Efficient Estimation for Dimension Reduction with Censored Survival Data

Ge Zhao, Yanyuan Ma, Wenbin Lu

Portland State University, Penn State University, North Carolina State University

Supplementary Material

S1 Proof of Proposition 1

The result of Γ_1 is obvious. To obtain Γ_2, let $\mathbf{h}(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \boldsymbol{\gamma}) = \partial \log \lambda_0(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}$, where $\lambda_0(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \boldsymbol{\gamma})$ is a submodel of $\lambda_0(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$. Hence,

$$\begin{aligned}
\frac{\partial \log f(\mathbf{X}, Z, \Delta)}{\partial \boldsymbol{\gamma}} &= \Delta \frac{\partial \log \lambda_0(Z, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \boldsymbol{\gamma})}{\boldsymbol{\gamma}} - \int_0^Z \frac{\partial \lambda_0(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} ds \\
&= \Delta \mathbf{h}(Z, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \boldsymbol{\gamma}) - \int_0^Z \mathbf{h}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \boldsymbol{\gamma}) \lambda_0(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) ds \\
&= \int_0^\infty \mathbf{h}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \boldsymbol{\gamma}) dM(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}).
\end{aligned}$$

Because $\lambda_0(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$ can be any positive function, $\mathbf{h}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \boldsymbol{\gamma})$ can be any function. We denote it $\mathbf{h}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$. This leads to the form of Γ_2.

Similar derivation leads to Γ_3. Specifically, to obtain Γ_3, let $\mathbf{h}(t, \mathbf{X}, \boldsymbol{\gamma}) = \partial \log \lambda_c(t, \mathbf{X}, \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}$, where $\lambda_c(t,^{\mathrm{T}} \mathbf{X}, \boldsymbol{\gamma})$ is a submodel of $\lambda_c(t, \mathbf{X})$. Hence,

$$\begin{aligned}
\frac{\partial \log f(\mathbf{X}, Z, \Delta)}{\partial \boldsymbol{\gamma}} &= (1-\Delta) \frac{\partial \log \lambda_c(Z, \mathbf{X}, \boldsymbol{\gamma})}{\boldsymbol{\gamma}} - \int_0^Z \frac{\partial \lambda_c(s, \mathbf{X}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} ds \\
&= (1-\Delta) \mathbf{h}(Z, \mathbf{X}, \boldsymbol{\gamma}) - \int_0^Z \mathbf{h}(s, \mathbf{X}, \boldsymbol{\gamma}) \lambda_c(s, \mathbf{X}) ds
\end{aligned}$$

$$= \int_0^\infty \mathbf{h}(s, \mathbf{X}, \boldsymbol{\gamma}) dM_c(s, \mathbf{X}).$$

Note that here, $M_c(t, \mathbf{X}) = N_c(t) - \int_0^t I(Z \geq s)\lambda_c(s, \mathbf{X})ds$, and despite of the discontinuity at $s = \tau$ for $\lambda_c(s, \mathbf{X})$, is still a martingale process (See Theorem 1.3.2 in Fleming and Harrington (1991)). A similar result was also established by Prentice and Kalbfleisch (2003) for a mixed discrete and continuous Cox regression model. Because $\lambda_c(t, \mathbf{X})$ can be any positive function, $\mathbf{h}(s, \mathbf{X}, \boldsymbol{\gamma})$ can be any function. We denote it $\mathbf{h}(s, \mathbf{X})$. This leads to the form of Γ_3.

It is easy to verify that $\Gamma_1 \perp \Gamma_2$ and $\Gamma_1 \perp \Gamma_3$, where $\perp$ stands for orthogonality. Because $C \perp\!\!\!\perp T \mid \mathbf{X}$, the martingale integrations associated with $M(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$ and $M_C(z, \mathbf{X})$ are also independent conditional on $\mathbf{X}$, hence $\Gamma_2 \perp \Gamma_3$. This completes the proof. $\square$

S2 Proof of Proposition 2

Denoting the score function in (2.2) at the true coefficient $\boldsymbol{\beta}_0$ as $\mathbf{S}_{\boldsymbol{\beta}_0}(\Delta, Z, \mathbf{X})$, we can verify that $\mathbf{S}_{\boldsymbol{\beta}_0}(\Delta, Z, \mathbf{X}) \perp \Gamma_1$ and $\mathbf{S}_{\boldsymbol{\beta}_0}(\Delta, Z, \mathbf{X}) \perp \Gamma_3$ due to the martingale properties. Thus to look for the efficient score, we only need to project $\mathbf{S}_{\boldsymbol{\beta}}(\Delta, Z, \mathbf{X})$ onto Γ_2 and calculate its residual. We search for $\mathbf{h}^*(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$ so that

$$\begin{aligned}
\mathbf{S}_{\text{eff}}(\Delta, Z, \mathbf{X}) &= \mathbf{S}_{\boldsymbol{\beta}_0}(\Delta, Z, \mathbf{X}) - \int_0^\infty \mathbf{h}^*(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) dM(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) \\
&= \int_0^\infty \left\{ \frac{\boldsymbol{\lambda}_{10}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \mathbf{X}_l - \mathbf{h}^*(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) \right\} dM(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})
\end{aligned}$$

is orthogonal to Γ_2. This entails that for any $\mathbf{h}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$,

$$\begin{aligned}
0 &= E\left[\int_0^\infty \mathbf{h}^{\mathrm{T}}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) dM(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) \int_0^\infty \left\{ \frac{\boldsymbol{\lambda}_{10}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \mathbf{X}_l - \mathbf{h}^*(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) \right\} \right. \\
&\quad \left. \times\, dM(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) \right] \\
&= E\left[\int_0^\infty \mathbf{h}^{\mathrm{T}}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) \left\{ \frac{\boldsymbol{\lambda}_{10}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \mathbf{X}_l - \mathbf{h}^*(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) \right\} Y(s) \lambda_0(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) ds \right].
\end{aligned}$$

By letting $\mathbf{h}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) = I(s = t)\mathbf{a}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$ for any $\mathbf{a}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$, we obtain that

$$\begin{aligned} 0 &= E\left[\left\{\frac{\boldsymbol{\lambda}_{10}(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \mathbf{X}_l - \mathbf{h}^*(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})\right\} Y(t)\lambda_0(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right] \\ &= E\left[\left\{\frac{\boldsymbol{\lambda}_{10}(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \mathbf{X}_l - \mathbf{h}^*(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})\right\} Y(t) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right]. \end{aligned}$$

Note that

$$\frac{E\left\{\mathbf{X}_l Y(t) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{Y(t) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}} = \frac{E\left\{\mathbf{X}_l S_c(t, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{S_c(t, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}$$

on $[0, \tau)$, and we simply set the ratio to

$$\frac{E\left\{\mathbf{X}_l S_c(t, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{S_c(t, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}} \equiv \frac{E\left\{\mathbf{X}_l S_c(\tau-, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{S_c(\tau-, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}$$

for $t \geq \tau$ so the relation hold on $[0, \infty)$. Note that the ratio can be defined as any function of $\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}$ for $t > \tau$ and it will not affect the following result because $dM(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) = 0$ for any $t > \tau$. This leads to

$$\mathbf{h}^*(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) = \frac{\boldsymbol{\lambda}_{10}(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0(t, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \frac{E\left\{\mathbf{X}_l S_c(t, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{S_c(t, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}.$$

Hence, the efficient score is

$$\mathbf{S}_{\text{eff}}(\Delta, Z, \mathbf{X}) = \int_0^\infty \frac{\boldsymbol{\lambda}_{10}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \left[\mathbf{X}_l - \frac{E\left\{\mathbf{X}_l S_c(s, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{S_c(s, \mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}\right] dM(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}).$$

□

S3 Proof of Lemma 1

For notation convenience, we prove the results for $d = 1$ and assume the first component of $\boldsymbol{\beta}$ is 1. We first establish the pointwise convergence results. The first four bias and variance results are obtained from the convergence property of the kernel estimation (Mack and Silverman, 1982; Einmahl and Mason, 2005) under conditions C1-C2. Specifically, to derive the first four results, we first establish the following preliminary conclusion for any $\mathbf{X}$ and $\boldsymbol{\beta}$ in a local neighborhood of $\boldsymbol{\beta}_0$,

$$\frac{1}{n}\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) = f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + O_p(n^{-1/2}h^{-1/2} + h^2), \quad \text{(S3.1)}$$

$$-\frac{1}{n}\sum_{j=1}^{n} K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) = f'_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + O_p(n^{-1/2}h^{-3/2} + h^2). \quad \text{(S3.2)}$$

To see this, we compute the absolute bias of the left hand side of (S3.1) as

$$\begin{aligned}
& \left| E\left\{\frac{1}{n}\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\} - f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right| \\
= & \left| \int \frac{1}{h} K\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}{h}\right) f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j) d\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j - f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right| \\
= & \left| \int K(u) f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu) du - f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right| \\
= & \left| \int K(u)\left\{ f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + f'_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) hu + \frac{1}{2} f''_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*) h^2 u^2 \right\} du \right. \\
& \left. - f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right| \\
\leq & \frac{h^2}{2} \sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}} |f''_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})| \int u^2 K(u) du,
\end{aligned}$$

where throughout the text, $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*$ is on the line connecting $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}+hu$, and the variance to be

$$\begin{aligned}
& \operatorname{var}\left\{\frac{1}{n}\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\} \\
= & \frac{1}{n} \operatorname{var} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \\
= & \frac{1}{n}\left[E K_h^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) - \left\{ E K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\}^2 \right] \\
= & \frac{1}{n}\left[\int \frac{1}{h^2} K^2\{(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})/h\} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j) d\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j - f^2_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + O(h^2)\right] \\
= & \frac{1}{nh}\int K^2(u) f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu) du - \frac{1}{n} f^2_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + O(h^2/n) \\
\leq & \frac{1}{nh} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \int K^2(u) du + \frac{h}{2n} \sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} |f''_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})| \int u^2 K^2(u) du \\
& \times + \frac{1}{n}|f^2_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})| + O(h^2/n).
\end{aligned}$$

Therefore, we have

$$\frac{1}{n}\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) = f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + O_p(n^{-1/2}h^{-1/2} + h^2)$$

for all $\boldsymbol{\beta}$ and for all h that satisfies Condition C2, under conditions C1, C2 and C4. Note that Condition C4 also holds for any $\boldsymbol{\beta}$ in a local neighborhood of $\boldsymbol{\beta}_0$ due to the continuity. The proof of the pointwise result related to (S3.2) is similar to that of (S3.1), hence we skip it.

Next we prove bias and variance related to (3.8) and skip (3.5), (3.6) and (3.7) because their proofs are similar. To this end, we analyze the absolute bias and variance of $\partial\widehat{E}\left\{\mathbf{X}Y(Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}\right\}/\partial\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$, and combine these to obtain (3.8). We have

$$\begin{aligned}
& E\left\{-\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j I(Z_j \geq Z)K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\} \\
= \; & E\left\{-\mathbf{X}_j I(Z_j \geq Z)K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\} \\
= \; & -\frac{1}{h}\int(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu, \mathbf{x}_{jl}^{\mathrm{T}})^{\mathrm{T}}K'(u)S_c(Z, \mathbf{x}_{jl})S(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu, \boldsymbol{\beta}) \\
& \times f_{\mathbf{X}_l|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}}(\mathbf{x}_{jl} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu, \boldsymbol{\beta})f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu)d\mathbf{x}_{jl}du \\
= \; & \int\frac{\partial}{\partial\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}\left\{(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \mathbf{x}_{jl}^{\mathrm{T}})^{\mathrm{T}}S_c(Z, \mathbf{x}_{jl})S(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})f_{\mathbf{X}_l|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}}(\mathbf{x}_{jl} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})\right. \\
& \left.\times f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\}d\mathbf{x}_{jl} \\
& -\frac{h^2}{6}\int\int\frac{\partial^3}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^3}\left\{(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \mathbf{x}_{jl}^{\mathrm{T}})^{\mathrm{T}}S_c(Z, \mathbf{x}_{jl})S(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*, \boldsymbol{\beta})\right. \\
& \left.\times\, f_{\mathbf{X}_l|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}}(\mathbf{x}_{jl} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*, \boldsymbol{\beta})f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*)\right\}d\mathbf{x}_{jl}u^3K'(u)du \\
= \; & \frac{\partial}{\partial\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} \\
& -\frac{h^2}{6}\int\frac{\partial^3}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^3}f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*)E\left\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*\right\}u^3K'(u)du.
\end{aligned}$$

Note that the third variable $\boldsymbol{\beta}$ in $f_{\mathbf{X}_l|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}}(\cdot)$ and $S(\cdot)$ indicates that the functional forms differ as $\boldsymbol{\beta}$ changes. Hence, the absolute bias is

$$\left|E\left\{-\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j I(Z_j \geq Z)K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\}\right.$$

$$
\begin{aligned}
& -\frac{\partial}{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}\Big| \\
= & \left| -\frac{h^2}{6} \int \frac{\partial^3}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^3} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*) E\left\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*\right\} u^3 K'(u) du \right| \\
\leq & \frac{h^2}{6} \sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} \left| \frac{\partial^3}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^3} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\left\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right\} \right| \left\{ 3 \int u^2 K(u) du \right\}.
\end{aligned}
$$

The variance is

$$
\begin{aligned}
& \operatorname{var}\left\{ -\frac{1}{n} \sum_{j=1}^{n} \mathbf{X}_j I(Z_j \geq Z) K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \right\} \\
= & \frac{1}{n} \Big[E\left\{ \mathbf{X}_j I(Z_j \geq Z) K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \right\} \left\{ \mathbf{X}_j I(Z_j \geq Z) K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \right\}^{\mathrm{T}} \\
& - \left\{ E\mathbf{X}_j I(Z_j \geq Z) K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \right\} \left\{ E\mathbf{X}_j I(Z_j \geq Z) K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \right\}^{\mathrm{T}} \Big] \\
= & \frac{1}{n} \int \frac{1}{h^3} (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu, \mathbf{x}_{jl}^{\mathrm{T}})^{\mathrm{T}} (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu, \mathbf{x}_{jl}^{\mathrm{T}}) S_c(Z, \mathbf{x}_{jl}) S(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu, \boldsymbol{\beta}) \\
& \times f_{\mathbf{X}_l \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}}(\mathbf{x}_{jl} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu, \boldsymbol{\beta}) f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + hu) d\mathbf{x}_j K'^2(u) du \\
& + O(1/n) \\
= & \frac{1}{nh^3} \int (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \mathbf{x}_{jl}^{\mathrm{T}})^{\mathrm{T}} (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \mathbf{x}_{jl}^{\mathrm{T}}) S_c(Z, \mathbf{x}_{jl}) S(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) \\
& \times f_{\mathbf{X}_l \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}}(\mathbf{x}_{jl} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) d\mathbf{x}_{jl} \int K'^2(u) du \\
& + \frac{1}{2nh^3} \frac{\partial^2}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^2} \iint (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*, \mathbf{x}_{jl}^{\mathrm{T}})^{\mathrm{T}} (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*, \mathbf{x}_{jl}^{\mathrm{T}}) S_c(Z, \mathbf{x}_{jl}) S(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*, \boldsymbol{\beta}) \\
& \times f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*) f_{\mathbf{X}_l \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}}(\mathbf{x}_{jl} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*, \boldsymbol{\beta}) d\mathbf{x}_{jl} h^2 u^2 K'^2(u) du + O(1/n) \\
\leq & \frac{1}{nh^3} \sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} \left| f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{\mathbf{X}_j \mathbf{X}_j^{\mathrm{T}} I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} \right| \int K'^2(u) du \\
& + \frac{1}{2nh} \sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*} \left| \frac{\partial^2}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^2} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*) E\{\mathbf{X}_j \mathbf{X}_j^{\mathrm{T}} I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*\} \right| \int u^2 K'^2(u) du \\
& + O(1/n).
\end{aligned}
$$

So

$$
\begin{aligned}
& \frac{1}{n} \sum_{j=1}^{n} \mathbf{X}_j I(Z_j \geq Z) K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \\
= & -\frac{\partial}{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} + O_p(n^{-1/2} h^{-3/2} + h^2). \qquad \text{(S3.3)}
\end{aligned}
$$

Following similar derivations, we have

$$\begin{aligned}
& E\left\{\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j I(Z_j \geq Z) K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\} \\
= \; & f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} \\
& + \frac{h^2}{2}\int \frac{\partial^2}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^2} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*) E\left\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*\right\} u^2 K(u) du
\end{aligned}$$

and the absolute bias is

$$\begin{aligned}
& \Bigg| E\left\{\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j I(Z_j \geq Z) K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\} \\
& - f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} \Bigg| \\
\leq \; & \frac{h^2}{2}\sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} \left| \frac{\partial^2}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^2} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\left\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right\} \right| \left\{\int u^2 K(u) du\right\}.
\end{aligned}$$

The variance term satisfies

$$\begin{aligned}
& \mathrm{var}\left\{-\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j I(Z_j \geq Z) K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\} \\
\leq \; & \frac{1}{nh}\sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} \left| f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{\mathbf{X}_j \mathbf{X}_j^{\mathrm{T}} I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} \right| \int K^2(u) du \\
& + \frac{h}{2n}\sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*} \left| \frac{\partial^2}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^2} f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*) E\{\mathbf{X}_j \mathbf{X}_j^{\mathrm{T}} I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*\} \right| \int u^2 K^2(u) du \\
& + O(1/n).
\end{aligned}$$

So

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j I(Z_j \geq Z) K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) &= f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} \\
& \quad + O_p(n^{-1/2}h^{-1/2} + h^2).
\end{aligned}$$

Finally, combining the results of (S3.1), (S3.2), (S3.3) and (S3.4), we have

$$\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} \widehat{E}\left\{\mathbf{X}Y(Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}\right\} \\
= \; & -\frac{\sum_{j=1}^{n}\mathbf{X}_j I(Z_j \geq Z) K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}
\end{aligned}$$

$$
\begin{aligned}
& + \frac{\{\sum_{j=1}^n \mathbf{X}_j I(Z_j \geq Z) K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}\{\sum_{j=1}^n K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}}{\{\sum_{j=1}^n K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}^2} \\
= & \frac{\partial f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}/\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + O_p(n^{-1/2}h^{-3/2} + h^2)}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + O_p(n^{-1/2}h^{-1/2} + h^2)} \\
& + \frac{\left[f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{\mathbf{X}_j I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} + O_p(n^{-1/2}h^{-1/2} + h^2)\right]}{f^2_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + O(n^{-1/2}h^{-1/2} + h^2)} \\
& \times \left[-f'_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + O_p(n^{-1/2}h^{-3/2} + h^2)\right] \\
= & \frac{\partial}{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} E\{\mathbf{X}Y(Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} + O_p\{(nh^3)^{-1/2} + h^2\}
\end{aligned}
$$

for all $\boldsymbol{\beta}$ and for all h that satisfies Condition C2.

Now we inspect the consistency of the Kaplan Meier estimator on the hazard function and its derivatives, i.e. (3.9) and (3.10). Similar to the proof of (3.8), we show (3.9) and (3.10) through analyzing their absolute biases and variances. Let $A = n^{-1}\sum_{j=1}^n I(Z_j \geq Z_i) K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) - f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}$.

$$
\begin{aligned}
\widehat{\lambda}(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) &= \sum_{i=1}^n K_b(Z_i - Z) \frac{\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\sum_{j=1}^n I(Z_j \geq Z_i) K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} \\
&= \frac{1}{n}\sum_{i=1}^n K_b(Z_i - Z) \frac{\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} + A} \\
&= \frac{1}{n}\sum_{i=1}^n \frac{K_b(Z_i - Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\{1 + O_p(A)\},
\end{aligned}
$$

We first inspect

$$
\frac{1}{n}\sum_{i=1}^n \frac{K_b(Z_i - Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}.
$$

Using the same technique in the proof of (3.8) we have

$$
\begin{aligned}
& E\left[\frac{1}{n}\sum_{i=1}^n \frac{K_b(Z_i - Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\right] \\
= & \lambda(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) + \frac{b^2\partial^2}{2\partial Z^2}\iint \frac{K(v)K(u)}{S(Z^*, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})} f(Z^*, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) v^2 dv du \\
& + \frac{h^2\partial^2}{2\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^2}\iint \frac{K(v)K(u)E\{S_c(Z, \mathbf{X}_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*\} f(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*, \boldsymbol{\beta})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) S(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) E\{S_c(Z, \mathbf{X}_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}} \\
& \times f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*) u^2 dv du,
\end{aligned}
$$

where throughout the text, Z^* is on the line connecting Z and $Z + bv$. Thus, the absolute bias is

$$\begin{aligned}
& \left| E\left\{ \frac{1}{n}\sum_{i=1}^{n} \frac{K_b(Z_i - Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}} \right\} - \lambda(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) \right| \\
\leq \;& b^2 \sup_{Z^*, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}} \left| \frac{\partial^2}{2\partial Z^2} \frac{f(Z^*, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})}{S(Z^*, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})} \right| \int v^2 K(v)dv \\
& + h^2 \sup_{Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*} \left| \frac{\partial^2}{2\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^2} \frac{E\{S_c(Z, \mathbf{X}_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*\} f(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*, \boldsymbol{\beta}) f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*)}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) S(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) E\{S_c(Z, \mathbf{X}_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}} \right| \\
& \times \int u^2 K(u)du \\
= \;& O(h^2 + b^2)
\end{aligned}$$

under conditions C1–C6. Following the same procedure, noting that $A = O_p\{(nh)^{-1/2} + h^2\}$ uniformly, we can show that

$$\frac{1}{n}\sum_{i=1}^{n} \frac{K_b(Z_i - Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}} O_p(A) = O_p\{h^2 + (nh)^{-1/2}\},$$

hence the bias of $\widehat{\lambda}(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})$ is of order $O_p\{(nh)^{-1/2} + h^2 + b^2\}$ uniformly. On the other hand, the variance of $\widehat{\lambda}(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})$ is

$$\begin{aligned}
& \mathrm{var}\left\{ \widehat{\lambda}(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) \right\} \\
= \;& \mathrm{var}\left[\frac{1}{n}\sum_{i=1}^{n} K_b(Z_i - Z) \frac{\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} + A} \right] \\
= \;& \mathrm{var}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{K_b(Z_i - Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}} \{1 + O_p(A)\} \right] \\
\leq \;& 2\mathrm{var}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{K_b(Z_i - Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}} \right] \\
& + 2\mathrm{var}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{K_b(Z_i - Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}} O_p(A) \right].
\end{aligned}$$

We inspect the first term first.

$$2\mathrm{var}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{K_b(Z_i - Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z \geq Z_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}} \right]$$

$$
\begin{aligned}
&= \frac{2}{n}\left(E\left[\frac{K_b(Z_i-Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z\geq Z_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\right]^2\right)+O(1/n)\\
&= \frac{2}{bhn}\iint \frac{f(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})K^2(v)K^2(u)}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})S^2(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})E\{S_c(Z,\mathbf{X}_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}dvdu\\
&\quad +\frac{b\partial^2}{nh\partial Z^2}\iint \frac{f(Z^*,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})K^2(v)K^2(u)}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})S^2(Z^*,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})E\{S_c(Z^*,\mathbf{X}_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}v^2dvdu\\
&\quad +\frac{h\partial^2}{nb\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^2}\iint \frac{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*)f(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*,\boldsymbol{\beta})E\{S_c(Z,\mathbf{X}_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*\}}{f^2_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})S^2(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})E\{S_c(Z,\mathbf{X}_i)^2\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\\
&\quad \times K^2(v)K^2(u)u^2dvdu+O(1/n)\\
&\leq \frac{2}{bhn}\frac{f(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})S^2(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})E\{S_c(Z,\mathbf{X}_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\left\{\int K^2(u)du\right\}^2\\
&\quad +\frac{b}{nh}\sup_{Z^*,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}\left|\frac{\partial^2}{\partial Z^2}\frac{f(Z^*,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})S^2(Z^*,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})E\{S_c(Z^*,\mathbf{X}_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\right|\\
&\quad \times\left\{\int K^2(u)u^2du\right\}\left\{\int K^2(u)du\right\}\\
&\quad +\frac{h}{nb}\sup_{Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*}\left|\frac{\partial^2}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})^2}\frac{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*)f(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*,\boldsymbol{\beta})E\{S_c(Z,\mathbf{X}_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^*\}}{f^2_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})S^2(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})E\{S_c(Z,\mathbf{X}_i)^2\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\right|\\
&\quad \times\left\{\int K^2(u)u^2du\right\}\left\{\int K^2(u)du\right\}+O(1/n)\\
&= O\{1/(nhb)+h/(nb)+b/(nh)+1/n\}\\
&= O\{1/(nhb)\}.
\end{aligned}
$$

For the second term

$$
\begin{aligned}
&2\mathrm{var}\left[\frac{1}{n}\sum_{i=1}^n\frac{K_b(Z_i-Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z\geq Z_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}O_p(A)\right]\\
\leq\ &2E\left(\frac{1}{n}\sum_{i=1}^n\left[\frac{K_b(Z_i-Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z\geq Z_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\right]^2\sup\{O_p^2(A)\}\right)\\
=\ &2E\left(\frac{1}{n}\sum_{i=1}^n\left[\frac{K_b(Z_i-Z)\Delta_i K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})E\{I(Z\geq Z_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\right]^2O_p\{(nh)^{-1}+h^4\}\right)\\
=\ &\left[\frac{2}{bhn}\frac{f(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})S^2(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})E\{S_c(Z,\mathbf{X}_i)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}}\left\{\int K^2(u)du\right\}^2\right.\\
&\left.+O(n^{-1}b^{-1}h+n^{-1}h^{-1}b)\right]O\{(nh)^{-1}+h^4\}\\
=\ &O\{(nh)^{-2}b^{-1}+n^{-1}h^3b^{-1}\}
\end{aligned}
$$

under conditions C1–C6. Summarizing the above results, the variance of $\widehat{\lambda}(Z,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})$ is of order $1/(nhb)$ for all $\boldsymbol\beta$ and for all h and b that satisfy Condition C2. Hence, we have the consistency of estimator $\widehat\lambda(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)$, specifically

$$\widehat\lambda(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)=\lambda(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)+O_p\{(nhb)^{-1/2}+h^2+b^2\}$$

under condition C1–C6.

Next we inspect the estimator for the first derivative of hazard function $\lambda(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)$. Let

$$\begin{aligned}
\widehat{\boldsymbol\lambda}_{11} &= -\sum_{i=1}^n K_b(Z_i-Z)\frac{\Delta_i K_h'(\boldsymbol\beta^{\rm T}\mathbf{X}_i-\boldsymbol\beta^{\rm T}\mathbf{X})}{\sum_j I(Z_j\ge Z)K_h(\boldsymbol\beta^{\rm T}\mathbf{X}_j-\boldsymbol\beta^{\rm T}\mathbf{X})}\\
\widehat{\boldsymbol\lambda}_{12} &= \sum_{i=1}^n K_b(Z_i-Z)\Delta_i K_h(\boldsymbol\beta^{\rm T}\mathbf{X}_i-\boldsymbol\beta^{\rm T}\mathbf{X})\frac{\sum_{j=1}^n I(Z_j\ge Z)K_h'(\boldsymbol\beta^{\rm T}\mathbf{X}_j-\boldsymbol\beta^{\rm T}\mathbf{X})}{\{\sum_{j=1}^n I(Z_j\ge Z)K_h(\boldsymbol\beta^{\rm T}\mathbf{X}_j-\boldsymbol\beta^{\rm T}\mathbf{X})\}^2}.
\end{aligned}$$

Then $\widehat{\boldsymbol\lambda}_1(Z,\boldsymbol\beta^{\rm T}\mathbf{X})=\widehat{\boldsymbol\lambda}_{11}+\widehat{\boldsymbol\lambda}_{12}$. Following similar procedures, we have

$$\begin{aligned}
\widehat{\boldsymbol\lambda}_{11} &= \frac{\partial\left[f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})\right]/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}}\\
&\quad+O_p\{(nbh^3)^{-1/2}+b^2+h^2\},\\
\widehat{\boldsymbol\lambda}_{12} &= -\frac{\partial\left[f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}\right]/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})S^2(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}}f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)\\
&\quad+O_p\{(nbh)^{-1/2}+b^2+h^2\}.
\end{aligned}$$

In addition, we have

$$\begin{aligned}
&\frac{\partial\left[f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})\right]/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}}\\
&-\frac{\partial\left[f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}\right]/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})S^2(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}}f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)\\
= &\frac{\partial f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)}+\frac{f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)\partial E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}}\\
&+\frac{f'_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)}{f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)}-\frac{f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)\partial E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)E\{S_c(Z,\mathbf{X}_i)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}}\\
&-\frac{f'_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)}{f_{\boldsymbol\beta^{\rm T}\mathbf{X}}(\boldsymbol\beta^{\rm T}\mathbf{X})S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)}-\frac{f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)\partial S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{S^2(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)}\\
= &\frac{\partial f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)}-\frac{f(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)\partial S(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)/\partial\boldsymbol\beta^{\rm T}\mathbf{X}}{S^2(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)}\\
= &\frac{\partial}{\partial\boldsymbol\beta^{\rm T}\mathbf{X}}\lambda(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta)\\
= &\boldsymbol\lambda_1(Z,\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta).
\end{aligned}$$

Summarizing the results above, the estimator $\widehat{\boldsymbol{\lambda}}_1(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})$ satisfies

$$\widehat{\boldsymbol{\lambda}}_1(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) = \boldsymbol{\lambda}_1(Z, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) + O_p\{(nbh^3)^{-1/2} + h^2 + b^2\}$$

for all $\boldsymbol{\beta}$ and for all h and b that satisfy Condition C2.

In order to handle the zero-denominator issue, we implement the trimmed estimators in (3.1), (3.2), (3.3) and (3.4). Here we prove that they achieve the same asymptotic properties as the usual estimators. Because they have very similar structures, we show the detailed proof of (3.3) only. For further reading about the trimmed kernel estimators, please see Appendix A.2 of Härdle and Stoker (1989). For notation simplicity, we let $\widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \equiv 1/n \sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$. The absolute bias of the trimmed estimator is given by

$$\begin{aligned}
&\left| E\left[\frac{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{\widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) > d_n\right\}\right] \right. \\
&\left. -E\{I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} \right| \\
\leq\; & \left| E\left[\frac{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{\widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) > d_n\right\}\right] \right. \\
&\quad \left. -E\left[\frac{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) > d_n\right\}\right] \right| \\
&+ \left| E\left[\frac{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) > d_n\right\}\right] \right. \\
&\quad \left. -E\left[\frac{n^{-1}\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) I(Z_j \geq Z)}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) > d_n\right\}\right] \right| \\
&+ \left| E\left[\frac{n^{-1}\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) I(Z_j \geq Z)}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) > d_n\right\}\right] \right. \\
&\left. -E\{I(Z_j \geq Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} \right|.
\end{aligned}$$

The first term satisfies

$$\begin{aligned}
&\left| E\left[\frac{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{\widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) > d_n\right\}\right] \right. \\
&\left. - E\left[\frac{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) > d_n\right\}\right] \right|
\end{aligned}$$

$$
\begin{aligned}
&\le \left| E\left[\frac{\sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X) I(Z_j \ge Z)}{\sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X)} \right.\right. \\
&\quad \left.\left. \times\ I\left\{ \widehat f(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n, f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) \le d_n \right\}\right]\right| \\
&\quad + \left| E\left[\frac{\sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X) I(Z_j \ge Z)}{\sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X)} \right.\right. \\
&\quad \left.\left. \times\ I\left\{ \widehat f(\boldsymbol\beta^{\mathrm T}\mathbf X) \le d_n, f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n \right\}\right]\right| \\
&\le \left| E\left[I\left\{ \widehat f(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n, f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) \le d_n \right\}\right]\right| \\
&\quad + \left| E\left[I\left\{ \widehat f(\boldsymbol\beta^{\mathrm T}\mathbf X) \le d_n, f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n \right\}\right]\right| \\
&\le \left| E\left[I\left\{ f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) \le d_n \right\}\right]\right| + \left| E\left[I\left\{ \widehat f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) \le d_n \right\}\right]\right| \\
&= O_p\{n^{-\epsilon} + h^2 + (nh)^{-1/2}\} \\
&= O_p\{h^2 + (nh)^{-1/2}\}.
\end{aligned}
$$

The second term is

$$
\begin{aligned}
&\left| E\left[\frac{\sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X) I(Z_j \ge Z)}{\sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X)} I\left\{ f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n \right\}\right] \right. \\
&\quad \left. - E\left[\frac{1/n \sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X) I(Z_j \ge Z)}{f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X)} I\left\{ f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n \right\}\right] \right| \\
&\le \left| E\left[\frac{1/n \sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X) I(Z_j \ge Z)}{f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) + O_p\{h^2 + (nh)^{-1/2}\}} I\left\{ f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n \right\}\right] \right. \\
&\quad \left. - E\left[\frac{1/n \sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X) I(Z_j \ge Z)}{f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X)} I\left\{ f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n \right\}\right] \right| \\
&\le E\left[\frac{1/n \sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X) I(Z_j \ge Z)}{f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X)} I\left\{ f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n \right\}\right] \\
&\quad \times O_p\{h^2 + (nh)^{-1/2}\} \\
&= O_p\{h^2 + (nh)^{-1/2}\}.
\end{aligned}
$$

The third term is

$$
\begin{aligned}
&\left| E\left[\frac{1/n \sum_{j=1}^n K_h(\boldsymbol\beta^{\mathrm T}\mathbf X_j - \boldsymbol\beta^{\mathrm T}\mathbf X) I(Z_j \ge Z)}{f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X)} I\left\{ f_{\boldsymbol\beta^{\mathrm T}\mathbf X}(\boldsymbol\beta^{\mathrm T}\mathbf X) > d_n \right\}\right] \right. \\
&\left. - E\{I(Z_j \ge Z) \mid \boldsymbol\beta^{\mathrm T}\mathbf X\} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| E\left[\frac{K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})I(Z_j\geq Z)}{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})>d_n\right\}\right]\right. \\
&\quad \left. -E\{I(Z_j\geq Z)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}\right| \\
&= E\left|E\left[I(Z_j\geq Z)I\left\{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})>d_n\right\}\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right]-E\{I(Z_j\geq Z)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}\right. \\
&\quad \left. +O_p(h^2)\right| \\
&= E\left[I(Z_j\geq Z)I\left\{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\leq d_n\right\}\right]+O_p(h^2) \\
&\leq E\left[I\left\{f_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\leq d_n\right\}\right]+O_p(h^2) \\
&= O_p\{n^{-\epsilon}+h^2+(nh)^{-1/2}\} \\
&= O_p\{h^2+(nh)^{-1/2}\}.
\end{aligned}
$$

It is easy to see the variance of this trimmed estimator,

$$
\mathrm{var}\left[\frac{\sum_{j=1}^n K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})I(Z_j\geq Z)}{\sum_{j=1}^n K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{\widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})>d_n\right\}\right]=O_p\{(nh)^{-1/2}\}.
$$

Summarizing the above result, we have

$$
\begin{aligned}
&\frac{\sum_{j=1}^n K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})I(Z_j\geq Z)}{\sum_{j=1}^n K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} I\left\{\widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})>d_n\right\} \\
=\ & E\{I(Z_j\geq Z)\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\}+O_p\{h^2+(nh)^{-1/2}\}.
\end{aligned}
$$

The above analysis illustrates that trimming can be used to bypass the zero-denominator issue so that the results in Theorem 1 and 2 still hold when Condition C4 is replaced by Condition C4′.

The above analysis establishes the bias and variance property of the nonparametric estimators at each $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}, Z$ for (3.5) to (3.10). To further obtain the uniform convergence results, we divide the region into smaller sets, bound the difference between the estimator and its mean by their boundary differences, and invoke the union law and Bernstein inequality. Because the techniques going from the pointwise convergence results to uniform results are similar for (3.5) to (3.8), we give detailed proof for (3.8) only. Because the domain of $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})$ is compact, we divide it into rectangular regions. In each region, the distance between a point $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\beta})$ in this region and the nearest grid point is less than n^{-2}. Note that we only need $N \leq Cn^{2p}$ grid points, where C is a constant. Let the grid points be $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_N$. For notation brevity, let $\widehat{\boldsymbol{\rho}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) = \partial \widehat{E}\{\mathbf{X}Y(Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}\}/\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ and $\boldsymbol{\rho}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) = \partial E\{\mathbf{X}Y(Z) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}\}/\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$. Then for any $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})$, there exists a $\boldsymbol{\kappa}_i$, $1 \leq i \leq N$, such that

$$\begin{aligned} |\widehat{\boldsymbol{\rho}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) - \boldsymbol{\rho}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})| &\leq |\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i) - \boldsymbol{\rho}(\boldsymbol{\kappa}_i)| + |\widehat{\boldsymbol{\rho}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) - \widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i)| \\ &\quad + |\boldsymbol{\rho}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) - \boldsymbol{\rho}(\boldsymbol{\kappa}_i)| \\ &\leq |\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i) - \boldsymbol{\rho}(\boldsymbol{\kappa}_i)| + C_1 n^{-2}/h^2, \end{aligned}$$

for an absolute constant C_1 under Conditions C1 and C6. Thus, for sufficiently large C,

$$\begin{aligned} & \mathrm{pr}(\sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\beta}} |\widehat{\boldsymbol{\rho}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}) - \boldsymbol{\rho}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta})| > 2C[h^2 + \{\log n (nh^3)^{-1}\}^{1/2}]) \\ \leq\ & \mathrm{pr}(\sup_{\boldsymbol{\kappa}_i} |\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i) - \boldsymbol{\rho}(\boldsymbol{\kappa}_i)| > 2C[h^2 + \{\log n (nh^3)^{-1}\}^{1/2}] - C_1 (nh)^{-2}) \\ \leq\ & \mathrm{pr}(\sup_{\boldsymbol{\kappa}_i} |\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i) - \boldsymbol{\rho}(\boldsymbol{\kappa}_i)| > C[h^2 + \{\log n (nh^3)^{-1}\}^{1/2}]) \end{aligned}$$

under Condition C2. Using Bernstein's inequality on each sum in the numerator and denominator of $\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i)$, under Conditions C1, C2, C4, C5, and C6, we have that for any $A > 0$,

$$\begin{aligned} & \mathrm{pr}[|\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i) - E\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i)| \geq A\{\log n/(nh^3)\}^{1/2}] \\ \leq\ & \exp\left\{\frac{-nA^2 \log n/(nh^3)}{C_2 h^{-3} + AC_2 (\log n)^{1/2} (nh^3)^{-1/2} h^{-2}}\right\} \\ \leq\ & \exp\left(\frac{-A^2 \log n}{2C_2}\right), \end{aligned}$$

where C_2 is a constant. This leads to

$$\begin{aligned} & \mathrm{pr}[\sup_{\boldsymbol{\kappa}_i} |\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i) - E\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i)| \geq A\{\log n/(nh^3)\}^{1/2}] \\ \leq\ & Cn^{2p} \exp\left(\frac{-A^2 \log n}{2C_2}\right) \\ =\ & C \exp\left[\{2p - A^2/(2C_2)\} \log n\right] \to 0 \end{aligned}$$

for all $A^2 > 2pC_2$. Combining the above results, we get that for $A_1 = \max(A, C)$,

$$
\begin{aligned}
& \mathrm{pr}(\sup_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x},\boldsymbol{\beta}} |\widehat{\boldsymbol{\rho}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}) - \boldsymbol{\rho}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta})| > 2A_1[h^2 + \{\log n(nh^3)^{-1}\}^{1/2}]) \\
\leq \ & \mathrm{pr}(\sup_{\boldsymbol{\kappa}_i} |\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i) - \boldsymbol{\rho}(\boldsymbol{\kappa}_i)| > A_1[h^2 + \{\log n(nh^3)^{-1}\}^{1/2}]) \\
\leq \ & \mathrm{pr}\{\sup_{\boldsymbol{\kappa}_i} |\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i) - E\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i)| > A_1 h^2\} \\
& + \mathrm{pr}([\sup_{\boldsymbol{\kappa}_i} |\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i) - E\widehat{\boldsymbol{\rho}}(\boldsymbol{\kappa}_i)| \geq A_1\{\log n(nh^3)^{-1}\}^{1/2}]) \\
\to \ & 0.
\end{aligned}
$$

The uniform convergence results concerning (3.9) and (3.10) are slightly different because these functions contain the additional component Z. Nevertheless, under Condition C6, the support of $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}, Z_i)$ or $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}, Z_j)$ is also bounded so we can similarly divide the region using $N \leq Cn^{2p+2}$ grid points while the distance of a point to the nearest grid point is less than n^{-2}. The rest of the analysis can then be similarly carried out as above, where we can establish the uniform convergence of the respective numerator and denominator terms, and hence their ratios.

□

S4 Proof of Theorem 1

$$
\frac{1}{n}\sum_{i=1}^{n} \Delta_i \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i, \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_n)}{\widehat{\lambda}(Z_i, \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_n)} \otimes \left[\mathbf{X}_{li} - \frac{\widehat{E}\left\{\mathbf{X}_{li} Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_n\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_n\right\}}\right] = \mathbf{0}.
$$

Under condition C3, there exists a subsequence of $\widehat{\boldsymbol{\beta}}_n, n = 1, 2, \dots,$ that converges. For notation simplicity, we still write $\widehat{\boldsymbol{\beta}}_n, n = 1, 2, \dots,$ as the subsequence that converges, and let the limit be $\boldsymbol{\beta}^*$.

We first have

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n)}{\widehat{\lambda}(Z_i,\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n)} \otimes \left[\mathbf{X}_{li} - \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n\right\}}\right] \\
= & \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)}{\widehat{\lambda}(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)} \otimes \left[\mathbf{X}_{li} - \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*\right\}}\right] \\
& + O_p(\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\|) \\
= & \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)}{\widehat{\lambda}(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)} \otimes \left[\mathbf{X}_{li} - \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*\right\}}\right] \\
& + o_p(1)
\end{aligned}
$$

where the first equality is because the first derivative of the summation with respect to $\boldsymbol{\beta}$ is bounded uniformly under conditions C1–C2 by Lemma 1. The last equality is because $\widehat{\boldsymbol{\beta}}_n$ converges to $\boldsymbol{\beta}^*$. Thus, for sufficiently large n, we have

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)}{\widehat{\lambda}(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)} \otimes \left[\mathbf{X}_{li} - \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*\right\}}\right] \\
& + o_p(1) \\
= & \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\boldsymbol{\lambda}_1(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*) + O_p\{(nbh^3)^{-1/2}(\log n)^{1/2} + h^2 + b^2\}}{\lambda(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*) + O_p\{(nbh)^{-1/2}(\log n)^{1/2} + h^2 + b^2\}} \\
& \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i\right\} + O_p\{(nh)^{-1/2}(\log n)^{1/2} + h^2\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i\right\} + O_p\{(nh)^{-1/2}(\log n)^{1/2} + h^2\}}\right] \\
& + o_p(1) \\
= & \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\boldsymbol{\lambda}_1(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)}{\lambda(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)} \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i\right\}}\right] \\
& + o_p(1)
\end{aligned}
$$

from a direct application of Lemma 1. In addition,

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\boldsymbol{\lambda}_1(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)}{\lambda(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)} \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i\right\}}\right] \\
= & E\left(\Delta \frac{\boldsymbol{\lambda}_1(Z,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X},\boldsymbol{\beta}^*)}{\lambda(Z,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X},\boldsymbol{\beta}^*)} \otimes \left[\mathbf{X}_{l} - \frac{E\left\{\mathbf{X}_{l}Y(Z) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}\right\}}{E\left\{Y(Z) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}\right\}}\right]\right) + o_p(1)
\end{aligned}
$$

under conditions C1–C2. Thus, for sufficient large n we have

$$
\begin{aligned}
\mathbf{0} &= \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n)}{\widehat{\lambda}(Z_i,\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n)} \otimes \left[\mathbf{X}_{li} - \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n\right\}}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\boldsymbol{\lambda}_1(Z_i,\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n)}{\lambda(Z_i,\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_n)} \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}_i\right\}}\right] + o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\boldsymbol{\lambda}_1(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)}{\lambda(Z_i,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}^*)} \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i\right\}}\right] + o_p(1) \\
&= E\left(\Delta \frac{\boldsymbol{\lambda}_1(Z,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X},\boldsymbol{\beta}^*)}{\lambda(Z,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X},\boldsymbol{\beta}^*)} \otimes \left[\mathbf{X}_l - \frac{E\left\{\mathbf{X}_l Y(Z) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}\right\}}{E\left\{Y(Z) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}\right\}}\right]\right) + o_p(1)
\end{aligned}
$$

under conditions C1–C2, and C3. Note that

$$
E\left(\Delta \frac{\boldsymbol{\lambda}_1(Z,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X},\boldsymbol{\beta}^*)}{\lambda(Z,\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X},\boldsymbol{\beta}^*)} \otimes \left[\mathbf{X}_l - \frac{E\left\{\mathbf{X}_l Y(Z) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}\right\}}{E\left\{Y(Z) \mid \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}\right\}}\right]\right)
$$

is a nonrandom quantity that does not depend on n, hence it is zero. Thus the uniqueness requirement in Condition C7 ensures that $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$.

We now show that the subsequence that converges includes all but a finite number of n. Assume this is not the case, then we can obtain an infinite sequence of $\widehat{\boldsymbol{\beta}}_n$ that do not converge to $\boldsymbol{\beta}^*$. As an infinite sequence in a compact set $\mathcal{B}$, we can thus obtain another subsequence that converges, say to $\boldsymbol{\beta}^{**} \neq \boldsymbol{\beta}^*$. Identical derivation as before then leads to $\boldsymbol{\beta}^{**} = \boldsymbol{\beta}_0$, which is a contradiction to $\boldsymbol{\beta}^{**} \neq \boldsymbol{\beta}^*$. Thus we conclude $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \to \mathbf{0}$ in probability when $n \to \infty$ under condition C1–C7. □

S5 Proof of Theorem 2

We first expand (2.4) as

$$
\begin{aligned}
\mathbf{0} &= n^{-1/2}\sum_{i=1}^{n}\Delta_i \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}})}{\widehat{\lambda}(Z_i,\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}})} \otimes \left[\mathbf{X}_{li} - \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}_i,\widehat{\boldsymbol{\beta}}\right\}}\right] \\
&= n^{-1/2}\sum_{i=1}^{n}\Delta_i \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0)}{\widehat{\lambda}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0)} \otimes \left[\mathbf{X}_{li} - \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}\right]
\end{aligned}
$$

(S5.1)

$$+\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)}\left(\Delta_i\frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta})}{\widehat{\lambda}(Z_i,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta})}\right.\right.$$
$$\left.\left.\left.\otimes\left[\mathbf{X}_{li}-\frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}\right\}}{\widehat{E}\left\{Y_i(Z_i)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}\right\}}\right]\right)\otimes\mathbf{X}_{li}^{\mathrm{T}}\right\}\right|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}$$

(S5.2)

$$\times\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0),$$

where $\widetilde{\boldsymbol{\beta}}$ is on the line connecting $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$.

We first consider (S5.2). Because of Theorem 1, and Lemma 1, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)}\left(\Delta_i\frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta})}{\widehat{\lambda}(Z_i,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta})}\right.\right.$$
$$\left.\left.\left.\otimes\left[\mathbf{X}_{li}-\frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}\right\}}{\widehat{E}\left\{Y_i(Z_i)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}\right\}}\right]\right)\otimes\mathbf{X}_{li}^{\mathrm{T}}\right\}\right|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}$$
$$=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial}{\partial(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}\left(\Delta_i\frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0)}{\widehat{\lambda}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0)}\right.\right.$$
$$\left.\left.\otimes\left[\mathbf{X}_{li}-\frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}{\widehat{E}\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}\right]\right)\otimes\mathbf{X}_{li}^{\mathrm{T}}\right\}$$
$$+o_p(1)$$
$$=-\frac{1}{n}\sum_{i=1}^{n}\left(\Delta_i\frac{\widehat{\boldsymbol{\lambda}}_1^{\otimes 2}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0)}{\widehat{\lambda}^2(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0)}\right.$$
$$\left.\otimes\left[\mathbf{X}_{li}-\frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}{\widehat{E}\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}\right]\otimes\mathbf{X}_{li}^{\mathrm{T}}\right)$$

(S5.3)

$$+\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{\widehat{\lambda}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0)}\frac{\partial}{\partial(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}$$
$$\times\left(\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0)\otimes\left[\mathbf{X}_{li}-\frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}{\widehat{E}\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}\right]\right)\otimes\mathbf{X}_{li}^{\mathrm{T}}$$

(S5.4)

$$+o_p(1).$$

Because of Lemma 1, (S5.3) converges uniformly in probability to

$$
\begin{aligned}
& -E\left(\int_0^\infty \frac{\boldsymbol{\lambda}_{10}^{\otimes 2}(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0^2(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \left[\mathbf{X}_l - \frac{E\left\{\mathbf{X}_l Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}\right] \otimes \mathbf{X}_l^{\mathrm{T}} dN(s)\right) \\
= & -E\left(\int_0^\infty \frac{\boldsymbol{\lambda}_{10}^{\otimes 2}(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0^2(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \left[\mathbf{X}_l - \frac{E\left\{\mathbf{X}_l Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}\right] \right.\\
& \left. \otimes \mathbf{X}_l^{\mathrm{T}} Y(s)\lambda_0(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}) ds\right) \\
= & -E\left(\int_0^\infty \frac{\boldsymbol{\lambda}_{10}^{\otimes 2}(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \left[\mathbf{X}_l - \frac{E\left\{\mathbf{X}_l Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}\right] \right.\\
& \left. \otimes \left[\mathbf{X}_l - \frac{E\left\{\mathbf{X}_l Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}\right]^{\mathrm{T}} Y(s) ds\right) \\
& -E\left(\int_0^\infty \frac{\boldsymbol{\lambda}_{10}^{\otimes 2}(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})}{\lambda_0(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})} \otimes \left[\mathbf{X}_l - \frac{E\left\{\mathbf{X}_l Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}{E\left\{Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}}\right] \right.\\
& \left. \otimes \frac{E\left\{\mathbf{X}_l Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}^{\mathrm{T}}}{E\left\{Y(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\right\}} Y(s) ds\right) \\
= & -E\{\mathbf{S}_{\mathrm{eff}}(\Delta, Z, \mathbf{X})^{\otimes 2}\},
\end{aligned}
$$

where the last equality is because the second term above is zero by first taking expectation conditional on $\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}$. Note that $\boldsymbol{\lambda}_1(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}_0) = \boldsymbol{\lambda}_{10}(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$ and $\lambda(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}_0) = \lambda_0(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X})$.

Similarly, from Lemma 1, the term in (S5.4) converges uniformly in probability to the limit of

$$
E\left\{\frac{\Delta_i}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)} \frac{\partial}{\partial(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}\left(\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0) \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li} Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}\right]\right) \otimes \mathbf{X}_{li}^{\mathrm{T}}\right\}.
$$

Now let $\widehat{\boldsymbol{\lambda}}_{1,-i}(Z,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}_0)$ be the leave-one-out version of $\widehat{\boldsymbol{\lambda}}_1(Z,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}_0)$, i.e. it is constructed the same as $\widehat{\boldsymbol{\lambda}}_1(Z,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X},\boldsymbol{\beta}_0)$ except that the ith observation is not used. Obviously,

$$
\begin{aligned}
& \frac{\Delta_i}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)} \frac{\partial}{\partial(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}\left(\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0) \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li} Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}\right]\right) \otimes \mathbf{X}_{li}^{\mathrm{T}} \\
- & \frac{\Delta_i}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)} \frac{\partial}{\partial(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}\left(\widehat{\boldsymbol{\lambda}}_{1,-i}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,\boldsymbol{\beta}_0) \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li} Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}\right]\right) \otimes \mathbf{X}_{li}^{\mathrm{T}} \\
= & \; o_p(1).
\end{aligned}
$$

Now let E_i mean taking expectation with respect to the ith observation conditional on all other observations, then

$$\begin{aligned}
& E_i\left\{\frac{\Delta_i}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i)}\frac{\partial}{\partial(\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i)}\left(\widehat{\boldsymbol{\lambda}}_{1,-i}(Z_i,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i,\boldsymbol{\beta}_0)\right.\right.\\
& \left.\left.\otimes\left[\mathbf{X}_{li}-\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}\right]\right)\otimes\mathbf{X}_{li}^{\rm T}\right\}\\
= & E_i\left\{\int\frac{1}{\lambda_0(s,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i)}\frac{\partial}{\partial(\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i)}\left(\widehat{\boldsymbol{\lambda}}_{1,-i}(s,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i,\boldsymbol{\beta}_0)\right.\right.\\
& \left.\left.\otimes\left[\mathbf{X}_{li}-\frac{E\left\{\mathbf{X}_{li}Y_i(s)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}{E\left\{Y_i(s)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}\right]\right)\otimes\mathbf{X}_{li}^{\rm T}dN_i(s)\right\}\\
= & E_i\left\{\frac{\partial}{\partial\boldsymbol{\beta}_0}\int\widehat{\boldsymbol{\lambda}}_{1,-i}(s,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i,\boldsymbol{\beta}_0)\otimes\left[\mathbf{X}_{li}-\frac{E\left\{\mathbf{X}_{li}Y_i(s)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}{E\left\{Y_i(s)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}\right]E\{Y_i(s)\mid\mathbf{X}_i\}ds\right\}\\
= & \frac{\partial}{\partial\boldsymbol{\beta}_0}E_i\left\{\int\widehat{\boldsymbol{\lambda}}_{1,-i}(s,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i,\boldsymbol{\beta}_0)\otimes\left[\mathbf{X}_{li}-\frac{E\left\{\mathbf{X}_{li}Y_i(s)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}{E\left\{Y_i(s)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}\right]Y_i(s)ds\right\}\\
= & \mathbf{0}.
\end{aligned}$$

Here, the last equality is because the integrand has expectation zero conditional on $\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i$ and all other observations, and the third last equality is because the expectation is with respect to $\mathbf{X}_i$, and does not involve $\boldsymbol{\beta}_0$. Therefore, the term in (S5.4) converges in probability uniformly to

$$E\left\{\frac{\Delta_i}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i)}\frac{\partial}{\partial(\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i)}\left(\widehat{\boldsymbol{\lambda}}_{1,-i}(Z_i,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i,\boldsymbol{\beta}_0)\otimes\left[\mathbf{X}_{li}-\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}\right]\right)\otimes\mathbf{X}_{li}^{\rm T}\right\}=0$$

Combining the results concerning (S5.3) and (S5.4), we thus have obtained that the expression in (S5.2) is $-E\{\mathbf{S}_{\rm eff}(\Delta,Z,\mathbf{X})^{\otimes 2}\}+o_p(1)$.

Next we decompose (S5.1) into

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^n\Delta_i\frac{\widehat{\boldsymbol{\lambda}}_1(Z_i,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i,\boldsymbol{\beta}_0)}{\widehat{\lambda}(Z_i,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i,\boldsymbol{\beta}_0)}\otimes\left[\mathbf{X}_{li}-\frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}{\widehat{E}\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i,\boldsymbol{\beta}_0\right\}}\right]\\
= & \mathbf{T}_1+\mathbf{T}_2+\mathbf{T}_3+\mathbf{T}_4, \qquad \text{(S5.5)}
\end{aligned}$$

where

$$\mathbf{T}_1 = n^{-1/2}\sum_{i=1}^n\Delta_i\frac{\boldsymbol{\lambda}_{10}(Z_i,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i)}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i)}\otimes\left[\mathbf{X}_{li}-\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\rm T}\mathbf{X}_i\right\}}\right],$$

$$
\begin{aligned}
\mathbf{T}_2 &= n^{-1/2}\sum_{i=1}^n \Delta_i \left\{ \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0)}{\widehat{\lambda}(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0)} - \frac{\boldsymbol{\lambda}_{10}(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)} \right\} \\
&\quad \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}} \right], \\
\mathbf{T}_3 &= n^{-1/2}\sum_{i=1}^n \Delta_i \frac{\boldsymbol{\lambda}_{10}(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)} \otimes \left[\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}} \right. \\
&\quad \left. - \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0\right\}} \right], \\
\mathbf{T}_4 &= n^{-1/2}\sum_{i=1}^n \Delta_i \left\{ \frac{\widehat{\boldsymbol{\lambda}}_1(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0)}{\widehat{\lambda}(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0)} - \frac{\boldsymbol{\lambda}_{10}(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)} \right\} \\
&\quad \otimes \left[\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}} - \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0\right\}}{\widehat{E}\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0\right\}} \right].
\end{aligned}
$$

First, note that

$$
\begin{aligned}
\mathbf{T}_2 &= n^{-1/2}\sum_{i=1}^n \int \left\{ \frac{\widehat{\boldsymbol{\lambda}}_1(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0)}{\widehat{\lambda}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0)} - \frac{\boldsymbol{\lambda}_{10}(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)} \right\} \\
&\quad \otimes \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li}Y_i(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}} \right] dN_i(s) \\
&= o_p\left(n^{-1/2}\sum_{i=1}^n \int \left[\mathbf{X}_{li} - \frac{E\left\{\mathbf{X}_{li}Y_i(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{E\left\{Y_i(s) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}} \right] Y_i(s)\lambda_0(s, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_{li}) ds \right) \\
&= o_p(1),
\end{aligned}
$$

where the last equality above is because the quantity inside the parenthesis is a mean zero normal random quantity of order $O_p(1)$. Further,

$$
\begin{aligned}
\mathbf{T}_3 &= n^{-1/2}\sum_{i=1}^n \Delta_i \frac{\boldsymbol{\lambda}_{10}(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)} \otimes \left(- \frac{\widehat{E}\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}} \right. \\
&\quad \left. + \frac{\widehat{E}\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i, \boldsymbol{\beta}_0\right\} E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{[E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}]^2} \right) + o_p(1) \\
&= n^{-3/2}\sum_{i=1}^n \sum_{j=1}^n \Delta_i \frac{\boldsymbol{\lambda}_{10}(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)} \otimes \left[- \frac{K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_{lj} I(Z_j \geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i) E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}} \right. \\
&\quad \left. + \frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\} K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j - \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i) I(Z_j \geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)[E\left\{Y_i(Z_i) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}]^2} \right] + o_p(1) \\
&= \mathbf{T}_{31} + \mathbf{T}_{32} + \mathbf{T}_{33} + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{T}_{31} &= n^{-1/2}\sum_{i=1}^{n}\Delta_i\frac{\boldsymbol{\lambda}_{10}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}\otimes E\left[-\frac{K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_{lj}I(Z_j\geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}\right.\\
&\quad\left.+\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)I(Z_j\geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)[E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}]^2}\mid\Delta_i,Z_i,\mathbf{X}_i\right]\\
\mathbf{T}_{32} &= n^{-1/2}\sum_{j=1}^{n}E\left(\Delta_i\frac{\boldsymbol{\lambda}_{10}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}\otimes\left[-\frac{K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_{lj}I(Z_j\geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}\right.\right.\\
&\quad\left.\left.+\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)I(Z_j\geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)[E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}]^2}\right]\mid\Delta_j,Z_j,\mathbf{X}_j\right)\\
\mathbf{T}_{33} &= -n^{1/2}E\left(\Delta_i\frac{\boldsymbol{\lambda}_{10}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}\otimes E\left[-\frac{K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_{lj}I(Z_j\geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}\right.\right.\\
&\quad\left.\left.+\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)I(Z_j\geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)[E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}]^2}\right]\right).
\end{aligned}
$$

Here we used U-statistic property in the last equality above. Now when $nh^4\to 0$,

$$
\begin{aligned}
\mathbf{T}_{31} &= n^{-1/2}\sum_{i=1}^{n}\Delta_i\frac{\boldsymbol{\lambda}_{10}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}\otimes\left[-\frac{E\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\}}{E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}\right.\\
&\quad\left.+\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}}{[E\left\{Y_i(Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i\right\}]^2}\right]+O(n^{1/2}h^2)\\
&= o_p(1).
\end{aligned}
$$

Thus, $\mathbf{T}_{33}=o_p(1)$ as well. To analyze $\mathbf{T}_{32}$,

$$
\begin{aligned}
\mathbf{T}_{32} &= n^{-1/2}\sum_{j=1}^{n}E\left(\Delta_i\frac{\boldsymbol{\lambda}_{10}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)}\right.\\
&\quad\otimes\left[-\frac{K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_{lj}I(Z_j\geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)E\left\{I(Z\geq Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}=\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,Z_i\right\}}\right.\\
&\quad\left.\left.+\frac{E\left\{\mathbf{X}_l I(Z\geq Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}=\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,Z_i\right\}K_h(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_j-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)I(Z_j\geq Z_i)}{f_{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)[E\left\{I(Z\geq Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}=\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i,Z_i\right\}]^2}\right]\mid\Delta_j,Z_j,\mathbf{X}_j\right)\\
&= n^{-1/2}\sum_{j=1}^{n}E\left(\Delta_i\frac{\boldsymbol{\lambda}_{10}(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_j)}{\lambda_0(Z_i,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_j)}\otimes\left[-\frac{\mathbf{x}_{lj}I(z_j\geq Z_i)}{E\left\{I(Z\geq Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}=\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_j,Z_i\right\}}\right.\right.\\
&\quad\left.\left.+\frac{E\left\{\mathbf{X}_l I(Z\geq Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}=\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_j,Z_i\right\}I(z_j\geq Z_i)}{[E\left\{I(Z\geq Z_i)\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}=\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_j,Z_i\right\}]^2}\right]\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i=\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_j\right)\\
&\quad+O_p(n^{1/2}h^2)\\
&= n^{-1/2}\sum_{j=1}^{n}E\left(\int_0^{z_j}\frac{\boldsymbol{\lambda}_{10}(s,\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_j)}{E\left\{S_c(s,\mathbf{X})\mid\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}=\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_j\right\}}\right.
\end{aligned}
$$

$$\otimes\left[\frac{E\left\{\mathbf{X}_l S_c(s,\mathbf{X}) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}=\boldsymbol\beta_0^{\rm T}\mathbf{x}_j\right\}}{E\left\{S_c(s,\mathbf{X}) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}=\boldsymbol\beta_0^{\rm T}\mathbf{x}_j\right\}}-\mathbf{x}_{lj}\right] S_c(s,\mathbf{X}_i)ds \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i=\boldsymbol\beta_0^{\rm T}\mathbf{x}_j\Bigg)$$
$$+O_p(n^{1/2}h^2)$$
$$= n^{-1/2}\sum_{j=1}^n\int Y_j(s)\lambda_0(s,\boldsymbol\beta_0^{\rm T}\mathbf{x}_j)\frac{\boldsymbol\lambda_{10}(s,\boldsymbol\beta_0^{\rm T}\mathbf{x}_j)}{\lambda_0(s,\boldsymbol\beta_0^{\rm T}\mathbf{x}_j)}\otimes\left[\frac{E\left\{\mathbf{X}_{lj}Y_j(s) \mid \boldsymbol\beta_0^{\rm T}\mathbf{x}_j\right\}}{E\left\{Y_j(s) \mid \boldsymbol\beta_0^{\rm T}\mathbf{x}_j\right\}}-\mathbf{x}_{lj}\right]ds$$
$$+O_p(n^{1/2}h^2).$$

When $nh^4\to 0$, plugging the results of $\mathbf{T}_1$ and $\mathbf{T}_{32}$ to (S5.5), we obtain that the expression in (S5.1) is

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^n\Delta_i\frac{\widehat{\boldsymbol\lambda}_1(Z_i,\boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0)}{\widehat\lambda(Z_i,\boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0)}\otimes\left[\mathbf{X}_{li}-\frac{\widehat E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0\right\}}{\widehat E\left\{Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0\right\}}\right]\\
=\ & n^{-1/2}\sum_{i=1}^n\int\frac{\boldsymbol\lambda_{10}(t,\boldsymbol\beta_0^{\rm T}\mathbf{X}_i)}{\lambda_0(t,\boldsymbol\beta_0^{\rm T}\mathbf{X}_i)}\otimes\left[\mathbf{X}_{li}-\frac{E\left\{\mathbf{X}_{li}Y_i(t) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i\right\}}{E\left\{Y_i(t) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i\right\}}\right]dM_i(t)+o_p(1)\\
=\ & n^{-1/2}\sum_{i=1}^n\mathbf{S}_{\rm eff}(\Delta_i,Z_i,\mathbf{X}_i)+o_p(1).
\end{aligned}$$

Finally, note that

$$\begin{aligned}
\mathbf{T}_4 =\ & n^{-1/2}\sum_{i=1}^n\Delta_i\left\{\frac{\widehat{\boldsymbol\lambda}_1(Z_i,\boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0)}{\widehat\lambda(Z_i,\boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0)}-\frac{\boldsymbol\lambda_{10}(Z_i,\boldsymbol\beta_0^{\rm T}\mathbf{X}_i)}{\lambda_0(Z_i,\boldsymbol\beta_0^{\rm T}\mathbf{X}_i)}\right\}\\
& \times\left[\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i\right\}}-\frac{\widehat E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0\right\}}{\widehat E\left\{Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0\right\}}\right]\\
=\ & o_p\left(n^{-1/2}\sum_{i=1}^n\Delta_i\left[\frac{E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i\right\}}{E\left\{Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i\right\}}-\frac{\widehat E\left\{\mathbf{X}_{li}Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0\right\}}{\widehat E\left\{Y_i(Z_i) \mid \boldsymbol\beta_0^{\rm T}\mathbf{X}_i,\boldsymbol\beta_0\right\}}\right]\right)\\
=\ & o_p\left(n^{-1/2}\sum_{i=1}^n\int Y_i(s)\lambda_0(s,\boldsymbol\beta_0^{\rm T}\mathbf{x}_i)\left[\frac{E\left\{\mathbf{X}_{li}Y_i(s) \mid \boldsymbol\beta_0^{\rm T}\mathbf{x}_i\right\}}{E\left\{Y_i(s) \mid \boldsymbol\beta_0^{\rm T}\mathbf{x}_i\right\}}-\mathbf{x}_{li}\right]ds\right)\\
& +o_p(n^{1/2}h^2)\\
=\ & o_p(1),
\end{aligned}$$

where the last equality is because the integrands have mean zero conditional on $\boldsymbol\beta_0^{\rm T}\mathbf{X}$, and the second last equality is obtained following the same derivation of $\mathbf{T}_3$. Using these results in (S5.1), combined with the results on (S5.2), it is now clear that the theorem holds. □

References

Einmahl, U. and D. M. Mason (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Annals of Statistics 22*, 1380–1403.

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

Härdle, W. and T. M. Stoker (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American statistical Association 84*, 986–995.

Mack, Y. P. and B. W. Silverman (1982). Weak and strong uniform consistency of kernel regression estimates. *Probability Theory and Related Fields 61*, 405–415.

Prentice, R. L. and J. D. Kalbfleisch (2003). Mixed discrete and continuous cox regression model. *Lifetime Data Analysis 9*, 195–210.

Table S1: Results of study 1, based on 1000 simulations with sample size 100. "bias" is $|\text{mean}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\beta}|$ of each component in $\boldsymbol{\beta}$, "sd" is the sample standard errors of the corresponding estimation. The last column is the mean and standard errors of the largest singular value of $\widehat{\mathbf{P}} - \mathbf{P}$.

		β_2	β_3	β_4	β_5	β_6	β_7	$\lambda_{\max}$
	true	0	-1	0	1	0	-1	
		No censoring						
Cox	bias	0.0138	0.0567	0.0044	0.0730	0.0007	0.0648	0.3033
	sd	0.2921	0.4441	0.2990	0.4507	0.3107	0.4515	0.0968
AFT	bias	0.6133	0.1859	0.0890	0.4496	0.0415	0.3152	0.3752
	sd	1.4697	2.3427	1.4635	2.5065	1.4187	2.3088	0.1161
hmave	bias	0.0001	0.0966	0.0166	0.0945	0.0008	0.1105	0.3291
	sd	0.3572	0.5363	0.3611	0.5310	0.3604	0.5304	0.1107
semi	bias	0.0056	0.0251	0.0026	0.0021	0.0031	0.0032	0.0980
	sd	0.2500	0.3037	0.2175	0.2821	0.2371	0.3019	0.0864
		20% censoring						
Cox	bias	0.0300	0.1293	0.0154	0.1191	0.0265	0.0783	0.3497
	sd	0.3720	0.5856	0.3906	0.6141	0.3701	0.5639	0.1079
AFT	bias	0.1517	1.2165	0.2537	0.6880	0.3937	0.6235	0.3836
	sd	1.5375	3.9783	1.4569	3.2176	1.5471	3.1561	0.1159
hmave	bias	0.0819	0.1767	0.0190	0.1379	0.0180	0.0913	0.3539
	sd	0.3983	0.6216	0.3889	0.6503	0.3955	0.6180	0.1123
semi	bias	0.0051	0.0150	0.0060	0.0217	0.0080	0.0015	0.1201
	sd	0.2627	0.3112	0.2737	0.3087	0.2406	0.3060	0.1187
		40% censoring						
Cox	bias	0.0149	0.2851	0.0468	0.1855	0.0152	0.2006	0.4390
	sd	0.7109	1.3647	0.7311	1.1360	0.6157	1.0867	0.1289
AFT	bias	0.4269	0.5728	0.3515	0.6415	0.2871	0.9728	0.4446
	sd	2.4898	5.2837	1.8935	3.4309	2.0395	4.8535	0.1335
hmave	bias	0.9139	1.1169	0.0180	0.2058	0.0100	0.2571	0.4871
	sd	1.4238	2.2052	0.7544	1.1944	0.8227	1.3427	0.1341
semi	bias	0.0199	0.0394	0.0080	0.0002	0.0065	0.0083	0.1457
	sd	0.3275	0.3795	0.2899	0.3510	0.2946	0.3560	0.1702

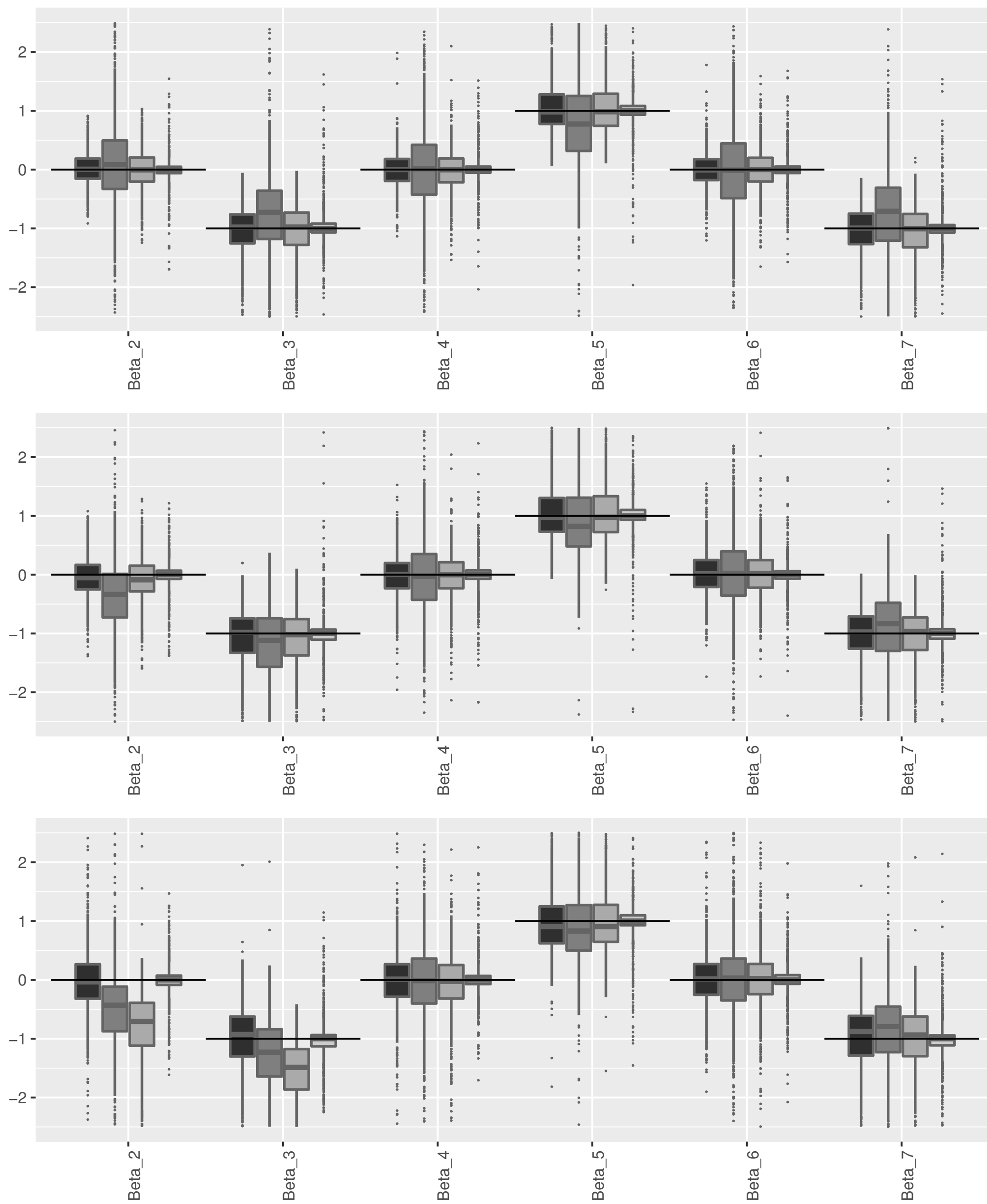

Figure S1: Boxplot of parameter estimation by different methods in study 1. First row: no censoring; Second row: 20% censoring rate; Third row: 40% censoring rate. Solid line: True $\boldsymbol{\beta}$. From left to right in each group: Cox, AFT, hmave, semiparametric.

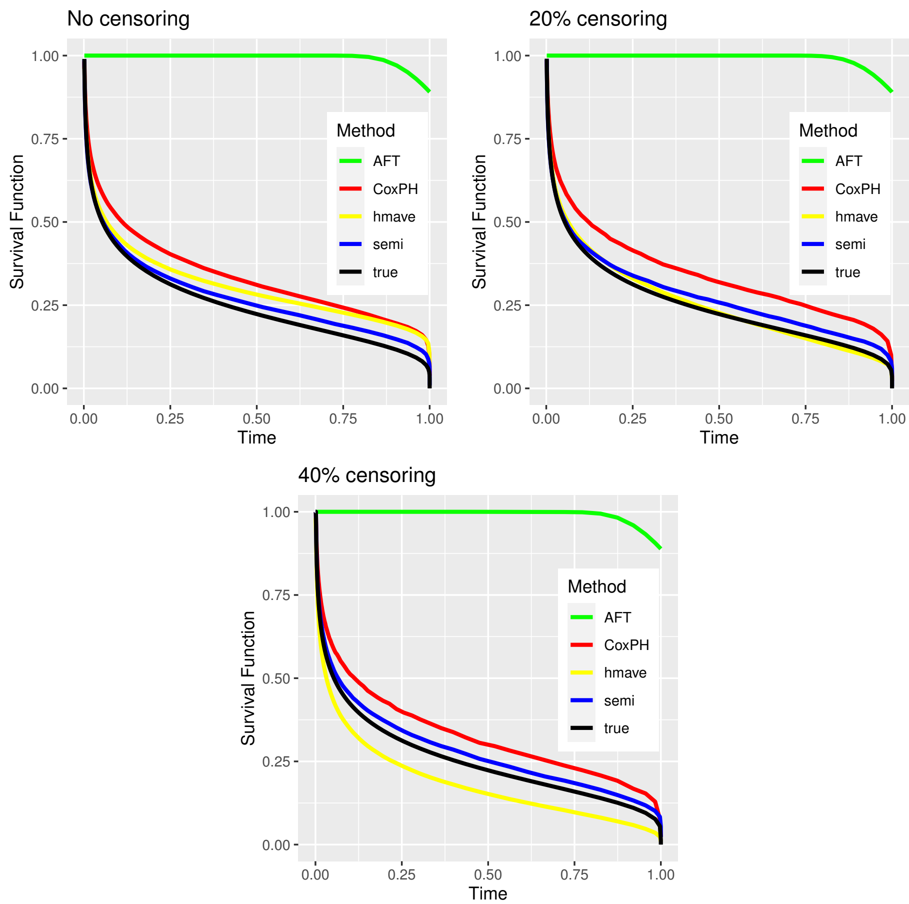

Figure S2: Average estimated survival functions by different methods at different censoring rates in study 1, where $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ is fixed at $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\overline{\mathbf{X}}$.

Table S2: Harrell's concordance index of study 1, based on 1000 simulations with sample size 100. "sd" is the sample standard errors of the concordance index.

Censoring		Cox	AFT	hmave	semi
0%	C-statistics	0.994	0.550	0.986	0.986
	sd	0.005	0.072	0.003	0.003
20%	C-statistics	0.996	0.415	0.981	0.985
	sd	0.002	0.089	0.003	0.003
40%	C-statistics	0.997	0.253	0.987	0.978
	sd	0.003	0.081	0.003	0.004

Table S3: Results of study 2, based on 1000 simulations with sample size 200. "bias" is $|\text{mean}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\beta}|$ of each component in $\boldsymbol{\beta}$, "sd" is the sample standard errors of the corresponding estimation. The last column is the mean and standard errors of the largest singular value of $\widehat{\mathbf{P}} - \mathbf{P}$.

		β_2	β_3	β_4	β_5	β_6	β_7	$\lambda_{\max}$
	true	1.3	−1.3	1	−0.5	0.5	−0.5	
					No censoring			
Cox	bias	0.0518	0.0437	0.0404	0.0020	0.0008	0.0232	0.2263
	sd	0.4444	0.4471	0.3983	0.3152	0.2990	0.3074	0.0643
AFT	bias	0.0208	0.0177	0.0144	0.0024	0.0018	0.0049	0.1472
	sd	0.2667	0.2563	0.2237	0.1818	0.1769	0.1865	0.0449
hmave	bias	0.0115	0.0128	0.0040	0.0033	0.0009	0.0043	0.1517
	sd	0.2708	0.2688	0.2328	0.1949	0.1851	0.1895	0.0458
semi	bias	0.0235	0.0426	0.0139	0.0183	0.0168	0.0131	0.0784
	sd	0.2356	0.2190	0.1982	0.1626	0.1635	0.1633	0.0408
					20% censoring			
Cox	bias	0.0880	0.0681	0.0691	0.0169	0.0380	0.0151	0.2357
	sd	0.4860	0.4734	0.4123	0.3124	0.3219	0.3222	0.0680
AFT	bias	0.0402	0.0304	0.0281	0.0049	0.0138	0.0013	0.1536
	sd	0.2833	0.2795	0.2455	0.1918	0.1861	0.1910	0.0460
hmave	bias	0.0283	0.0269	0.0268	0.0049	0.0055	0.0043	0.1648
	sd	0.2912	0.2974	0.2416	0.2339	0.2478	0.2398	0.0514
semi	bias	0.0356	0.0599	0.0081	0.0195	0.0294	0.0130	0.1170
	sd	0.2946	0.3119	0.2636	0.2103	0.2019	0.2095	0.0448
					40% censoring			
Cox	bias	0.1062	0.1054	0.0915	0.0387	0.0586	0.0201	0.1996
	sd	0.5761	0.5699	0.4965	0.3687	0.3698	0.3478	0.0701
AFT	bias	0.0409	0.0440	0.0217	0.0247	0.0149	0.0136	0.1634
	sd	0.3024	0.2978	0.2579	0.2073	0.2105	0.2050	0.0499
hmave	bias	0.0708	0.0757	0.0571	0.0304	0.0333	0.0248	0.2027
	sd	0.4225	0.4192	0.3748	0.2876	0.2745	0.2760	0.0639
semi	bias	0.0641	0.0681	0.0127	0.0054	0.0210	0.0410	0.1445
	sd	0.3372	0.3202	0.3047	0.2627	0.2366	0.2606	0.0468

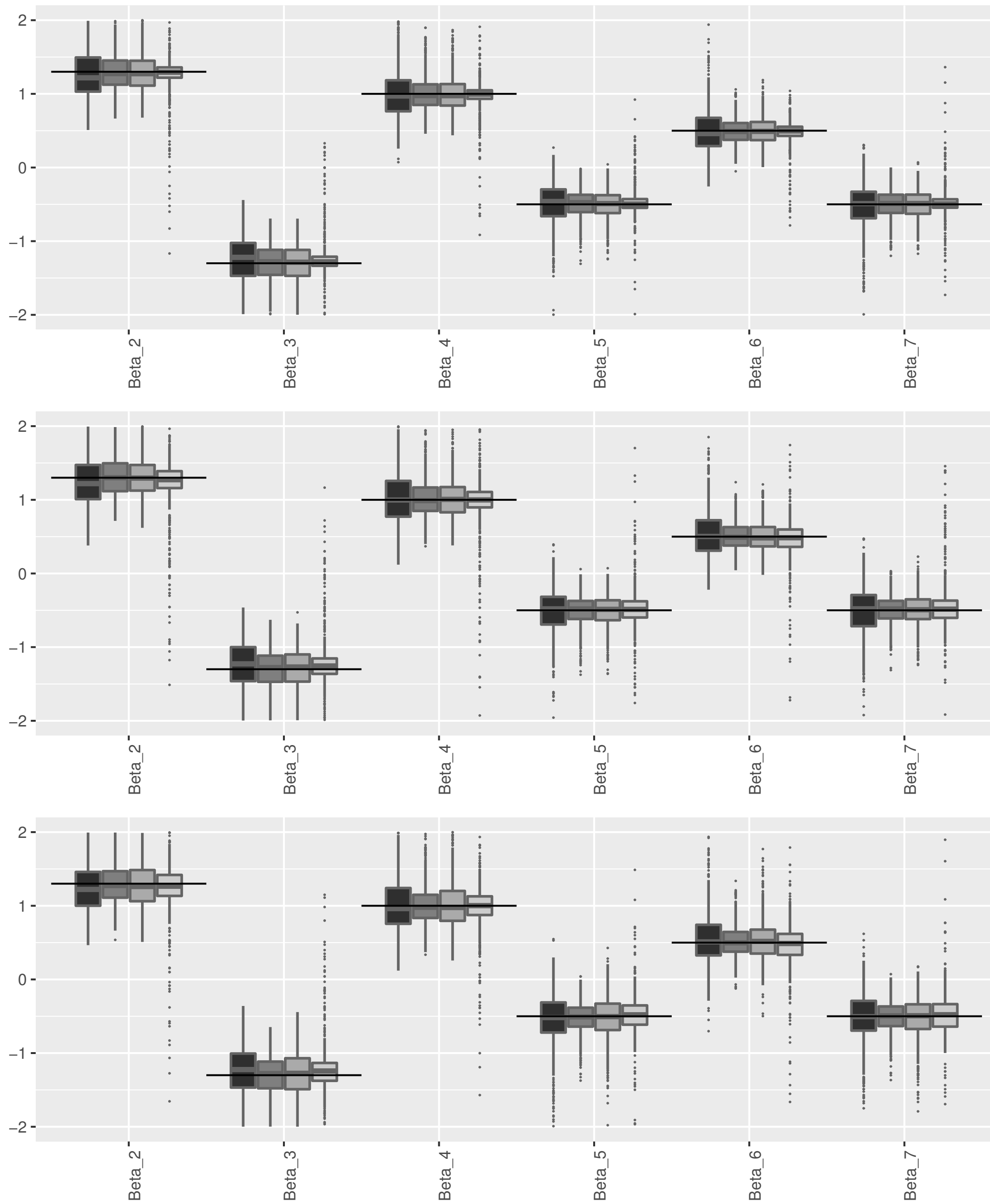

Figure S3: Boxplot of parameter estimation by different methods of study 2. First row: no censoring; Second row: 20% censoring rate; Third row: 40% censoring rate. Solid line: True $\boldsymbol{\beta}$. From left to right in each group: Cox, AFT, hmave, semiparametric.

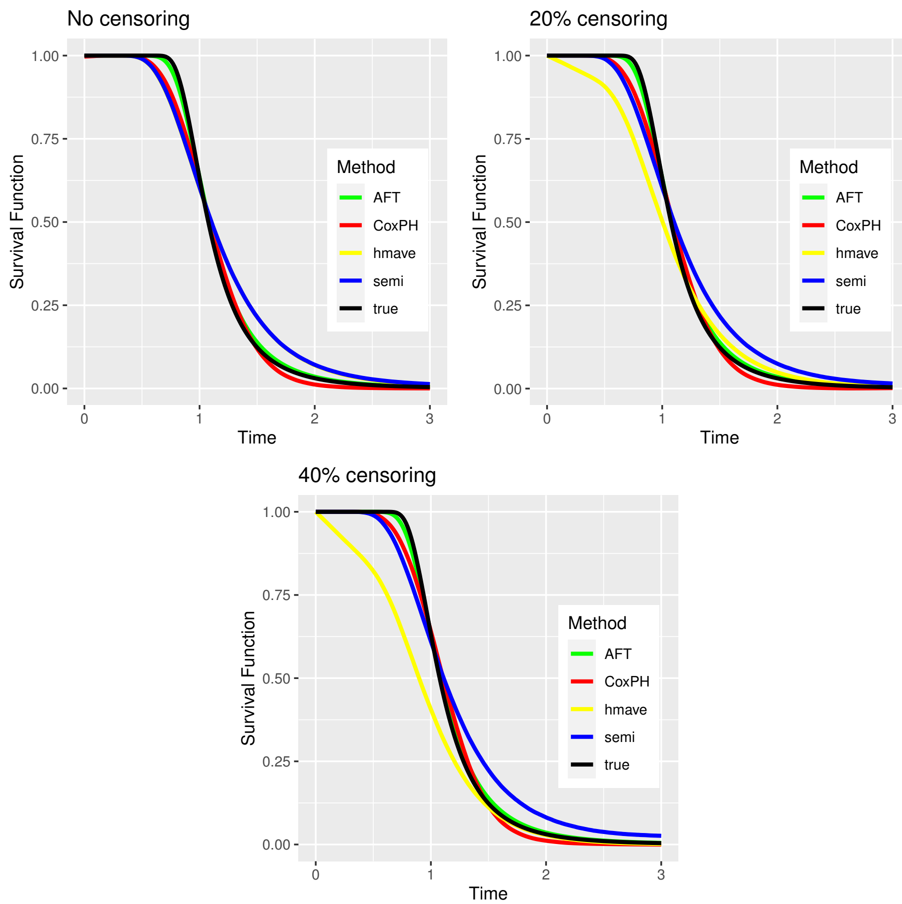

Figure S4: Estimated survival function by different methods at different censoring rates of study 2, where $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ is fixed at $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\overline{\mathbf{X}}$.

Table S4: Harrell's concordance index of study 2, based on 1000 simulations with sample size 200. "sd" is the sample standard errors of the concordance index.

Censoring		Cox	AFT	hmave	semi
0%	C-statistics	0.899	0.971	0.955	0.955
	sd	0.054	0.036	0.010	0.010
20%	C-statistics	0.877	0.992	0.961	0.952
	sd	0.076	0.004	0.015	0.017
40%	C-statistics	0.810	0.999	0.927	0.901
	sd	0.104	0.002	0.032	0.041

Table S5: Results of study 3, based on 1000 simulations with sample size 200. "bias" is $|\text{mean}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\beta}|$ of each component in $\boldsymbol{\beta}$, "sd" is the sample standard errors of the corresponding estimation. The last column is the mean and standard errors of the largest singular value of $\widehat{\mathbf{P}} - \mathbf{P}$.

		β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	$\lambda_{\max}$
	true	-0.6	0	-0.3	-0.1	0	0.1	0.3	0	0.6	
						No censoring					
Cox	bias	0.1409	0.1355	0.5010	0.2523	0.2152	0.0645	0.3055	0.3495	0.8551	0.9336
	sd	5.9041	5.7703	6.0618	7.5865	6.9145	7.8148	5.6968	7.3566	5.2515	0.0779
AFT	bias	0.9970	1.1378	1.1105	0.7777	1.1407	3.2320	0.0519	1.3635	0.3193	0.8755
	sd	7.8813	8.0801	7.9166	8.3889	7.0144	6.5110	7.7910	7.8402	6.6598	0.2551
hmave	bias	0.6192	0.3443	0.1297	0.3207	0.0109	0.1889	0.1482	0.0185	0.2722	0.8624
	sd	5.3614	6.6516	5.9474	5.0169	4.6656	4.0464	4.7291	4.6429	4.6929	0.2438
semi	bias	0.0033	0.0007	0.0009	0.0039	0.0006	0.0001	0.0061	0.0005	0.0119	0.2943
	sd	0.1474	0.1465	0.1502	0.1512	0.1539	0.1479	0.1475	0.1521	0.1495	0.0500
						20% censoring					
Cox	bias	0.5226	0.0787	0.3027	0.0912	0.1738	0.0888	0.4557	0.2642	0.6556	0.8652
	sd	7.5486	6.2857	6.4019	6.9874	5.6022	6.8008	7.3245	6.2297	5.5681	0.2621
AFT	bias	1.7606	0.7222	1.4912	0.8368	0.8282	0.5411	0.7437	0.7244	0.4102	0.8856
	sd	4.5490	4.9955	4.6505	4.9782	5.2053	4.8893	5.0968	5.5299	5.1731	0.2407
hmave	bias	0.4018	0.1721	0.2905	0.2639	0.4307	0.2403	0.0752	0.2478	0.3978	0.8748
	sd	7.0157	7.6713	7.6211	6.4765	6.2337	7.4385	6.2330	5.9572	6.0522	0.2240
semi	bias	0.0062	0.0058	0.0004	0.0042	0.0088	0.0137	0.0072	0.0063	0.0010	0.2930
	sd	0.1513	0.1510	0.1529	0.1538	0.1506	0.1539	0.1461	0.1525	0.1481	0.0528
						40% censoring					
Cox	bias	0.3980	0.2558	0.3340	0.0155	0.1067	0.3787	0.0534	0.7641	0.5346	0.8377
	sd	6.0678	8.6683	9.8791	8.5798	10.271	10.487	10.806	9.2136	7.4471	0.2658
AFT	bias	4.6917	2.2959	1.5812	1.6128	4.8147	4.3596	4.1690	8.7963	4.6270	0.8330
	sd	6.6751	7.8759	7.6864	7.9714	6.3581	8.5983	6.7791	7.6934	8.6455	0.2963
hmave	bias	0.3073	0.2527	0.3905	0.2035	0.0036	0.4633	0.0904	0.1348	0.4520	0.8374
	sd	6.4744	7.3740	11.650	6.5892	6.9468	6.4143	10.415	9.6963	8.0583	0.2660
semi	bias	0.0040	0.0081	0.0055	0.0107	0.0064	0.0024	0.0043	0.0016	0.0079	0.2987
	sd	0.1536	0.1547	0.1549	0.1529	0.1542	0.1499	0.1513	0.1443	0.1549	0.0603

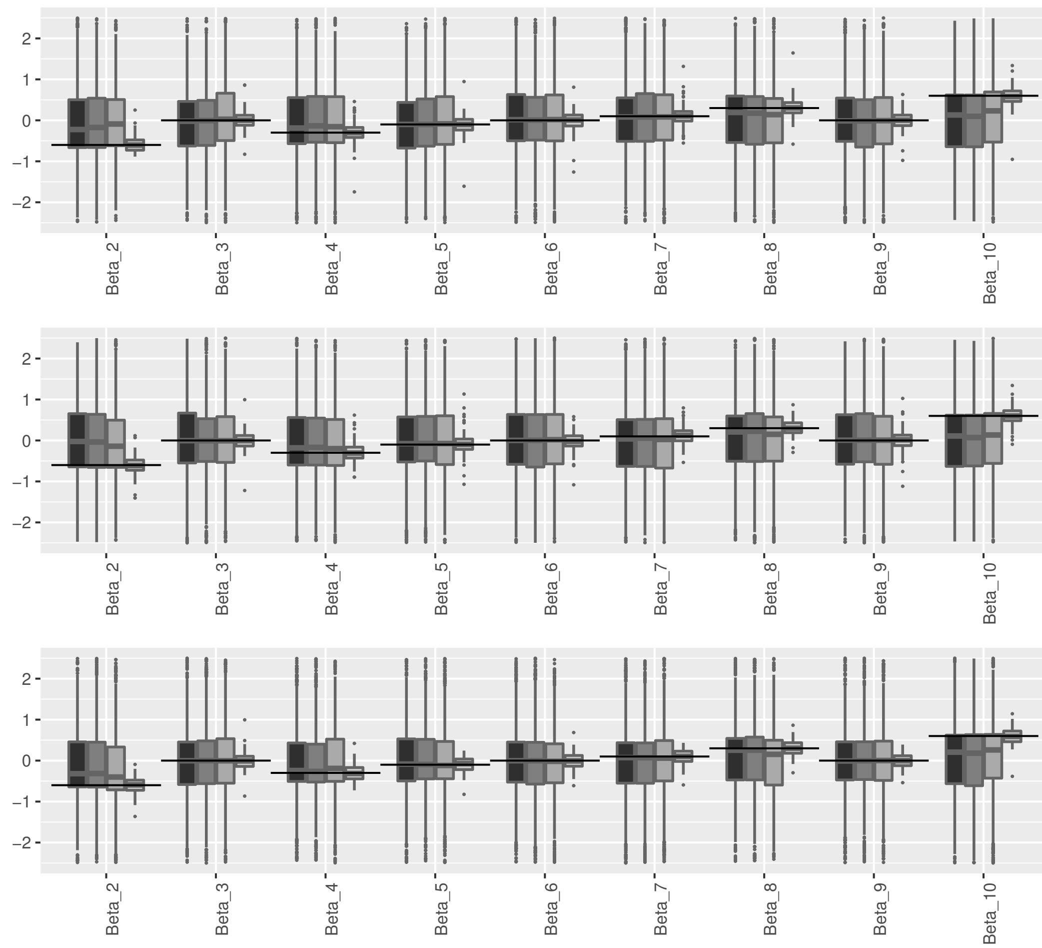

Figure S5: Boxplot of parameter estimation by different methods of study 3. First row: no censoring; Second row: 20% censoring rate; Third row: 40% censoring rate. Solid line: True $\boldsymbol{\beta}$. From left to right in each group: Cox, AFT, hmave, semiparametric.

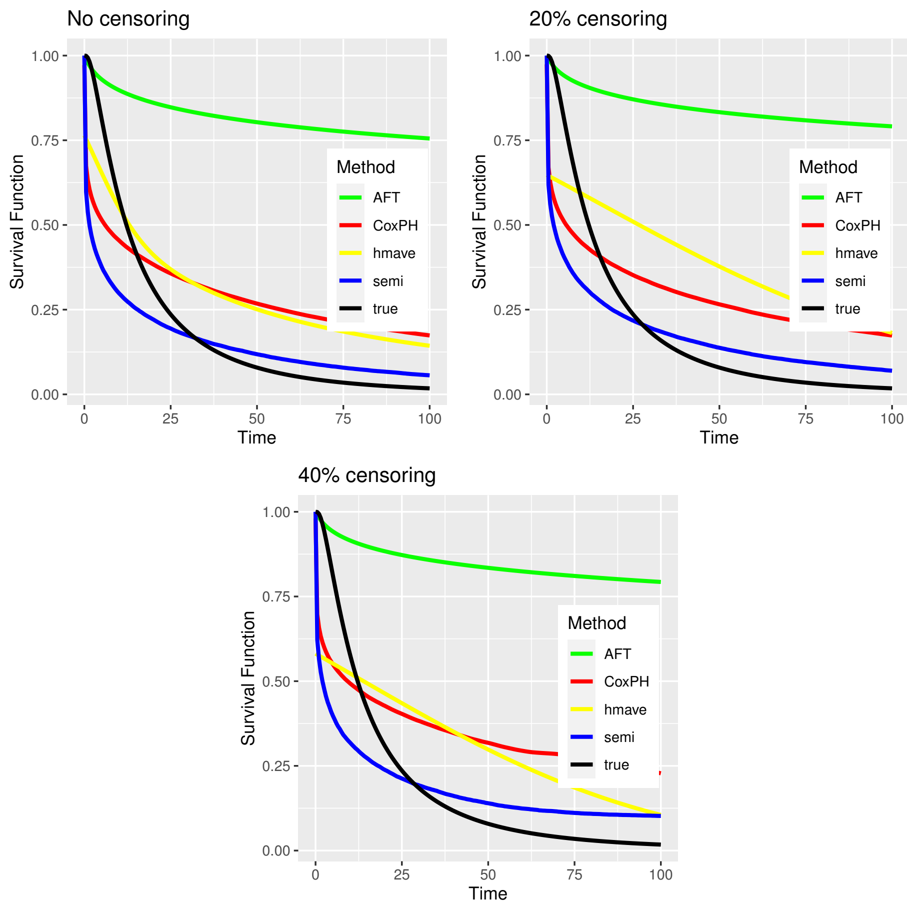

Figure S6: Estimated survival function by different methods at different censoring rates of study 3, where $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ is fixed at $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\overline{\mathbf{X}}$.

Table S6: Harrell's concordance index of study 3, based on 1000 simulations with sample size 200. "sd" is the sample standard errors of the concordance index.

Censoring		Cox	AFT	hmave	semi
0%	C-statistics	0.499	0.900	0.986	0.958
	sd	0.023	0.254	0.036	0.016
20%	C-statistics	0.499	0.903	0.989	0.994
	sd	0.026	0.256	0.005	0.001
40%	C-statistics	0.498	0.915	0.989	0.990
	sd	0.029	0.249	0.006	0.002

Table S7: Results of study 4, based on 1000 simulations with sample size 200. "bias" is $|\text{mean}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\beta}|$ of each component in $\boldsymbol{\beta}$, "sd" is the sample standard errors of the corresponding estimation. The last column is the mean and standard errors of the largest singular value of $\widehat{\mathbf{P}} - \mathbf{P}$.

		$\beta_{3,1}$	$\beta_{4,1}$	$\beta_{5,1}$	$\beta_{6,1}$	$\beta_{3,2}$	$\beta_{4,2}$	$\beta_{5,2}$	$\beta_{6,2}$	$\lambda_{\max}$
	true	2.75	-0.75	-1	2.0	-3.125	-1.125	1.0	-2.0	
		No censoring								
hmave	bias	2.7700	0.1199	0.9817	2.3104	2.7002	0.9829	1.1085	1.6572	0.9218
	sd	6.4580	8.1845	5.8410	6.1477	5.8180	8.8380	6.2825	5.6126	0.0960
semi	bias	0.0090	0.0179	0.0066	0.0105	0.0142	0.0041	0.0029	0.0051	0.0788
	sd	0.3293	0.1781	0.1718	0.2380	0.3167	0.1809	0.1517	0.2309	0.0946
		20% censoring								
hmave	bias	2.9199	0.5533	1.2512	1.8446	2.4579	1.1224	0.7592	1.8994	0.9273
	sd	7.2250	9.8458	8.6384	7.8197	11.622	13.251	11.758	11.751	0.1038
semi	bias	0.0710	0.0098	0.0229	0.0549	0.0560	0.0027	0.0185	0.0389	0.1430
	sd	0.5451	0.3699	0.2798	0.4449	0.6172	0.3884	0.3387	0.4752	0.1780
		40% censoring								
hmave	bias	2.6564	1.4719	1.6397	2.3155	2.9895	1.8283	1.1251	1.4566	0.9173
	sd	14.464	26.730	9.8537	12.913	12.505	31.240	9.4431	12.806	0.1076
semi	bias	0.0529	0.0235	0.0260	0.0310	0.0849	0.0293	0.0365	0.0538	0.1438
	sd	0.7350	0.3982	0.3396	0.5753	0.7750	0.4048	0.3942	0.5257	0.1737

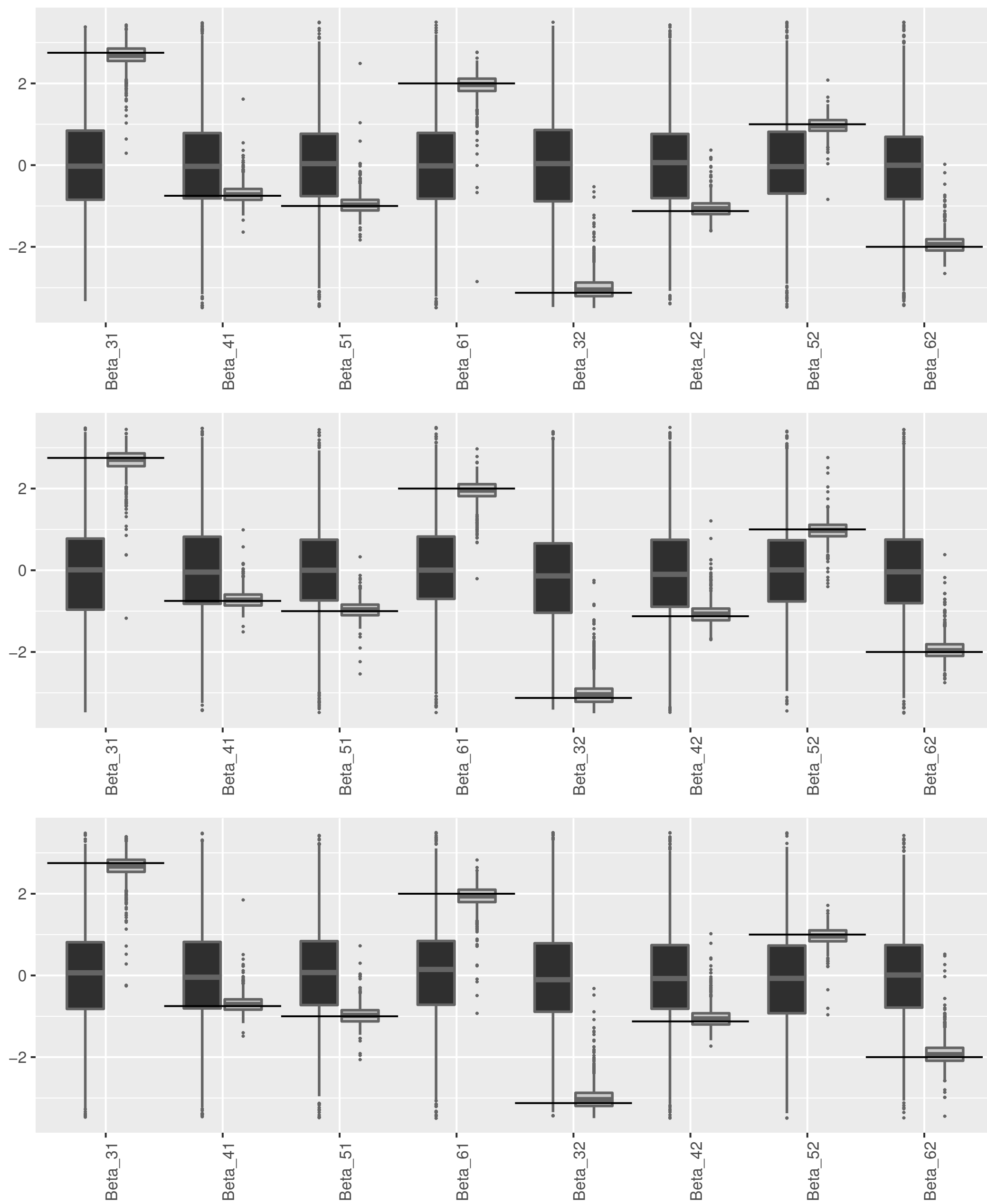

Figure S7: Boxplot of hmave and the semiparametric methods of study 4. First row: no censoring; Second row: 20% censoring rate; Third row: 40% censoring rate. Solid line: True $\boldsymbol{\beta}$. From left to right in each group: Cox, AFT, hmave, semiparametric.

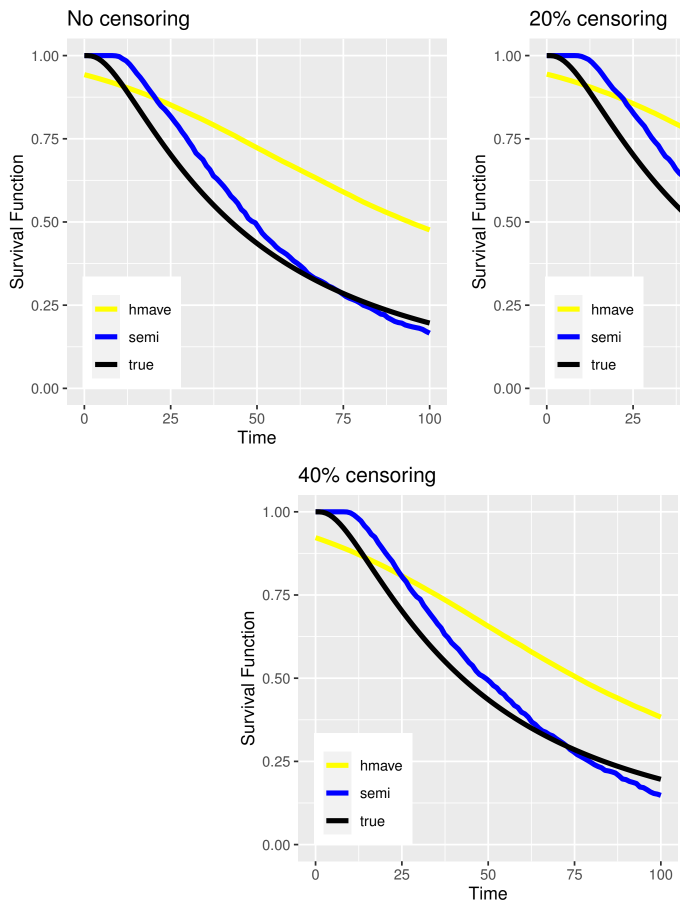

Figure S8: Estimated survival function by different methods at different censoring rates of study 4.

Table S8: Harrell's concordance index of study 4, based on 1000 simulations with sample size 200. "sd" is the sample standard errors of the concordance index.

Censoring		hmave	semi
0%	C-statistics	0.903	0.997
	sd	0.201	0.001
20%	C-statistics	0.920	0.997
	sd	0.130	0.001
40%	C-statistics	0.913	0.993
	sd	0.139	0.001

Table S9: Index misspecification on study 4. $d = 2$ is the correct index, $d = 1$ and $d = 3$ are misspecified indices.

Parameters	d=1	d=2		d=3		
$\boldsymbol{\beta}_{1\cdot}$	1	1	0	1	0	0
$\boldsymbol{\beta}_{2\cdot}$	-0.248(4.08)	0	1	0	1	0
$\boldsymbol{\beta}_{3\cdot}$	-0.696(4.36)	2.665(0.34)	-3.011(0.40)	0	0	1
$\boldsymbol{\beta}_{4\cdot}$	-1.179(4.53)	-0.714(0.20)	-1.052(0.26)	2.750(0.08)	-3.124(0.08)	-0.000(0.08)
$\boldsymbol{\beta}_{5\cdot}$	2.585(4.69)	-0.968(0.20)	0.969(0.23)	-0.744(0.08)	-1.124(0.08)	0.003(0.08)
$\boldsymbol{\beta}_{6\cdot}$	-4.078(4.15)	1.940(0.26)	-1.927(0.29)	-0.996(0.08)	0.998(0.08)	-0.002(0.08)

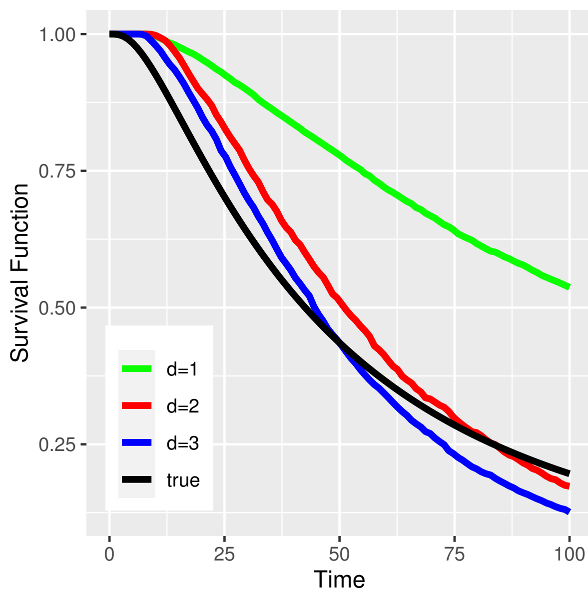

Figure S9: Estimation of the survival functions based on the misspecified d of study 4 at 20% censoring rate.

Table S10: Results of study 5, based on 1000 simulations with sample size 100, 500, 1000 respectively. “bias” is $|\text{mean}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\beta}|$ of each component in $\boldsymbol{\beta}$, “sd” is the sample standard errors of the corresponding estimation, “$\widehat{\sigma}$” is the mean of the estimated standard errors of $\widehat{\boldsymbol{\beta}}$ component, “95%” is the sample coverage of the 95% confidence intervals.

	β_{31} 2.75	β_{41} -0.75	β_{51} -1	β_{61} 2.0	β_{32} -3.125	β_{42} -1.125	β_{52} 1.0	β_{62} -2.0
	$n = 100$							
bias	0.3995	0.5031	0.2066	0.3799	0.5515	0.5349	0.1757	0.3395
sd	0.5760	0.4236	0.4673	0.5608	0.6163	0.4376	0.4772	0.5377
$\widehat{\sigma}$	0.3868	0.3188	0.3312	0.3427	0.3956	0.3131	0.3331	0.3602
95%	0.7100	0.6577	0.8051	0.7034	0.6416	0.6292	0.8089	0.7414
	$n = 500$							
bias	0.1790	0.1258	0.0714	0.1338	0.2100	0.1489	0.07386	0.1340
sd	0.2741	0.1714	0.2177	0.2380	0.2979	0.1897	0.2202	0.2244
$\widehat{\sigma}$	0.1585	0.1371	0.1644	0.1659	0.2683	0.2179	0.2538	0.2558
95%	0.6663	0.8022	0.8298	0.7566	0.8127	0.8773	0.9125	0.8764
	$n = 1000$							
bias	0.0611	0.0492	0.0188	0.0423	0.0695	0.0467	0.0209	0.0448
sd	0.1951	0.1451	0.1555	0.1538	0.1867	0.1433	0.1650	0.1711
$\widehat{\sigma}$	0.1062	0.1113	0.1134	0.1190	0.1823	0.1712	0.1749	0.1740
95%	0.8060	0.8830	0.8783	0.8621	0.9268	0.9705	0.9515	0.9287

Table S11: The estimated coefficients, standard errors and p-value in AIDS data.

	$\widehat{\boldsymbol{\beta}}_{2,1}$	$\widehat{\boldsymbol{\beta}}_{3,1}$	$\widehat{\boldsymbol{\beta}}_{4,1}$	$\widehat{\boldsymbol{\beta}}_{5,1}$	$\widehat{\boldsymbol{\beta}}_{6,1}$	$\widehat{\boldsymbol{\beta}}_{7,1}$	$\widehat{\boldsymbol{\beta}}_{8,1}$	$\widehat{\boldsymbol{\beta}}_{9,1}$	$\widehat{\boldsymbol{\beta}}_{10,1}$	$\widehat{\boldsymbol{\beta}}_{11,1}$	$\widehat{\boldsymbol{\beta}}_{12,1}$	$\widehat{\boldsymbol{\beta}}_{13,1}$
est	0.115	-0.002	0.093	0.088	-0.090	0.231	0.003	-0.178	0.058	-0.031	0.201	0.156
std	0.039	0.039	0.039	0.037	0.043	0.046	0.036	0.046	0.035	0.042	0.033	0.038
p-value	0.003	0.965	0.017	0.017	0.036	0.001	0.928	0.001	0.100	0.457	0.001	0.001

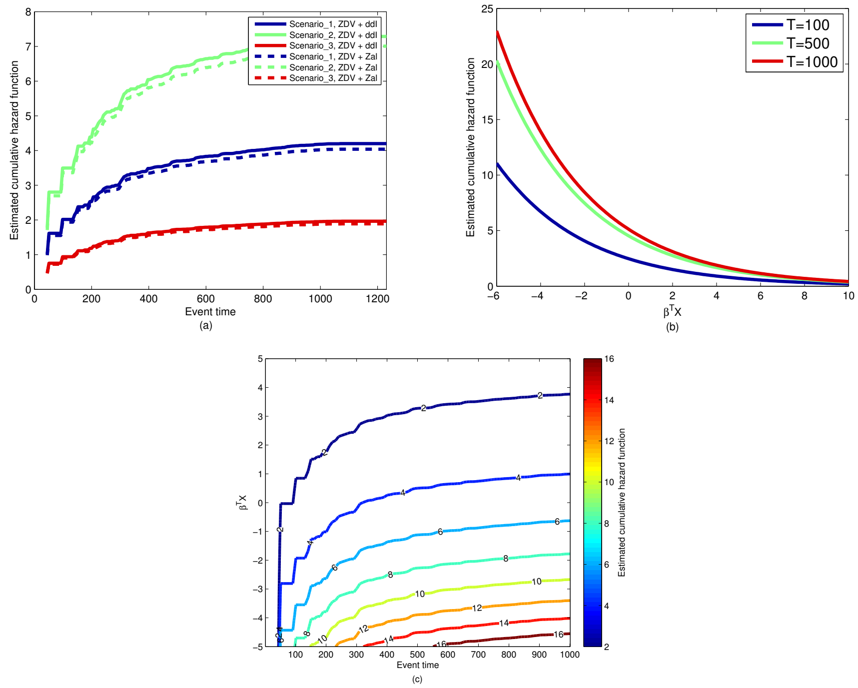

Figure S10: Estimated cumulative hazard function $\widehat{\Lambda}$ in AIDS data. (a). Comparisons of $\widehat{\Lambda}$ as a function of t between treatments ZDV+ddl and ZDV+Zal when other covariates are fixed at three indices. (b). $\widehat{\Lambda}$ as a function of $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}$ at $T = 100, 500, 1000$. (c). Contour plot of $\widehat{\Lambda}$ as a function of T and $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}$.

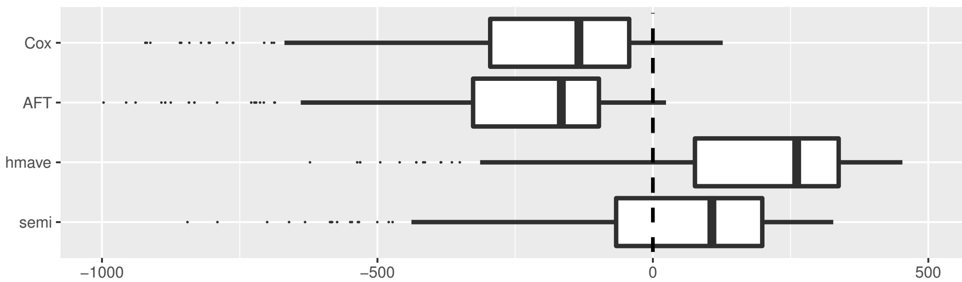

Figure S11: Residuals of predicted survival time by different methods.