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 $\Gamma_1$ is obvious. To obtain $\Gamma_2$, let $h(t, \beta_0^T X, \gamma) = \partial \log \lambda_0(t, \beta_0^T X, \gamma) / \partial \gamma$, where $\lambda_0(t, \beta_0^T X, \gamma)$ is a submodel of $\lambda_0(t, \beta_0^T X)$. Hence,

$$
\frac{\partial \log f(X, Z, \Delta)}{\partial \gamma} = \Delta \frac{\partial \log \lambda_0(Z, \beta_0^T X, \gamma)}{\gamma} - \int_0^Z \frac{\partial \lambda_0(s, \beta_0^T X, \gamma)}{\partial \gamma} ds
$$

$$
= \Delta h(Z, \beta_0^T X, \gamma) - \int_0^Z h(s, \beta_0^T X, \gamma) \lambda_0(s, \beta_0^T X) ds
$$

$$
= \int_0^\infty h(s, \beta_0^T X, \gamma) dM(s, \beta_0^T X).
$$

Because $\lambda_0(t, \beta_0^T X)$ can be any positive function, $h(s, \beta_0^T X, \gamma)$ can be any function. We denote it $h(s, \beta_0^T X)$. This leads to the form of $\Gamma_2$.

Similar derivation leads to $\Gamma_3$. Specifically, to obtain $\Gamma_3$, let $h(t, X, \gamma) = \partial \log \lambda_c(t, X, \gamma) / \partial \gamma$, where $\lambda_c(t, X, \gamma)$ is a submodel of $\lambda_c(t, x)$. Hence,

$$
\frac{\partial \log f(X, Z, \Delta)}{\partial \gamma} = (1 - \Delta) \frac{\partial \log \lambda_c(Z, X, \gamma)}{\gamma} - \int_0^Z \frac{\partial \lambda_c(s, X, \gamma)}{\partial \gamma} ds
$$

$$
= (1 - \Delta) h(Z, X, \gamma) - \int_0^Z h(s, X, \gamma) \lambda_c(s, X) ds
$$
Note that here, $M_c(t, X) = N_c(t) - \int_0^t I(Z \geq s) \lambda_c(s, X) ds$, and despite of the discontinuity at $s = \tau$ for $\lambda_c(t, X)$, 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 $\lambda_c(t, X)$ can be any positive function, $h(s, X, \gamma)$ can be any function. We denote it $h(s, X)$. This leads to the form of $\Gamma_3$.

It is easy to verify that $\Gamma_1 \perp \Gamma_2$ and $\Gamma_1 \perp \Gamma_3$, where $\perp$ stands for orthogonality. Because $C \perp T | X$, the martingale integrations associated with $M(t, \beta_0^T X)$ and $M_C(z, X)$ are also independent conditional on $X$, hence $\Gamma_2 \perp \Gamma_3$. This completes the proof. 

\section*{S2 Proof of Proposition 2}

Denoting the score function in (2.2) at the true coefficient $\beta_0$ as $S_{\beta_0}(\Delta, Z, X)$, we can verify that $S_{\beta_0}(\Delta, Z, X) \perp \Gamma_1$ and $S_{\beta_0}(\Delta, Z, X) \perp \Gamma_3$ due to the martingale properties. Thus to look for the efficient score, we only need to project $S_{\beta}(\Delta, Z, X)$ onto $\Gamma_2$ and calculate its residual.

We search for $h^*(s, \beta_0^T X)$ so that

$$S_{\text{eff}}(\Delta, Z, X) = S_{\beta_0}(\Delta, Z, X) - \int_0^\infty h^*(s, \beta_0^T X) dM(s, \beta_0^T X)$$

is orthogonal to $\Gamma_2$. This entails that for any $h(s, \beta_0^T X)$,

$$0 = E \left[ \int_0^\infty h^T(s, \beta_0^T X) dM(s, \beta_0^T X) \int_0^\infty \left\{ \frac{\lambda_{10}(s, \beta_0^T X)}{\lambda_0(s, \beta_0^T X)} \otimes X_i - h^*(s, \beta_0^T X) \right\} dM(s, \beta_0^T X) \right]$$

$$= E \left[ \int_0^\infty h^T(s, \beta_0^T X) \left\{ \frac{\lambda_{10}(s, \beta_0^T X)}{\lambda_0(s, \beta_0^T X)} \otimes X_i - h^*(s, \beta_0^T X) \right\} Y(s) \lambda_0(s, \beta_0^T X) ds \right].$$
By letting \( h(s, \beta_0^T X) = I(s = t)a(\beta_0^T X) \) for any \( a(\beta_0^T X) \), we obtain that

\[
0 = E \left[ \left( \frac{\lambda_{10}(t, \beta_0^T X)}{\lambda_0(t, \beta_0^T X)} \otimes X_i - h^*(t, \beta_0^T X) \right) Y(t) \lambda_0(t, \beta_0^T X) \right] \beta_0^T X
\]

\[
= E \left[ \left( \frac{\lambda_{10}(t, \beta_0^T X)}{\lambda_0(t, \beta_0^T X)} \otimes X_i - h^*(t, \beta_0^T X) \right) Y(t) \right] \beta_0^T X.
\]

Note that

\[
E \left\{ X_l S_{c}(t, X) \right\} = E \left\{ Y(t) \right\} = E \left\{ S_{c}(t, X) \right\}
\]

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

\[
E \left\{ X_i S_{c}(t, X) \right\} = E \left\{ Y(t) \right\} = E \left\{ S_{c}(t, X) \right\}
\]

for \( t \geq \tau \) so the relation hold on \([0, \infty)\). Note that the ratio can be defined as any function of \( \beta_0^T X \) for \( t > \tau \) and it will not affect the following result because \( dM(t, \beta_0^T X) = 0 \) for any \( t > \tau \).

This leads to

\[
h^*(t, \beta_0^T X) = \frac{\lambda_{10}(t, \beta_0^T X)}{\lambda_0(t, \beta_0^T X)} \otimes E \left\{ X_i S_{c}(t, X) \right\} \beta_0^T X.
\]

Hence, the efficient score is

\[
S_{eff}(\Delta, Z, X) = \int_{0}^{\infty} \lambda_{10}(s, \beta_0^T X) \otimes \left[ X_i - \frac{E \left\{ X_i S_{c}(s, X) \right\} \beta_0^T X}{E \left\{ S_{c}(s, X) \right\} \beta_0^T X} \right] dM(s, \beta_0^T X).
\]

S3 Proof of Lemma 1

For notation convenience, we prove the results for \( d = 1 \) and assume the first component of \( \beta \) 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 \( X \) and \( \beta \) in a local neighborhood of \( \beta_0 \),

\[
\frac{1}{n} \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) = f_{\beta^T X}(\beta^T X) + O_p(n^{-1/2}h^{-1/2} + h^2), \tag{S3.1}
\]

\[
-\frac{1}{n} \sum_{j=1}^{n} K_h'(\beta^T X_j - \beta^T X) = f_{\beta^T X}'(\beta^T X) + O_p(n^{-1/2}h^{-3/2} + h^2). \tag{S3.2}
\]

To see this, we compute the absolute bias of the left hand side of \( \text{[S3.1]} \) as

\[
\left| E \left\{ \frac{1}{n} \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) \right\} - f_{\beta^T X}(\beta^T X) \right| = \left| \int \frac{1}{h} K \left( \frac{\beta^T X_j - \beta^T X}{h} \right) f_{\beta^T X}(\beta^T x_j) d\beta^T x_j - f_{\beta^T X}(\beta^T X) \right| = \left| \int K(u)f_{\beta^T X}(\beta^T X + hu)du - f_{\beta^T X}(\beta^T X) \right| = \left| \int K(u) \left\{ f_{\beta^T X}(\beta^T X) + f_{\beta^T X}(\beta^T X)hu + \frac{1}{2} f_{\beta^T X}'(\beta^T X^+)h^2u^2 \right\} du - f_{\beta^T X}(\beta^T X) \right| \leq \frac{h^2}{2} \sup_{\beta^T X} |f_{\beta^T X}''(\beta^T X)| \int u^2 K(u)du,
\]

where throughout the text, \( \beta^T X^+ \) is on the line connecting \( \beta^T X \) and \( \beta^T X + hu \), and the variance to be

\[
\text{var} \left\{ \frac{1}{n} \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) \right\} = \frac{1}{n} \text{var} K_h(\beta^T X_j - \beta^T X) = \frac{1}{n} \left[ EK_h^2(\beta^T X_j - \beta^T X) - \left\{ EK_h(\beta^T X_j - \beta^T X) \right\}^2 \right] = \frac{1}{n} \left[ \int \frac{1}{h^2} K^2(\beta^T x_j - \beta^T X)/h) f_{\beta^T X}(\beta^T x_j)d\beta^T x_j - f_{\beta^T X}^2(\beta^T X) + O(h^2) \right] = \frac{1}{nh} \int K^2(u)f_{\beta^T X}(\beta^T X + hu)du - \frac{1}{n} f_{\beta^T X}^2(\beta^T X) + O(h^2/n) \leq \frac{1}{nh} f_{\beta^T X}(\beta^T X) \int K^2(u)du + \frac{h}{2n} \sup_{\beta^T X} |f_{\beta^T X}''(\beta^T X)| \int u^2 K^2(u)du \times + \frac{1}{n} |f_{\beta^T X}'(\beta^T X)| + O(h^2/n).
Therefore, we have
\[
\frac{1}{n} \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) = f_{\beta^TX}(\beta^T X) + O_p(n^{-1/2}h^{-1/2} + h^2)
\]
for all \(\beta\) and for all \(h\) that satisfies Condition \(\text{C2}\) under conditions \(\text{C1, C2, and C4}\). Note that
Condition \(\text{C4}\) also holds for any \(\beta\) in a local neighborhood of \(\beta_0\) 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
\[
\hat{E} \left\{ XY(Z) \mid \beta^T X, \beta \right\} / \partial \beta^T X,\]
and combine these to obtain \((3.8)\). We have
\[
E \left\{ -\frac{1}{n} \sum_{j=1}^{n} X_j I(Z_j \geq Z) K'_h(\beta^T X_j - \beta^T X) \right\} = E \left\{ -X_j I(Z_j \geq Z) K'_h(\beta^T X_j - \beta^T X) \right\} = -\frac{1}{h} \int (\beta^T X + hu, x_{ji}^T) K'(u) S_c(Z, x_{ji}) S(Z, \beta^T X + hu, \beta) \times f_{X_i | \beta^TX, \beta}(x_{ji}) | \beta^T X + hu, \beta) f_{\beta^TX}(\beta^T X + hu) dx_{ji} du
\]
\[
= \int \frac{\partial}{\partial \beta^TX} \left\{ (\beta^T X, x_{ji}^T)^T S_c(Z, x_{ji}) S(Z, \beta^T X, \beta) f_{X_i | \beta^TX, \beta}(x_{ji}) | \beta^T X, \beta) \right\} dx_{ji}
\]
\[
- \frac{h^2}{3} \int \frac{\partial^3}{\partial (\beta^T X)^3} \left\{ (\beta^T X, x_{ji}^T)^T S_c(Z, x_{ji}) S(Z, \beta^T X^*, \beta) \times f_{X_i | \beta^TX, \beta}(x_{ji}) | \beta^T X^*, \beta) f_{\beta^TX}(\beta^T X^*) \right\} du dx_{ji} u^3 K'(u) du
\]
\[
= \frac{\partial}{\partial \beta^TX} f_{\beta^TX}(\beta^T X) E \{ X_j I(Z_j \geq Z) \mid \beta^T X\} - \frac{h^2}{3} \int \frac{\partial^3}{\partial (\beta^T X)^3} f_{\beta^TX}(\beta^T X^*) E \{ X_j I(Z_j \geq Z) \mid \beta^T X^*\} u^3 K'(u) du.
\]
Note that the third variable \(\beta\) in \(f_{X_i | \beta^TX, \beta}(\cdot)\) and \(S(\cdot)\) indicates that the functional forms differ
as \(\beta\) changes. Hence, the absolute bias is
\[
\left| E \left\{ -\frac{1}{n} \sum_{j=1}^{n} X_j I(Z_j \geq Z) K'_h(\beta^T X_j - \beta^T X) \right\} \right|
\]
The variance is

\[
\text{var} \left\{ -\frac{1}{n} \sum_{j=1}^{n} X_j I(Z_j \geq Z) K'_n(\beta^T X_j - \beta^T X) \right\}
\]

\[
= \frac{1}{n} \left[ E \left\{ X_j I(Z_j \geq Z) K'_n(\beta^T X_j - \beta^T X) \right\} \{ X_j I(Z_j \geq Z) K'_n(\beta^T X_j - \beta^T X) \}^T \right.
- \left. \left\{ E X_j I(Z_j \geq Z) K'_n(\beta^T X_j - \beta^T X) \right\} \{ E X_j I(Z_j \geq Z) K'_n(\beta^T X_j - \beta^T X) \}^T \right]
\]

\[
= \frac{1}{n} \int \frac{1}{h^3} (\beta^T X + hu, x^T_j) (\beta^T X + hu, x^T_j) S(Z, x_j) S(Z, \beta^T X + hu, \beta) \times f_{X_j | \beta^T X, \beta}(x_j | \beta^T X + hu, \beta) f_{\beta^T X}(\beta^T X + hu) dx_j K'^2(u) du + O(1/n)
\]

\[
= \frac{1}{nh^3} \int (\beta^T X, x^T_j) (\beta^T X, x^T_j) S(Z, x_j) S(Z, \beta^T X, \beta) \times f_{X_j | \beta^T X, \beta}(x_j | \beta^T X, \beta) f_{\beta^T X}(\beta^T X) dx_j \int K'^2(u) du 
+ \frac{1}{2nh^3} \frac{\partial^2}{\partial(\beta^T X)^2} \int (\beta^T X, x^T_j) (\beta^T X, x^T_j) S(Z, x_j) S(Z, \beta^T X, \beta) \times f_{\beta^T X}(\beta^T X) f_{X_j | \beta^T X, \beta}(x_j | \beta^T X, \beta) dx_j h^2 u^2 K'^2(u) du + O(1/n)
\]

\[
\leq \frac{1}{nh^3} \sup_{\beta^T X} \left| f_{\beta^T X}(\beta^T X) E\{ X_j X^T_j I(Z_j \geq Z) | \beta^T X \} \right| \int K'^2(u) du 
+ \frac{1}{2nh^3} \sup_{\beta^T X} \left| \frac{\partial^2}{\partial(\beta^T X)^2} f_{\beta^T X}(\beta^T X) E\{ X_j X^T_j I(Z_j \geq Z) | \beta^T X^* \} \right| \int u^2 K'^2(u) du
\]

\[
+ O(1/n).
\]

So

\[
- \frac{1}{n} \sum_{j=1}^{n} X_j I(Z_j \geq Z) K'_n(\beta^T X_j - \beta^T X)
\]

\[
= - \frac{\partial}{\partial \beta^T X} f_{\beta^T X}(\beta^T X) E\{ X_j I(Z_j \geq Z) | \beta^T X \} + O_p(n^{-1/2} h^{-3/2} + h^2). \tag{3.3}
\]
Following similar derivations, we have

\[
E \left\{ \frac{1}{n} \sum_{j=1}^{n} X_j I(Z_j \geq Z) K_h(\beta^T X_j - \beta^T X) \right\} = f_{\beta^T X}(\beta^T X) E\{X_j I(Z_j \geq Z) | \beta^T X\} \\
+ \frac{h^2}{2} \int \frac{\partial^2}{\partial(\beta^T X)^2} f_{\beta^T X}(\beta^T X') E \left\{ X_j I(Z_j \geq Z) | \beta^T X' \right\} u^2 K(u) du
\]

and the absolute bias is

\[
\left| E \left\{ \frac{1}{n} \sum_{j=1}^{n} X_j I(Z_j \geq Z) K_h(\beta^T X_j - \beta^T X) \right\} - f_{\beta^T X}(\beta^T X) E\{X_j I(Z_j \geq Z) | \beta^T X\} \right| \leq \frac{h^2}{2} \sup_{\beta^T X} \left| \frac{\partial^2}{\partial(\beta^T X)^2} f_{\beta^T X}(\beta^T X') E \left\{ X_j I(Z_j \geq Z) | \beta^T X' \right\} \right| \int u^2 K^2(u) du .
\]

The variance term satisfies

\[
\text{var} \left\{ \frac{1}{n} \sum_{j=1}^{n} X_j I(Z_j \geq Z) K_h(\beta^T X_j - \beta^T X) \right\} \leq \frac{1}{nh} \sup_{\beta^T X} \left| f_{\beta^T X}(\beta^T X) E\{X_j I(Z_j \geq Z) | \beta^T X\} \right| \int K^2(u) du \\
+ \frac{h}{2n} \sup_{\beta^T X} \left| \frac{\partial^2}{\partial(\beta^T X)^2} f_{\beta^T X}(\beta^T X') E \left\{ X_j I(Z_j \geq Z) | \beta^T X' \right\} \right| \int u^2 K^2(u) du \\
+ O(1/n).
\]

So

\[
\frac{1}{n} \sum_{j=1}^{n} X_j I(Z_j \geq Z) K_h(\beta^T X_j - \beta^T X) = f_{\beta^T X}(\beta^T X) E\{X_j I(Z_j \geq Z) | \beta^T X\} \\
+ O_p(n^{-1/2} h^{-1/2} + h^2).
\]

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

\[
\frac{\partial}{\partial\beta^T X} \hat{E} \left\{ XY(Z) | \beta^T X, \beta \right\} = -\frac{\sum_{j=1}^{n} X_j I(Z_j \geq Z) K_h(\beta^T X_j - \beta^T X)}{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X)}
\]
Using the same technique in the proof of (3.8) we have for all $\beta$.

\begin{align*}
\frac{\{\sum_{i=1}^{n} X_i I(Z_i \geq Z) K_h(\beta^T X_i - \beta^T X)\} \{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X)\}}{\{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X)\}^2}
&= \frac{\partial f_{\beta^T X}(\beta^T X) E\{X_i I(Z_i \geq Z) | \beta^T X\} / \partial \beta^T X + O_p(n^{-1/2} h^{-3/2} + h^2)}{f_{\beta^T X}(\beta^T X) + O_p(n^{-1/2} h^{-1/2} + h^2)}
+ \frac{[f_{\beta^T X}(\beta^T X) E\{X_i I(Z_i \geq Z) | \beta^T X\} + O_p(n^{-1/2} h^{-3/2} + h^2)]}{f_{\beta^T X}(\beta^T X) + O_p(n^{-1/2} h^{-1/2} + h^2)}
\times [-f_{\beta^T X}(\beta^T X) + O_p(n^{-1/2} h^{-3/2} + h^2)]
\frac{\partial}{\partial \beta^T X} E\{XY(Z) | \beta^T X\} + O_p\left((nh^3)^{-1/2} + h^2\right)
\end{align*}

for all $\beta$ 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 $A = n^{-1} \sum_{j=1}^{n} I(Z_i \geq Z) K_h(\beta^T X_j - \beta^T X) - f_{\beta^T X}(\beta^T X) E\{I(Z \geq Z_i) | \beta^T X\}$.

\begin{align*}
\hat{\lambda}(Z, \beta^T X, \beta) &= \frac{\sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X)}{\sum_{j=1}^{n} I(Z_j \geq Z_i) K_h(\beta^T X_j - \beta^T X)}
\frac{\sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X)}{f_{\beta^T X}(\beta^T X) E\{I(Z \geq Z_i) | \beta^T X\} + A}
= \frac{\sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X)}{f_{\beta^T X}(\beta^T X) E\{I(Z \geq Z_i) | \beta^T X\} + A}
\end{align*}

We first inspect

\begin{align*}
\frac{1}{n} \sum_{i=1}^{n} \frac{K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X)}{f_{\beta^T X}(\beta^T X) E\{I(Z \geq Z_i) | \beta^T X\}}
\end{align*}

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

\begin{align*}
E \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X)}{f_{\beta^T X}(\beta^T X) E\{I(Z \geq Z_i) | \beta^T X\}} \right]
= \lambda(Z, \beta^T X, \beta) + \frac{h^2 \beta^2}{20Z^2} \int \int \frac{K(v)K(u)}{S(Z^*, \beta^T X, \beta)} f(Z^*, \beta^T X, \beta) v^2 du dv
+ \frac{h^2 \beta^2}{20(\beta^T X)^2} \int \int \frac{K(v)K(u)E\{S(Z, X_i) | \beta^T X^*, \beta\} f(Z, \beta^T X^*, \beta)}{f_{\beta^T X}(\beta^T X) S(Z, \beta^T X, \beta) E\{S(Z, X_i) | \beta^T X\}} \times f_{\beta^T X}(\beta^T X^*) u^2 du dv,
\end{align*}
where throughout the text, \( Z^* \) is on the line connecting \( Z \) and \( Z + bv \). Thus, the absolute bias is

\[
E \left\{ \frac{1}{n} \sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X) \right\} - \lambda(Z, \beta^T X, \beta)
\]

\[
\leq b^2 \sup_{Z^*, \beta^T X} \left| \frac{\partial^2 f(Z^*, \beta^T X, \beta)}{2\partial Z^2} \right| \int u^2 K(u) du
\]

\[
+ h^2 \sup_{Z^*, \beta^T X} \left| \frac{\partial^2 E(S_i(Z, X_i) | \beta^T X^* f_{\beta^T X}(\beta^T X^*)}{2\partial(\beta^T X)^2} \right| \int u^2 K(u) du
\]

\[
= O(h^2 + b^2)
\]

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

\[
\frac{1}{n} \sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X) f_{\beta^T X}(\beta^T X) E[I(Z \geq Z_i) | \beta^T X] O_p(A) = O_p\{(nh)^{-1/2} + h^2\},
\]

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

\[
\text{var} \left\{ \hat{\lambda}(Z, \beta^T X, \beta) \right\}
\]

\[
= \text{var} \left[ \frac{1}{n} \sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X) f_{\beta^T X}(\beta^T X) E[I(Z \geq Z_i) | \beta^T X] + A \right]
\]

\[
= \text{var} \left[ \frac{1}{n} \sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X) f_{\beta^T X}(\beta^T X) E[I(Z \geq Z_i) | \beta^T X] \right] \{1 + O_p(A)\}
\]

\[
\leq 2\text{var} \left[ \frac{1}{n} \sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X) f_{\beta^T X}(\beta^T X) E[I(Z \geq Z_i) | \beta^T X] \right]
\]

\[
+ 2\text{var} \left[ \frac{1}{n} \sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X) f_{\beta^T X}(\beta^T X) E[I(Z \geq Z_i) | \beta^T X] \right] O_p(A)
\]

We inspect the first term first.

\[
2\text{var} \left[ \frac{1}{n} \sum_{i=1}^{n} K_h(Z_i - Z) \Delta_i K_h(\beta^T X_i - \beta^T X) f_{\beta^T X}(\beta^T X) E[I(Z \geq Z_i) | \beta^T X] \right]
\]
For the second term under conditions C1–C6. Summarizing the above results, the variance of $\hat{\lambda}(Z, \beta^T \mathbf{X}, \beta) = \lambda^2$ is of

$$
\begin{align*}
&= \frac{2}{n} \left( E \left[ \frac{K_h(Z_i - Z) \Delta K_h(\beta^T \mathbf{X}_i - \beta^T \mathbf{X})}{\beta^T \mathbf{X}_i} \right] \right) + O(1/n) \\
&= \frac{2}{bhn} \int \int f_{\beta^T \mathbf{X}}(\beta^T \mathbf{X})K^2(v)K^2(u) dv du \\
&+ \frac{\hat{b}^2}{nh} \int \int f_{\beta^T \mathbf{X}}(\beta^T \mathbf{X})S^2(Z, \beta^T \mathbf{X}, \beta)E\{S_{c}(Z, \mathbf{X}_i)|\beta^T \mathbf{X}\} dv du \\
&+ \frac{\hat{h}^2}{nh} \int \int f_{\beta^T \mathbf{X}}(\beta^T \mathbf{X})f(Z, \beta^T \mathbf{X}, \beta)E\{S_{c}(Z, \mathbf{X}_i)|\beta^T \mathbf{X}\} \times K^2(v)K^2(u)u^2 du + O(1/n) \\
&= O\{1/(nhb) + h/(nb) + b/(nh) + 1/n\} \\
&= O\{1/(nhb)\}.
\end{align*}
$$

For the second term

$$
\begin{align*}
&= 2 \text{var} \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{K_h(Z_i - Z) \Delta K_h(\beta^T \mathbf{X}_i - \beta^T \mathbf{X})}{\beta^T \mathbf{X}_i} O_p(A) \right] \\
&\leq 2 E \left( \frac{1}{n} \sum_{i=1}^{n} \frac{K_h(Z_i - Z) \Delta K_h(\beta^T \mathbf{X}_i - \beta^T \mathbf{X})}{\beta^T \mathbf{X}_i} \right)^2 O_p(\binom{h^2}{A}) \\
&= 2 E \left( \frac{1}{n} \sum_{i=1}^{n} \frac{K_h(Z_i - Z) \Delta K_h(\beta^T \mathbf{X}_i - \beta^T \mathbf{X})}{\beta^T \mathbf{X}_i} \right)^2 O_p((nh)^{-1} + h^4) \\
&= 2 \int \int \frac{f(Z, \beta^T \mathbf{X}, \beta)}{bhn} \frac{f_{\beta^T \mathbf{X}}(\beta^T \mathbf{X})S^2(Z, \beta^T \mathbf{X}, \beta)E\{S_{c}(Z, \mathbf{X}_i)|\beta^T \mathbf{X}\}}{\beta^T \mathbf{X}_i} \times K^2(u)du \times K^2(v)du + O(n^{-1}b^{-1}h + n^{-1}h^{-1}b) O((nh)^{-1} + h^4) \\
&= O\{(nh)^{-2}b^{-1} + n^{-1}h^{-1}b^{-1}\}
\end{align*}
$$

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

$$\hat{\lambda}(Z, \beta^T X, \beta) = \lambda(Z, \beta^T X, \beta) + O_p\{(n_{hh})^{-1/2} + h^2 + b^2\}$$

under condition C1–C6.

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

$$\hat{\lambda}_{11} = -\sum_{i=1}^n K_h(Z_i - Z) \frac{\Delta_i \lambda_i'(\beta^T X_i - \beta^T X)}{\sum_j I(Z_j \geq Z) K_h(\beta^T X_j - \beta^T X)}$$

$$\hat{\lambda}_{12} = -\sum_{i=1}^n K_h(Z_i - Z) \Lambda_i K_h(\beta^T X_i - \beta^T X) \frac{\sum_{j=1}^n I(Z_j \geq Z) \lambda_i'(\beta^T X_j - \beta^T X)}{(\sum_{j=1}^n I(Z_j \geq Z) K_h(\beta^T X_j - \beta^T X))^2}.$$

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

$$\hat{\lambda}_{11} = \frac{\partial}{\partial \beta} \left[ f(Z, \beta^T X, \beta) E\{S_i(Z, X_i) \mid \beta^T X\} f_{\beta^T X}(\beta^T X) S(Z, \beta^T X, \beta) E\{S_i(Z, X_i) \mid \beta^T X\} \right] / \partial \beta^T X + O_p\{(n_{hh})^{-1/2} + b^2 + h^2\},$$

$$\hat{\lambda}_{12} = \frac{\partial}{\partial \beta} \left[ f_{\beta^T X}(\beta^T X) S(Z, \beta^T X, \beta) E\{S_i(Z, X_i) \mid \beta^T X\} \right] / \partial \beta^T X \frac{f(Z, \beta^T X, \beta)}{f_{\beta^T X}(\beta^T X) S^2(Z, \beta^T X, \beta) E\{S_i(Z, X_i) \mid \beta^T X\}} + O_p\{(n_{hh})^{-1/2} + b^2 + h^2\}.$$

In addition, we have

$$\frac{\partial}{\partial \beta} \left[ f(Z, \beta^T X, \beta) E\{S_i(Z, X_i) \mid \beta^T X\} f_{\beta^T X}(\beta^T X) S(Z, \beta^T X, \beta) E\{S_i(Z, X_i) \mid \beta^T X\} \right] / \partial \beta^T X = \frac{\partial f(Z, \beta^T X, \beta)}{\partial \beta^T X} \frac{\partial E\{S_i(Z, X_i) \mid \beta^T X\}}{\partial \beta^T X} + \frac{f(Z, \beta^T X, \beta)}{S(Z, \beta^T X, \beta)} \frac{\partial E\{S_i(Z, X_i) \mid \beta^T X\}}{\partial \beta^T X}$$
Summarizing the results above, the estimator \( \hat{\lambda}_1(Z, \beta^T X, \beta) \) satisfies

\[
\hat{\lambda}_1(Z, \beta^T X, \beta) = \lambda_1(Z, \beta^T X, \beta) + O_p\left\{ (nh^3)^{-1/2} + h^2 + b^2 \right\}
\]

for all \( \beta \) and for all \( h \) and \( b \) that satisfy Condition (7).

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 H"ardle and Stoker (1989). For notation simplicity, we let \( \hat{f}(\beta^T X) \equiv 1/n \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) \). The absolute bias of the trimmed estimator is given by

\[
E \left[ \frac{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X)} I \left\{ \hat{f}(\beta^T X) > d_n \right\} \right] - E\{I(Z_j \geq Z) \mid \beta^T X \} \\
\leq E \left[ \frac{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X)} I \left\{ \hat{f}(\beta^T X) > d_n \right\} \right] - E \left[ \frac{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X)} I \left\{ f_{\beta^T X}(\beta^T X) > d_n \right\} \right] \\
+ E \left[ \frac{n^{-1} \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\int_{\beta^T X} f_{\beta^T X}(\beta^T X) \, d\beta} I \left\{ f_{\beta^T X}(\beta^T X) > d_n \right\} \right] - E\{I(Z_j \geq Z) \mid \beta^T X \}.
\]

The first term satisfies

\[
E \left[ \frac{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X)} I \left\{ \hat{f}(\beta^T X) > d_n \right\} \right] - E \left[ \frac{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X)} I \left\{ f_{\beta^T X}(\beta^T X) > d_n \right\} \right]
\]
\[ \leq E \left[ \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z) \right] \]
\[ \times I \left\{ \hat{f}(\beta^T X) > d_n, f_{\beta^T X}(\beta^T X) \leq d_n \right\} \]
\[ + E \left[ \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z) \right] \]
\[ \times I \left\{ \hat{f}(\beta^T X) \leq d_n, f_{\beta^T X}(\beta^T X) > d_n \right\} \]
\[ \leq E \left[ I \left\{ \hat{f}(\beta^T X) > d_n, f_{\beta^T X}(\beta^T X) \leq d_n \right\} \right] \]
\[ + E \left[ I \left\{ \hat{f}(\beta^T X) \leq d_n, f_{\beta^T X}(\beta^T X) > d_n \right\} \right] \]
\[ = O_p\{n^{-\epsilon} + h^2 + (nh)^{-1/2}\} \]
\[ = O_p\{h^2 + (nh)^{-1/2}\}. \]

The second term is

\[ \left| E \left[ \frac{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X)} I \left\{ \hat{f}_{\beta^T X}(\beta^T X) > d_n \right\} \right] \right| \]
\[ \leq E \left[ \frac{1/n \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\hat{f}_{\beta^T X}(\beta^T X)} \right] \left\{ \hat{f}_{\beta^T X}(\beta^T X) > d_n \right\} \]
\[ - E \left[ \frac{1/n \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\hat{f}_{\beta^T X}(\beta^T X) + O_p\{h^2 + (nh)^{-1/2}\}} \right] \left\{ \hat{f}_{\beta^T X}(\beta^T X) > d_n \right\} \]
\[ \leq E \left[ \frac{1/n \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\hat{f}_{\beta^T X}(\beta^T X)} \right] \left\{ \hat{f}_{\beta^T X}(\beta^T X) > d_n \right\} \]
\[ \times O_p\{h^2 + (nh)^{-1/2}\}. \]

The third term is

\[ \left| E \left[ \frac{1/n \sum_{j=1}^{n} K_h(\beta^T X_j - \beta^T X) I(Z_j \geq Z)}{\hat{f}_{\beta^T X}(\beta^T X)} I \left\{ \hat{f}_{\beta^T X}(\beta^T X) > d_n \right\} \right] \right| \]
\[ - E \left( I(Z_j \geq Z) \right) \left| \hat{f}(\beta^T X) \right| \]
\[ \begin{align*}
&= E \left[ \frac{K_h(\beta^T X_j - \beta^T X)I(Z_j \geq Z)}{f_{\beta^T X}(\beta^T X)} I \left\{ f_{\beta^T X}(\beta^T X) > d_n \right\} \right] \\
&\quad - E(I(Z_j \geq Z) | \beta^T X) \\
&= E \left[ E \left[ I(Z_j \geq Z) I \left\{ f_{\beta^T X}(\beta^T X) > d_n \right\} | \beta^T X \right] - E \left\{ I(Z_j \geq Z) | \beta^T X \right\} \right] \\
&\quad + O_p(h^2) \\
&= E \left[ I(Z_j \geq Z) I \left\{ f_{\beta^T X}(\beta^T X) \leq d_n \right\} \right] + O_p(h^2) \\
&\leq E \left[ I \left\{ f_{\beta^T X}(\beta^T X) \leq d_n \right\} \right] + O_p(h^2) \\
&= O_p\{n^{-\epsilon} + h^2 + (nh)^{-1/2}\} \\
&= O_p\{h^2 + (nh)^{-1/2}\}.
\end{align*} \]

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

\[ \text{var} \left[ \frac{\sum_{j=1}^n K_h(\beta^T X_j - \beta^T X)I(Z_j \geq Z)}{\sum_{j=1}^n K_h(\beta^T X_j - \beta^T X)} I \left\{ \hat{f}(\beta^T X) > d_n \right\} \right] = O_p\{(nh)^{-1/2}\}. \]

Summarizing the above result, we have

\[ \frac{\sum_{j=1}^n K_h(\beta^T X_j - \beta^T X)I(Z_j \geq Z)}{\sum_{j=1}^n K_h(\beta^T X_j - \beta^T X)} I \left\{ \hat{f}(\beta^T X) > d_n \right\} = E \left\{ I(Z_j \geq Z) | \beta^T X \right\} + O_p\{h^2 + (nh)^{-1/2}\}. \]

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 \( \beta^T X, \beta, Z \) 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 \( (\beta^T X, \beta) \) is compact, we
divide it into rectangular regions. In each region, the distance between a point \((\beta^T x, \beta)\) in this region and the nearest grid point is less than \(n^{-2}\). Note that we only need \(N \leq C n^{2p}\) grid points, where \(C\) is a constant. Let the grid points be \(\kappa_1, \ldots, \kappa_N\). For notation brevity, let \(\hat{p}(\beta^T X, \beta) = \partial E[XY(Z) \mid \beta^T X, \beta]/\partial (\beta^T X)\) and \(p(\beta^T X, \beta) = \partial E[XY(Z) \mid \beta^T X, \beta]/\partial (\beta^T X)\). Then for any \((\beta^T X, \beta)\), there exists a \(\kappa_i, 1 \leq i \leq N\), such that

\[
|\hat{p}(\beta^T X, \beta) - p(\beta^T X, \beta)| \leq |\hat{p}(\kappa_i) - p(\kappa_i)| + |\hat{p}(\beta^T X, \beta) - \hat{p}(\kappa_i)| + |p(\beta^T X, \beta) - p(\kappa_i)| \\
\leq |\hat{p}(\kappa_i) - p(\kappa_i)| + C_1 n^{-2}/h^2,
\]

for an absolute constant \(C_1\) under Conditions \([C1]\) and \([C6]\). Thus, for sufficiently large \(C\),

\[
\text{pr}(\sup_{\beta^T X, \beta} |\hat{p}(\beta^T X, \beta) - p(\beta^T X, \beta)| > 2C[h^2 + \{\log(nh^3)^{-1}\}^{1/2}]) \\
\leq \text{pr}(\sup_{\kappa_i} |\hat{p}(\kappa_i) - p(\kappa_i)| > 2C[h^2 + \{\log(nh^3)^{-1}\}^{1/2}] - C_1(nh)^{-2}) \\
\leq \exp\left\{-\frac{nA^2 \log n}{2C_2 (nh^3)^{-1/2}h^{-2}}\right\},
\]

under Condition \([C2]\). Using Bernstein’s inequality on each sum in the numerator and denominator of \(\hat{p}(\kappa_i)\), under Conditions \([C1]\) \([C2]\) \([C4]\) \([C5]\) and \([C6]\) we have that for any \(A > 0\),

\[
\text{pr}[|\hat{p}(\kappa_i) - E\hat{p}(\kappa_i)| \geq A\{\log n/(nh^3)^{1/2}\}] \\
\leq \exp\left\{-\frac{nA^2 \log n}{2C_2 (nh^3)^{-1/2}h^{-2}}\right\},
\]

where \(C_2\) is a constant. This leads to

\[
\text{pr}[\sup_{\kappa_i} |\hat{p}(\kappa_i) - E\hat{p}(\kappa_i)| \geq A\{\log n/(nh^3)^{1/2}\}] \\
\leq C n^{2p} \exp\left\{-\frac{A^2 \log n}{2C_2}\right\} \\
= C \exp\left\{\{2p - A^2/(2C_2)\}\log n\right\} \to 0
\]
for all $A^2 > 2pC_2$. Combining the above results, we get that for $A_1 = \max(A, C)$,

$$
\begin{align*}
\Pr\left( \sup_{\beta^T \mathbf{X}, \beta} |\hat{\rho}(\beta^T \mathbf{X}, \beta) - \rho(\beta^T \mathbf{X}, \beta)| > 2A_1[h^2 + \{\log(nh^3)^{-1}\}]^{1/2} \right) \\
\leq \Pr\left( \sup_{\kappa_i} |\hat{\rho}(\kappa_i) - \rho(\kappa_i)| > A_1[h^2 + \{\log(nh^3)^{-1}\}]^{1/2} \right) \\
\leq \Pr\left( \sup_{\kappa_i} |\hat{\rho}(\kappa_i) - E\hat{\rho}(\kappa_i)| > A_1h^2 \right) \\
+ \Pr\left( \sup_{\kappa_i} |E\hat{\rho}(\kappa_i)| \geq A_1\{\log(nh^3)^{-1}\}^{1/2} \right) \\
\to 0.
\end{align*}
$$

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 $(\beta^T \mathbf{X}, \beta, Z_i)$ or $(\beta^T \mathbf{X}, \beta, Z_j)$ 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.

\[ \square \]

### S4 Proof of Theorem 1

$$
\begin{align*}
\frac{1}{n} \sum_{i=1}^{n} \Delta_i \frac{\hat{\lambda}(Z_i, \hat{\beta}_n^T \mathbf{X}_i, \hat{\beta}_n)}{\hat{\lambda}(Z_i, \hat{\beta}_n^T \mathbf{X}_i, \hat{\beta}_n)} \otimes \left[ \mathbf{X}_i - \frac{E\{Y_i(Z_i) \mid \hat{\beta}_n^T \mathbf{X}_i, \hat{\beta}_n\}}{E\{Y_i(Z_i) \mid \hat{\beta}_n^T \mathbf{X}_i, \hat{\beta}_n\}} \right] = 0.
\end{align*}
$$

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

We first have

\[
\frac{1}{n} \sum_{i=1}^{n} \Delta \frac{\lambda_i(Z_i, \beta^T \beta_n X_i, \beta_n)}{\lambda(Z_i, \beta^T \beta_n X_i, \beta_n)} \odot \left[ X_{li} - \frac{E \{X_{li} Y_i(Z_i) \mid \beta^T \beta_n X_i, \beta_n\}}{E \{Y_i(Z_i) \mid \beta^T \beta_n X_i, \beta_n\}} \right]
= \frac{1}{n} \sum_{i=1}^{n} \Delta \frac{\lambda_i(Z_i, \beta^T \beta^* X_i, \beta^*)}{\lambda(Z_i, \beta^T \beta^* X_i, \beta^*)} \odot \left[ X_{li} - \frac{E \{X_{li} Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}}{E \{Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}} \right]
+ O_p(\|\beta_n - \beta^*\|)
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \Delta \frac{\lambda_i(Z_i, \beta^T \beta^* X_i, \beta^*)}{\lambda(Z_i, \beta^T \beta^* X_i, \beta^*)} \odot \left[ X_{li} - \frac{E \{X_{li} Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}}{E \{Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}} \right]
+ o_p(1)
\]

where the first equality is because the first derivative of the summation with respect to \( \beta \) is bounded uniformly under conditions \([C1, C2] \) by Lemma \([C] \). The last equality is because \( \hat{\beta}_n \) converges to \( \beta^* \). Thus, for sufficiently large \( n \), we have

\[
\frac{1}{n} \sum_{i=1}^{n} \Delta \frac{\lambda_i(Z_i, \beta^T \beta^* X_i, \beta^*)}{\lambda(Z_i, \beta^T \beta^* X_i, \beta^*)} \odot \left[ X_{li} - \frac{E \{X_{li} Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}}{E \{Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}} \right]
+ o_p(1)
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \Delta \frac{\lambda_i(Z_i, \beta^T \beta^* X_i, \beta^*)}{\lambda(Z_i, \beta^T \beta^* X_i, \beta^*)} \odot \left[ X_{li} - \frac{E \{X_{li} Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}}{E \{Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}} \right]
+ o_p(1)
\]

from a direct application of Lemma \([C] \). In addition,

\[
\frac{1}{n} \sum_{i=1}^{n} \Delta \frac{\lambda_i(Z_i, \beta^T \beta^* X_i, \beta^*)}{\lambda(Z_i, \beta^T \beta^* X_i, \beta^*)} \odot \left[ X_{li} - \frac{E \{X_{li} Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}}{E \{Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}} \right]
+ o_p(1)
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \Delta \frac{\lambda_i(Z_i, \beta^T \beta^* X_i, \beta^*)}{\lambda(Z_i, \beta^T \beta^* X_i, \beta^*)} \odot \left[ X_{li} - \frac{E \{X_{li} Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}}{E \{Y_i(Z_i) \mid \beta^T \beta^* X_i, \beta^*\}} \right]
+ o_p(1)
\]
under conditions $\text{C1}\text{C2}$. Thus, for sufficient large $n$ we have

$$0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda_i(Z_i, \beta_n^T X_i, \beta_n)}{\lambda(Z_i, \beta_n^T X_i, \beta_n)} \odot \left[ X_{ii} - \frac{E \left\{ X_{ii} Y_i(Z_i) \mid \beta_n^T X_i, \beta_n \right\}}{E \left\{ Y_i(Z_i) \mid \beta_n^T X_i, \beta_n \right\}} \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \Delta_i \lambda_i(Z_i, \beta_n^T X_i, \beta_n) \odot \left[ X_{ii} - \frac{E \left\{ X_{ii} Y_i(Z_i) \mid \beta_n^T X_i \right\}}{E \left\{ Y_i(Z_i) \mid \beta_n^T X_i \right\}} \right] + o_p(1)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \Delta_i \frac{\lambda_i(Z_i, \beta^T X_i, \beta^*)}{\lambda(Z_i, \beta^T X_i, \beta^*)} \odot \left[ X_{ii} - \frac{E \left\{ X_{ii} Y_i(Z_i) \mid \beta^T X_i \right\}}{E \left\{ Y_i(Z_i) \mid \beta^T X_i \right\}} \right] + o_p(1)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \Delta_i \frac{\lambda_i(Z_i, \beta^T X_i, \beta^*)}{\lambda(Z_i, \beta^T X_i, \beta^*)} \odot \left[ X_{ii} - \frac{E \left\{ X_{ii} Y_i(Z_i) \mid \beta^T X_i \right\}}{E \left\{ Y_i(Z_i) \mid \beta^T X_i \right\}} \right] + o_p(1)$$

under conditions $\text{C1}\text{C2}$ and $\text{C3}$. Note that

$$E \left( \frac{\Delta_i \lambda_i(Z, \beta^T X_i, \beta^*)}{\lambda(Z, \beta^T X_i, \beta^*)} \odot \left[ X_{ii} - \frac{E \left\{ X_{ii} Y_i(Z) \mid \beta^T X \right\}}{E \left\{ Y(Z) \mid \beta^T X \right\}} \right] \right)$$

is a nonrandom quantity that does not depend on $n$, hence it is zero. Thus the uniqueness requirement in Condition $\text{C7}$ ensures that $\beta^* = \beta_0$.

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 $\hat{\beta}_n$ that do not converge to $\beta^*$. As an infinite sequence in a compact set $\mathcal{B}$, we can thus obtain another subsequence that converges, say to $\beta^{**} \neq \beta^*$. Identical derivation as before then leads to $\beta^{**} = \beta_0$, which is a contradiction to $\beta^{**} \neq \beta^*$. Thus we conclude $\hat{\beta} - \beta_0 \to 0$ in probability when $n \to \infty$ under condition $\text{C1}\text{C7}$.

**S5  Proof of Theorem 2**

We first expand (2.4) as

$$0 = n^{-1/2} \sum_{i=1}^{n} \Delta_i \frac{\lambda_i(Z_i, \beta_n^T X_i, \beta)}{\lambda(Z_i, \beta_n^T X_i, \beta)} \odot \left[ X_{ii} - \frac{E \left\{ X_{ii} Y_i(Z_i) \mid \beta_n^T X_i, \beta \right\}}{E \left\{ Y_i(Z_i) \mid \beta_n^T X_i, \beta \right\}} \right]$$

$$= n^{-1/2} \sum_{i=1}^{n} \Delta_i \frac{\lambda_i(Z_i, \beta_0^T X_i, \beta_0)}{\lambda(Z_i, \beta_0^T X_i, \beta_0)} \odot \left[ X_{ii} - \frac{E \left\{ X_{ii} Y_i(Z_i) \mid \beta_0^T X_i, \beta_0 \right\}}{E \left\{ Y_i(Z_i) \mid \beta_0^T X_i, \beta_0 \right\}} \right]$$
where \( \tilde{\beta} \) is on the line connecting \( \beta_0 \) and \( \hat{\beta} \).

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

\[
\begin{align*}
\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\partial}{\partial (\beta^T X_i)} \left( \Delta_i \hat{\lambda}_i(Z_i, \beta^T X_i, \beta) \right) \right. \\
\times \left[ X_i - \frac{\hat{E} \{ X_i, Y_i(Z_i) \mid \beta^T X_i, \beta \}}{\hat{E} \{ Y_i(Z_i) \mid \beta^T X_i, \beta \}} \right] \right. \\
\left. \left. \times \hat{E} \{ X_i \mid \beta^T X_i, \beta \} \right\} \right|_{\beta=\hat{\beta}} \\
\times \sqrt{n}(\hat{\beta} - \beta_0),
\end{align*}
\]

(S5.2)
Because of Lemma 1, (S5.3) converges uniformly in probability to

\[-E \left( \int_0^\infty \frac{X_{10}^{(2)}(s, \beta^T_0 X)}{\lambda_0(s, \beta^T_0 X)} \odot X_i - \frac{E \left\{ X_i Y(s) \mid \beta^T_0 X \right\}}{E \left\{ Y(s) \mid \beta^T_0 X \right\}} \right) \odot X^T_i dN(s) \]

\[= -E \left( \int_0^\infty \frac{X_{10}^{(2)}(s, \beta^T_0 X)}{\lambda_0(s, \beta^T_0 X)} \odot X_i - \frac{E \left\{ X_i Y(s) \mid \beta^T_0 X \right\}}{E \left\{ Y(s) \mid \beta^T_0 X \right\}} \right) \odot X^T_i Y(s) \lambda_0(s, \beta^T_0 X) ds \]

\[= -E \left( \int_0^\infty \frac{X_{10}^{(2)}(s, \beta^T_0 X)}{\lambda_0(s, \beta^T_0 X)} \odot X_i - \frac{E \left\{ X_i Y(s) \mid \beta^T_0 X \right\}}{E \left\{ Y(s) \mid \beta^T_0 X \right\}} \right) \odot X^T_i \left( \frac{X_i Y(s) \mid \beta^T_0 X}{E \left\{ Y(s) \mid \beta^T_0 X \right\}} \right)^T Y(s) ds \]

\[= -E \left( \frac{X_{10}^{(2)}(s, \beta^T_0 X)}{\lambda_0(s, \beta^T_0 X)} \odot X_i - \frac{E \left\{ X_i Y(s) \mid \beta^T_0 X \right\}}{E \left\{ Y(s) \mid \beta^T_0 X \right\}} \right) \odot \left( \frac{X_i Y(s) \mid \beta^T_0 X}{E \left\{ Y(s) \mid \beta^T_0 X \right\}} \right)^T Y(s) ds \]

\[= -E \{ S_{\text{eff}}(\Delta, Z, X)^{\otimes 2} \}, \]

where the last equality is because the second term above is zero by first taking expectation conditional on \( \beta^T_0 X \). Note that \( \lambda_1(s, \beta^T_0 X, \beta_0) = \lambda_{10}(s, \beta^T_0 X) \) and \( \lambda(s, \beta^T_0 X, \beta_0) = \lambda_0(s, \beta^T_0 X) \).

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

\[E \left\{ \frac{\Delta_i}{\lambda_0(Z, \beta^T_0 X_i)} \frac{\partial}{\partial (\beta^T_0 X_i)} \left( \hat{\lambda}_1(Z, \beta^T_0 X_i, \beta_0) \odot \left[ X_{iL} - \frac{E \left\{ X_{iL} Y(Z) \mid \beta^T_0 X_i \right\}}{E \left\{ Y(Z) \mid \beta^T_0 X_i \right\}} \right] \right) \right\} \odot X^T_i \]

Now let \( \hat{\lambda}_{1,-i}(Z, \beta_0^T X_i, \beta_0) \) be the leave-one-out version of \( \hat{\lambda}_1(Z, \beta^T_0 X_i, \beta_0) \), i.e., it is constructed the same as \( \hat{\lambda}_1(Z, \beta^T_0 X_i, \beta_0) \) except that the ith observation is not used. Obviously,

\[\frac{\Delta_i}{\lambda_0(Z, \beta^T_0 X_i)} \frac{\partial}{\partial (\beta^T_0 X_i)} \left( \hat{\lambda}_1(Z, \beta^T_0 X_i, \beta_0) \odot \left[ X_{iL} - \frac{E \left\{ X_{iL} Y(Z) \mid \beta^T_0 X_i \right\}}{E \left\{ Y(Z) \mid \beta^T_0 X_i \right\}} \right] \right) \odot X^T_i \]

\[= \frac{\Delta_i}{\lambda_0(Z, \beta^T_0 X_i)} \frac{\partial}{\partial (\beta^T_0 X_i)} \left( \hat{\lambda}_{1,-i}(Z, \beta_0^T X_i, \beta_0) \odot \left[ X_{iL} - \frac{E \left\{ X_{iL} Y(Z) \mid \beta_0^T X_i \right\}}{E \left\{ Y(Z) \mid \beta_0^T X_i \right\}} \right] \right) \odot X^T_i \]

\[= o_p(1). \]

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

\[ E_i \left\{ \frac{\Delta_1}{\lambda_0(Z_i, \beta_0^T X_i)} \frac{\partial}{\partial (\beta_0^T X_i)} \left( \hat{\lambda}_{1,i}(Z_i, \beta_0^T X_i, \beta_0) \right) \right\} \]

\[ \times X_{li} - E \left\{ \frac{X_{li} Y_i(Z_i) | \beta_0^T X_i}{E\{Y_i(Z_i) | \beta_0^T X_i\}} \right\} \right\} \otimes X_i^T \right\} \]

\[ = E_i \left\{ \int \frac{1}{\lambda_0(s, \beta_0^T X_i)} \frac{\partial}{\partial (\beta_0^T X_i)} \left( \hat{\lambda}_{1,i}(s, \beta_0^T X_i, \beta_0) \right) \right\}

\[ \times X_{li} - E \left\{ \frac{X_{li} Y_i(s) | \beta_0^T X_i}{E\{Y_i(s) | \beta_0^T X_i\}} \right\} \right\} \otimes X_i^T dN_i(s) \right\} \]

\[ = E_i \left\{ \frac{\partial}{\partial \beta_0} \int \hat{\lambda}_{1,i}(s, \beta_0^T X_i, \beta_0) \right\}

\[ \times X_{li} - E \left\{ \frac{X_{li} Y_i(s) | \beta_0^T X_i}{E\{Y_i(s) | \beta_0^T X_i\}} \right\} \right\} Y_i(s) ds \right\} \]

\[ = 0. \]

Here, the last equality is because the integrand has expectation zero conditional on $\beta_0^T X_i$ and all other observations, and the third last equality is because the expectation is with respect to $X_i$, and does not involve $\beta_0$. Therefore, the term in (S5.4) converges in probability uniformly to

\[ E \left\{ \frac{\Delta_1}{\lambda_0(Z_i, \beta_0^T X_i)} \frac{\partial}{\partial (\beta_0^T X_i)} \left( \hat{\lambda}_{1,i}(Z_i, \beta_0^T X_i, \beta_0) \right) \right\} \times X_{li} - E \left\{ \frac{X_{li} Y_i(Z_i) | \beta_0^T X_i}{E\{Y_i(Z_i) | \beta_0^T X_i\}} \right\} \right\} \otimes X_i^T = 0 \]

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

Next we decompose (S5.1) into

\[ n^{-1/2} \sum_{i=1}^{n} \Delta_i \frac{i}{\hat{\lambda}(Z_i, \beta_0^T X_i, \beta_0)} \times \left[ X_{li} - E \left\{ \frac{X_{li} Y_i(Z_i) | \beta_0^T X_i}{E\{Y_i(Z_i) | \beta_0^T X_i\}} \right\} \right] = T_1 + T_2 + T_3 + T_4, \quad \text{(S5.5)} \]

where

\[ T_1 = n^{-1/2} \sum_{i=1}^{n} \Delta_i \frac{i}{\lambda_0(Z_i, \beta_0^T X_i)} \times \left[ X_{li} - E \left\{ \frac{X_{li} Y_i(Z_i) | \beta_0^T X_i}{E\{Y_i(Z_i) | \beta_0^T X_i\}} \right\} \right], \]
\[ T_2 = n^{-1/2} \sum_{i=1}^{n} \Delta_i \left\{ \frac{\hat{\lambda}_i(Z_i, \beta^T_0 X_i, \beta_0)}{\lambda(Z_i, \beta^T_0 X_i, \beta_0)} - \frac{\lambda_{10}(Z_i, \beta^T_0 X_i)}{\lambda_0(Z_i, \beta^T_0 X_i)} \right\} \]
\[ \otimes \left[ X_{it} - \frac{E \{ X_{it} Y_i(Z_i) \mid \beta^T_0 X_i \}}{E \{ Y_i(Z_i) \mid \beta^T_0 X_i \}} \right], \]
\[ T_3 = n^{-1/2} \sum_{i=1}^{n} \Delta_i \left( \frac{\hat{\lambda}_i(Z_i, \beta^T_0 X_i, \beta_0)}{\lambda_0(Z_i, \beta^T_0 X_i, \beta_0)} - \frac{\lambda_{10}(Z_i, \beta^T_0 X_i)}{\lambda_0(Z_i, \beta^T_0 X_i)} \right) \]
\[ \otimes \left[ \frac{E \{ X_{it} Y_i(Z_i) \mid \beta^T_0 X_i \}}{E \{ Y_i(Z_i) \mid \beta^T_0 X_i \}} - \frac{E \{ X_{it} Y_i(Z_i) \mid \beta^T_0 X_i, \beta_0 \}}{E \{ Y_i(Z_i) \mid \beta^T_0 X_i, \beta_0 \}} \right], \]
\[ T_4 = n^{-1/2} \sum_{i=1}^{n} \Delta_i \left( \frac{\hat{\lambda}_i(Z_i, \beta^T_0 X_i, \beta_0)}{\lambda_0(Z_i, \beta^T_0 X_i, \beta_0)} - \frac{\lambda_{10}(Z_i, \beta^T_0 X_i)}{\lambda_0(Z_i, \beta^T_0 X_i)} \right) \]
\[ \otimes \left[ \frac{E \{ X_{it} Y_i(Z_i) \mid \beta^T_0 X_i \}}{E \{ Y_i(Z_i) \mid \beta^T_0 X_i \}} - \frac{E \{ X_{it} Y_i(Z_i) \mid \beta^T_0 X_i, \beta_0 \}}{E \{ Y_i(Z_i) \mid \beta^T_0 X_i, \beta_0 \}} \right]. \]

First, note that

\[ T_2 = n^{-1/2} \sum_{i=1}^{n} \int \left\{ \frac{\hat{\lambda}_i(s, \beta^T_0 X_i, \beta_0)}{\lambda_0(s, \beta^T_0 X_i, \beta_0)} - \frac{\lambda_{10}(s, \beta^T_0 X_i)}{\lambda_0(s, \beta^T_0 X_i)} \right\} \]
\[ \otimes \left[ X_{it} - \frac{E \{ X_{it} Y_i(s) \mid \beta^T_0 X_i \}}{E \{ Y_i(s) \mid \beta^T_0 X_i \}} \right] dN_i(s) \]
\[ = o_p \left( n^{-1/2} \sum_{i=1}^{n} \int \left[ X_{it} - \frac{E \{ X_{it} Y_i(s) \mid \beta^T_0 X_i \}}{E \{ Y_i(s) \mid \beta^T_0 X_i \}} \right] Y_i(s) \lambda_0(s, \beta^T_0 X_i) ds \right) \]
\[ = o_p(1), \]

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,

\[ T_3 = n^{-1/2} \sum_{i=1}^{n} \Delta_i \left( \frac{\hat{\lambda}_i(Z_i, \beta^T_0 X_i, \beta_0)}{\lambda_0(Z_i, \beta^T_0 X_i, \beta_0)} - \frac{\lambda_{10}(Z_i, \beta^T_0 X_i)}{\lambda_0(Z_i, \beta^T_0 X_i)} \right) \]
\[ \otimes \left[ \frac{E \{ X_{it} Y_i(Z_i) \mid \beta^T_0 X_i \}}{E \{ Y_i(Z_i) \mid \beta^T_0 X_i \}} - \frac{E \{ X_{it} Y_i(Z_i) \mid \beta^T_0 X_i, \beta_0 \}}{E \{ Y_i(Z_i) \mid \beta^T_0 X_i, \beta_0 \}} \right], \]
\[ T_4 = \frac{E \{ X_{it} Y_i(Z_i) \mid \beta^T_0 X_i, \beta_0 \}}{E \{ Y_i(Z_i) \mid \beta^T_0 X_i, \beta_0 \}}, \]
\[ T_3 = T_{31} + T_{32} + T_{33} + o_p(1), \]
where

\[ T_{31} = n^{-1/2} \sum_{i=1}^{n} \Delta_i \lambda_{10}(Z_i, \beta_0^T X_i) \otimes E \left[ -\frac{K_h(\beta_0 X_j - \beta_0^T X_i)X_j I(Z_j \geq Z_i)}{f_{\beta_0^T X}(\beta_0^T X_i)E \{ Y_i(Z_i) \mid \beta_0^T X_i \}} \right] \]

\[ + E \{ X_i Y_i(Z_i) \mid \beta_0^T X_i \} K_h(\beta_0^T X_j - \beta_0^T X_i) X_j I(Z_j \geq Z_i) \frac{1}{f_{\beta_0^T X}(\beta_0^T X_i)E \{ Y_i(Z_i) \mid \beta_0^T X_i \}^2} | \Delta_i, Z_i, X_i \]

\[ T_{32} = n^{-1/2} \sum_{j=1}^{n} E \left( \Delta_i \lambda_{10}(Z_i, \beta_0^T X_i) \otimes \left[ -\frac{K_h(\beta_0^T X_j - \beta_0^T X_i)X_j I(Z_j \geq Z_i)}{f_{\beta_0^T X}(\beta_0^T X_i)E \{ Y_i(Z_i) \mid \beta_0^T X_i \}} \right] \right) \]

\[ + E \{ X_i Y_i(Z_i) \mid \beta_0^T X_i \} K_h(\beta_0^T X_j - \beta_0^T X_i) X_j I(Z_j \geq Z_i) \frac{1}{f_{\beta_0^T X}(\beta_0^T X_i)E \{ Y_i(Z_i) \mid \beta_0^T X_i \}^2} | \Delta_j, Z_j, X_j \]

\[ T_{33} = -n^{-1/2} E \left( \Delta_i \lambda_{10}(Z_i, \beta_0^T X_i) \otimes E \left[ -\frac{K_h(\beta_0^T X_j - \beta_0^T X_i)X_j I(Z_j \geq Z_i)}{f_{\beta_0^T X}(\beta_0^T X_i)E \{ Y_i(Z_i) \mid \beta_0^T X_i \}} \right] \right) \]

\[ + E \{ X_i Y_i(Z_i) \mid \beta_0^T X_i \} K_h(\beta_0^T X_j - \beta_0^T X_i) X_j I(Z_j \geq Z_i) \frac{1}{f_{\beta_0^T X}(\beta_0^T X_i)E \{ Y_i(Z_i) \mid \beta_0^T X_i \}^2} | \Delta_i, Z_i, X_i \]
When \( nh^4 \to 0 \), plugging the results of \( T_1 \) and \( T_{32} \) to \( (S5.5) \), we obtain that the expression in \( (S5.1) \) is

\[
\begin{align*}
&= n^{-1/2} \sum_{i=1}^{n} \int Y_i(s)\lambda_0(s) \frac{\Lambda_{10}(s, \beta_0^T n X_i) + \hat{E} \{ X_i Y_i(s) \} - \hat{E} \{ X_i \} \beta_0^T n X_i}{\lambda_0(s) \beta_0^T n X_i} \, ds \\
&\quad + \mathcal{O}_p(n^{1/2}h^2) \\
&= n^{-1/2} \sum_{i=1}^{n} \Delta_i \left\{ \frac{\hat{X}_i}{\lambda(Z_i, \beta_0^T n X_i, \beta_0)} - \frac{\Lambda_{10}(Z_i, \beta_0^T n X_i)}{\lambda_0(Z_i, \beta_0^T n X_i)} \right\} \\
&\qquad \times \left[ \frac{E \{ X_i Y_i(Z_i) \} \beta_0^T n X_i}{E \{ Y_i(Z_i) \} \beta_0^T n X_i} - \frac{\hat{E} \{ X_i Y_i(Z_i) \} \beta_0^T n X_i}{\hat{E} \{ Y_i(Z_i) \} \beta_0^T n X_i} \right] \\
&\quad + \mathcal{O}_p(n^{1/2}h^2) \\
&= \mathcal{O}_p(1),
\end{align*}
\]

where the last equality is because the integrands have mean zero conditional on \( \beta_0^T n X \), and the second last equality is obtained following the same derivation of \( T_3 \). Using these results in \( (S5.1) \), combined with the results on \( (S5.2) \), it is now clear that the theorem holds. \( \square \)
References


Table S1: Results of study 1, based on 1000 simulations with sample size 100. “bias” is $|\text{mean}(\hat{\beta}) - \beta|$ of each component in $\beta$. “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 $\hat{P} - P$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>$\lambda_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>No censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox bias</td>
<td>0.0138</td>
<td>0.0567</td>
<td>0.0044</td>
<td>0.0730</td>
<td>0.0007</td>
<td>0.0648</td>
<td>0.3033</td>
</tr>
<tr>
<td>sd</td>
<td>0.2921</td>
<td>0.4441</td>
<td>0.2990</td>
<td>0.4507</td>
<td>0.3107</td>
<td>0.4515</td>
<td>0.0968</td>
</tr>
<tr>
<td>AFT bias</td>
<td>0.6133</td>
<td>0.1859</td>
<td>0.0890</td>
<td>0.4496</td>
<td>0.0415</td>
<td>0.3152</td>
<td>0.3752</td>
</tr>
<tr>
<td>sd</td>
<td>1.4697</td>
<td>2.3427</td>
<td>1.4635</td>
<td>2.5065</td>
<td>1.4187</td>
<td>2.3088</td>
<td>0.1161</td>
</tr>
<tr>
<td>hmave bias</td>
<td>0.0001</td>
<td>0.0966</td>
<td>0.0166</td>
<td>0.0945</td>
<td>0.0008</td>
<td>0.1105</td>
<td>0.3291</td>
</tr>
<tr>
<td>sd</td>
<td>0.3572</td>
<td>0.5363</td>
<td>0.3611</td>
<td>0.5310</td>
<td>0.3604</td>
<td>0.5304</td>
<td>0.1107</td>
</tr>
<tr>
<td>semi bias</td>
<td>0.0056</td>
<td>0.0251</td>
<td>0.0026</td>
<td>0.0021</td>
<td>0.0031</td>
<td>0.0032</td>
<td>0.0980</td>
</tr>
<tr>
<td>sd</td>
<td>0.2500</td>
<td>0.3037</td>
<td>0.2175</td>
<td>0.2821</td>
<td>0.2371</td>
<td>0.3019</td>
<td>0.0864</td>
</tr>
</tbody>
</table>

|          | 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 $\beta$. From left to right in each group: Cox, AFT, lmave, semiparametric.
<table>
<thead>
<tr>
<th>Method</th>
<th>No censoring</th>
<th>20% censoring</th>
<th>40% censoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CoxPH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hmave</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>semi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>true</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Censoring</th>
<th>C-statistics</th>
<th>Cox</th>
<th>AFT</th>
<th>hmave</th>
<th>semi</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td></td>
<td>0.994</td>
<td>0.550</td>
<td>0.986</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.005</td>
<td>0.072</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td>0.996</td>
<td>0.415</td>
<td>0.981</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.002</td>
<td>0.089</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td>0.997</td>
<td>0.253</td>
<td>0.987</td>
<td>0.978</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.003</td>
<td>0.081</td>
<td>0.003</td>
<td>0.004</td>
</tr>
</tbody>
</table>
Table S3: Results of study 2, based on 1000 simulations with sample size 200. “bias” is $|\text{mean}(\hat{\beta}) - \beta|$ of each component in $\beta$, “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 $\hat{P} - P$.

<table>
<thead>
<tr>
<th>true</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>$\lambda_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.3</td>
<td>-1.3</td>
<td>1</td>
<td>-0.5</td>
<td>0.5</td>
<td>-0.5</td>
<td></td>
</tr>
<tr>
<td>Cox bias</td>
<td>0.0518</td>
<td>0.0437</td>
<td>0.0404</td>
<td>0.0020</td>
<td>0.0008</td>
<td>0.0232</td>
<td>0.2263</td>
</tr>
<tr>
<td>sd</td>
<td>0.4444</td>
<td>0.4471</td>
<td>0.3983</td>
<td>0.3152</td>
<td>0.2990</td>
<td>0.3074</td>
<td>0.0643</td>
</tr>
<tr>
<td></td>
<td>0.0208</td>
<td>0.0177</td>
<td>0.0144</td>
<td>0.0024</td>
<td>0.0018</td>
<td>0.0049</td>
<td>0.1472</td>
</tr>
<tr>
<td>sd</td>
<td>0.2667</td>
<td>0.2563</td>
<td>0.2237</td>
<td>0.1818</td>
<td>0.1769</td>
<td>0.1865</td>
<td>0.0449</td>
</tr>
<tr>
<td>hmove bias</td>
<td>0.0115</td>
<td>0.0128</td>
<td>0.0040</td>
<td>0.0033</td>
<td>0.0009</td>
<td>0.0043</td>
<td>0.1517</td>
</tr>
<tr>
<td>sd</td>
<td>0.2708</td>
<td>0.2688</td>
<td>0.2328</td>
<td>0.1949</td>
<td>0.1851</td>
<td>0.1895</td>
<td>0.0458</td>
</tr>
<tr>
<td>semi bias</td>
<td>0.0235</td>
<td>0.0426</td>
<td>0.0139</td>
<td>0.0183</td>
<td>0.0168</td>
<td>0.0131</td>
<td>0.0784</td>
</tr>
<tr>
<td>sd</td>
<td>0.2356</td>
<td>0.2190</td>
<td>0.1982</td>
<td>0.1626</td>
<td>0.1635</td>
<td>0.1633</td>
<td>0.0408</td>
</tr>
<tr>
<td>No censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20% censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox bias</td>
<td>0.0880</td>
<td>0.0681</td>
<td>0.0691</td>
<td>0.0169</td>
<td>0.0380</td>
<td>0.0151</td>
<td>0.2357</td>
</tr>
<tr>
<td>sd</td>
<td>0.4860</td>
<td>0.4734</td>
<td>0.4123</td>
<td>0.3124</td>
<td>0.3219</td>
<td>0.3222</td>
<td>0.0680</td>
</tr>
<tr>
<td></td>
<td>0.0402</td>
<td>0.0304</td>
<td>0.0281</td>
<td>0.0049</td>
<td>0.0138</td>
<td>0.0013</td>
<td>0.1536</td>
</tr>
<tr>
<td>sd</td>
<td>0.2833</td>
<td>0.2795</td>
<td>0.2455</td>
<td>0.1918</td>
<td>0.1861</td>
<td>0.1910</td>
<td>0.0460</td>
</tr>
<tr>
<td>hmove bias</td>
<td>0.0283</td>
<td>0.0269</td>
<td>0.0268</td>
<td>0.0049</td>
<td>0.0055</td>
<td>0.0043</td>
<td>0.1648</td>
</tr>
<tr>
<td>sd</td>
<td>0.2912</td>
<td>0.2974</td>
<td>0.2416</td>
<td>0.2339</td>
<td>0.2478</td>
<td>0.2398</td>
<td>0.0514</td>
</tr>
<tr>
<td>semi bias</td>
<td>0.0356</td>
<td>0.0599</td>
<td>0.0081</td>
<td>0.0195</td>
<td>0.0294</td>
<td>0.0130</td>
<td>0.1170</td>
</tr>
<tr>
<td>sd</td>
<td>0.2946</td>
<td>0.3119</td>
<td>0.2636</td>
<td>0.2103</td>
<td>0.2019</td>
<td>0.2095</td>
<td>0.0448</td>
</tr>
<tr>
<td></td>
<td>40% censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox bias</td>
<td>0.1062</td>
<td>0.1054</td>
<td>0.0915</td>
<td>0.0387</td>
<td>0.0586</td>
<td>0.0201</td>
<td>0.1996</td>
</tr>
<tr>
<td>sd</td>
<td>0.5761</td>
<td>0.5699</td>
<td>0.4965</td>
<td>0.3687</td>
<td>0.3698</td>
<td>0.3478</td>
<td>0.0701</td>
</tr>
<tr>
<td></td>
<td>0.0409</td>
<td>0.0440</td>
<td>0.0217</td>
<td>0.0247</td>
<td>0.0149</td>
<td>0.0136</td>
<td>0.1634</td>
</tr>
<tr>
<td>sd</td>
<td>0.3024</td>
<td>0.2978</td>
<td>0.2579</td>
<td>0.2073</td>
<td>0.2105</td>
<td>0.2050</td>
<td>0.0499</td>
</tr>
<tr>
<td>hmove bias</td>
<td>0.0708</td>
<td>0.0757</td>
<td>0.0571</td>
<td>0.0304</td>
<td>0.0333</td>
<td>0.0248</td>
<td>0.2027</td>
</tr>
<tr>
<td>sd</td>
<td>0.4225</td>
<td>0.4192</td>
<td>0.3748</td>
<td>0.2876</td>
<td>0.2745</td>
<td>0.2760</td>
<td>0.0639</td>
</tr>
<tr>
<td>semi bias</td>
<td>0.0641</td>
<td>0.0681</td>
<td>0.0127</td>
<td>0.0054</td>
<td>0.0210</td>
<td>0.0410</td>
<td>0.1445</td>
</tr>
<tr>
<td>sd</td>
<td>0.3372</td>
<td>0.3202</td>
<td>0.3047</td>
<td>0.2627</td>
<td>0.2366</td>
<td>0.2606</td>
<td>0.0468</td>
</tr>
</tbody>
</table>
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 $\beta$. From left to right in each group: Cox, AFT, lmave, semiparametric.
Figure S4: Estimated survival function by different methods at different censoring rates of study 2, where $\beta^T X$ is fixed at $\hat{\beta}^T X$. 
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.

<table>
<thead>
<tr>
<th>Censoring</th>
<th>Cox C-statistics</th>
<th>AFT C-statistics</th>
<th>hmave C-statistics</th>
<th>semi C-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.899</td>
<td>0.971</td>
<td>0.955</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>0.054</td>
<td>0.036</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>20%</td>
<td>0.877</td>
<td>0.992</td>
<td>0.961</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>0.076</td>
<td>0.004</td>
<td>0.015</td>
<td>0.017</td>
</tr>
<tr>
<td>40%</td>
<td>0.810</td>
<td>0.999</td>
<td>0.927</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>0.104</td>
<td>0.002</td>
<td>0.032</td>
<td>0.041</td>
</tr>
</tbody>
</table>
Table S5: Results of study 3, based on 1000 simulations with sample size 200. "bias" is $|\text{mean}(\hat{\beta}) - \beta|$ of each component in $\beta$. "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 $\hat{P} - P$. 

<table>
<thead>
<tr>
<th>true</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>$\beta_8$</th>
<th>$\beta_9$</th>
<th>$\beta_{10}$</th>
<th>$\lambda_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox bias</td>
<td>0.1409</td>
<td>0.1355</td>
<td>0.5010</td>
<td>0.2523</td>
<td>0.2152</td>
<td>0.0645</td>
<td>0.3055</td>
<td>0.3495</td>
<td>0.8551</td>
<td>0.9336</td>
</tr>
<tr>
<td>sd</td>
<td>5.9041</td>
<td>5.7703</td>
<td>6.0618</td>
<td>7.5865</td>
<td>6.9145</td>
<td>7.8148</td>
<td>5.6968</td>
<td>7.3566</td>
<td>5.2515</td>
<td>0.0779</td>
</tr>
<tr>
<td>AFT bias</td>
<td>0.9970</td>
<td>1.1378</td>
<td>1.1105</td>
<td>0.7777</td>
<td>1.1407</td>
<td>3.2320</td>
<td>0.0519</td>
<td>1.3635</td>
<td>0.3193</td>
<td>0.8755</td>
</tr>
<tr>
<td>sd</td>
<td>7.8813</td>
<td>8.0801</td>
<td>7.9166</td>
<td>8.3889</td>
<td>7.0144</td>
<td>6.5110</td>
<td>7.7910</td>
<td>7.8402</td>
<td>6.6598</td>
<td>0.2551</td>
</tr>
<tr>
<td>hmave bias</td>
<td>0.6192</td>
<td>0.3443</td>
<td>0.1297</td>
<td>0.3207</td>
<td>0.0109</td>
<td>0.1899</td>
<td>0.0011</td>
<td>0.0005</td>
<td>0.0119</td>
<td>0.2943</td>
</tr>
<tr>
<td>sd</td>
<td>5.3614</td>
<td>6.5161</td>
<td>5.9474</td>
<td>5.0169</td>
<td>4.6656</td>
<td>4.0464</td>
<td>4.7291</td>
<td>4.6429</td>
<td>4.6929</td>
<td>0.2438</td>
</tr>
<tr>
<td>semi bias</td>
<td>0.0033</td>
<td>0.0007</td>
<td>0.0009</td>
<td>0.0039</td>
<td>0.0006</td>
<td>0.0001</td>
<td>0.0061</td>
<td>0.0005</td>
<td>0.0119</td>
<td>0.2943</td>
</tr>
<tr>
<td>sd</td>
<td>0.1474</td>
<td>0.1465</td>
<td>0.1502</td>
<td>0.1512</td>
<td>0.1539</td>
<td>0.1479</td>
<td>0.1475</td>
<td>0.1521</td>
<td>0.1495</td>
<td>0.0500</td>
</tr>
<tr>
<td>20% censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox bias</td>
<td>0.5226</td>
<td>0.0787</td>
<td>0.3027</td>
<td>0.0912</td>
<td>0.1738</td>
<td>0.0888</td>
<td>0.4557</td>
<td>0.2642</td>
<td>0.6556</td>
<td>0.8652</td>
</tr>
<tr>
<td>sd</td>
<td>7.5486</td>
<td>6.2857</td>
<td>6.4019</td>
<td>6.9874</td>
<td>5.6022</td>
<td>6.8008</td>
<td>7.3245</td>
<td>6.2297</td>
<td>5.5681</td>
<td>0.2621</td>
</tr>
<tr>
<td>AFT bias</td>
<td>1.7606</td>
<td>0.7222</td>
<td>1.4912</td>
<td>0.8368</td>
<td>0.8282</td>
<td>0.5411</td>
<td>0.7437</td>
<td>0.7244</td>
<td>0.4102</td>
<td>0.8856</td>
</tr>
<tr>
<td>sd</td>
<td>4.5490</td>
<td>4.9955</td>
<td>4.6505</td>
<td>4.9782</td>
<td>5.2053</td>
<td>4.8935</td>
<td>5.0968</td>
<td>5.5299</td>
<td>5.1731</td>
<td>0.2407</td>
</tr>
<tr>
<td>hmave bias</td>
<td>0.4018</td>
<td>0.1721</td>
<td>0.2905</td>
<td>0.2639</td>
<td>0.4307</td>
<td>0.2403</td>
<td>0.0752</td>
<td>0.2478</td>
<td>0.3978</td>
<td>0.8748</td>
</tr>
<tr>
<td>sd</td>
<td>7.0157</td>
<td>7.6713</td>
<td>7.6211</td>
<td>6.4765</td>
<td>6.2337</td>
<td>7.4385</td>
<td>6.2330</td>
<td>5.9572</td>
<td>6.0522</td>
<td>0.2240</td>
</tr>
<tr>
<td>semi bias</td>
<td>0.0062</td>
<td>0.0058</td>
<td>0.0004</td>
<td>0.0042</td>
<td>0.0088</td>
<td>0.0137</td>
<td>0.0072</td>
<td>0.0063</td>
<td>0.0010</td>
<td>0.2930</td>
</tr>
<tr>
<td>sd</td>
<td>0.1513</td>
<td>0.1510</td>
<td>0.1529</td>
<td>0.1538</td>
<td>0.1506</td>
<td>0.1539</td>
<td>0.1461</td>
<td>0.1525</td>
<td>0.1481</td>
<td>0.0528</td>
</tr>
<tr>
<td>40% censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox bias</td>
<td>0.3980</td>
<td>0.2558</td>
<td>0.3340</td>
<td>0.0155</td>
<td>0.1067</td>
<td>0.3787</td>
<td>0.0534</td>
<td>0.7641</td>
<td>0.5346</td>
<td>0.8377</td>
</tr>
<tr>
<td>AFT bias</td>
<td>4.6917</td>
<td>2.2959</td>
<td>1.5812</td>
<td>1.6128</td>
<td>4.8147</td>
<td>4.3596</td>
<td>4.1690</td>
<td>8.7963</td>
<td>4.6270</td>
<td>0.8330</td>
</tr>
<tr>
<td>hmave bias</td>
<td>0.3073</td>
<td>0.2527</td>
<td>0.3905</td>
<td>0.2035</td>
<td>0.0036</td>
<td>0.4633</td>
<td>0.0904</td>
<td>0.1348</td>
<td>0.4520</td>
<td>0.8374</td>
</tr>
<tr>
<td>semi bias</td>
<td>0.0040</td>
<td>0.0081</td>
<td>0.0055</td>
<td>0.0107</td>
<td>0.0064</td>
<td>0.0024</td>
<td>0.0043</td>
<td>0.0016</td>
<td>0.0079</td>
<td>0.2987</td>
</tr>
<tr>
<td>sd</td>
<td>0.1536</td>
<td>0.1547</td>
<td>0.1549</td>
<td>0.1529</td>
<td>0.1542</td>
<td>0.1499</td>
<td>0.1513</td>
<td>0.1443</td>
<td>0.1549</td>
<td>0.0603</td>
</tr>
</tbody>
</table>
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 $\beta$. From left to right in each group: Cox, AFT, hmaxe, semiparametric.
Figure S6: Estimated survival function by different methods at different censoring rates of study 3, where $\beta^T X$ is fixed at $\hat{\beta}^T X$. 
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.

<table>
<thead>
<tr>
<th>Censoring</th>
<th>C-statistics</th>
<th>Cox</th>
<th>AFT</th>
<th>hmade</th>
<th>semi</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>C-statistics</td>
<td>0.499</td>
<td>0.900</td>
<td>0.986</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.023</td>
<td>0.254</td>
<td>0.036</td>
<td>0.016</td>
</tr>
<tr>
<td>20%</td>
<td>C-statistics</td>
<td>0.499</td>
<td>0.903</td>
<td>0.989</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.026</td>
<td>0.256</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>40%</td>
<td>C-statistics</td>
<td>0.498</td>
<td>0.915</td>
<td>0.989</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.029</td>
<td>0.249</td>
<td>0.006</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Table S7: Results of study 4, based on 1000 simulations with sample size 200. “bias” is $|\text{mean}(\hat{\beta}) - \beta|$ of each component in $\beta$, “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 $\hat{P} - P$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_{3,1}$</th>
<th>$\beta_{4,1}$</th>
<th>$\beta_{5,1}$</th>
<th>$\beta_{6,1}$</th>
<th>$\beta_{3,2}$</th>
<th>$\beta_{4,2}$</th>
<th>$\beta_{5,2}$</th>
<th>$\beta_{6,2}$</th>
<th>$\lambda_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>2.75</td>
<td>-0.75</td>
<td>-1</td>
<td>2.0</td>
<td>-3.125</td>
<td>-1.125</td>
<td>1.0</td>
<td>-2.0</td>
<td></td>
</tr>
<tr>
<td>No censoring</td>
<td>2.3104</td>
<td>2.7002</td>
<td>0.9829</td>
<td>1.1085</td>
<td>1.6572</td>
<td>0.9218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>semi</td>
<td>0.0090</td>
<td>0.0179</td>
<td>0.0066</td>
<td>0.0105</td>
<td>0.0142</td>
<td>0.0041</td>
<td>0.0029</td>
<td>0.0051</td>
<td>0.0788</td>
</tr>
<tr>
<td>sd</td>
<td>0.3293</td>
<td>0.1781</td>
<td>0.1718</td>
<td>0.2380</td>
<td>0.3167</td>
<td>0.1809</td>
<td>0.1517</td>
<td>0.2309</td>
<td>0.0946</td>
</tr>
<tr>
<td>20% censoring</td>
<td>2.9199</td>
<td>0.5533</td>
<td>1.2512</td>
<td>1.8446</td>
<td>2.4579</td>
<td>1.1224</td>
<td>0.7592</td>
<td>1.8994</td>
<td>0.9273</td>
</tr>
<tr>
<td>semi</td>
<td>0.0710</td>
<td>0.0098</td>
<td>0.0229</td>
<td>0.0549</td>
<td>0.0560</td>
<td>0.0027</td>
<td>0.0185</td>
<td>0.0389</td>
<td>0.1430</td>
</tr>
<tr>
<td>sd</td>
<td>0.5451</td>
<td>0.3699</td>
<td>0.2798</td>
<td>0.4449</td>
<td>0.6172</td>
<td>0.3884</td>
<td>0.3387</td>
<td>0.4752</td>
<td>0.1780</td>
</tr>
<tr>
<td>40% censoring</td>
<td>2.6564</td>
<td>1.4719</td>
<td>1.6397</td>
<td>2.3155</td>
<td>2.9895</td>
<td>1.8283</td>
<td>1.1251</td>
<td>1.4566</td>
<td>0.9173</td>
</tr>
<tr>
<td>semi</td>
<td>0.0529</td>
<td>0.0235</td>
<td>0.0260</td>
<td>0.0310</td>
<td>0.0849</td>
<td>0.0293</td>
<td>0.0365</td>
<td>0.0538</td>
<td>0.1438</td>
</tr>
<tr>
<td>sd</td>
<td>0.7350</td>
<td>0.3982</td>
<td>0.3396</td>
<td>0.5753</td>
<td>0.7750</td>
<td>0.4048</td>
<td>0.3942</td>
<td>0.5257</td>
<td>0.1737</td>
</tr>
</tbody>
</table>
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 $\beta$. 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.

<table>
<thead>
<tr>
<th>Censoring</th>
<th>C-statistics</th>
<th>hmave</th>
<th>semi</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>C-statistics</td>
<td>0.903</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.201</td>
<td>0.001</td>
</tr>
<tr>
<td>20%</td>
<td>C-statistics</td>
<td>0.920</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.130</td>
<td>0.001</td>
</tr>
<tr>
<td>40%</td>
<td>C-statistics</td>
<td>0.913</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.139</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Table S9: Index misspecification on study 4. $d = 2$ is the correct index, $d = 1$ and $d = 3$ are misspecified indices.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$d=1$</th>
<th>$d=2$</th>
<th>$d=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.248(4.08)</td>
<td>0</td>
<td>-0.248(4.08)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.696(4.36)</td>
<td>2.665(0.34)</td>
<td>-3.011(0.40)</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>-1.179(4.53)</td>
<td>-0.714(0.20)</td>
<td>-1.052(0.26)</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>2.585(4.69)</td>
<td>-0.968(0.20)</td>
<td>0.969(0.23)</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-4.078(4.15)</td>
<td>1.940(0.26)</td>
<td>-1.927(0.29)</td>
</tr>
</tbody>
</table>

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 $|\text{mean}(\hat{\beta}) - \beta|$ of each component in $\beta$, “sd” is the sample standard errors of the corresponding estimation, “$\hat{\sigma}$” is the mean of the estimated standard errors of $\hat{\beta}$ component, “95%” is the sample coverage of the 95% confidence intervals.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_{31}$</th>
<th>$\beta_{41}$</th>
<th>$\beta_{51}$</th>
<th>$\beta_{61}$</th>
<th>$\beta_{32}$</th>
<th>$\beta_{42}$</th>
<th>$\beta_{52}$</th>
<th>$\beta_{62}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td>2.75</td>
<td>-0.75</td>
<td>-1</td>
<td>2.0</td>
<td>-3.125</td>
<td>-1.125</td>
<td>1.0</td>
<td>-2.0</td>
</tr>
<tr>
<td>bias</td>
<td>0.3995</td>
<td>0.5031</td>
<td>0.2066</td>
<td>0.3799</td>
<td>0.5515</td>
<td>0.5349</td>
<td>0.1757</td>
<td>0.3395</td>
</tr>
<tr>
<td>sd</td>
<td>0.5760</td>
<td>0.4236</td>
<td>0.4673</td>
<td>0.5608</td>
<td>0.6163</td>
<td>0.4376</td>
<td>0.4772</td>
<td>0.5377</td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.3868</td>
<td>0.3188</td>
<td>0.3312</td>
<td>0.3427</td>
<td>0.3956</td>
<td>0.3131</td>
<td>0.3331</td>
<td>0.3602</td>
</tr>
<tr>
<td>95%</td>
<td>0.7100</td>
<td>0.6577</td>
<td>0.8051</td>
<td>0.7034</td>
<td>0.6416</td>
<td>0.6292</td>
<td>0.8089</td>
<td>0.7414</td>
</tr>
<tr>
<td>$n = 500$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.1790</td>
<td>0.1258</td>
<td>0.0714</td>
<td>0.1338</td>
<td>0.2100</td>
<td>0.1489</td>
<td>0.07386</td>
<td>0.1340</td>
</tr>
<tr>
<td>sd</td>
<td>0.2741</td>
<td>0.1714</td>
<td>0.2177</td>
<td>0.2380</td>
<td>0.2979</td>
<td>0.1897</td>
<td>0.2202</td>
<td>0.2244</td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.1585</td>
<td>0.1371</td>
<td>0.1644</td>
<td>0.1659</td>
<td>0.2683</td>
<td>0.2179</td>
<td>0.2538</td>
<td>0.2558</td>
</tr>
<tr>
<td>95%</td>
<td>0.6663</td>
<td>0.8022</td>
<td>0.8298</td>
<td>0.7566</td>
<td>0.8127</td>
<td>0.8773</td>
<td>0.9125</td>
<td>0.8764</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.0611</td>
<td>0.0492</td>
<td>0.0188</td>
<td>0.0423</td>
<td>0.0695</td>
<td>0.0467</td>
<td>0.0209</td>
<td>0.0448</td>
</tr>
<tr>
<td>sd</td>
<td>0.1951</td>
<td>0.1451</td>
<td>0.1555</td>
<td>0.1538</td>
<td>0.1867</td>
<td>0.1433</td>
<td>0.1650</td>
<td>0.1711</td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.1062</td>
<td>0.1113</td>
<td>0.1134</td>
<td>0.1190</td>
<td>0.1823</td>
<td>0.1712</td>
<td>0.1749</td>
<td>0.1740</td>
</tr>
<tr>
<td>95%</td>
<td>0.8060</td>
<td>0.8830</td>
<td>0.8783</td>
<td>0.8621</td>
<td>0.9268</td>
<td>0.9705</td>
<td>0.9515</td>
<td>0.9287</td>
</tr>
</tbody>
</table>
Table S11: The estimated coefficients, standard errors and p-value in AIDS data.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}_{2,1}$</th>
<th>$\hat{\beta}_{3,1}$</th>
<th>$\hat{\beta}_{4,1}$</th>
<th>$\hat{\beta}_{5,1}$</th>
<th>$\hat{\beta}_{6,1}$</th>
<th>$\hat{\beta}_{7,1}$</th>
<th>$\hat{\beta}_{8,1}$</th>
<th>$\hat{\beta}_{9,1}$</th>
<th>$\hat{\beta}_{10,1}$</th>
<th>$\hat{\beta}_{11,1}$</th>
<th>$\hat{\beta}_{12,1}$</th>
<th>$\hat{\beta}_{13,1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>est</td>
<td>0.115</td>
<td>-0.002</td>
<td>0.093</td>
<td>0.088</td>
<td>-0.090</td>
<td>0.231</td>
<td>0.003</td>
<td>-0.178</td>
<td>0.058</td>
<td>-0.031</td>
<td>0.201</td>
<td>0.156</td>
</tr>
<tr>
<td>std</td>
<td>0.039</td>
<td>0.039</td>
<td>0.039</td>
<td>0.037</td>
<td>0.043</td>
<td>0.046</td>
<td>0.036</td>
<td>0.046</td>
<td>0.035</td>
<td>0.042</td>
<td>0.033</td>
<td>0.038</td>
</tr>
<tr>
<td>p-value</td>
<td>0.003</td>
<td>0.965</td>
<td>0.017</td>
<td>0.017</td>
<td>0.036</td>
<td>0.001</td>
<td>0.928</td>
<td>0.001</td>
<td>0.100</td>
<td>0.457</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Figure S10: Estimated cumulative hazard function $\hat{\Lambda}$ in AIDS data. (a). Comparisons of $\hat{\Lambda}$ as a function of $t$ between treatments ZDV+ddl and ZDV+Zal when other covariates are fixed at three indices. (b). $\hat{\Lambda}$ as a function of $\hat{\beta}^T X$ at $T = 100, 500, 1000$. (c). Contour plot of $\hat{\Lambda}$ as a function of $T$ and $\hat{\beta}^T X$. 
Figure S11: Residuals of predicted survival time by different methods.