# Supplementary Material

# Versatile Estimation in Censored Single-Index Hazards Regression

Chin-Tsang Chiang, Shao-Hsuan Wang, and Ming-Yueh Huang
Institute of Applied Mathematical Sciences
National Taiwan University, Taipei 10617, Taiwan, R.O.C.

January 19, 2017

## Supplementary Material (I)

### Notations

The notations used in Sections 2.3 and 3.2 and the proofs in Appendices A and B are briefly collected as follows:

$$
\begin{aligned}
H_{k_1k_2k_3}(t, z_t, v) &= E[\lambda^{k_1}(t, Z_{t\beta_0})Y^{k_2}(t)(\widetilde{Z}_t - \widetilde{z}_t)^{\otimes k_3}|Z_{t\beta} = v]f_{Z_{t\beta}}(v), k_1, k_2, k_3 = 0, 1, 2,\\
\Lambda^{[\ell]}(t, z_{H_t}, \beta) &= \sum_{\ell_1=0}^{\ell}(-1)^{\ell+\ell_1}\int_0^t \frac{H_{(1-\ell_1)(1-\ell+\ell_1)0}(u, z_u, z_{u\beta})(\partial_v H_{(1-\ell+\ell_1)(1-\ell_1)1}(u, z_u, z_{u\beta}))^{\ell}}{H_{010}^{1+\ell}(u, z_u, z_{u\beta}))}du,\\
S^{[\ell]}(t, Z_{H_t}, \beta) &= \exp(-\Lambda^{[0]}(t, z_{H_t}, \beta))(-\Lambda^{[1]}(t, z_{H_t}, \beta))^{\ell}, \quad \ell = 0, 1,\\
J_1(t, \nu) &= E[(I(Z_{\beta_0} \le \nu) - \mathcal{A}(t; \beta_0))(1 - S(t, Z, \beta_0))], J_2(t, \nu) = E[(I(Z_{\beta_0} > \nu) - \mathcal{A}(t; \beta_0))S(t, Z, \beta_0)],\\
\eta(t) &= E\left[(J_1(t, Z_{\beta_0}) - J_2(t, Z_{\beta_0}))S(t, Z_{\beta_0})\int_0^t(\Lambda^{[1]}(t, Z, \beta_0) - \Lambda^{[1]}(u, Z, \beta_0))d_uF_C(u|Z)\right],\\
\xi_0 &= -V_2^{-1}\int_0^{\tau}\{N(t; S, \beta_0) - (1 - S(t, Z, \beta_0))\}S^{[1]}(t, Z, \beta_0)dW(t),\\
\phi(t, z_{\beta_0}) &= \int_0^t \frac{Y(u)\lambda(u, z_{\beta_0})du}{H_{010}(u, z, z_{\beta_0})} - \frac{N(t)}{H_{010}(X, z, z_{\beta_0})},\\
\xi_0(t) &= \eta^{\top}(t)\xi_0, \quad \xi_1(t) = N(t; S, \beta_0)J_1(t, Z_{\beta_0}) + (1 - N(t; S, \beta_0))J_2(t, Z_{\beta_0}),\\
\xi_2(t) &= (J_1(t, Z_{\beta_0}) - J_2(t, Z_{\beta_0}))S(t, Z, \beta_0)\int_0^t(\phi(u, Z_{\beta_0}) - \phi(t, Z_{\beta_0}))d_uF_C(u, Z_{\beta_0}),\\
\zeta_1(t) &= E[\xi_1(t)|Z_{\beta_0}], \quad \zeta_2(t) = (J_1(t, Z_{\beta_0}) - J_2(t, Z_{\beta_0}))S(t, Z, \beta_0)\phi(t, Z_{\beta_0}),
\end{aligned}
$$

For notational clarity, we further define

$$\widehat{\Lambda}_h^{[0]}(t,z_{H_t},\beta) = \int_0^t \frac{d_u\widehat{H}_{1,h}(u,z_u,\beta)}{H_{010}(u,z_u,z_{u\beta})}$$

$$\text{and } \widehat{\Lambda}_h^{[1]}(t,z_{H_t},\beta) = \sum_{i=1}^n\sum_{\ell=0}^1(-1)^\ell\int_0^t \frac{\partial_\upsilon^\ell H_{01\ell}(u,z_u,z_{u\beta})\partial_\beta^{1-\ell}K_h(Z_{iu\beta}-z_{u\beta})}{nH_{010}^2(u,z_u,z_{u\beta})}du$$

as the surrogates for $\widehat{\Lambda}_h(t,z_{H_t},\beta)$ and $\partial_\beta\widehat{\Lambda}_h(t,z_{H_t},\beta)$, respectively. Moreover, let $H_0^{[k]}(t,z_t,\upsilon) = \partial_\upsilon^k H_{01k}(t,z_t,\upsilon)$ and $\bar{A}_{c,h}^{[k]}(\cdot) = \partial_\beta^k\bar{A}_h(\cdot) - A^{[k]}(\cdot)$ for any generic estimator $\bar{A}_h(\cdot)$ in which the target function of $\partial_\beta^k\bar{A}_h(\cdot)$ is denoted by $A^{[k]}(\cdot)$. Throughout the rest of this article, a possible value of $(z_t,\beta,t)$ is restricted in $\mathcal{Z}_t^{\delta_t}\times\mathcal{B}_n\times[0,\tau]$.

## Appendix A.

For the derivation of the main results, some technical lemmas are established in this appendix.

**Lemma 1.** *Under assumptions A1-A3,*

$$\sup_{(z_t,\beta,t)}\|\widehat{H}_{0c,h}^{[k]}(t,z_t,z_{t\beta})\| = o\left(\sqrt{\frac{\ln n}{nh^{2k+1}}}\right) + O(h^q) \ \ a.s., k=0,1,2. \tag{1}$$

*Proof.* Let $\mathcal{F}_k = \{cK_q^{(k)}(a_1^\top Z_0+b)(Z_0-a_2)^{\otimes k} : a_1,a_2\in\mathbb{R}^{d-1}, b,c\in\mathbb{R}\}$ and $\mathcal{F}_{\ell k} = \{K_q^{(k)}((Z_{t\beta}-z_{t\beta})/h)Y^\ell(t)(\widetilde{Z}_t-\widetilde{z}_t)^{\otimes k}/h^{1+k} : z_t\in\mathcal{Z}_t^{\delta_t},\beta\in\mathcal{B}_n,t\in[0,\tau]\}$, $\ell=0,1$, $k=0,1,2$. Apparently, $\mathcal{F}_{0k}$ has a smaller VC-dimension than $\mathcal{F}_k$ in which its VC-index is $(d+1)$. It is easy to conclude from Lemma 2.12 in Pakes and Pollard (1989) that $\mathcal{F}_{0k}$ is also Euclidean. Combining with the fact that $\{Y(t) : t\in[0,\tau]\}$ is a VC-class, the Euclidean class $\mathcal{F}_{1k}$ is further ascertained by an application of Lemma 2.14 in Pakes and Pollard (1989). Thus, the following property can be implied by Theorem II.37 in Pollard (1984):

$$\sup_{(z_t,\beta,t)}\|\partial_\beta^k\widehat{H}_{0,h}(t,z_t,\beta) - E[\partial_\beta^k\widehat{H}_{0,h}(t,z_t,\beta)]\| = O\left(\sqrt{\frac{\ln n}{nh^{2k+1}}}\right) \ \ a.s. \tag{2}$$

Moreover, a simple calculation yields that

$$E[\partial_\beta^k\widehat{H}_{0,h}(t,z_t,\beta)] = \partial_\upsilon^k H_{01k}(t,z_t,z_{t\beta}) + h\int_0^t \partial_\upsilon^{k+1}H_{01k}(u,z_u,z_{u\beta}^*)u^{k+1}K_{q,h}^{(k)}(u)du, \tag{3}$$

where $z_{t\beta}^*$ lies between $z_{t\beta}$ and $z_{t\beta}+hu$. By assumption A3 and integration by parts, one has

$$\begin{aligned}&\sup_{(z_t,\beta,t)}\|E[\partial_\beta^k\widehat{H}_{0,h}(t,z_t,\beta)] - \partial_\upsilon^k H_{01k}(t,z_t,z_{t\beta})\|\\ &= \sup_{(z_t,\beta,t)}\left\|\int_0^t(\partial_\upsilon^{k+1}H_{01k}(u,z_u,z_{u\beta}^*) - \partial_\upsilon^{k+1}H_{01k}(u,z_u,z_{u\beta}))hu^{k+1}K_{q,h}^{(k)}(u)du\right\| = O(h^q).\end{aligned} \tag{4}$$

Together with (2), (1) is directly obtained. □

**Lemma 2.** *Under assumptions A1-A4,*

$$\sup_{(z_{H_t},\beta,t)}\left|\widehat{\Lambda}^{[0]}_{c,h}(t,z_{H_t},\beta)-\left(\widehat{\Lambda}^{[0]}_{0c,h}(t,z_{H_t},\beta)-\int_0^t \widehat{H}^{[0]}_{0c,h}(u,z_u,z_{u\beta})\frac{H_{110}(u,z_u,z_{u\beta})}{H^2_{010}(u,z_u,z_{u\beta})}du\right)\right|=o_p\left(\frac{1}{\sqrt{n}}\right) \quad (5)$$

$$\textit{and} \sup_{(z_{H_t},\beta,t)}|\widehat{\Lambda}^{[\ell]}_{c,h}(t,z_{H_t},\beta)|=o\left(\sqrt{\frac{\ln n}{nh^{2\ell+1}}}\right)+O(h^q)\ \ a.s., \ell=0,1. \quad (6)$$

*Proof.* A simple application of Taylor expansion for $\widehat{\Lambda}_h(t,z_{H_t},\beta)$ gives

$$\widehat{\Lambda}_h(t,z_{H_t},\beta)=\widehat{\Lambda}^{[0]}_h(t,z_{H_t},\beta)-\int_0^t \widehat{H}^{[0]}_{0c,h}(u,z_u,z_{u\beta})\frac{d_u\widehat{H}_{1,h}(u,z_u,\beta)}{H^2_{010}(u,z_u,z_{u\beta})}+r_n(t,z_{H_t},\beta), \quad (7)$$

where $r_n(t,z_{H_t},\beta)=2\sum_{i=1}^n\int_0^t K_h(Z_{iu\beta}-z_{u\beta})\widehat{H}^{[0]2}_{0c,h}(u,z_u,z_{u\beta})dN_i(u)/(n\widehat{H}^{*3}_{0,h}(u,z_u,z_{u\beta}))$ with $\widehat{H}^*_{0,h}(u,z_u,z_{u\beta})$ lying on the line segment between $\widehat{H}_{0,h}(u,z_u,\beta)$ and $H_{010}(u,z_u,z_{u\beta})$. By assumptions A3 and A4, we can also derive that

$$\sup_{(z_{H_t},\beta,t)}\left\|\int_0^t \frac{\widehat{H}^{[0]}_{0c,h}(u,z_u,z_{u\beta})}{H^2_{010}(u,z_u,z_{u\beta})}[d_u\widehat{H}_{1,h}(u,z_{H_u},\beta)-H_{110}(u,z_u,z_{u\beta})du]\right\|=o_p\left(\frac{1}{\sqrt{n}}\right)$$

$$\text{and} \sup_{(z_{H_t},\beta,t)}|r_n(t,z_{H_t},\beta)|=o_p\left(\frac{1}{\sqrt{n}}\right). \quad (8)$$

By subtracting $\Lambda^{[0]}(t,z_{H_t},\beta)$ from both sides of (7), (5) is a direct result of (7)-(8). As for the proof of (6), it can be shown in a similar way to Lemma 1. □

**Lemma 3.** *Under assumptions A1-A4,*

$$\sup_{(z_{H_t},\beta,t)}\left|\widehat{S}^{[0]}_{c,h}(t,z_{H_t},\beta)+S^{[0]}(t,z_{H_t},\beta)\left(\widehat{\Lambda}^{[0]}_{c,h}(t,z_{H_t},\beta)-\int_0^t \widehat{H}^{[0]}_{0c,h}(u,z_u,z_{u\beta})\frac{H_{110}(u,z_u,z_{u\beta})}{H^2_{010}(u,z_u,z_{u\beta})}du\right)\right|=o_p\left(\frac{1}{\sqrt{n}}\right) \quad (9)$$

$$\textit{and} \sup_{(z_{H_t},\beta,t)}|\widehat{S}^{[\ell]}_{c,h}(t,z_{H_t},\beta)|=o\left(\sqrt{\frac{\ln n}{nh^{2\ell+1}}}\right)+O(h^q)\ \ a.s., \ell=0,1. \quad (10)$$

*Proof.* By substituting the kernel-weighted version of the Kaplan-Meier estimator into Theorem 3.2.3 in Fleming and Harrington (1991), we have

$$\frac{\widehat{S}_h(t,z_{H_t},\beta)}{S^{[0]}(t,z_{H_t},\beta)}=1-\int_0^t\frac{\widehat{S}_h(u-,z_{H_{u-}},\beta)}{S^{[0]}(u,z_{H_u},\beta)}d_u\widehat{\Lambda}^{[0]}_{c,h}(u,z_{H_u},\beta). \quad (11)$$

It follows from (11) that

$$\begin{aligned}\frac{\widehat{S}^{[0]}_{c,h}(t,z_{H_t},\beta)}{S^{[0]}(t,z_{H_t},\beta)} &= -\int_0^t\frac{S^{[0]}(u-,z_{H_{u-}},\beta)}{S^{[0]}(u,z_{H_u},\beta)}d_u\widehat{\Lambda}^{[0]}_{c,h}(u,z_{H_u},\beta)-\int_0^t\frac{S^{[0]}_{c,h}(u-,z_{H_{u-}},\beta)}{S^{[0]}(u,z_{H_u},\beta)}d_u\widehat{\Lambda}^{[0]}_{c,h}(u,z_{H_u},\beta)\\ &= -\widehat{\Lambda}^{[0]}_{c,h}(t,z_{H_t},\beta)+\int_0^t\int_0^{u-}\frac{\widehat{S}_h(v-,z_{H_{v-}},\beta)}{S^{[0]}(v,z_{H_v},\beta)}d_v\widehat{\Lambda}^{[0]}_{c,h}(v,z_{H_v},\beta)d_u\widehat{\Lambda}^{[0]}_{c,h}(u,z_{H_u},\beta). \quad (12)\end{aligned}$$

Coupled with (6) in Lemma 2, the following property can be derived by parallelling with the proof steps from line 11 of page 161 to line 22 of page 162 in Du and Akritas (2002):

$$\sup_{(z_{H_t},\beta,t)}\left|\int_0^t\int_0^{u-}\frac{\widehat{S}_h(v-,z_{H_{v-}},\beta)}{S^{[0]}(v,z_{H_v},\beta)}d_v\widehat{\Lambda}^{[0]}_{c,h}(v,z_{H_v},\beta)d_u\widehat{\Lambda}^{[0]}_{c,h}(u,z_{H_u},\beta)\right| = o_p\left(\frac{1}{\sqrt{n}}\right). \tag{13}$$

From (12)-(13), Lemma 1, and (5) in Lemma 2, (9) is obtained and the assertion in (10) holds for $\ell = 0$. A straightforward calculation and an application of Lemma 1 further yield that

$$\frac{\partial_\beta\widehat{S}_h(t,z_{H_t},\beta)}{\widehat{S}_h(t,z_{H_t},\beta)} = -\frac{1}{n}\sum_{i=1}^n\sum_{\ell=0}^1\int_0^t\frac{(-1)^\ell\partial_\beta^{1-\ell}K_{q,h}(Z_{iu\beta}-z_{u\beta})\partial_\beta^\ell\widehat{H}_{0,h}(u,z_u,\beta)dN_i(u)}{\widehat{H}_{0,h}(u,z_u,\beta)(\widehat{H}_{0,h}(u,z_u,\beta)-\frac{1}{n}\sum_{j=1}^n I(X_j=u)K_{q,h}(Z_{ju\beta}-z_{u\beta}))}$$

$$= \frac{1}{n}\sum_{i=1}^n\sum_{\ell_1,\ell_2\in\{0,1\}}(-1)^{1+\ell_1}\int_0^t\frac{\partial_\beta^{1-\ell_1}K_{q,h}(Z_{iu\beta}-z_{u\beta})\partial_\beta^{\ell_1}\widehat{H}_{0,h}(u,z_u,\beta)(2\widehat{H}^{[0]}_{0c,h}(u,z_u,z_{u\beta}))^{\ell_2}dN_i(u)}{H_{010}^{2+\ell_2}(u,z_u,z_{u\beta})} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

$$= -\Lambda^{[1]}(t,z_{H_t},\beta) - \widehat{\Lambda}^{[1]}_{c,h}(t,z_{H_t},\beta) + \sum_{\ell_1,\ell_2\in\{0,1\}}(-1)^{1+\ell_1}(1+\ell_2)\int_0^t\frac{\widehat{H}^{[1-\ell_2]}_{0c,h}(u,z_u,z_{u\beta})}{H_{010}^{2+\ell_2}(u,z_u,z_{u\beta})}(\partial_v^{\ell_1}H_{10\ell_1}(u,z_u,z_{u\beta}))^{1-\ell_2}$$

$$\cdot\,(H_{\ell_1(1-\ell_1)0}(u,z_u,z_{u\beta})\partial_v H_{(1-\ell_1)\ell_1 1}(u,z_u,z_{u\beta}))^{\ell_2}du + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{14}$$

uniformly in $(z_{H_t},\beta,t)$. Finally, the assertion in (10) is ascertained for $\ell = 1$ by mean of Lemma 1. □

## Appendix B.

**Proof of Theorem 1.** By assumption A5, (10) in Lemma 3, and applying the Taylor expansion technique, we have

$$PS_h(\beta) = PS^{[0]}(\beta) + r_n(\beta), \tag{15}$$

where $PS^{[0]}(\beta) = \mathbb{P}_n\xi(\beta)$ and $\sup_{\beta\in\mathcal{B}_n}\|r_n(\beta)\| = o_p(1)$. The Lipschitz continuity of $\xi(\beta)$ in $\beta$, which is ensured by assumption A3, further implies that the class $\{\xi(\beta):\beta\in\mathcal{B}_n\}$ is Euclidean (cf. Lemma 2.13 in Pakes and Pollard (1989)). Moreover, it follows from Lemma 2.8 in Pakes and Pollard (1989) that

$$\sup_{\beta\in\mathcal{B}_n}\|PS^{[0]}(\beta) - E[\xi(\beta)]\| = o_p(1). \tag{16}$$

Together with (15), one has

$$\sup_{\beta\in\mathcal{B}_n}\|PS_h(\beta) - E[\xi(\beta)]\| = o_p(1). \tag{17}$$

In view of assumptions A3 and A6, $\beta_0$ is the unique root of $E[\xi(\beta)] = 0$ locally. Thus, the consistency of $\widehat{\beta}$ can be ascertained by (17) and Theorem 2.10 in Kosorok (2008).

By the first-order Taylor expansion of $PS_1(\widehat{\beta})$ at $\beta_0$, it yields that

$$\sqrt{n}PS_h(\beta_0) + V_2[\sqrt{n}(\widehat{\beta} - \beta_0)] + R_n[\sqrt{n}(\widehat{\beta} - \beta_0)] = 0, \tag{18}$$

where $R_n = \partial_\beta PS_h(\widehat{\beta}^*) - V_2$ and $\widehat{\beta}^*$ lies on the line segment between $\widehat{\beta}$ and $\beta_0$. Further,

$$\begin{aligned}
1 &= P(-\sqrt{n}PS_h(\beta_0) = V_2[\sqrt{n}(\widehat{\beta} - \beta_0)] + R_n[\sqrt{n}(\widehat{\beta} - \beta_0)]) \\
&= P((I_{d-1} + V_2^{-1}R_n)[\sqrt{n}(\widehat{\beta} - \beta_0)] = -V_2^{-1}[\sqrt{n}PS_h(\beta_0)]) \\
&\leq P\left(\sqrt{n}(\widehat{\beta} - \beta_0) = -\left[\sum_{k=0}^{\infty}(V_2^{-1}R_n)^k\right] V_2^{-1}[\sqrt{n}PS_h(\beta_0)], \|V_2^{-1}R_n\| < \frac{1}{2}\right) + P\left(\|V_2^{-1}R_n\| \geq \frac{1}{2}\right) \\
&\leq P\left(\sqrt{n}\|\widehat{\beta} - \beta_0\| \leq \frac{\|V_2^{-1}\|\|\sqrt{n}PS_h(\beta_0)\|}{1 - \|V_2^{-1}R_n\|}, \|V_2^{-1}R_n\| < \frac{1}{2}\right) + P\left(\|V_2^{-1}R_n\| \geq \frac{1}{2}\right) \\
&\leq P\left(\|\sqrt{n}(\widehat{\beta} - \beta_0)\| \leq 2\|V_2^{-1}\|\|\sqrt{n}PS_h(\beta_0)\|, \|V_2^{-1}R_n\| < \frac{1}{2}\right) + P\left(\|V_2^{-1}R_n\| \geq \frac{1}{2}\right) \\
&\leq P\left(\|\sqrt{n}(\widehat{\beta} - \beta_0)\| \leq M\right) + P\left(M \leq 2\|V_2^{-1}\|\|\sqrt{n}PS_h(\beta_0)\|\right) + P\left(\|R_n\| \geq \frac{\|V_2\|}{2}\right) \quad \forall M > 0.
\end{aligned} \tag{19}$$

As a result, the pseudo score function can be decomposed as

$$\sqrt{n}(PS_h(\beta_0) - PS^{[0]}(\beta_0)) = \sum_{k=1}^{5} \frac{\sqrt{n}}{n(n-1)} \sum_{i \neq j} \int_0^\tau \varphi_{ki}(t)\psi_{kij}(t)dW_{in}(t) + \varphi_{6i}(t)d\eta_{ij}(t) + o_p(1), \tag{20}$$

where $\varphi_{ki}(t)$'s are some bounded smooth functions,

$$\begin{aligned}
\psi_{mij}(t) &= \partial_\beta^{m-1} K_{q,h}(Z_{jt\beta_0} - Z_{it\beta_0})Y_j(t) - H_0^{[m-1]}(t, Z_{it}, Z_{it\beta_0}), \quad m = 1, 2, \\
\psi_{3ij}(t) &= \int_0^t \frac{K_{q,h}(Z_{ju\beta_0} - Z_{iu\beta_0})dN_j(u)}{H_{010}(u, Z_{iu}, Z_{iu\beta_0})} - \Lambda^{[0]}(t, Z_{H_{it}}, \beta_0), \quad \psi_{4ij}(t) = I(X_i = t)K_{q,h}(Z_{jt\beta_0} - Z_{it\beta_0}), \\
\text{and } \psi_{5ij}(t) &= \sum_{\ell=0}^{1}(-1)^\ell \int_0^t \frac{H_0^{[\ell]}(u, Z_{iu}, Z_{iu\beta_0})\partial_\beta^{1-\ell} K_{q,h}(Z_{ju\beta_0} - Z_{iu\beta_0})dN_j(u)}{H_{010}^2(u, Z_{iu}, Z_{iu\beta_0})} - \Lambda^{[1]}(t, Z_{H_{it}}, \beta_0).
\end{aligned}$$

By $E[H_0^{[1]}(t, Z_t, Z_{t\beta_0})|Z_{t\beta_0}] = 0$, $E[\Lambda^{[1]}(t, Z_{H_t}, \beta_0)|Z_{t\beta_0}] = 0$, and Lemmas 1-2, it can be shown that $E[\varphi_{ki}(t)|Z_{iH_t}] = 0$ and $E[\psi_{kij}(t)|Z_{iH_t}] = O(h^q)$, $k = 1, \ldots, 5$. These facts together with $E[\varphi_{6i}(t)|Z_{iH_t}] = 0$ and assumption A5 assure that the summands in (20) are all $o_p(1)$ and, hence,

$$\sqrt{n}(PS_h(\beta_0) - PS^{[0]}(\beta_0)) = o_p(1). \tag{21}$$

An application of the multivariate central limit theorem to $PS^{[0]}(\beta_0)$ further enables us to have

$$\sqrt{n}PS_h(\beta_0) \xrightarrow{d} N(0, V_1). \tag{22}$$

By (10) in Lemma 3, there exists a continuous function $V(\beta)$ of $\beta$ such that $V(\beta_0) = V_2$ and

$$\sup_{\beta \in \mathcal{B}_n} \|\partial_\beta PS_h(\beta) - V(\beta)\| = o_p(1). \tag{23}$$

For any $\varepsilon > 0$, there also exists an integer $n_\varepsilon$ such that $\beta_0 \in \mathcal{B}_n$ and $\sup_{\beta\in\mathcal{B}_n} \|V(\beta) - V_2\| < \varepsilon/3$ whenever $n \geq n_\varepsilon$. Together with (18) and (23), we can thus derive that

$$\begin{aligned} &P\left(\sup_{\beta\in\mathcal{B}_n} \|\partial_\beta PS_h(\beta) - \partial_\beta PS_h(\beta_0)\| > \varepsilon\right) \\ &\leq P\left(\sup_{\beta\in\mathcal{B}_n} \|\partial_\beta PS_h(\beta) - V(\beta)\| > \frac{\varepsilon}{3}\right) + P\left(\|V(\beta_0) - \partial_\beta PS_h(\beta_0)\| > \frac{\varepsilon}{3}\right) \to 0 \text{ as } n \to \infty. \end{aligned} \tag{24}$$

Moreover, it is implied by (22) and (24) that

$$\liminf_{n\to\infty} P\left(\|V_2^{-1}\|\|\sqrt{n}PS_h(\beta_0)\| < \frac{M_\varepsilon}{2\|V_2^{-1}\|}\right) > 1 - \varepsilon \text{ for some } M_\varepsilon > 0 \tag{25}$$

and

$$\begin{aligned} P\left(\|R_n\| > \frac{\|V_2\|}{2}\right) &\leq P\left(\|\partial_\beta PS_h(\widehat{\beta}^*) - \partial_\beta PS_h(\beta_0)\| > \frac{\|V_2\|}{4}\right) + P\left(\|\partial_\beta PS_h(\beta_0) - V_2\| > \frac{\|V_2\|}{4}\right) \\ &\leq P\left(\sup_{\beta\in\mathcal{B}_{\frac{\|V_2\|}{2}}} \|\partial_\beta PS_h(\beta) - \partial_\beta PS_h(\beta_0)\| > \frac{\|V_2\|}{4}\right) + P\left(\widehat{\beta} \notin \mathcal{B}_{\frac{\|V_2\|}{2}}\right) \\ &\quad + P\left(\|\partial_\beta PS_h(\beta_0) - V_2\| > \frac{\|V_2\|}{4}\right) \to 0 \text{ as } n \to \infty. \end{aligned} \tag{26}$$

Substituting (25)-(26) into (19), one has

$$\liminf_{n\to\infty} P(\sqrt{n}\|\widehat{\beta} - \beta_0\| \leq M_\varepsilon) > 1 - \varepsilon, \text{ i.e. } \sqrt{n}\|\widehat{\beta} - \beta_0\| = O_p(1), \tag{27}$$

and (18) can be simplified as

$$\sqrt{n}(\widehat{\beta} - \beta_0) = -V_2^{-1}\sqrt{n}PS_h(\beta_0)(1 + o_p(1)). \tag{28}$$

From (22) and (28), the proof for the asymptotic normality of $\widehat{\beta}$ is completed.

**Proof of Theorem 2.** By the decomposition of $(\widetilde{S}_\varsigma(t, z_{H_t}, \widehat{\beta}) - S(t, z_{H_t}, \beta_0))$ and the first-order Taylor expansion of $\widetilde{S}_\varsigma(t, z_{H_t}, \widehat{\beta})$ with respect to $\beta_0$, one has

$$\begin{aligned} &\sqrt{n\varsigma}(\widetilde{S}_\varsigma(t, z_{H_t}, \widehat{\beta}) - S(t, z_{H_t}, \beta_0)) \\ &= \sqrt{\varsigma}\partial_\beta \widetilde{S}_\varsigma(t, z_{H_t}, \beta^*)[\sqrt{n}(\widehat{\beta} - \beta_0)] + \sqrt{n\varsigma}(\widetilde{S}_\varsigma(t, z_{H_t}, \beta_0) - S(t, z_{H_t}, \beta_0)) \end{aligned} \tag{29}$$

with $\beta^*$ lying between $\widehat{\beta}$ and $\beta_0$. It follows from Lemma 3 and (27) that (29) can be further simplified as

$$\sqrt{n\varsigma}(\widetilde{S}_\varsigma(t, z_{H_t}, \widehat{\beta}) - S(t, z_{H_t}, \beta_0)) = \sqrt{n\varsigma}\mathbb{P}_n\Psi_n(t, z_{H_t})(1 + \gamma_n(t, z_{H_t})), \tag{30}$$

where $\sup_{(z_{H_t},t)}|\gamma_n(t,z_{H_t})| = o_p(1)$. Further, the class $\{\varsigma\Psi_n(t,z_{H_t}) : z_t \in \mathcal{Z}_t^{\delta_t}, t \in [0,\tau]\}$ is easily shown to be Euclidean and satisfies the conditions UBV and LUBV in Rio (1994). By Corollary 1.1 in Rio (1994), there exists a sequence of centered Gaussian processes $G_n(t,z_{H_t})$ with continuous sample paths and $E[G_n(t_1,z_{H_{t_1}})G_n(t_2,z_{H_{t_2}})] = \varsigma E[\Psi_n(t_1,z_{H_{t_1}})\Psi_n(t_2,z_{H_{t_2}})]$ such that

$$\sup_{(z_{H_t},t)}|\sqrt{n\varsigma}(\mathbb{P}_n\Psi_n(t,z_{H_t}) - E[\Psi_n(t,z_{H_t})]) - G_n(t,z_{H_t})| = O\left(\sqrt{\frac{\ln n}{n^{(1-2\kappa_1)/2}}}\right) \ a.s. \tag{31}$$

Subtracting $(\sqrt{n\varsigma}E[\Psi_n(t,z_{H_t})] - G_n(t,z_{H_t}))$ from both sides of (30), the asymptotic Gaussian process of $\widetilde{S}_\varsigma(t,z,\widehat{\beta})$ is thus ascertained.

**Proof of Theorem 3.** From (3.2) in Section 3, $\bar{\mathcal{A}}(t;\widehat{\beta}) - \mathcal{A}(t;\beta_0)$ can be expressed as

$$\frac{1}{n(n-1)\bar{\mathcal{A}}_0(t;\widehat{\beta})}\sum_{i\neq j}(I(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}}) - \mathcal{A}(t;\beta_0))N_i(t;\widetilde{S}_\varrho,\widehat{\beta})(1 - N_j(t;\widetilde{S}_\varrho,\widehat{\beta})). \tag{32}$$

It follows from (10) in Lemma 3 and Theorem 1 that

$$\frac{\widetilde{S}_\varrho(t,z,\widehat{\beta}) - S(t,z,\beta_0)}{S(t,z,\beta_0)} = \frac{\widetilde{S}_\varrho(t,z,\beta_0) - S(t,z,\beta_0)}{S(t,z,\beta_0)} + \Lambda^{[1]}(t,z,\beta_0)(\widehat{\beta} - \beta_0) + r_{1n}(t,z) \tag{33}$$

with $\sup_{(z,t)}|r_{1n}(t,z)| = o_p(n^{-1/2})$. Together with

$$\sup_{(z_{\beta_0},t)}\left|\frac{\widetilde{S}_\varrho(t,z,\beta_0) - S(t,z,\beta_0)}{S(t,z,\beta_0)} - \frac{1}{n}\sum_{i=1}^n \phi_i(t,Z_{i\beta_0})K_{2,\varrho}(Z_{i\beta_0} - z_{\beta_0})\right| = o_p\left(\frac{1}{\sqrt{n}}\right), \tag{34}$$

which is a simplified version of Lemma 3, it can be derived that

$$\frac{1}{n(n-1)}\sum_{i\neq j}(I(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}}) - \mathcal{A}_t(\beta_0))N_i(t;\widetilde{S}_\varrho,\widehat{\beta})(1 - N_j(t;\widetilde{S}_\varrho,\widehat{\beta})) = \sum_{k=1}^2 U_{kn}(t;\widehat{\beta}) + (\widehat{\beta} - \beta_0)U_{3n}(t) + r_{2n}(t), \tag{35}$$

where

$$\sup_t |r_{2n}(t)| = o_p(n^{-1/2}), \quad U_{1n}(t;\beta) = \frac{1}{n(n-1)}\sum_{i\neq j}(I(Z_{i\beta} > Z_{j\beta}) - \mathcal{A}(t;\beta_0)N_i(t;S,\beta_0)(1 - N_j(t;S,\beta_0)),$$

$$U_{2n}(t;\beta) = \frac{1}{n^2(n-1)}\sum_{k=1}^n\sum_{i\neq j}B_{ij}(t;\beta)(1 - V_i(t))\frac{S(t,Z_i,\beta_0)}{S(X_i,Z_i,\beta_0)}[\phi_k(X_i,Z_{i\beta_0}) - \phi_k(t,Z_{i\beta_0})]K_{2,\varrho}(Z_{k\beta_0} - Z_{i\beta_0}),$$

$$\text{and } U_{3n}(t) = \frac{1}{n(n-1)}\sum_{i\neq j}B_{ij}(t;\widehat{\beta})(1 - V_i(t))\frac{S(t,Z_i,\beta_0)}{S(X_i,Z_i,\beta_0)}(\Lambda^{[1]}(X_i,Z_i,\beta_0) - \Lambda^{[1]}(t,Z_i,\beta_0))$$

with $B_{ij}(t;\beta) = (I(Z_{i\beta} > Z_{j\beta}) - \mathcal{A}(t;\beta_0))(1 - N_j(t;S,\beta_0)) - (I(Z_{j\beta} > Z_{i\beta}) - \mathcal{A}(t;\beta_0))N_j(t;S,\beta_0)$.

Substituting the kernels of the above U-statistics into the proof steps of Theorem 4 in Sherman (1993), we also have

$$U_{kn}(t;\beta) = U_{kn}(t;\beta_0) + \frac{1}{\sqrt{n}}(\beta - \beta_0)^\top\mathcal{N}_n(t) + \frac{1}{2}(\beta - \beta_0)^\top\mathcal{V}_t(\beta - \beta_0) + o_p(|\beta - \beta_0|^2) + o_p\left(\frac{1}{\sqrt{n}}\right), k = 1,2, \tag{36}$$

uniformly in $(\beta, t)$ with $\mathcal{N}_n(t)$ converging to a mean zero Gaussian process and $\mathcal{V}(t)$ being a non-degenerate matrix function of $t$, which is justified to be bounded by assumption A6. Thus, (35) can be simplified as

$$\frac{1}{n(n-1)}\sum_{i\neq j}(I(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}})-\mathcal{A}(t;\beta_0))N_i(t;\widetilde{S}_\varrho,\widehat{\beta})(1-N_j(t;\widetilde{S}_\varrho,\widehat{\beta})) = \sum_{k=1}^{2}U_{kn}(t;\beta_0)+(\widehat{\beta}-\beta_0)U_{3n}(t)+r_{3n}(t), \quad (37)$$

where $\sup_t |r_{3n}(t)| = o_p(n^{-1/2})$. Obviously, both $U_{1n}(t;\beta_0)$ and $U_{2n}(t;\beta_0)$ are centralized $U$-processes. By assumption A3, the class of each corresponding kernel function indexed by $t$ is Euclidean. It is implied by Corollary 4 in Sherman (1994) that

$$U_{1n}(t;\beta_0) = \frac{1}{n}\sum_{i=1}^{n}\xi_{1i}(t) + o_p(n^{-1/2}) \text{ and } U_{2n}(t;\beta_0) = \frac{1}{n}\sum_{i=1}^{n}\xi_{2i}(t) + O(\varrho^2) + o_p(n^{-1/2}) \quad (38)$$

uniformly in $t$. A straightforward calculation further leads to

$$\sup_t |U_{3n}(t) - \eta(t)| = o_p(1) \text{ and } \sup_t |\bar{\mathcal{A}}_0(t;\widehat{\beta}) - \mathcal{A}_0(t;\beta_0)| = o_p(1). \quad (39)$$

Combining (32), (37)-(39), and the *i.i.d.* approximation $\sum_{i=1}^{n}\xi_{0i}/n$ of $(\widehat{\beta}-\beta_0)$, the proof for the consistency of $\bar{\mathcal{A}}(t;\widehat{\beta})$ is thus completed. Moreover, the limiting Gaussian process of $\sqrt{n}(\bar{\mathcal{A}}(t;\widehat{\beta}) - \mathcal{A}(t;\beta_0))$ can be ascertained by applying the functional central limit theorem and the Slutsky's theorem. As for the uniform consistency and asymptotic Gaussian process of $\breve{\mathcal{A}}(t;\widehat{\beta})$, the proof is similar and is omitted in the interest of brevity.

# Supplementary Material (II)

Let $E_i(t;h) = N_i(t;S,\beta_0) - (1-\widehat{S}_h^{-i}(t,Z_{iH_t},\beta_0))$, $\mathrm{PSE}_i(t;h) = \widehat{S}_h(t,Z_{iH_t},\widehat{\beta}) - \widehat{S}_h(t,Z_{iH_t},\beta_0)$, and $\mathrm{SE}_i(t;h,\beta) = \widehat{S}_h(t,Z_{iH_t},\beta) - S(t,Z_{iH_t},\beta_0)$, $i=1,\ldots,n$. Moreover, $\mathrm{PSE}_i^{-i}(t;h)$ and $\mathrm{SE}_i^{-i}(t;h,\beta)$ are defined as $\mathrm{PSE}_i(t;h)$ and $\mathrm{SE}_i(t;h,\beta)$ with the $i$th unit being removed in computation.

## S1. Convergence Rate of $\widehat{h}$ for the PILSE

After some algebraic manipulations, $PS_{\widehat{\beta}}(h)$ can be decomposed into

$$\begin{aligned} PS_{\widehat{\beta}}(h) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} E_i(t;h)\partial_h\widehat{S}_h^{-i}(t,Z_{iH_t},\beta_0)dW_{in}(t) + \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} E_i(t;h)\partial_h\mathrm{PSE}_i^{-i}(t;h)dW_{in}(t) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{(N_i(t;\widehat{S}_h,\widehat{\beta}) - N_i(t;S,\beta_0)) + \mathrm{PSE}_i^{-i}(t;h)\}\partial_h\widehat{S}_h^{-i}(t,Z_{iH_t},\beta_0)dW_{in}(t) \\ &\triangleq I(h) + II(h) + III(h). \end{aligned} \quad (40)$$

Moreover, $I(h)$ can be re-expressed as

$$
\begin{aligned}
I(h) &= \partial_h \left( \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} (\mathrm{SE}_i^{-i}(t;h,\beta_0))^2 dW_{in}(t) + \frac{2}{n} \sum_{i=1}^{n} \int_0^{\tau} E_i(t;h)\mathrm{SE}_i^{-i}(t;h,\beta_0) dW_{in}(t) \right) \\
&\triangleq \partial_h (\widehat{\mathrm{MISE}}(h) + I_0(h)). \qquad (41)
\end{aligned}
$$

Let $\widehat{S}_{c,h}^{-i}(t, Z_{iH_t}, \beta_0) = \dfrac{-S(t, Z_{iH_t}, \beta_0)}{n} \sum_{j \neq i} \int_0^t \dfrac{K_{q,h}(Z_{ju\beta_0} - Z_{iu\beta_0})}{H_{010}(u, Z_{iu}, Z_{iu\beta_0})} (dN_j(u) - Y_j(u)\lambda(u, Z_{iu\beta_0})du)$,

$$
\widetilde{\mathrm{MISE}}(h) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} (\widehat{S}_{c,h}^{-i}(t, Z_{iH_t}, \beta_0))^2 dW_i(t), \text{ and } \widetilde{I}_0(h) = \frac{2}{n} \sum_{i=1}^{n} \int_0^{\tau} E_i(t;h) \widehat{S}_{c,h}^{-i}(t, Z_{iH_t}, \beta_0) dW_i(t).
$$

By assumption A5 and the *i.i.d.* approximation of $\widehat{S}_h(t, z_{H_t}, \beta_0)$ in Lemma 3, one has

$$
\sup_h \left| \frac{\widehat{\mathrm{MISE}}(h) - \widetilde{\mathrm{MISE}}(h)}{\widetilde{\mathrm{MISE}}(h)} \right| = o_p(1) \text{ and } \sup_h \left| \frac{I_0(h) - \widetilde{I}_0(h)}{\widetilde{I}_0(h)} \right| = o_p(1). \qquad (42)
$$

It is implied by Theorem 2 in Marron and Härdle (1986) that

$$
\sup_h \left| \frac{\widetilde{\mathrm{MISE}}(h) - \mathrm{AMISE}(h)}{\mathrm{AMISE}(h)} \right| = o_p(1), \qquad (43)
$$

where AMISE$(h)$ is the asymptotic mean integrated square error of $\widehat{S}_h(t, z_{H_t}, \beta_0)$ with

$$
\begin{aligned}
\mathrm{AMISE}(h) &= h^{2q} \int_0^{\tau} E[b^2(Z_{H_t}) dW(t)] + n^{-1}h^{-1} \int_0^{\tau} E[V(Z_{H_t}) dW(t)], \qquad (44) \\
b(z_{H_t}) &= \frac{-S(t, z_{H_t}, \beta_0)}{q!} \left( \int u^q K_q(u) du \right) \int_0^t \frac{\sum_{\ell=0}^{1} (-1)^{\ell} \partial_v^q H_{\ell 10}(u, Z_u, Z_{u\beta_0}) \lambda^{1-\ell}(u, z_{u\beta_0})}{H_{010}(u, z_u, z_{u\beta_0})} du, \\
\text{and } V(z_{H_t}) &= S^2(t, z_{H_t}, \beta_0) \left( \int K_q^2(u) du \right) \left\{ \int_0^t \frac{H_{110}(u, z_u, z_{u\beta_0})}{H_{010}^2(u, z_u, z_{u\beta_0})} du - (\int_0^t \lambda(u, z_{u\beta_0}) du)^2 \right\}.
\end{aligned}
$$

Coupled with (43), we have

$$
\widetilde{\mathrm{MISE}}(h) = O_p(h^{2q} + n^{-1}h^{-1}). \qquad (45)
$$

By Lemma 4 in Härdle and Marron (1985), the following property is directly obtained:

$$
\sup_h \left| \frac{\widetilde{I}_0(h)}{\widetilde{\mathrm{MISE}}(h)} \right| = o_p(1). \qquad (46)
$$

As a consequence of (42) and (45)-(46), the convergence rates of $I(h)$ in (40) can be derived as

$$
I(h) = O_p(h^{2q-1} + n^{-1}h^{-2}). \qquad (47)
$$

For the convergence rates of $II(h)$ and $III(h)$, an application of the Taylor expansion yields that

$$
\begin{aligned}
N_i(t; \widehat{S}_h, \widehat{\beta}) - N_i(t; S, \beta_0) &= \frac{(1 - V_i(t))}{S(X_i, Z_{iH_{X_i}}, \beta_0)} \left\{ \frac{S(t, Z_{iH_t}, \beta_0)}{S(X_i, Z_{iH_{X_i}}, \beta_0)} \mathrm{SE}_i(X_i; h, \widehat{\beta}) - \mathrm{SE}_i(t; h, \widehat{\beta}) \right\} \\
&\quad + r_{1n}(t, Z_{iH_t}) \mathrm{SE}_i^2(X_i; h, \widehat{\beta}) \qquad (48)
\end{aligned}
$$

and

$$\begin{aligned}\mathrm{SE}_i(t;h,\widehat{\beta}) &= \mathrm{PSE}_i(t;h)+\mathrm{SE}_i(t;h,\beta_0)\\ &= \partial_\beta \widehat{S}_h(t,Z_{iH_t},\beta_0)(\widehat{\beta}-\beta_0)(1+r_{2n}(t,Z_{iH_t})\|\widehat{\beta}-\beta_0\|)+\mathrm{SE}_i(t;h,\beta_0), \qquad (49)\end{aligned}$$

where $\sup_{(Z_{H_t},t)}|r_{kn}(t,Z_{H_t})|=O_p(1)$, $k=1,2$. From the proof of Theorem 1, one has

$$\widehat{\beta}-\beta_0=O_p(n^{-1/2})+O_p(h^{2q}+n^{-1/2}h^{1/2}+n^{-1}h^{-5/2}). \qquad (50)$$

Moreover, it is ensured by (A.10) in Lemma 3 that

$$\sup_{(Z_{H_t},t)}|\widehat{S}^{[1]}_{c,h}(t,z_{H_t},\beta_0)|=o_p(1) \text{ and } \sup_{(Z_{H_t},t)}|\mathrm{SE}_i(t;h,\beta_0)|=O_p(h+n^{-1/2}h^{-1/2}). \qquad (51)$$

Substituting (50)-(51) into (49), the convergence rates of $\mathrm{PSE}_i(t;h)$ and $\mathrm{SE}_i(t;h,\widehat{\beta})$ in (49) are shown to be

$$\mathrm{PSE}_i(t;h)=O_p(h^{2q}+n^{-1/2}h^{-1/2}+n^{-1}h^{-5/2}) \text{ and } \mathrm{SE}_i(t;h,\widehat{\beta})=O_p(h^{q}+n^{-1/2}h^{1/2}+n^{-1}h^{-5/2}), \qquad (52)$$

respectively, uniformly over $(Z_{H_t},t)$, which imply that $(N_i(t,\widehat{S}_h,\widehat{\beta})-N_i(t;S,\beta_0))$ in (48) satisfies

$$\sup_{(Z_{H_t},t)}|N_i(t,\widehat{S}_h,\widehat{\beta})-N_i(t;S,\beta_0)|=O_p(h^{q}+n^{-1/2}h^{1/2}+n^{-1}h^{-5/2}). \qquad (53)$$

By using the equality $\partial_h\widehat{S}^{-i}_h(t,Z_{iH_t},\beta_0)=\partial_h\mathrm{SE}^{-i}_i(t;h,\beta_0)$ and replacing $(\mathrm{SE}_i(t;h,\beta_0),\mathrm{PSE}_i(t;h))$ with $(\mathrm{SE}^{-i}_i(t;h,\beta_0),\mathrm{PSE}^{-i}_i(t;h))$ in (51)-(52), we can further derive that

$$II(h)=O_p(n^{-1/2}h^{2q-1}+n^{-1}h^{-1/2}+n^{-3/2}h^{-7/2}) \qquad (54)$$

$$\text{and} \quad III(h)=O_p(h^{6q-1}+n^{-1/2}h^{4q-1}+n^{-1}h^{2q-1}+n^{-2}h^{-2}+n^{-3/2}h^{2q-4}+n^{-3}h^{-15/2}+n^{-5/2}h^{-5}). \qquad (55)$$

From (47) and (54)-(55), $II(h)$ and $III(h)$ are thus shown to be negligible compared to $I(h)$. As a result, the chosen bandwidth $\widehat{h}$ has the convergence rate of $O_p(n^{-1/(1+2q)})$.

## S2. Optimality of $\widetilde{\varsigma}$ for the Survival Predictor

A simple application of the Cauchy-Schwartz inequality in $CV_1(\varsigma)$ leads to

$$\begin{aligned}CV_1(\varsigma)&=\left\{\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau E_i^2(t;\varsigma)dt+\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau(\mathrm{PSE}^{-i}_i(t;\varsigma))^2dt+\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau(N_i(t;\widehat{S}_{\widehat{h}},\widetilde{\beta})-N_i(t;S,\beta_0))^2dt\right\}\\ &\cdot(1+o_p(1))\triangleq CV_{11}(\varsigma)+CV_{12}(\varsigma)+CV_{13}(\varsigma). \qquad (56)\end{aligned}$$

Same with the derivation for $I(h)$ in **S1** with $W_{in}(t)=t$, we can obtain that

$$CV_{11}(\varsigma)=\mathrm{AMISE}(\varsigma)(1+o_p(1)) \text{ for } q=2. \qquad (57)$$

By using $\widehat{h} = O_p(n^{-1/9})$ in $\widehat{S}_{\widehat{h}}(t, z_{H_t}, \widehat{\beta})$ and replacing $PSE_i(t;h)$ with $PSE_i^{-i}(t;\varsigma)$ in (52), it is ensured that

$$CV_{12}(\varsigma) = O_p(n^{-1}). \tag{58}$$

Moreover, from (53) and $\widehat{S}_{\widehat{h}}(t, z_{H_t}, \widehat{\beta}) - S(t, z_{H_t}, \beta_0) = O_p(n^{-4/9})$, one has

$$CV_{13}(\varsigma) = O_p(n^{-8/9}). \tag{59}$$

Finally, substituting (57)-(59) into (56), it yields that

$$CV_1(\varsigma) \;=\; \text{AMISE}(\varsigma)(1 + o_p(1)) \tag{60}$$

$$\text{and, hence, } \widetilde{\varsigma} \;=\; \arg\min_{\varsigma} \text{AMISE}(\varsigma)(1 + o_p(1)) = \frac{\int_0^\tau E[V(Z_{H_t})]dt}{4\int_0^\tau E[b^2(Z_{H_t})]dt} n^{-1/5}(1 + o_p(1)) \text{ for } q = 2. \tag{61}$$

## S3. Asymptotic Properties of the PMLE

Since the bandwidth selection for $\bar{\beta}$ with the second-order kernel function is infeasible and the resulting PMLEs with the specifications of the second and fourth order kernel functions in $\ell_{h_{12}}(\beta)$ have the same asymptotic behavior, we focus on establishing the consistency and asymptotic normality of $\bar{\beta}$ with the fourth-order kernel function in $\ell_{h_{12}}(\beta)$. The following notations and assumptions are further introduced:

$$\lambda^{[k]}(t, z_t, \beta) = \partial_v^k H_{10k}(t, z_t, z_{t\beta}), \quad \ell^{[0]}(\beta) = \mathbb{P}_n(\delta \ln \lambda(X, Z_{X\beta_0}) + \ln S^{[0]}(X, Z_{H_X}, \beta)),$$

$$\ell^{[1]}(\beta) = \mathbb{P}_n\left(\frac{\lambda^{[1]}(X, Z_X, \beta_0)}{\lambda(X, Z_{X\beta_0})}\delta + \frac{S^{[1]}(X, Z_{H_X}, \beta_0)}{S(X, Z_{H_X}, \beta_0)}\right), \quad V_5 = E\left[\left(\frac{\lambda^{[1]}(X, Z_X, \beta_0)}{\lambda(X, Z_{X\beta_0})}\delta + \frac{S^{[1]}(X, Z_{H_X}, \beta_0)}{S(X, Z_{H_X}, \beta_0)}\right)^{\otimes 2}\right],$$

$$\text{and} \quad V_6 = E\left[\sum_{\ell=0}^{1}(-1)^\ell \frac{(\lambda^{[2-\ell]}(X, Z_X, \beta_0))^{\otimes(1+\ell)}}{\lambda^{1+\ell}(X, Z_{X\beta_0})}\delta + (-1)^\ell \frac{(S^{[2-\ell]}(X, Z_{H_X}, \beta_0))^{\otimes(1+\ell)}}{S^{1+\ell}(X, Z_{H_X}, \beta_0)}\right].$$

A4$'$. $h_k = h_{0k}n^{-\delta_2}$ for $\delta_2 \in (1/16, 1/6)$ and some positive constant $h_{0k}$, $k = 1, 2$.

A5$'$. $V_6$ is non-singular.

**Theorem S3.1.** Suppose that assumptions A1-A3 and A4$'$-A5$'$ are satisfied. Then,

$$\bar{\beta} \xrightarrow{p} \beta_0 \text{ and } \sqrt{n}(\bar{\beta} - \beta_0) \xrightarrow{d} N(0, V_6^{-1}V_5V_6^{-1}). \tag{62}$$

*Proof.* Let $\widehat{\lambda}^{[\ell]}_{c,h_{12}}(t, z_t, \beta) = \partial_\beta \widehat{\lambda}_{h_{12}}(t, z_{t\beta}) - \lambda^{[\ell]}(t, z_t, \beta)$, $\ell = 0, 1$, $h_{max} = \max\{h_1, h_2\}$, and $h_{min} = \min\{h_1, h_2\}$. It follows from Lemma 2 that

$$\sup_{(z_{H_t}, \beta, t)} \left|\widehat{\lambda}^{[0]}_{c,h_{12}}(t, z_{H_t}, \beta) - \int_0^\tau K_{q,h_2}(t-u)\left(d\widehat{\Lambda}^{[0]}_{0c,h_{12}}(t, z_{H_t}, \beta) - \widehat{H}^{[0]}_{0c,h}(u, z_u, z_{u\beta})\frac{H_{110}(u, z_u, z_{u\beta})}{H^2_{010}(u, z_u, z_{u\beta})}du\right)\right|$$

$$= o_p\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad \sup_{(z_{H_t}, \beta, t)} |\widehat{\lambda}^{[\ell]}_{c,h_{12}}(t, z_{H_t}, \beta)| = o\left(\sqrt{\frac{\ln n}{n h_{min}^{2\ell+2}}}\right) + O(h^4_{max}) \ a.s. \tag{63}$$

By the Taylor expansion technique, (A.10) in Lemma 3, and (63), one has

$$\sup_{\beta\in\mathcal{B}_n}|\ell_{h_{12}}(\beta)-\ell^{[0]}(\beta)|=o_p(1). \tag{64}$$

Further, the class $\{\delta\ln\lambda(X,Z_{X\beta})+\ln S^{[0]}(X,Z_{H_X},\beta):\beta\in\mathcal{B}_n\}$ can be shown to be Euclidean by means of assumption A3. It is further implied by Lemma 2.8 in Pakes and Pollard (1989) that

$$\sup_{\beta\in\mathcal{B}_n}|\ell^{[0]}(\beta)-E[\ell^{[0]}(\beta)]|=o_p(1). \tag{65}$$

Coupled with (64), we can conclude that

$$\sup_{\beta\in\mathcal{B}_n}|\ell_{h_{12}}(\beta)-E[\ell^{[0]}(\beta)]|=o_p(1). \tag{66}$$

In view of assumptions A3 and A5$'$, $\beta_0$ can be shown to be the unique maximizer of $E[\ell^{[0]}(\beta)]$ locally and, thus, the consistency of $\bar\beta$ to $\beta_0$ can be ascertained by Theorem 2.12 in Kosorok (2008).

An application of the first-order Taylor expansion of $\partial_\beta\ell(\bar\beta)$ at $\bar\beta=\beta_0$ yields that

$$\sqrt{n}\partial_\beta\ell_{h_{12}}(\beta_0)+\partial^2_\beta\ell_{h_{12}}(\beta^*)\sqrt{n}(\bar\beta-\beta_0)=0, \tag{67}$$

where $\beta^*$ lies on the line segment between $\bar\beta$ and $\beta_0$. The pseudo score function $\partial_\beta\ell(\beta_0)$ is further derived as

$$\sqrt{n}(\partial_\beta\ell_{h_{12}}(\beta_0)-\ell^{[1]}(\beta_0))=\sum_{m=1}^{7}\frac{\sqrt{n}}{n(n-1)}\sum_{i\neq j}\varphi_{mi}\psi_{mij}+o_p(1) \tag{68}$$

with

$$\psi_{6ij}=\int_0^\tau\frac{K_{q,h_1}(Z_{ju\beta_0}-Z_{iu\beta_0})K_{q,h_2}(X_i-u)dN_j(u)}{H_{010}(u,Z_{iu},Z_{iu\beta_0})}-\lambda(X_i,Z_{iX_i\beta_0}),$$

$$\psi_{7ij}=\sum_{\ell=0}^{1}(-1)^\ell\int_0^\tau\frac{\partial_v^\ell H_{01\ell}(X_i,Z_{iX_i},Z_{iX_i\beta_0})\partial_\beta^{1-\ell}K_{q,h_1}(Z_{ju\beta_0}-Z_{iu\beta_0})K_{q,h_2}(X_i-u)dN_j(u)}{H^2_{010}(X_i,Z_{iX_i},Z_{iX_i\beta_0})}-\lambda^{[1]}(X_i,Z_{iX_i},\beta_0).$$

After some algebraic manipulations, $E[\varphi_{mi}|Z_{iH_{X_i}}]=0$ and $E[\psi_{mij}|Z_{iH_{X_i}}]=O(h^2_{max})$, $m=1,\ldots,7$, can be obtained. These facts together with assumption A4$'$ enable us to simplify (68) as

$$\sqrt{n}(\partial_\beta\ell_{h_{12}}(\beta_0)-\ell^{[1]}(\beta_0))=o_p(1). \tag{69}$$

It is ensured by an application of the multivariate central limit theorem to $\sqrt{n}\ell^{[1]}(\beta_0)$ that

$$\sqrt{n}\partial_\beta\ell_{h_{12}}(\beta_0)\xrightarrow{d}N(0,V_5). \tag{70}$$

From (A.10) in Lemma 3 and (63), there exists a continuous function $V(\beta)$ of $\beta$ such that

$$\sup_{\beta\in\mathcal{B}_n}\|\partial^2_\beta\ell_{h_{12}}(\beta)-V(\beta)\|=o_p(1)\text{ and }V(\beta_0)=V_6. \tag{71}$$

Moreover, the argument for (27)) can be used to show that

$$\sqrt{n}(\bar\beta-\beta_0)=O_p(1). \tag{72}$$

Substituting (70)-(72) into (67), the asymptotic normality of $\bar\beta$ is thus ascertained. □

## S4. Convergence Rate of $\widetilde{\varrho}$ for the AUC Estimator

It follows from a simple decomposition that

$$CV_2(\varrho) = CV_{21} + CV_{22}(\varrho) + CV_{23}(\varrho), \tag{73}$$

where

$$CV_{21} = \frac{1}{n(n-1)} \sum_{i \neq j} \int_0^\tau (I(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}}) N_i(t; \widehat{S}_{\widehat{h}}, \widehat{\beta})(1 - N_j(t; \widehat{S}_{\widehat{h}}, \widehat{\beta})) - \mathcal{A}_1(t; \beta_0))^2 dt,$$

$$CV_{22}(\varrho) = \frac{1}{n(n-1)} \sum_{i \neq j} \int_0^\tau (\mathcal{A}_1(t; \beta_0) - \bar{\mathcal{A}}_1^{-ij}(t; \widehat{\beta})^2 dt,$$

$$CV_{23}(\varrho) = \frac{1}{n(n-1)} \sum_{i \neq j} \int_0^\tau (I(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}}) N_i(t; \widehat{S}_{\widehat{h}}, \widehat{\beta})(1 - N_j(t; \widehat{S}_{\widehat{h}}, \widehat{\beta})) - \mathcal{A}_1(t; \beta_0))(\mathcal{A}_1(t; \beta_0) - \bar{\mathcal{A}}_1^{-ij}(t; \widehat{\beta}) dt.$$

Clearly, the first term $CV_{21}$ on the right-hand side of (73) is independent of $\varrho$. From the proof of Theorem 4, one immediately has

$$CV_{22}(\varrho) = O_p(\varrho^4 + (n\varrho)^{-3/4}). \tag{74}$$

The $\sqrt{n}$-consistency of $\widehat{\beta}$, (53), and (B.29) further imply that

$$\begin{aligned} & CV_{23}(\varrho)(1 + o_p(1)) \\ &= \frac{1}{n(n-1)} \sum_{i \neq j} \int_0^\tau (I(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}}) N_i(t; \widehat{S}_{\widehat{h}}, \widehat{\beta})(1 - N_j(t; \widehat{S}_{\widehat{h}}, \widehat{\beta})) - \mathcal{A}_1(t; \beta_0))(\mathcal{A}_1(t; \beta_0) - \bar{\mathcal{A}}_1^{-ij}(t; \beta_0) dt \\ &= \frac{1}{n(n-1)} \sum_{i \neq j} \int_0^\tau (I(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}}) N_i(t; , S, \beta_0)(1 - N_j(t; S, \beta_0)) - \mathcal{A}_1(t; \beta_0))(\mathcal{A}_1(t; \beta_0) - \bar{\mathcal{A}}_1^{-ij}(t; \beta_0) dt. \end{aligned} \tag{75}$$

Moreover, the dominant term of $CV_{23}(\varrho)$ in (75) can be expressed as the fourth-order U-statistic

$$\frac{1}{n^2(n-1)^2} \sum_{i_1 \neq j_1} \sum_{i_2 \neq j_2} \kappa_{i_1 j_1 i_2 j_2}(\varrho)$$

with $\kappa_{i_1 j_1 i_2 j_2}(\varrho) = \int_0^\tau (I(Z_{i_1\beta_0} > Z_{j_1\beta_0}) N_{i_1}(t; S, \beta_0)(1 - N_{j_1}(t; S, \beta_0)) - \mathcal{A}_1(t; \beta_0))(\mathcal{A}_1(t; \beta_0) - I(Z_{i_2\beta_0} > Z_{j_2\beta_0})\ N_{i_2}^{-i_1}(t; S, \beta_0)(1 - N_{j_2}^{-j_1}(t; S, \beta_0)))dt$. After some tedious calculations, we get

$$\frac{1}{n} \sum_{i_1=1}^n E[\kappa_{i_1 j_1 i_2 j_2}(\varrho) | X_{i_1}, \delta_{i_1}, Z_{i_1}] = O_p(\varrho^2 n^{-1/2} + n^{-1}\varrho^{-1/2}), \quad \frac{1}{n} \sum_{i_2=1}^n E[\kappa_{i_1 j_1 i_2 j_2}(\varrho) | X_{i_2}, \delta_{i_2}, Z_{i_2}] = 0,$$

$$\frac{1}{n} \sum_{j_1=1}^n E[\kappa_{i_1 j_1 i_2 j_2}(\varrho) | X_{j_1}, \delta_{j_1}, Z_{j_1}] = O_p(\varrho^2 n^{-1/2} + n^{-1}\varrho^{-1/2}), \text{ and } \frac{1}{n} \sum_{j_2=1}^n E[\kappa_{i_1 j_1 i_2 j_2}(\varrho) | X_{j_2}, \delta_{j_2}, Z_{j_2}] = 0.$$

The above properties imply that the first-order Hájek projection of the U-statistic is $O_p(n^{-1/2}\varrho^2 + n^{-1}\varrho^{-1/2})$. By applying the maximal inequality in Sherman (1994), it is further ensured that the difference between the U-statistic and its first-order Hájek projection is $O_p(n^{-1}\varrho^{-1})$. Combining these results, the convergence rate of $CV_{23}(\varrho)$ is obtained as follows:

$$CV_{23}(\varrho) = O_p(\varrho^2 n^{-1/2} + n^{-1}\varrho^{-1}). \tag{76}$$

Substituting (74) and (76) into (73), one also has

$$CV_2(\varrho) = T_n + O_p(\varrho^2 + (n\varrho)^{-3/4}) \tag{77}$$

with $T_n$ being independent of $\varrho$. Thus, the minimizer $\widetilde{\varrho}$ is $O_p(n^{-3/11})$ and the proof is completed.

## REFERENCES


Du, Y. and Akritas, M. G. (2002). Uniform strong representation of the conditional Kaplan- Meier process. *Mathematical Methods of Statistics.* **11**, 152-182.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis.* Wiley, New York.

Härdle, W. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Annals of Statistics.* **13**, 1465-1481.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer, New York.

Marron, J. S. and Härdle, W. (1986). Random approximations to some measures of accuracy in nonparametric curve estimation. *Journal of Multivariate Analysis.* **20**, 91-113.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica.* **57**, 1027-1057.

Pollard, D. (1984). *Convergence of stochastic processes.* Springer, New York.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica.* **61**, 123-137.

Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Annals of Statistics.* **22**, 439-459.

Rio, E. (1994). Local invariance principles and their application to density estimation. *Probability Theory and Related Fields.* **98**, 21-45.

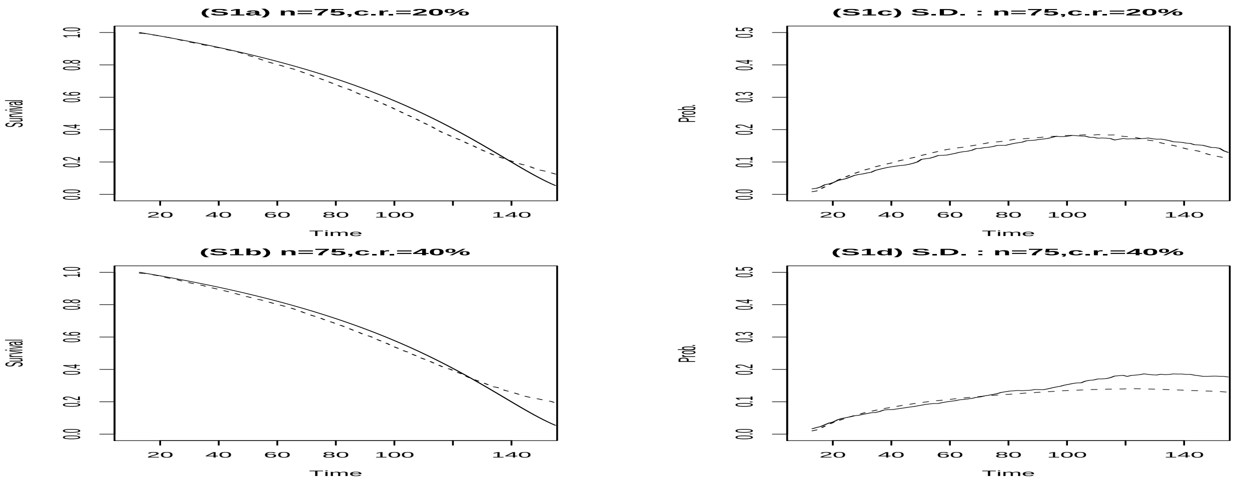


Figure S.1: (S1a)-(S1b) The survival functions (solid curves) conditioning on $Z_{\beta_0} = SI_{0.25}$ and the estimated conditional survival functions (dashed curve). (S1c)-(S1d) The standard deviation curves (solid curves) and the bootstrap standard error curves (dashed curves) of the estimated conditional survival functions.

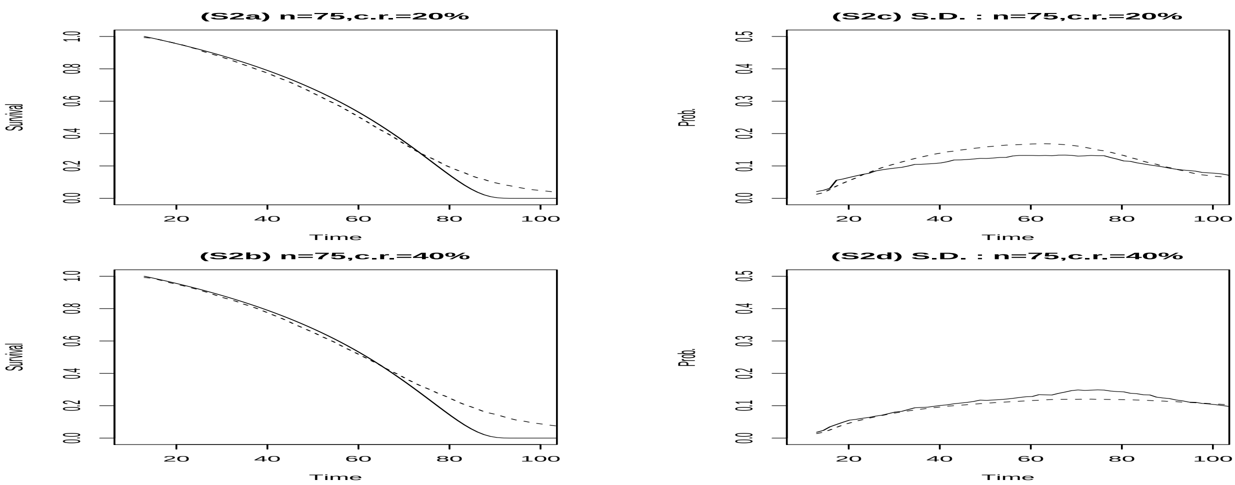


Figure S.2: (S2a)-(S2b) The survival functions (solid curves) conditioning on $Z_{\beta_0} = SI_{0.5}$ and the estimated conditional survival functions (dashed curves). (S2c)-(S2d) The standard deviation curves (solid curves) and the bootstrap standard error curves (dashed curves) of the estimated conditional survival functions.

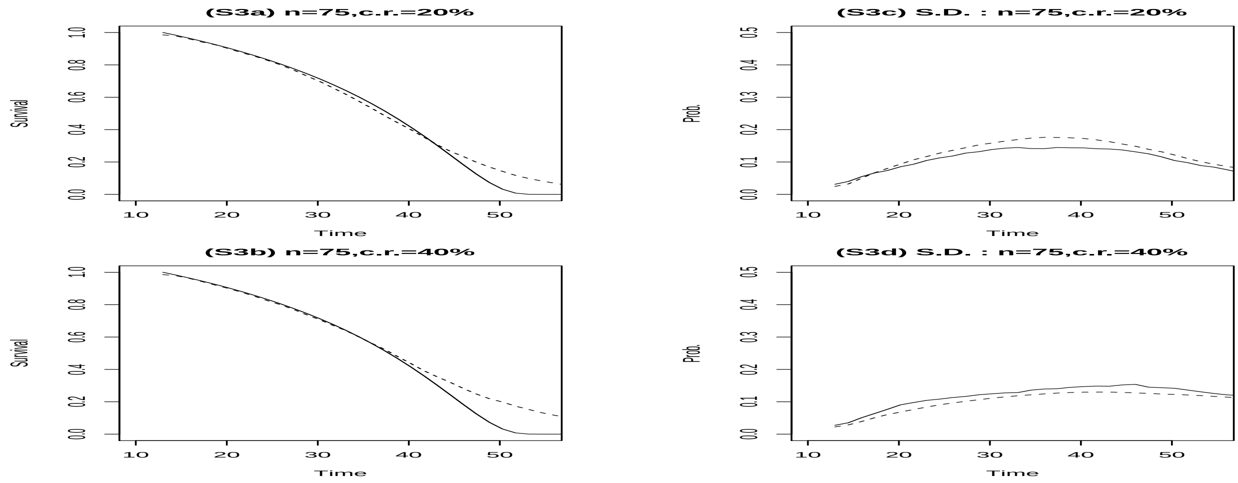


Figure S.3: (S3a)-(S3b) The survival functions (solid curves) conditioning on $Z_{\beta_0} = SI_{0.75}$ and the estimated conditional survival functions (dashed curves). (S3c)-(S3d) The standard deviation curves (solid curves) and the bootstrap standard error curves (dashed curves) of the estimated conditional survival functions.

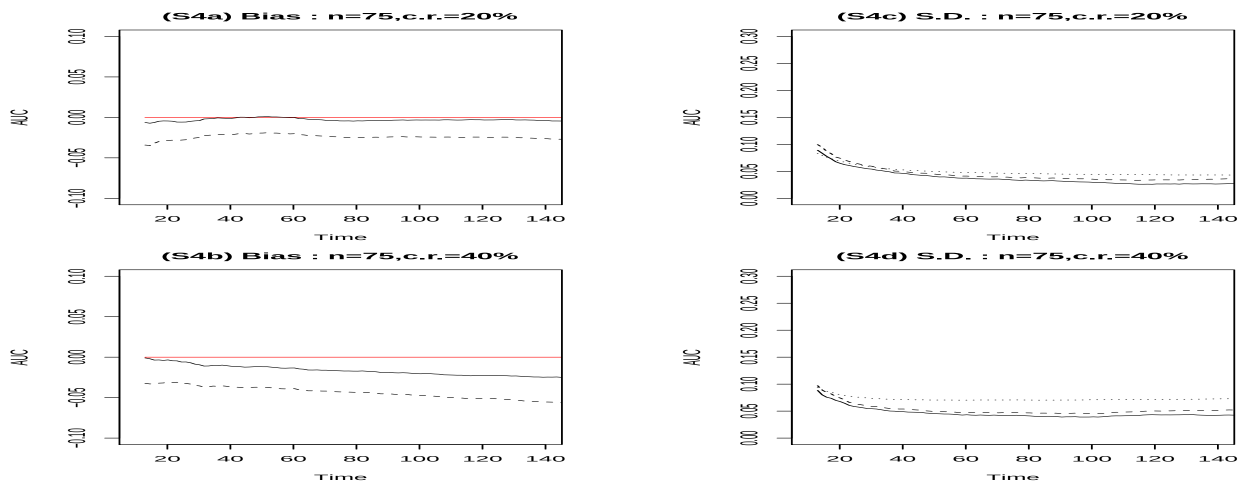


Figure S.4: (S4a)-(S4b) The bias curves (solid and dashed curves) of $\bar{\mathcal{A}}_{\varsigma t}(\widetilde{\beta}_h)$ and $\breve{\mathcal{A}}_{\varsigma t}(\widetilde{\beta}_h)$ with a horizontal line at zero (grey line). (S4c)-(S4d) The standard deviation curves (solid curve and dotted curves) of $\bar{\mathcal{A}}_{\varsigma t}(\widetilde{\beta}_h)$ and $\breve{\mathcal{A}}_{\varsigma t}(\widetilde{\beta}_h)$ and the bootstrap standard error curves (dotted-dashed curves) of $\bar{\mathcal{A}}_{\varsigma t}(\widetilde{\beta}_h)$.