal{P}_l\varsigma(\hat{\Lambda}_l,\hat{F}_l;X^{(l)})[w_l]+o_p(1)\rightsquigarrow N(0,\sigma_l^2).$$

Since $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ are independent, it then follows that $\tilde{U}_n\rightsquigarrow N(0,(\sigma_1^2/p_1+\sigma_2^2/p_2))$.

(ii) Set

$$\varphi_n^{(l)}(\Lambda, F; X; \tilde{Y}, \tilde{\Delta})[w] = 1_{\{\tilde{\Delta}=1\}} \frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y}, \infty; X)[w]}{G_n^{(l)}(\tilde{Y})} - \int_0^{\tilde{Y}} \frac{\tilde{\varphi}_{\Lambda,F}(u, \infty; X)[w]}{G_n^{(l)}(u)(1-F(u))} dF(u)$$
$$+ \int_0^{\tilde{Y}\wedge Y} \frac{\tilde{\varphi}_{\Lambda,F}(u, Y; X)[w]}{G_n^{(l)}(u)(1-F(u))} dF(u) - 1_{\{\tilde{\Delta}=1, Y>\tilde{Y}\}} \frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y}, Y; X)[w]}{G_n^{(l)}(\tilde{Y})},$$

where $G_n^{(l)}(s) = \mathbb{P}_{n_l} 1_{\{\tilde{Y}^{(l)} \geq s\}}$. Then

$$\begin{aligned}
\hat{\sigma}_l^2 - \sigma_l^2 =& \mathbb{P}_{n_l} \left[\left\{ \mathbb{P}_n \varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)}] + \psi(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[w_n^{(l)}] \right\}^2 \right. \\
& \left. - \left\{ \mathcal{P}\varphi_l(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)}] + \psi(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[w_n^{(l)}] \right\}^2 \right] \\
& + \mathbb{P}_{n_l} \left[\mathcal{P}\varphi_l(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)}] + \psi(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[w_n^{(l)}] \right]^2 \\
& - \mathcal{P} \left[\mathcal{P}\varphi_l(\Lambda_0, F_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_l] + \psi(\Lambda_0, F_l; X^{(l)})[w_l] \right]^2.
\end{aligned}$$

Note that

$$\begin{aligned}
& \mathbb{P}_{n_l} \left[\left\{ \mathbb{P}_n \varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)}] + \psi(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[w_n^{(l)}] \right\}^2 \right. \\
& \left. - \left\{ \mathcal{P}\varphi_l(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)}] + \psi(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[w_n^{(l)}] \right\}^2 \right] \\
=& \frac{1}{n} \sum_{i=1}^n \left[\left\{ \mathbb{P}_n \varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] - \mathcal{P}\varphi_l(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] \right\} \right. \\
& \times \left\{ \mathbb{P}_n \varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] + \mathcal{P}\varphi_l(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] \right. \\
& \left. \left. + \ 2\psi(\hat{\Lambda}_l, \hat{F}_l; X_i^{(l)})[w_n^{(l)}] \right\} \right].
\end{aligned}$$

For each i, we obtain

$$\begin{aligned}
&\mathbb{P}_n\varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] - \mathcal{P}\varphi_l(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] \\
=&(\mathbb{P}_n - \mathcal{P})\varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] \\
+&\mathcal{P}\left(\varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] - \varphi_l(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}]\right).
\end{aligned}$$

Since $\sup_{s\in[0,\tau]} |G_n^{(l)}(s) - G_l(s)| = O_p(n^{-1/2})$, we get

$$\mathcal{P}\left(\varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] - \varphi_l(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}]\right) = o_p(1).$$

Similar to the proof of Lemma 2, under Conditions (C2), (C4), (C6) and (C10),

$$\left\{\varphi_n^{(l)}(\Lambda, F; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] : \Lambda \in \Phi,\ F \in \mathcal{F},\ d_2(F, F_l) \le \delta,\ \Lambda \text{ is uniformly bounded}\right\}$$

is Donsker and it is Glivenko-Cantelli. It follows that

$$(\mathbb{P}_n - \mathcal{P})\varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}_i^{(l)}, \tilde{\Delta}_i^{(l)})[w_n^{(l)}] = o_p(1)$$

and the first term of $\hat{\sigma}_l^2 - \sigma_l^2$ is $o_p(1)$. Denote

$$\upsilon_l(\Lambda, F; \tilde{Y}, \tilde{\Delta}, X)[w] = \mathcal{P}\varphi_l(\Lambda, F; X^{(l)}; \tilde{Y}, \tilde{\Delta})[w] + \psi(\Lambda, F; X)[w].$$

To verify that the second term of $\hat{\sigma}_l^2 - \sigma_l^2$ is $o_p(1)$, we only need to prove

$$E[|\upsilon_l^2(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}, X^{(l)})[w_n^{(l)}] - \upsilon_l^2(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}, X^{(l)})[w_l]|] = o(1).$$

By the definition of $w_n^{(l)}(u,t)$, we obtain

$$
\begin{aligned}
&\left|\left\{w_n^{(l)}(u-t_1)-w_n^{(l)}(u-t_2)\right\}-\{w_l(u-t_1)-w_l(u-t_2)\}\right| \\
=&\left|\{(h_n(u-t_1)-h(u-t_1))-(h_n(u-t_2)-h(u-t_2))\}\left(\frac{n_l}{n}+\frac{n_r}{n}\frac{\hat{f}_r(u)}{\hat{f}_l(u)}\right)\right. \\
&\left.+\left(\frac{n_l}{n}+\frac{n_r}{n}\frac{\hat{f}_r(u)}{\hat{f}_l(u)}-p_r\frac{f_r(u)}{f_l(u)}-p_l\right)\{h(u-t_1)-h(u-t_2)\}\right| \\
\leq&|(h_n(u-t_1)-h(u-t_1))-(h_n(u-t_2)-h(u-t_2))|\left|\frac{n_l}{n}+\frac{n_r}{n}\frac{\hat{f}_r(u)}{\hat{f}_l(u)}\right| \\
&+\left|\frac{n_l}{n}+\frac{n_r}{n}\frac{\hat{f}_r(u)}{\hat{f}_l(u)}-p_r\frac{f_r(u)}{f_l(u)}-p_l\right||h(u-t_1)-h(u-t_2)| \\
\leq&c\left\{|h_n(u-t_1)-h(u-t_1)|+|h_n(u-t_2)-h(u-t_2)|+\left|\frac{\hat{f}_r(u)}{\hat{f}_l(u)}-\frac{f_r(u)}{f_l(u)}\right|\right\}
\end{aligned}
\tag{S7.11}
$$

with probability 1 for some constant c, where $l,r=1,2$ and $l\neq r$. According to Theorem 2.2 of Földes, Rejtő, and Winter (1981), we have

$$
\sup_u\left|\frac{\hat{f}_r(u)}{\hat{f}_l(u)}-\frac{f_r(u)}{f_l(u)}\right|\xrightarrow{a.s.}0. \tag{S7.12}
$$

Furthermore, by the definition of $\varphi_l(\Lambda,F;X;Y,\Delta)[w]$, $\bar{\varphi}_{\Lambda,F}(u;X)[w]$ and $\tilde{\varphi}_{\Lambda,F}(u,a;X)[w]$, after some calculations, we have

$$
|\upsilon_l(\Lambda_0,F_l;\tilde{Y}^{(l)},\tilde{\Delta}^{(l)},X^{(l)})[w_n^{(l)}+w_l]|\lesssim N^{(l)}(T_{K^{(l)}}^{(l)})+\Lambda_0(0) \tag{S7.13}
$$

and

$$
\begin{aligned}
&|\upsilon_l(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}, X^{(l)})[w_n^{(l)} - w_l]| \\
\lesssim &\left(N^{(l)}(T_{K^{(l)}}^{(l)}) + \Lambda_0(\tau)\right) \sum_{j=1}^{K^{(l)}} \left[\Delta^{(l)} \left|\triangle w_{n,j}^{(l)}(Y^{(l)}) - \triangle w_{l,j}(Y^{(l)})\right|\right. \\
&\left. + (1-\Delta^{(l)}) \frac{\int_{Y^{(l)}}^{\infty} \left|\triangle w_{n,j}^{(l)}(u) - \triangle w_{l,j}(u)\right| dF_l(u)}{1 - F_l(Y^{(l)})}\right] \\
&+|\mathcal{P}\varphi_l(\Lambda_0, F_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)} - w_l]|
\end{aligned}
\tag{S7.14}
$$

with probability 1. Combining (S7.11)–(S7.14) and using the Cauchy-Schwarz inequality, we obtain

$$
\begin{aligned}
&E|\upsilon_l^2(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}; X^{(l)})[w_n^{(l)}] - \upsilon_l^2(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}; X^{(l)})[w_l]| \\
\lesssim &E\left[\left(N^{(l)}(T_K^{(l)}) + \Lambda_0(\tau)\right)^2 \sum_{j=1}^{K^{(l)}} \left\{\Delta^{(l)} \left|\triangle w_{n,j}^{(l)}(Y^{(l)}) - \triangle w_{l,j}(Y^{(l)})\right|\right.\right. \\
&\left.\left. + (1-\Delta^{(l)}) \frac{\int_{Y^{(l)}}^{\infty} \left|\triangle w_{n,j}^{(l)}(u) - \triangle w_{l,j}(u)\right| dF_l(u)}{1 - F_l(Y^{(l)})}\right\}\right] \\
&+E\left[\left(N^{(l)}(\tau) + \Lambda_0(\tau)\right) |\mathcal{P}\varphi_l(\Lambda_0, F_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)} - w_l]|\right] \\
\lesssim &E\left[|\mathcal{P}\varphi_l(\Lambda_0, F_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)} - w_l]|\right] + ||h_n^{(l)} - h_l||_{L_2(\mu_l)} \\
\lesssim &||h_n^{(l)} - h_l||_{L_2(\mu_l)} \to 0.
\end{aligned}
$$

□

S8 Two-Sample Test with Crossed Distribution Functions of Terminal Event Times

At this section, we conducted the simulation studies for two groups with crossed distribution functions of the terminal event times. We generated $U_i^{(1)}$ and $U_i^{(2)}$ from $6 + \exp(1)$ and $6 + \text{weibull}(1, 2)$, respectively. Then the distribution functions of the terminal event times in the two groups were $F_1(u) = 1 - \exp(-u)$ and $F_2(u) = 1 - \exp(-u^2)$, which crossed over on $u \in [0, \tau]$. The censoring times $C_i^{(1)}$ and $C_i^{(2)}$ were from $6 + \kappa_1 \exp(2)$ and $6 + \kappa_2 \exp(2)$, where $(\kappa_1, \kappa_2) = (0.748, 0.801)$ and $(0.334, 0.421)$ to result in 40% and 60% censoring rate, respectively. We still considered two cases with sample size $n_1 = n_2 = 100$, 150 or 200. For Case 1, we set $\Lambda_1(s) = s$ and $\Lambda_2(s) = \beta s$ with $\beta = 1, 1.1, 1.2, 1.3$. For Case 2, we considered partly overlapped two functions in the following two scenarios by

$$\text{Scenario 1} : \Lambda_1(s) = s \text{ and } \Lambda_2(s) = 1_{\{s \geq 10-\beta\}} s + 1_{\{s < 10-\beta\}} \frac{10s}{s+\beta},$$

$$\text{Scenario 2} : \Lambda_1(s) = s \text{ and } \Lambda_2(s) = 1_{\{s \geq 10-\beta\}} \frac{10s}{s+\beta} + 1_{\{s < 10-\beta\}} s$$

with $\beta = 3, 4, 5, 6$.

Figure 1 displays the true mean functions in Case 2 for the two scenarios with different values of β, which reveals that Λ_1 and Λ_2 overlap on $[\tau - \beta, \tau]$ in Scenario 1 and on $[0, \tau - \beta]$ in Scenario 2. We also considered the three

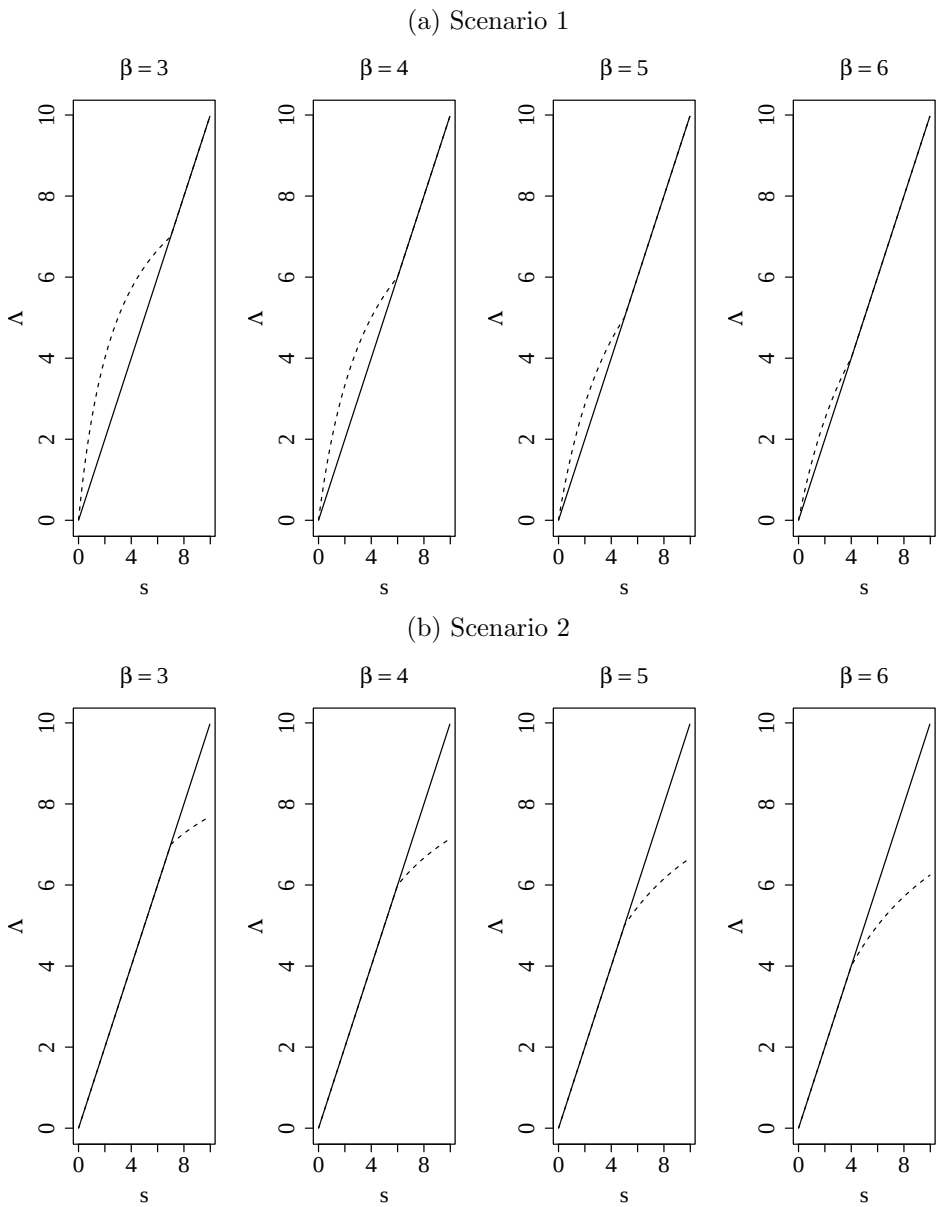

Figure 1: Discrepancy of the mean functions for Case 2.

weight processes $h_n^{(j)}(t), j = 1, \cdots, 3$ in Subsection 5.2, and the simulation results were evaluated with 1000 replications.

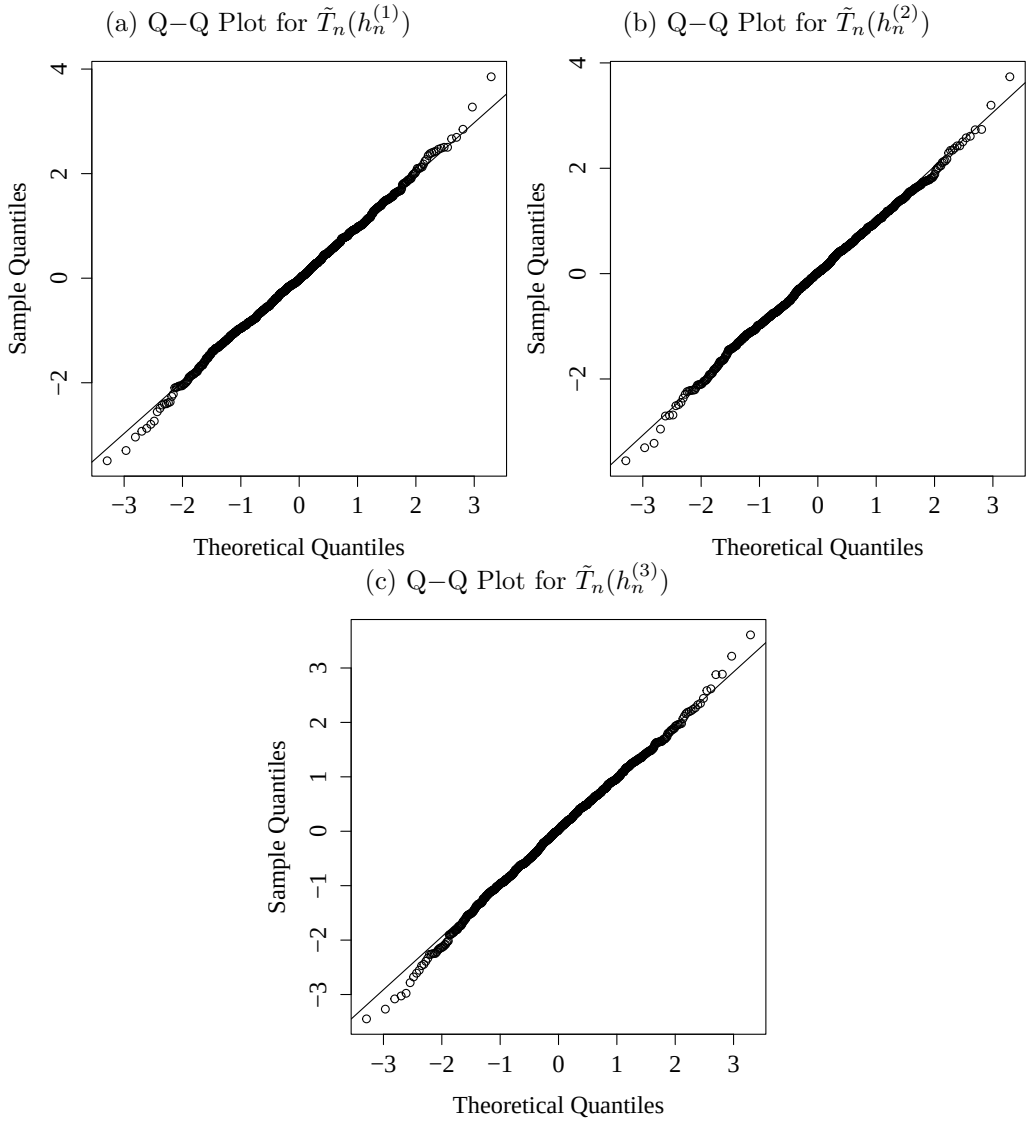

Figure 2: Q-Q plots for n_1=n_2=200, β=1, and censoring rate = 60% when the distribution functions of the terminal event time for the two groups cross over.

We also used the quantile plots against the standard normal distribution to justify the normality of the test statistics. Figure 2 shows the plot for

Table 1: Simulation results of the two-sample tests with different weights for Case 1 when the distribution functions of the terminal event time for the two groups cross over.

	Censoring rate 40%			Censoring rate 60%		
β	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$
			n_1=n_2=100			
1	0.057	0.057	0.059	0.066	0.060	0.062
1.1	0.268	0.235	0.267	0.253	0.249	0.251
1.2	0.713	0.683	0.710	0.730	0.690	0.731
1.3	0.956	0.944	0.960	0.953	0.948	0.950
			n_1=n_2=150			
1	0.059	0.041	0.059	0.056	0.056	0.055
1.1	0.368	0.340	0.369	0.375	0.336	0.376
1.2	0.869	0.835	0.869	0.878	0.860	0.878
1.3	0.998	0.995	0.998	0.996	0.993	0.996
			n_1=n_2=200			
1	0.055	0.050	0.051	0.054	0.052	0.054
1.1	0.468	0.422	0.473	0.450	0.421	0.450
1.2	0.960	0.946	0.959	0.955	0.940	0.953
1.3	0.999	0.998	0.999	0.999	0.999	0.999

Table 2: Simulation results of the two-sample tests with different weights for Scenario 1 in Case 2 when the distribution functions of the terminal event time for the two groups cross over.

	Censoring rate 40%			Censoring rate 60%		
β	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$
			n_1=n_2=100			
3	0.970	1.000	0.828	0.976	0.999	0.873
4	0.817	0.957	0.693	0.856	0.966	0.777
5	0.446	0.523	0.402	0.491	0.562	0.454
6	0.188	0.153	0.177	0.213	0.179	0.210
			n_1=n_2=150			
3	0.994	1.000	0.943	0.996	1.000	0.971
4	0.925	0.991	0.860	0.955	0.994	0.909
5	0.627	0.701	0.567	0.659	0.732	0.612
6	0.248	0.201	0.235	0.277	0.220	0.255
			n_1=n_2=200			
3	0.998	1.000	0.984	1.000	1.000	0.991
4	0.973	0.997	0.925	0.986	0.998	0.961
5	0.717	0.797	0.666	0.753	0.829	0.715
6	0.288	0.252	0.271	0.305	0.262	0.295

Table 3: Simulation results of the two-sample tests with different weights for Scenario 2 in Case 2 when the distribution functions of the terminal event time for the two groups cross over.

	Censoring rate 40%			Censoring rate 60%		
β	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$
			n_1=n_2=100			
3	0.073	0.057	0.078	0.081	0.077	0.078
4	0.309	0.390	0.274	0.295	0.380	0.263
5	0.829	0.941	0.750	0.796	0.922	0.726
6	0.986	1.000	0.973	0.987	0.999	0.960
			n_1=n_2=150			
3	0.074	0.073	0.069	0.076	0.072	0.075
4	0.432	0.562	0.387	0.406	0.525	0.362
5	0.923	0.981	0.891	0.924	0.983	0.881
6	0.998	1.000	0.993	0.981	1.000	0.994
			n_1=n_2=200			
3	0.060	0.060	0.060	0.060	0.056	0.059
4	0.513	0.656	0.462	0.462	0.613	0.412
5	0.980	0.996	0.958	0.973	0.997	0.946
6	1.000	1.000	1.000	1.000	1.000	0.999

$n_1 = n_2 = 200$, $\beta = 1$ and censoring rate 60% with the three weight processes in Case 1, and the plots for other situations are similar and not presented here. Tables 1–3 summarize the sizes and powers of the test with the three weight processes at significance level 0.05 for Cases 1 and 2, respectively. For Case 1, the sizes and powers yield similar conclusions to those in Subsection 5.2 and 5.3 even for a high censoring rate, say 60%. For Case 2, the simulation results reveal that the less the overlap, the higher power the test has, regardless of where the overlap locates. The test power increases as the sample size increases for all values of β. For the three weight processes, the powers with $h_n^{(3)}$ are the smallest when $\beta = 3$, 4, and

5 in Scenario 1; while the powers with $h_n^{(2)}$ are the largest when $\beta = 4$, 5, and 6 in Scenario 2.

S9 Real Data Analysis for Data from 2002

In this section, we treated the survey in 2002 as the baseline survey and analyzed all the 967 individuals who were still alive and contacted in the survey in 2005 among the 4362 elders. We still divided them into three groups according to their living area, which yields 346 elders living in urban, 319 elders living in rural, and 302 elders living in both areas. The censoring rate was 30.61%.

Figure 3 also justifies that urban-area residents may experience more serious illnesses than rural-area residents and dual-area residents. The patterns of the estimated mean functions for the four groups are fundamentally the same as those shown in Figure 5 in Section 6. The log-rank test revealed that the survival functions of death for any two groups were not significantly different, where p-value= 0.2 for the test between urban-area residents and rural-area residents, p-value= 0.6 for the test between urban-area residents and dual-area residents, and p-value= 0.4 for the test between rural-area residents and dual-area residents. Figure 4 shows the KM estimates for the survival functions of death. Therefore, we used the statistic in Theorem

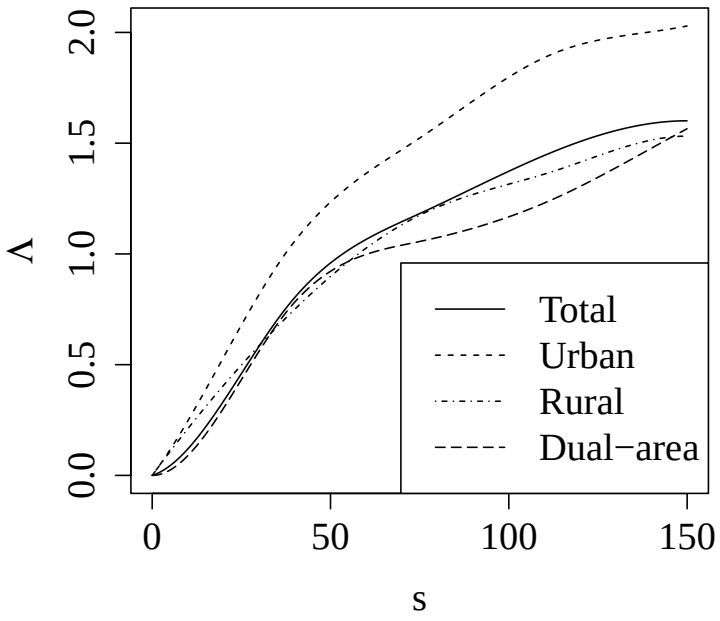

Figure 3: The estimates of the mean function for the CLHLS data from 2002.

4 to test the null hypotheses $H_0^{(1)}$: $\Lambda_U(s) = \Lambda_R(s)$, $H_0^{(2)}$: $\Lambda_U(s) = \Lambda_D(s)$, and $H_0^{(3)}$: $\Lambda_R(s) = \Lambda_D(s)$. Table 4 presents the test results, which yield the same conclusions as those for the whole data from 1998. However, since the sample size (only 967 individuals had at least one observation after 2002) is largely reduced, the powers for detecting the group differences as shown in Table 4 are significantly lost as expected.

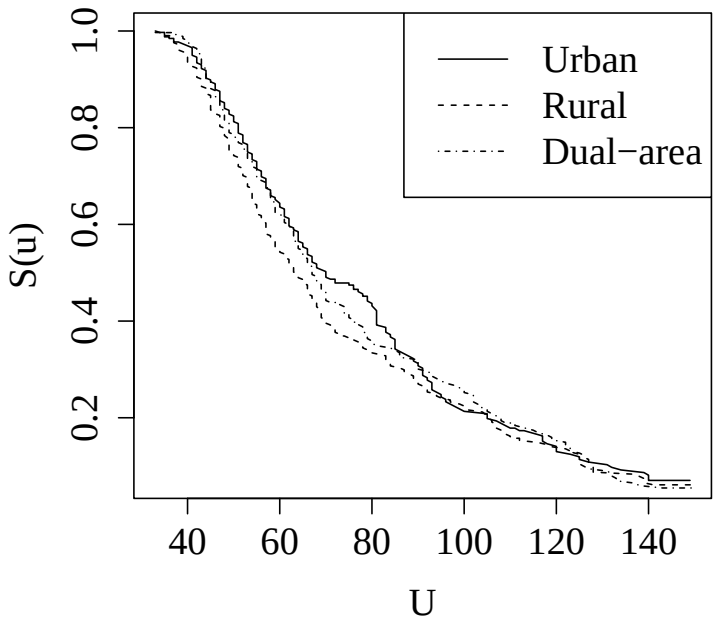

Figure 4: The KM estimates of the survival function for the CLHLS data from 2002.

Bibliography

Akritas, M. G. (2000). The central limit theorem under censoring. *Bernoulli* **6**, 1109–1120.

Földes, A., Rejtő, L., and Winter, B. B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data, II: Estimation of density and failure rate. *Periodica Mathematica Hungarica* **12**, 15–29.

He, X., and Shi, P. (1994). Convergence rate of B-spline estimators of

Table 4: Two-sample test results with the three weights for the CLHLS data from 2002.

	$h_n^{(1)}$	$h_n^{(2)}$	$h_n^{(3)}$
	$H_0^{(1)}$: $\Lambda_U(s) = \Lambda_R(s)$		
U_n	209.299	3.239	3.747
σ_n	97.905	1.669	1.847
T_n	2.138	1.941	2.028
p-value	0.033**	0.052*	0.042**
	$H_0^{(2)}$: $\Lambda_U(s) = \Lambda_D(s)$		
U_n	200.766	2.494	3.062
σ_n	100.329	1.678	2.024
T_n	2.001	1.486	1.513
p-value	0.045**	0.137	0.130
	$H_0^{(3)}$: $\Lambda_R(s) = \Lambda_D(s)$		
U_n	−15.854	−0.735	−0.742
σ_n	89.800	1.562	1.524
T_n	− 0.177	−0.470	−0.487
p-value	0.860	0.638	0.626

$T_n(h)$: the observed value of the test statistic with different weight functions;

* represents significance level of 0.1; ** represents significance level of 0.05

nonparametric conditional quantile functions. *Journal of Nonparametric Statistics* **3**, 299–308.

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

Lu, M., Zhang, Y., and Huang, J. (2009). Semiparametric estiamtion methods for panel count data using monotone B-splines. *Journal of the American Statistical Association* **104**, 1060–1070.

Shen, X., and Wong, W. H. (1994). Convergence rate of sieve estimates. *Annals of Statistics* **14**, 590–606.

van der Vaart, A. W. (1996). New donsker classes. *Annals of Probability* **24**, 862–878.

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

van der Vaart, A. W., and Wellner, J. A. (1996). *Weak convergence and empirical processes.* New York: Springer-Verlag.

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