Web Appendices for “Addressing Extreme Propensity Scores in Estimating Counterfactual Survival Functions via the Overlap Weights”

Chao Cheng, Fan Li, Laine E. Thomas, and Fan Li*

Web Appendix A: Validity of $\hat{\Delta}_h^I(t)$ and $\hat{\Delta}_h^{II}(t)$

Part (a): Validity of $\hat{\Delta}_h^I(t)$

We first verify that $\hat{S}_h^{I(1)}(t) = \frac{\sum_{i=1}^n w_i A_i \delta_i \mathbb{I}(U_i \geq t)/K_c^{(1)}(U_i, \boldsymbol{X}_i)}{\sum_{i=1}^n w_i A_i \delta_i / K_c^{(1)}(U_i, \boldsymbol{X}_i)}$ is a consistent estimator for $S_h^{(1)}(t)$. Specifically, leveraging the law of large numbers, we can show that

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \frac{w_i A_i \delta_i \mathbb{I}(U_i \geq t)}{K_c^{(1)}(U_i, \boldsymbol{X}_i)} =& E\left[\frac{h(\boldsymbol{X}) A \mathbb{I}(U \geq t, T \leq C)}{e(\boldsymbol{X}) K_c^{(1)}(U, \boldsymbol{X})}\right] + o_p(1) = E\left[E\left(\frac{h(\boldsymbol{X}) A \mathbb{I}(U \geq t, T \leq C)}{e(\boldsymbol{X}) K_c^{(1)}(U, \boldsymbol{X})} | \boldsymbol{X}\right)\right] + o_p(1) \\
=& E\left[\frac{h(\boldsymbol{X}) E[A \mathbb{I}(T \geq t, C \geq T) | \boldsymbol{X}]}{e(\boldsymbol{X}) P(C^{(1)} \geq T | \boldsymbol{X})}\right] + o_p(1) \\
=& E\left[\frac{h(\boldsymbol{X}) E[A \mathbb{I}(T^{(1)} \geq t) \mathbb{I}(C^{(1)} \geq T^{(1)}) | \boldsymbol{X}]}{e(\boldsymbol{X}) P(C^{(1)} \geq T^{(1)} | \boldsymbol{X})}\right] + o_p(1) \quad (\because T = T^{(1)}, C = C^{(1)} \text{ if } A = 1) \\
=& E\left[\frac{h(\boldsymbol{X}) E[A|\boldsymbol{X}] E[\mathbb{I}(T^{(1)} \geq t) | \boldsymbol{X}] E[\mathbb{I}(C^{(1)} \geq T^{(1)}) | \boldsymbol{X}]}{e(\boldsymbol{X}) P(C^{(1)} \geq T^{(1)} | \boldsymbol{X})}\right] + o_p(1) \quad (\because T^{(1)} \perp A | \boldsymbol{X}, T^{(1)} \perp C^{(1)} | \boldsymbol{X}, A) \\
=& E\left[\frac{h(\boldsymbol{X}) e(\boldsymbol{X}) S^{(1)}(t|\boldsymbol{X}) P(C^{(1)} \geq T^{(1)} | \boldsymbol{X})}{e(\boldsymbol{X}) P(C^{(1)} \geq T^{(1)} | \boldsymbol{X})}\right] + o_p(1). \\
=& E\left[h(\boldsymbol{X}) S^{(1)}(t|\boldsymbol{X})\right] + o_p(1),
\end{aligned}
$$

where $o_p(1)$ represents a vanishing term that converges to zero in probability when $n \to \infty$. The denominator of $\hat{S}_h^{I(1)}(t)$ can be represented as

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \frac{w_i A_i \delta_i}{K_c^{(1)}(U_i, \boldsymbol{X}_i)} =& E\left[\frac{h(\boldsymbol{X}) A \mathbb{I}(T \leq C)}{e(\boldsymbol{X}) K_c^{(1)}(U, \boldsymbol{X})}\right] + o_p(1) = E\left[\frac{h(\boldsymbol{X}) E[A \mathbb{I}(T \leq C) | \boldsymbol{X}]}{e(\boldsymbol{X}) P(C^{(1)} \geq T | \boldsymbol{X})}\right] + o_p(1) \\
=& E\left[\frac{h(\boldsymbol{X}) E[A \mathbb{I}(T^{(1)} \leq C^{(1)}) | \boldsymbol{X}]}{e(\boldsymbol{X}) P(C^{(1)} \geq T^{(1)} | \boldsymbol{X})}\right] + o_p(1) \quad (\because T = T^{(1)}, C = C^{(1)} \text{ if } A = 1) \\
=& E\left[\frac{h(\boldsymbol{X}) E[A|\boldsymbol{X}] E[\mathbb{I}(T^{(1)} \leq C^{(1)}) | \boldsymbol{X}]}{e(\boldsymbol{X}) P(C^{(1)} \geq T^{(1)} | \boldsymbol{X})}\right] + o_p(1) \quad (\because T^{(1)} \perp A | \boldsymbol{X}, T^{(1)} \perp C^{(1)} | \boldsymbol{X}, A) \\
=& E\left[\frac{h(\boldsymbol{X}) e(\boldsymbol{X}) P(C^{(1)} \geq T^{(1)} | \boldsymbol{X})}{e(\boldsymbol{X}) P(C^{(1)} \geq T^{(1)} | \boldsymbol{X})}\right] + o_p(1) = E[h(\boldsymbol{X})] + o_p(1).
\end{aligned}
$$

*Correspondence to Fan Li, PhD, Department of Biostatistics, Yale School of Public Health, Suite 200, Room 229, 135 College Street, New Haven, Connecticut 06510, USA (email: fan.f.li@yale.edu).

It follows that

$$\hat{S}_h^{I(1)}(t) = \frac{\sum_{i=1}^n \frac{w_i A_i \delta_i \mathbb{I}(U_i \geq t)}{K_c^{(1)}(U_i, \boldsymbol{X}_i)}}{\sum_{i=1}^n \frac{w_i A_i \delta_i}{K_c^{(1)}(U_i, \boldsymbol{X}_i)}} = \frac{E\left[h(\boldsymbol{X}) S^{(1)}(t|\boldsymbol{X})\right]}{E[h(\boldsymbol{X})]} + o_p(1) = S_h^{(1)}(t) + o_p(1).$$

Using a similar strategy, we can show that

$$\hat{S}_h^{I(0)}(t) = \frac{\sum_{i=1}^n \frac{w_i (1-A_i) \delta_i \mathbb{I}(U_i \geq t)}{K_c^{(0)}(U_i, \boldsymbol{X}_i)}}{\sum_{i=1}^n \frac{w_i (1-A_i) \delta_i}{K_c^{(0)}(U_i, \boldsymbol{X}_i)}} = \frac{E\left[h(\boldsymbol{X}) S^{(0)}(t|\boldsymbol{X})\right]}{E[h(\boldsymbol{X})]} + o_p(1) = S_h^{(0)}(t) + o_p(1).$$

To summarize,

$$\hat{\Delta}_h^I(t) = \hat{S}_h^{I(1)}(t) - \hat{S}_h^{I(0)}(t) = S_h^{(1)}(t) - S_h^{(0)}(t) + o_p(1) = \Delta_h(t) + o_p(1);$$

that is, $\hat{\Delta}_h^I(t)$ is a consistent estimator for $\Delta_h(t)$.

Part (b): Validity of $\hat{\Delta}_h^{II}(t)$

First, we can verify that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \frac{w_i A_i \mathbb{I}(U_i \geq t)}{K_c^{(1)}(t, \boldsymbol{X}_i)} =& E\left[\frac{h(\boldsymbol{X}) A \mathbb{I}(U \geq t)}{e(\boldsymbol{X}) K_c^{(1)}(t, \boldsymbol{X})}\right] + o_p(1) \\
=& E\left[E\left(\frac{h(\boldsymbol{X}) A \mathbb{I}(T \geq t, C \geq t)}{e(\boldsymbol{X}) P(C^{(1)} \geq t|\boldsymbol{X})}|\boldsymbol{X}\right)\right] + o_p(1) \\
=& E\left[\frac{h(\boldsymbol{X}) E[A \mathbb{I}(T^{(1)} \geq t, C^{(1)} \geq t)|\boldsymbol{X}]}{e(\boldsymbol{X}) P(C^{(1)} \geq t|\boldsymbol{X})}\right] + o_p(1) \quad (\because T = T^{(1)}, C = C^{(1)} \text{ if } A = 1) \\
=& E\left[\frac{h(\boldsymbol{X}) E[A|\boldsymbol{X}] E[\mathbb{I}(T^{(1)} \geq t)|\boldsymbol{X}] E[\mathbb{I}(C^{(1)} \geq t)|\boldsymbol{X}]}{e(\boldsymbol{X}) P(C^{(1)} \geq t|\boldsymbol{X})}\right] + o_p(1) \quad (\because T^{(1)} \perp A|\boldsymbol{X}, T^{(1)} \perp C^{(1)}|\boldsymbol{X}, A) \\
=& E\left[\frac{h(\boldsymbol{X}) e(\boldsymbol{X}) S^{(1)}(t|\boldsymbol{X}) P(C^{(1)} \geq t|\boldsymbol{X})}{e(\boldsymbol{X}) P(C^{(1)} \geq t|\boldsymbol{X})}\right] + o_p(1) \\
=& E\left[h(\boldsymbol{X}) S^{(1)}(t|\boldsymbol{X})\right] + o_p(1),
\end{aligned}$$

and

$$\frac{1}{n}\sum_{i=1}^n w_i A_i = E\left[\frac{h(\boldsymbol{X}) A}{e(\boldsymbol{X})}\right] + o_p(1) = E\left[\frac{h(\boldsymbol{X}) E[A|\boldsymbol{X}]}{e(\boldsymbol{X})}\right] + o_p(1) = E[h(\boldsymbol{X})] + o_p(1).$$

It follows that

$$\hat{S}_h^{II(1)}(t) = \frac{\frac{1}{n}\sum_{i=1}^n \frac{w_i A_i \mathbb{I}(U_i \geq t)}{K_c^{(1)}(t, \boldsymbol{X}_i)}}{\frac{1}{n}\sum_{i=1}^n w_i A_i} = \frac{E[h(\boldsymbol{X}) S^{(1)}(t|\boldsymbol{X})]}{E[h(\boldsymbol{X})]} + o_p(1) = S_h^{(1)}(t) + o_p(1).$$

Similarly, we can show that

$$\hat{S}_h^{II(0)}(t) = \frac{\frac{1}{n}\sum_{i=1}^n \frac{w_i (1-A_i) \mathbb{I}(U_i \geq t)}{K_c^{(0)}(t, \boldsymbol{X}_i)}}{\frac{1}{n}\sum_{i=1}^n w_i (1-A_i)} = S_h^{(0)}(t) + o_p(1).$$

In summary, we have $\hat{\Delta}_h^{II}(t) = \hat{S}_h^{II(1)}(t) - \hat{S}_h^{II(0)}(t) = S_h^{(1)}(t) - S_h^{(0)}(t) + o_p(1) = \Delta_h(t) + o_p(1)$; that is, $\hat{\Delta}_h^{II}(t)$ is a consistent estimator for $\Delta_h(t)$.

Web Appendix B: Optimality of the Overlap Weighting in Asymptotic Efficiency

In this appendix, we provide two results that under certain homoscedasticity conditions, OW still achieves the smallest asymptotic pointwise variance for estimating $\Delta_h(t)$ among the class of balancing weights, for both types of PS weighting estimators, $\hat{\Delta}_h^I(t)$ and $\hat{\Delta}_h^{II}(t)$.

Result 1. *(Optimal Type I estimator) If the variance of the pseudo-outcome* $\frac{\mathbb{I}(T_i^{(a)}\geq t)}{\sqrt{K_c^{(a)}(T_i^{(a)},\boldsymbol{X}_i)}}$ *is homoskedastic across both treatment groups, i.e.,*

$$Var\left(\frac{\mathbb{I}(T_i^{(1)}\geq t)}{\sqrt{K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)}}\Big|\boldsymbol{X}_i\right) = Var\left(\frac{\mathbb{I}(T_i^{(0)}\geq t)}{\sqrt{K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)}}\Big|\boldsymbol{X}_i\right) = v,$$

for some constant $v > 0$, then the OW with $\hat{\Delta}_{OW}^I(t)$ gives the smallest asymptotic variance for the Type I weighted estimator $\hat{\Delta}_h^I(t)$ among all $h(\boldsymbol{X}_i)$.

Proof. First recall that

$$\begin{aligned}
\hat{\Delta}_h^I(t) &= \frac{\sum_{i=1}^n w_iA_i\delta_i\mathbb{I}(U_i\geq t)/K_c^{(1)}(U_i,\boldsymbol{X}_i)}{\sum_{i=1}^n w_iA_i\delta_i/K_c^{(1)}(U_i,\boldsymbol{X}_i)} - \frac{\sum_{i=1}^n w_i(1-A_i)\delta_i\mathbb{I}(U_i\geq t)/K_c^{(0)}(U_i,\boldsymbol{X}_i)}{\sum_{i=1}^n w_i(1-A_i)\delta_i/K_c^{(0)}(U_i,\boldsymbol{X}_i)}\\
&= \frac{\sum_{i=1}^n w_iA_i\delta_i\mathbb{I}(T_i\geq t)/K_c^{(1)}(T_i,\boldsymbol{X}_i)}{\sum_{i=1}^n w_iA_i\delta_i/K_c^{(1)}(T_i,\boldsymbol{X}_i)} - \frac{\sum_{i=1}^n w_i(1-A_i)\delta_i\mathbb{I}(T_i\geq t)/K_c^{(0)}(T_i,\boldsymbol{X}_i)}{\sum_{i=1}^n w_i(1-A_i)\delta_i/K_c^{(0)}(T_i,\boldsymbol{X}_i)}, \qquad (1)\\
&= \frac{\sum_{i=1}^n w_iA_i\delta_i\mathbb{I}(T_i^{(1)}\geq t)/K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)}{\sum_{i=1}^n w_iA_i\delta_i/K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)} - \frac{\sum_{i=1}^n w_i(1-A_i)\delta_i\mathbb{I}(T_i^{(0)}\geq t)/K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)}{\sum_{i=1}^n w_i(1-A_i)\delta_i/K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)},
\end{aligned}$$

where the second equality holds because $U_i = T_i$ if $\delta_i = \mathbb{I}(T_i < C_i) = 1$ and the third equality holds because of the consistency assumption that $T_i = T_i^{(a)}$ if $A_i = a$. Conditional on the sample $\tilde{\boldsymbol{X}} = \{\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\}$, $\tilde{\boldsymbol{A}} = \{A_1,\ldots,A_n\}$, and $\tilde{\boldsymbol{\delta}} = \{\delta_1,\ldots,\delta_n\}$, only $\mathbb{I}(T_i^{(1)}\geq t)$ and $\mathbb{I}(T_i^{(0)}\geq t)$ are random in (1), so the variance of $\hat{\Delta}_h^I(t)$ is

$$\begin{aligned}
\mathrm{Var}\left(\hat{\Delta}_h^I(t)|\tilde{\boldsymbol{X}},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{\delta}}\right) =& \frac{\sum_{i=1}^n \frac{w_i^2A_i\delta_i\mathrm{Var}\left(\mathbb{I}(T_i^{(1)}\geq t)|\delta_i,A_i,\boldsymbol{X}_i\right)}{K_c^{(1)^2}(T_i^{(1)},\boldsymbol{X}_i)}}{\left\{\sum_{i=1}^n w_iA_i\delta_i/K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)\right\}^2} + \frac{\sum_{i=1}^n \frac{w_i(1-A_i)\delta_i\mathrm{Var}\left(\mathbb{I}(T_i^{(0)}\geq t)|\delta_i,A_i,\boldsymbol{X}_i\right)}{K_c^{(0)^2}(T_i^{(0)},\boldsymbol{X}_i)}}{\left\{\sum_{i=1}^n w_i(1-A_i)\delta_i/K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)\right\}^2}\\
=& \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)^2A_i\delta_i\mathrm{Var}\left(\mathbb{I}(T_i^{(1)}\geq t)|\delta_i,A_i,\boldsymbol{X}_i\right)}{e(\boldsymbol{X}_i)^2K_c^{(1)^2}(T_i^{(1)},\boldsymbol{X}_i)}}{\left\{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)A_i\delta_i}{e(\boldsymbol{X}_i)K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)}\right\}^2} + \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)^2(1-A_i)\delta_i\mathrm{Var}\left(\mathbb{I}(T_i^{(0)}\geq t)|\delta_i,A_i,\boldsymbol{X}_i\right)}{e(\boldsymbol{X}_i)^2K_c^{(0)^2}(T_i^{(0)},\boldsymbol{X}_i)}}{\left\{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)(1-A_i)\delta_i}{(1-e(\boldsymbol{X}_i))K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)}\right\}^2}\\
=& \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)^2A_i\delta_i}{e(\boldsymbol{X}_i)^2K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)}\mathrm{Var}\left(\frac{\mathbb{I}(T_i^{(1)}\geq t)}{\sqrt{K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)}}|\delta_i,A_i,\boldsymbol{X}_i\right)}{\left\{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)A_i\delta_i}{e(\boldsymbol{X}_i)K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)}\right\}^2} + \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)^2(1-A_i)\delta_i}{e(\boldsymbol{X}_i)^2K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)}\mathrm{Var}\left(\frac{\mathbb{I}(T_i^{(0)}\geq t)}{\sqrt{K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)}}|\delta_i,A_i,\boldsymbol{X}_i\right)}{\left\{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)(1-A_i)\delta_i}{(1-e(\boldsymbol{X}_i))K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)}\right\}^2}\\
=& \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)^2A_i\delta_i}{e(\boldsymbol{X}_i)^2K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)}\mathrm{Var}\left(\frac{\mathbb{I}(T_i^{(1)}\geq t)}{\sqrt{K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)}}|\boldsymbol{X}_i\right)}{\left\{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)A_i\delta_i}{e(\boldsymbol{X}_i)K_c^{(1)}(T_i^{(1)},\boldsymbol{X}_i)}\right\}^2} + \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)^2(1-A_i)\delta_i}{e(\boldsymbol{X}_i)^2K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)}\mathrm{Var}\left(\frac{\mathbb{I}(T_i^{(0)}\geq t)}{\sqrt{K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)}}|\boldsymbol{X}_i\right)}{\left\{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)(1-A_i)\delta_i}{(1-e(\boldsymbol{X}_i))K_c^{(0)}(T_i^{(0)},\boldsymbol{X}_i)}\right\}^2},
\end{aligned}$$

where the last equality holds in the above equation because $\mathrm{Var}\left(\mathbb{I}(T_i^{(a)}\geq t)|\delta_i,A_i,\boldsymbol{X}_i\right) = \mathrm{Var}\left(\mathbb{I}(T_i^{(a)}\geq t)|A_i,\boldsymbol{X}_i\right)$ by noticing the assumption $T_i^{(a)} \perp C_i^{(a)}|A_i,\boldsymbol{X}_i$ and then $\mathrm{Var}\left(\mathbb{I}(T_i^{(a)}\geq t)|A_i,\boldsymbol{X}_i\right) = \mathrm{Var}\left(\mathbb{I}(T_i^{(a)}\geq t)|\boldsymbol{X}_i\right)$ by applying the assumption $T_i^{(a)} \perp A_i | \boldsymbol{X}_i$. Hereafter, we denote $\text{Var}\left(\frac{\mathbb{I}(T_i^{(a)} \geq t)}{\sqrt{K_c^{(a)}(T_i^{(a)}, \boldsymbol{X}_i)}} | \boldsymbol{X}_i\right)$ as $v_a(\boldsymbol{X}_i)$ for $a = 0, 1$, which leads to

$$\text{Var}\left(\hat{\Delta}_h^I(t) | \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{\delta}}\right) = \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)^2 A_i \delta_i}{e(\boldsymbol{X}_i)^2 K_c^{(1)}(T_i^{(1)}, \boldsymbol{X}_i)} v_1(\boldsymbol{X}_i)}{\left\{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i) A_i \delta_i}{e(\boldsymbol{X}_i) K_c^{(1)}(T_i^{(1)}, \boldsymbol{X}_i)}\right\}^2} + \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)^2 (1-A_i) \delta_i}{e(\boldsymbol{X}_i)^2 K_c^{(0)}(T_i^{(0)}, \boldsymbol{X}_i)} v_0(\boldsymbol{X}_i)}{\left\{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)(1-A_i)\delta_i}{(1-e(\boldsymbol{X}_i)) K_c^{(0)}(T_i^{(0)}, \boldsymbol{X}_i)}\right\}^2}$$

Averaging the above first over the distribution of δ (using $E[\delta_i / K_c^{(1)}(T_i^{(1)}, \boldsymbol{X}_i) | A_i = 1] = E[\delta_i / K_c^{(0)}(T_i^{(0)}, \boldsymbol{X}_i) | A_i = 0] = 1$), next over the distribution of A (using $E[A_i / e(\boldsymbol{X}_i)] = E[(1 - A_i)/(1 - e(\boldsymbol{X}_i))] = 1$), and then over the distribution of $\boldsymbol{X}$, and again applying Slutsky's theorem, we have that

$$n \times \text{Var}\left(\hat{\Delta}_h^I(t) | \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{\delta}}\right) \rightarrow \frac{\int \left(\frac{v_1(\boldsymbol{X})}{e(\boldsymbol{X})} + \frac{v_0(\boldsymbol{X})}{1-e(\boldsymbol{X})}\right) h(\boldsymbol{X})^2 f(\boldsymbol{X}) \mu(d\boldsymbol{X})}{\left(\int h(\boldsymbol{X}) f(\boldsymbol{X}) \mu(d\boldsymbol{X})\right)^2},$$

where $f(\boldsymbol{X})$ is the population density function of $\boldsymbol{X}$. If the pseudo-outcome is homoskedastic, i.e., $v_1(\boldsymbol{X}) = v_0(\boldsymbol{X}) = v$, then the above formula simplifies to

$$\begin{aligned} n \times \text{Var}\left(\hat{\Delta}_h^I(t) | \boldsymbol{X}, \boldsymbol{A}, \boldsymbol{\delta}\right) &\rightarrow \frac{v \int \left(\frac{1}{e(\boldsymbol{X})} + \frac{1}{1-e(\boldsymbol{X})}\right) h(\boldsymbol{X})^2 f(\boldsymbol{X}) \mu(d\boldsymbol{X})}{\left(\int h(\boldsymbol{X}) f(\boldsymbol{X}) \mu(d\boldsymbol{X})\right)^2} \\ &= v / C_h \int \frac{h(\boldsymbol{X})^2 f(\boldsymbol{X})}{e(\boldsymbol{X})(1 - e(\boldsymbol{X}))} \mu(d\boldsymbol{X}), \end{aligned} \tag{2}$$

where $C_h = \left(\int h(\boldsymbol{X}) f(\boldsymbol{X}) \mu(d\boldsymbol{X})\right)^2$. Then, by applying the Cauchy–Schwarz inequality, we have that

$$\begin{aligned} C_h &= \left(\int \frac{h(\boldsymbol{X})}{\sqrt{e(\boldsymbol{X})(1 - e(\boldsymbol{X}))}} \sqrt{e(\boldsymbol{X})(1 - e(\boldsymbol{X}))} f(\boldsymbol{X}) \mu(d\boldsymbol{X})\right)^2 \\ &\leq \int \frac{h(\boldsymbol{X})^2}{e(\boldsymbol{X})(1 - e(\boldsymbol{X}))} f(\boldsymbol{X}) \mu(d\boldsymbol{X}) \times \int e(\boldsymbol{X})(1 - e(\boldsymbol{X})) f(\boldsymbol{X}) \mu(d\boldsymbol{X}). \end{aligned}$$

The above equality is achieved when $\frac{h(\boldsymbol{X})}{\sqrt{e(\boldsymbol{X})(1-e(\boldsymbol{X}))}} \propto \sqrt{e(\boldsymbol{X})(1 - e(\boldsymbol{X}))}$, or equivalently $h(\boldsymbol{X}) \propto e(\boldsymbol{X})(1 - e(\boldsymbol{X}))$. Finally, result 1 follows by directly applying the above to the right-hand side of (2).

□

Result 2. *(Optimal Type II estimator) Define $U^{(a)} = min(T^{(a)}, C^{(a)})$ as the right-censored survival outcome that would have been observed under treatment and control assignment if $a = 1$ and 0, respectively. If the variance of the pseudo-outcome $\frac{\mathbb{I}(U_i^{(a)} \geq t)}{K_c^{(a)}(t, \boldsymbol{X}_i)}$ is homoskedastic across both treatment groups, i.e.,*

$$Var\left(\frac{\mathbb{I}(U_i^{(1)} \geq t)}{K_c^{(1)}(t, \boldsymbol{X}_i)} | \boldsymbol{X}_i\right) = Var\left(\frac{\mathbb{I}(U_i^{(0)} \geq t)}{K_c^{(0)}(t, \boldsymbol{X}_i)} | \boldsymbol{X}_i\right) = c,$$

for some constant $c > 0$, then the OW with $\hat{\Delta}_{OW}^{II}(t)$ gives the smallest asymptotic variance for the Type II weighted estimator $\hat{\Delta}_h^{II}(t)$ among all $h(\boldsymbol{X}_i)$.

Proof. This proof is analogous to the proof for result 1. First notice that

$$\begin{aligned}
\hat{\Delta}_h^{II}(t) &= \frac{\sum_{i=1}^n w_i A_i \mathbb{I}(U_i \geq t)/K_c^{(1)}(t, \boldsymbol{X}_i)}{\sum_{i=1}^n w_i A_i} - \frac{\sum_{i=1}^n w_i (1-A_i) \mathbb{I}(U_i \geq t)/K_c^{(0)}(t, \boldsymbol{X}_i)}{\sum_{i=1}^n w_i (1-A_i)} \\
&= \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i) A_i \mathbb{I}(U_i \geq t)}{e(\boldsymbol{X}_i) K_c^{(1)}(t,\boldsymbol{X}_i)}}{\sum_{i=1}^n h(\boldsymbol{X}_i) A_i / e(\boldsymbol{X}_i)} - \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)(1-A_i)\mathbb{I}(U_i \geq t)}{(1-e(\boldsymbol{X}_i)) K_c^{(0)}(t,\boldsymbol{X}_i)}}{\sum_{i=1}^n h(\boldsymbol{X}_i)(1-A_i)/(1-e(\boldsymbol{X}_i))} \\
&= \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i) A_i \mathbb{I}(U_i^{(1)} \geq t)}{e(\boldsymbol{X}_i) K_c^{(1)}(t,\boldsymbol{X}_i)}}{\sum_{i=1}^n h(\boldsymbol{X}_i) A_i / e(\boldsymbol{X}_i)} - \frac{\sum_{i=1}^n \frac{h(\boldsymbol{X}_i)(1-A_i)\mathbb{I}(U_i^{(0)} \geq t)}{(1-e(\boldsymbol{X}_i)) K_c^{(0)}(t,\boldsymbol{X}_i)}}{\sum_{i=1}^n h(\boldsymbol{X}_i)(1-A_i)/(1-e(\boldsymbol{X}_i))},
\end{aligned} \tag{3}$$

where the last equality holds by noticing $\mathbb{I}(U_i \geq t) = \mathbb{I}(U_i^{(a)} \geq t)$ under $A_i = a$. Conditional on the sample $\tilde{\boldsymbol{X}} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ and $\tilde{\boldsymbol{A}} = \{A_1, \ldots, A_n\}$, only $\mathbb{I}(U_i^{(1)} \geq t)$ and $\mathbb{I}(U_i^{(0)} \geq t)$ are random in (3), so the variance of $\hat{\Delta}_h^{II}(t)$ is

$$\begin{aligned}
\mathrm{Var}\left(\hat{\Delta}_h^{II}(t) | \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{A}}\right) &= \frac{\sum_{i=1}^n \frac{h^2(\boldsymbol{X}_i) A_i}{e^2(\boldsymbol{X}_i)} \mathrm{Var}\left(\frac{\mathbb{I}(U_i^{(1)} \geq t)}{K_c^{(1)}(t,\boldsymbol{X}_i)} | \boldsymbol{X}_i, A_i\right)}{\left\{\sum_{i=1}^n h(\boldsymbol{X}_i) A_i / e(\boldsymbol{X}_i)\right\}^2} + \frac{\sum_{i=1}^n \frac{h^2(\boldsymbol{X}_i)(1-A_i)}{(1-e(\boldsymbol{X}_i))^2} \mathrm{Var}\left(\frac{\mathbb{I}(U_i^{(0)} \geq t)}{K_c^{(0)}(t,\boldsymbol{X}_i)} | \boldsymbol{X}_i, A_i\right)}{\left\{\sum_{i=1}^n h(\boldsymbol{X}_i)(1-A_i)/(1-e(\boldsymbol{X}_i))\right\}^2} \\
&= \frac{\sum_{i=1}^n \frac{h^2(\boldsymbol{X}_i) A_i}{e^2(\boldsymbol{X}_i)} \mathrm{Var}\left(\frac{\mathbb{I}(U_i^{(1)} \geq t)}{K_c^{(1)}(t,\boldsymbol{X}_i)} | \boldsymbol{X}_i\right)}{\left\{\sum_{i=1}^n h(\boldsymbol{X}_i) A_i / e(\boldsymbol{X}_i)\right\}^2} + \frac{\sum_{i=1}^n \frac{h^2(\boldsymbol{X}_i)(1-A_i)}{(1-e(\boldsymbol{X}_i))^2} \mathrm{Var}\left(\frac{\mathbb{I}(U_i^{(0)} \geq t)}{K_c^{(0)}(t,\boldsymbol{X}_i)} | \boldsymbol{X}_i\right)}{\left\{\sum_{i=1}^n h(\boldsymbol{X}_i)(1-A_i)/(1-e(\boldsymbol{X}_i))\right\}^2} \\
&= \frac{\sum_{i=1}^n \frac{h^2(\boldsymbol{X}_i) A_i}{e^2(\boldsymbol{X}_i)} c_1(\boldsymbol{X}_i)}{\left\{\sum_{i=1}^n h(\boldsymbol{X}_i) A_i / e(\boldsymbol{X}_i)\right\}^2} + \frac{\sum_{i=1}^n \frac{h^2(\boldsymbol{X}_i)(1-A_i)}{(1-e(\boldsymbol{X}_i))^2} c_0(\boldsymbol{X}_i)}{\left\{\sum_{i=1}^n h(\boldsymbol{X}_i)(1-A_i)/(1-e(\boldsymbol{X}_i))\right\}^2},
\end{aligned}$$

where $c_a(\boldsymbol{X}_i) = \mathrm{Var}\left(\frac{\mathbb{I}(U_i^{(a)} \geq t)}{K_c^{(a)}(t,\boldsymbol{X}_i)} | \boldsymbol{X}_i\right)$. Averaging the above first over the distribution of A (using $E[A_i/e(\boldsymbol{X}_i)] = E[(1-A_i)/(1-e(\boldsymbol{X}_i))] = 1$), and then over the distribution of $\boldsymbol{X}$, and again applying Slutsky's theorem, we have that

$$n \times \mathrm{Var}\left(\hat{\Delta}_h^{II}(t) | \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{A}}\right) \rightarrow \frac{\int \left(\frac{c_1(\boldsymbol{X})}{e(\boldsymbol{X})} + \frac{c_0(\boldsymbol{X})}{1-e(\boldsymbol{X})}\right) h(\boldsymbol{X})^2 f(\boldsymbol{X}) \mu(d\boldsymbol{X})}{\left(\int h(\boldsymbol{X}) f(\boldsymbol{X}) \mu(d\boldsymbol{X})\right)^2},$$

where $f(\boldsymbol{X})$ is the population density function of $\boldsymbol{X}$. If the pseudo-outcome is homoskedastic, i.e., $c_1(\boldsymbol{X}) = c_0(\boldsymbol{X}) = c$, then the above formula simplifies to

$$\begin{aligned}
n \times \mathrm{Var}\left(\hat{\Delta}_h^{I}(t) | \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{A}}\right) &\rightarrow \frac{c \int \left(\frac{1}{e(\boldsymbol{X})} + \frac{1}{1-e(\boldsymbol{X})}\right) h(\boldsymbol{X})^2 f(\boldsymbol{X}) \mu(d\boldsymbol{X})}{\left(\int h(\boldsymbol{X}) f(\boldsymbol{X}) \mu(d\boldsymbol{X})\right)^2} \\
&= c/C_h \int \frac{h(\boldsymbol{X})^2 f(\boldsymbol{X})}{e(\boldsymbol{X})(1-e(\boldsymbol{X}))} \mu(d\boldsymbol{X}),
\end{aligned} \tag{4}$$

where $C_h = \left(\int h(\boldsymbol{X}) f(\boldsymbol{X}) \mu(d\boldsymbol{X})\right)^2$. Then, by applying the Cauchy–Schwarz inequality, we have that

$$\begin{aligned}
C_h &= \left(\int \frac{h(\boldsymbol{X})}{\sqrt{e(\boldsymbol{X})(1-e(\boldsymbol{X}))}} \sqrt{e(\boldsymbol{X})(1-e(\boldsymbol{X}))} f(\boldsymbol{X}) \mu(d\boldsymbol{X})\right)^2 \\
&\leq \int \frac{h(\boldsymbol{X})^2}{e(\boldsymbol{X})(1-e(\boldsymbol{X}))} f(\boldsymbol{X}) \mu(d\boldsymbol{X}) \times \int e(\boldsymbol{X})(1-e(\boldsymbol{X})) f(\boldsymbol{X}) \mu(d\boldsymbol{X}).
\end{aligned}$$

The above equality is achieved when $\frac{h(\boldsymbol{X})}{\sqrt{e(\boldsymbol{X})(1-e(\boldsymbol{X}))}} \propto \sqrt{e(\boldsymbol{X})(1-e(\boldsymbol{X}))}$, or equivalently $h(\boldsymbol{X}) \propto e(\boldsymbol{X})(1-e(\boldsymbol{X}))$. Finally, result 1 follows by directly applying the above to the right-hand side of (4).

□

Web Appendix C: Variance estimation for $\hat{\Delta}_h^I(t)$

We will derive the variance estimator for $\hat{\Delta}_h^I(t)$ based on the empirical sandwich method, when the PS and censoring process are estimated by a logistic regression and Weibull regression, respectively. The derivation consists of three components. In part (a) and (b), we will derive the estimating equation for the PS and censoring process model, respectively. Then, in part (c), we will finally propose the variance estimator for $\hat{\Delta}_h^I(t)$.

Part (a) Propensity score

We use a logistic model $e(\boldsymbol{X}_i;\boldsymbol{\beta}) = P(A_i = 1|\boldsymbol{X}_i) = \frac{1}{1+\exp(-\boldsymbol{X}_i^T\boldsymbol{\beta})}$ to describe the propensity score. The estimating equation for $\boldsymbol{\beta}$ is

$$\begin{aligned}\mathbf{0} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial e(\boldsymbol{X}_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\left(e(\boldsymbol{X}_i;\boldsymbol{\beta})(1-e(\boldsymbol{X}_i;\boldsymbol{\beta}))\right)^{-1}[A_i - e(\boldsymbol{X}_i;\boldsymbol{\beta})] \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{X}_i[A_i - e(\boldsymbol{X}_i;\boldsymbol{\beta})].\end{aligned}$$

Now, we expand the above score equation around the true parameter $\boldsymbol{\beta}$ leading to

$$\begin{aligned}\mathbf{0} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{X}_i\left(A_i - e(\boldsymbol{X}_i;\boldsymbol{\beta})\right) + \frac{1}{\sqrt{n}}\frac{\partial\left\{\sum_{i=1}^{n}\boldsymbol{X}_i\left(A_i - e(\boldsymbol{X}_i;\boldsymbol{\beta})\right)\right\}}{\partial\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{X}_i\left(A_i - e(\boldsymbol{X}_i;\boldsymbol{\beta})\right) + \left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^T e(\boldsymbol{X}_i;\boldsymbol{\beta})\left(1-e(\boldsymbol{X}_i;\boldsymbol{\beta})\right)\right\}\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + o_p(1) \\ \Longrightarrow \sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) &= \boldsymbol{E}_{\beta\beta}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{X}_i\left(A_i - e(\boldsymbol{X}_i;\boldsymbol{\beta})\right) + o_p(1),\end{aligned}$$

where $\boldsymbol{E}_{\beta\beta} = -\frac{1}{n}\sum_{i=1}^{n} e(\boldsymbol{X}_i;\boldsymbol{\beta})\left(1-e(\boldsymbol{X}_i;\boldsymbol{\beta})\right)\boldsymbol{X}_i\boldsymbol{X}_i^T$.

Part (b) Censoring process

We consider the following parametric Weibull regression for the censoring time C:

$$K_c^{(a)}(t|\boldsymbol{X}_i) = P(C_i \geq t|\boldsymbol{X}_i, A_i = a) = \exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_a}t^{\gamma_a}\right),$$

where γ_a is a treatment-specific scale parameter and $\boldsymbol{\theta}_a$ is treatment-specific coefficients associated with covariates $\boldsymbol{X}$. The hazard function is $h_i^{(a)}(t|\boldsymbol{X}_i, A_i = a) = e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_a}\gamma_a t^{\gamma_a - 1}$. The unknown parameters, γ_a and $\boldsymbol{\theta}_a$, for $a = 1, 0$, are estimated through all subjects from the treatment and control group, respectively. Then the log-likelihood for the Weibull regression based on the observed outcome ($U = \min\{T, C\}, \delta = \mathbb{I}(T \leq C)$) is

$$\begin{aligned}l(\boldsymbol{\theta}_a, \gamma_a) &= \log\left(\prod_{i\in\text{Group } a} h_i(U_i|\boldsymbol{X}_i, A_i = a)^{1-\delta_i}S(U_i|\boldsymbol{X}_i, A_i = a)\right) \\ &= \sum_{i=1}^{n}\mathbb{I}(A_i = a)\left\{(1-\delta_i)(\boldsymbol{X}_i^T\boldsymbol{\theta}_a + \log\gamma_a + (\gamma_a - 1)\log U_i) - e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_a}U_i^{\gamma_a}\right\},. \\ &\propto \sum_{i=1}^{n}\mathbb{I}(A_i = a)\left\{(1-\delta_i)(\boldsymbol{X}_i^T\boldsymbol{\theta}_a + \log\gamma_a + \gamma_a\log U_i) - e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_a}U_i^{\gamma_a}\right\},\end{aligned}$$

Therefore the estimating equations for $\boldsymbol{\theta}_a$ and γ_a, respectively, are

$$\begin{cases} \mathbf{0} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}(A_i = a) \left\{ \boldsymbol{X}_i \left((1-\delta_i) - U_i^{\gamma_a} e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_a} \right) \right\} \\ 0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}(A_i = a) \left\{ (1-\delta_i) \left(\frac{1}{\gamma_a} + \log U_i \right) - U_i^{\gamma_a} \log U_i e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_a} \right\} \end{cases}.$$

Now, we expand the estimating equation for $\boldsymbol{\theta}_a$ around the true parameter $\boldsymbol{\theta}_a$ leading to

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_a - \boldsymbol{\theta}_a) = \boldsymbol{E}_{\theta_a\theta_a}^{(a)-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}(A_i = a) \left\{ \boldsymbol{X}_i \left((1-\delta_i) - U_i^{\gamma_a} e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_a} \right) \right\} + o_p(1),$$

where $\boldsymbol{E}_{\theta_a\theta_a}^{(a)} = -\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i = a) \left\{ U_i^{\gamma_a} e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_a} \boldsymbol{X}_i \boldsymbol{X}_i \right\}$. Similarly, we have the following property for $\hat{\gamma}_a$ by expanding the estimating equation for γ_a:

$$\sqrt{n}(\hat{\gamma}_a - \gamma_a) = E_{\gamma_a\gamma_a}^{(a)-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}(A_i = a) \left\{ (1-\delta_i) \left(\frac{1}{\gamma_a} + \log U_i \right) - U_i^{\gamma_a} \log U_i e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_a} \right\} + o_p(1),$$

where $E_{\gamma_a\gamma_a}^{(a)} = -\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i = a) \left\{ \frac{1-\delta_i}{\gamma_a^2} + U_i^{\gamma_a} (\log U_i)^2 e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_a} \right\}$.

Part (c) variance estimation for $\hat{\Delta}_h^I(t)$

Recall the weighted estimator $\hat{\Delta}_h^I(t)$:

$$\hat{\Delta}_h^I(t) = \frac{\sum_{i=1}^{n} \frac{\hat{w}_i A_i \delta_i \mathbb{I}(U_i \geq t)}{\hat{K}_c^{(1)}(U_i|\boldsymbol{X}_i)}}{\sum_{i=1}^{n} \frac{\hat{w}_i A_i \delta_i}{\hat{K}_c^{(1)}(U_i|\boldsymbol{X}_i)}} - \frac{\sum_{i=1}^{n} \frac{\hat{w}_i (1-A_i) \delta_i \mathbb{I}(U_i \geq t)}{\hat{K}_c^{(0)}(U_i|\boldsymbol{X}_i)}}{\sum_{i=1}^{N} \frac{\hat{w}_i (1-A_i) \delta_i}{\hat{K}_c^{(0)}(U_i|\boldsymbol{X}_i)}} = \hat{S}_h^{I(1)}(t) - \hat{S}_h^{I(0)}(t),$$

where $\hat{w}_i = \frac{h(\boldsymbol{X}_i)}{\hat{e}(\boldsymbol{X}_i)}$ for treated units and $\hat{w}_i = \frac{h(\boldsymbol{X}_i)}{1-\hat{e}(\boldsymbol{X}_i)}$ for control units. Noting that $\hat{K}_c^{(1)}(U_i|\boldsymbol{X}_i) = \exp\left(-e^{\boldsymbol{X}_i^T \hat{\boldsymbol{\theta}}_1} U_i^{\hat{\gamma}_1} \right)$, $\hat{S}_h^{I(1)}(t)$ can be viewed as the solution of the following estimating equation

$$0 = \sum_{i=1}^{n} \frac{\hat{w}_i A_i \delta_i (\mathbb{I}(U_i \geq t) - \hat{S}_h^{I(1)}(t))}{\exp\left(-e^{\boldsymbol{X}_i^T \hat{\boldsymbol{\theta}}_1} U_i^{\hat{\gamma}_1} \right)}$$

We can expand it around $S_h^{(1)}(t) = \frac{\mathbb{E}\left[h(\boldsymbol{X}) \mathbb{I}(T^{(1)} \geq t) \right]}{\mathbb{E}[h(\boldsymbol{X})]}$, the true propensity score, and the true censoring function to obtain

$$\begin{aligned} \sqrt{n}(\hat{S}_h^{I(1)}(t) - S_h^{(1)}(t)) = & \frac{E_h^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \frac{w_i A_i \delta_i (\mathbb{I}(U_i \geq t) - S_h^{(1)}(t))}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_1} U_i^{\gamma_1} \right)} + E_h^{-1} \boldsymbol{H}_{\theta_1}^{(1)T} \sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) \\ & + E_h^{-1} H_{\gamma_1}^{(1)} \sqrt{n}(\hat{\gamma}_1 - \gamma_1) + E_h^{-1} \boldsymbol{H}_{\beta}^{(1)T} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1), \end{aligned} \quad (5)$$

where

$$\begin{aligned}
E_h &= \frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{X}_i),\\
\boldsymbol{H}_{\theta_1}^{(1)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{w_i A_i \delta_i(\mathbb{I}(U_i \geq t) - \tau_1)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_1} U_i^{\gamma_1}\right)} U_i^{\gamma_1}\exp(\boldsymbol{X}_i^T\boldsymbol{\theta}_1)\boldsymbol{X}_i,\\
H_{\gamma_1}^{(1)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{w_i A_i \delta_i(\mathbb{I}(U_i \geq t) - \tau_1)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_1} U_i^{\gamma_1}\right)} U_i^{\gamma_1}\log U_i \exp(\boldsymbol{X}_i^T\boldsymbol{\theta}_1),\\
\boldsymbol{H}_{\beta}^{(1)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{A_i \delta_i(\mathbb{I}(U_i \geq t) - \tau_1)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_1} U_i^{\gamma_1}\right)} \frac{\partial w_i}{\partial \boldsymbol{\beta}}.
\end{aligned}$$

Similarly, we can view $\hat{S}_h^{I(0)}(t)$ as the solution of the following estimating equation

$$0 = \sum_{i=1}^{n} \frac{\hat{w}_i(1 - A_i)\delta_i(\mathbb{I}(U_i \geq t) - \hat{S}_h^{I(0)}(t))}{\exp\left(-e^{\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}}_0} U_i^{\hat{\gamma}_0}\right)},$$

and then expand it around $S_h^{(0)}(t) = \frac{\mathbb{E}\left[h(\boldsymbol{X})\mathbb{I}(T^{(0)}\geq t)\right]}{\mathbb{E}[h(\boldsymbol{X})]}$, the true propensity score, and the true censoring function to obtain

$$\begin{aligned}
\sqrt{n}(\hat{S}_h^{I(0)}(t) - S_h^{(0)}(t)) =& \frac{E_h^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \frac{w_i(1 - A_i)\delta_i(\mathbb{I}(U_i \geq t) - \tau_0)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0} U_i^{\gamma_0}\right)} + E_h^{-1}\boldsymbol{H}_{\theta_0}^{(0)T}\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)\\
&+ E_h^{-1}H_{\gamma_0}^{(0)}\sqrt{n}(\hat{\gamma}_0 - \gamma_0) + E_h^{-1}\boldsymbol{H}_{\beta}^{(0)T}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1),
\end{aligned} \tag{6}$$

where

$$\begin{aligned}
\boldsymbol{H}_{\theta_0}^{(0)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{w_i(1 - A_i)\delta_i(\mathbb{I}(U_i \geq t) - \tau_0)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0} U_i^{\gamma_0}\right)} U_i^{\gamma_0}\exp(\boldsymbol{X}_i^T\boldsymbol{\theta}_0)\boldsymbol{X}_i\\
H_{\gamma_0}^{(0)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{w_i(1 - A_i)\delta_i(\mathbb{I}(U_i \geq t) - \tau_0)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0} U_i^{\gamma_0}\right)} U_i^{\gamma_0}\log U_i \exp(\boldsymbol{X}_i^T\boldsymbol{\theta}_0)\\
\boldsymbol{H}_{\beta}^{(0)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{(1 - A_i)\delta_i(\mathbb{I}(U_i \geq t) - \tau_0)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0} U_i^{\gamma_0}\right)} \frac{\partial w_i}{\partial \boldsymbol{\beta}}
\end{aligned}$$

Then, we combine (5) and (6) to obtain the following influence function of $\hat{\Delta}_h^I(t)$:

$$
\begin{aligned}
&\sqrt{n}(\hat{\Delta}_h^I(t) - \Delta_h(t)) \\
=&\frac{E_h^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{w_i A_i \delta_i (\mathbb{I}(U_i \geq t) - \hat{S}_h^{I(1)}(t))}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_1} U_i^{\gamma_1}\right)} - \frac{w_i (1 - A_i) \delta_i (\mathbb{I}(U_i \geq t) - \hat{S}_h^{I(0)}(t))}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_0} U_i^{\gamma_0}\right)} \right\} \\
&+ E_h^{-1} \boldsymbol{H}_{\theta_1}^{(1)T} \sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) - E_h^{-1} \boldsymbol{H}_{\theta_0}^{(0)T} \sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0) \\
&+ E_h^{-1} H_{\gamma_1}^{(1)} \sqrt{n}(\hat{\gamma}_1 - \gamma_1) - E_h^{-1} H_{\gamma_0}^{(0)} \sqrt{n}(\hat{\gamma}_0 - \gamma_0) \\
&+ E_h^{-1} \left\{ \boldsymbol{H}_{\beta}^{(1)} - \boldsymbol{H}_{\beta}^{(0)} \right\}^T \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1) \\
=&\frac{E_h^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \underbrace{\frac{w_i A_i \delta_i (\mathbb{I}(U_i \geq t) - \hat{S}_h^{I(1)}(t))}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_1} U_i^{\gamma_1}\right)} - \frac{w_i (1 - A_i) \delta_i (\mathbb{I}(U_i \geq t) - \hat{S}_h^{I(0)}(t))}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_0} U_i^{\gamma_0}\right)}}_{\text{denoted by } I_{\Delta,i}} \right. \\
&+ \underbrace{\boldsymbol{H}_{\theta_1}^{(1)^T} \boldsymbol{E}_{\theta_1\theta_1}^{(1)^{-1}} \left\{ A_i \boldsymbol{X}_i \left((1 - \delta_i) - U_i^{\gamma_1} e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_1} \right) \right\} - \boldsymbol{H}_{\theta_0}^{(0)^T} \boldsymbol{E}_{\theta_0\theta_0}^{(0)^{-1}} \left\{ (1 - A_i) \boldsymbol{X}_i \left((1 - \delta_i) - U_i^{\gamma_0} e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_0} \right) \right\}}_{\text{denoted by } I_{\boldsymbol{\theta},i}} \\
&+ \underbrace{H_{\gamma_1}^{(1)} E_{\gamma_1\gamma_1}^{(1)^{-1}} A_i \left\{ (1 - \delta_i) \left(\frac{1}{\gamma_1} + \log U_i \right) - U_i^{\gamma_1} \log U_i e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_1} \right\} - H_{\gamma_0}^{(0)} E_{\gamma_0\gamma_0}^{(0)^{-1}} (1 - A_i) \left\{ (1 - \delta_i) \left(\frac{1}{\gamma_0} + \log U_i \right) - U_i^{\gamma_0} \log U_i e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_0} \right\}}_{\text{denoted by } I_{\boldsymbol{\gamma},i}} \\
&\left. + \underbrace{(\boldsymbol{H}_{\beta}^{(1)} - \boldsymbol{H}_{\beta}^{(0)})^T \boldsymbol{E}_{\beta\beta}^{-1} \boldsymbol{X}_i \left(A_i - e(\boldsymbol{X}_i; \boldsymbol{\beta}) \right)}_{\text{denoted by } I_{\boldsymbol{\beta},i}} \right\} + o_p(1) \\
=&\frac{E_h^{-1}}{\sqrt{n}} \sum_{i=1}^{n} (I_{\Delta,i} + I_{\boldsymbol{\theta},i} + I_{\boldsymbol{\gamma},i} + I_{\boldsymbol{\beta},i}) + o_p(1)
\end{aligned}
$$

The empirical sandwich method uses the sample variance of the influence function

$$
\frac{1}{n^2 E_h^2} \sum_{i=1}^{n} \hat{I}_i^2 = \frac{1}{n^2 E_h^2} \sum_{i=1}^{n} (\hat{I}_{\Delta,i} + \hat{I}_{\boldsymbol{\theta},i} + \hat{I}_{\boldsymbol{\gamma},i} + \hat{I}_{\boldsymbol{\beta},i})^2
$$

to estimate $\mathrm{Var}(\hat{\Delta}_h^I(t))$, where $E_h = \frac{1}{n} \sum_{i=1}^{n} h(\boldsymbol{X}_i)$, and $\hat{I}_{\Delta,i}$, $\hat{I}_{\boldsymbol{\theta},i}$, $\hat{I}_{\boldsymbol{\gamma},i}$, and $\hat{I}_{\boldsymbol{\beta},i}$ are $I_{\Delta,i}$, $I_{\boldsymbol{\theta},i}$, $I_{\boldsymbol{\gamma},i}$, and $I_{\boldsymbol{\beta},i}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, $\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}_1$, $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_0$, $\gamma_1 = \hat{\gamma}_1$, and $\gamma_0 = \hat{\gamma}_0$, respectively.

Web Appendix D: Variance estimation for $\hat{\Delta}_h^{II}(t)$

Recall the weighted estimator $\hat{\Delta}_h^{II}(t)$:

$$\hat{\Delta}_h^{II}(t) = \frac{\sum_{i=1}^n \frac{\hat{w}_i A_i \mathbb{I}(U_i \geq t)}{\hat{K}_c^{(1)}(u|\boldsymbol{X}_i)}}{\sum_{i=1}^n \hat{w}_i A_i} - \frac{\sum_{i=1}^n \frac{\hat{w}_i (1-A_i) \mathbb{I}(U_i \geq t)}{\hat{K}_c^{(0)}(u|\boldsymbol{X}_i)}}{\sum_{i=1}^N w_i (1-A_i)} = \hat{S}_h^{II(1)}(t) - \hat{S}_h^{II(0)}(t),$$

where $\hat{w}_i = \frac{h(\boldsymbol{X}_i)}{\hat{e}(\boldsymbol{X}_i)}$ for treated units and $\hat{w}_i = \frac{h(\boldsymbol{X}_i)}{1-\hat{e}(\boldsymbol{X}_i)}$ for control units. Next we derive the asymptotic variance. Specifically, $\hat{S}_h^{II(1)}(t)$ can be viewed as the solution of the following estimating equation

$$0 = \sum_{i=1}^n \frac{\hat{w}_i A_i \mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T \hat{\boldsymbol{\theta}}_1} t^{\hat{\gamma}_1}\right)} - w_i A_i \hat{S}_h^{II(1)}(t)$$

We can expand it around $S_h^{(1)}(t) = \frac{\mathbb{E}\left[h(\boldsymbol{X})\mathbb{I}(T^{(1)} \geq t)\right]}{\mathbb{E}[h(\boldsymbol{X})]}$, the true propensity score, and the true censoring function to obtain

$$\begin{aligned}\sqrt{n}(\hat{S}_h^{II(1)}(t) - S_h^{(1)}(t)) =& \frac{E_h^{-1}}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{w_i A_i \mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_1} t^{\gamma_1}\right)} - w_i A_i S_h^{(1)}(t) \right\} + E_h^{-1} \boldsymbol{H}_{\theta_1}^{(1)T} \sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) \\ &+ E_h^{-1} H_{\gamma_1}^{(1)} \sqrt{n}(\hat{\gamma}_1 - \gamma_1) + E_h^{-1} \boldsymbol{H}_{\beta}^{(1)T} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1),\end{aligned}$$

where

$$\begin{aligned} E_h &= \frac{1}{n} \sum_{i=1}^n h(\boldsymbol{X}_i), \\ \boldsymbol{H}_{\theta_1}^{(1)} &= \frac{1}{n} \sum_{i=1}^n \frac{w_i A_i \mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_1} t^{\gamma_1}\right)} t^{\gamma_1} \exp(\boldsymbol{X}_i^T \boldsymbol{\theta}_1) \boldsymbol{X}_i, \\ H_{\gamma_1}^{(1)} &= \frac{1}{n} \sum_{i=1}^n \frac{w_i A_i \mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_1} t^{\gamma_1}\right)} t^{\gamma_1} \log t \exp(\boldsymbol{X}_i^T \boldsymbol{\theta}_1), \\ \boldsymbol{H}_{\beta}^{(1)} &= \frac{1}{n} \sum_{i=1}^n \frac{A_i \mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_1} t^{\gamma_1}\right)} \frac{\partial w_i}{\partial \boldsymbol{\beta}} - \tau_1 A_i \frac{\partial w_i}{\partial \boldsymbol{\beta}}. \end{aligned}$$

Similarly, the estimating equation for $\hat{S}_h^{(0)}(t)$ is

$$0 = \sum_{i=1}^n \frac{\hat{w}_i (1-A_i) \mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T \hat{\boldsymbol{\theta}}_0} t^{\hat{\gamma}_0}\right)} - w_i (1-A_i) \hat{S}_h^{II(0)}(t).$$

Then we can expand it around $S_h^{(0)}(t) = \frac{\mathbb{E}\left[h(\boldsymbol{X})\mathbb{I}(T^{(0)} \geq t)\right]}{\mathbb{E}[h(\boldsymbol{X})]}$, the true propensity score, and the true censoring function to obtain

$$\begin{aligned}\sqrt{n}(\hat{S}_h^{II(0)}(t) - S_h^{(0)}(t)) =& \frac{E_h^{-1}}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{w_i (1-A_i) \mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T \boldsymbol{\theta}_0} t^{\gamma_0}\right)} - w_i (1-A_i) S_h^{(0)}(t) \right\} + E_h^{-1} \boldsymbol{H}_{\theta_0}^{(1)T} \sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0) \\ &+ E_h^{-1} H_{\gamma_0}^{(1)} \sqrt{n}(\hat{\gamma}_0 - \gamma_0) + E_h^{-1} \boldsymbol{H}_{\beta}^{(0)T} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1),\end{aligned}$$

where

$$
\begin{aligned}
\boldsymbol{H}_{\theta_0}^{(0)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{w_i(1-A_i)\mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0} t^{\gamma_0}\right)} t^{\gamma_0}\exp(\boldsymbol{X}_i^T\boldsymbol{\theta}_0)\boldsymbol{X}_i,\\
H_{\gamma_0}^{(0)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{w_i(1-A_i)\mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0} t^{\gamma_0}\right)} t^{\gamma_0}\log u\exp(\boldsymbol{X}_i^T\boldsymbol{\theta}_0),\\
\boldsymbol{H}_{\beta}^{(0)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{(1-A_i)\mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0} t^{\gamma_0}\right)} \frac{\partial w_i}{\partial \boldsymbol{\beta}} - \tau_0(1-A_i)\frac{\partial w_i}{\partial \boldsymbol{\beta}}.
\end{aligned}
$$

To summarize, we have that

$$
\begin{aligned}
&\sqrt{n}(\hat{\Delta}_h^{II}(t) - \Delta_h(t))\\
=&\frac{E_h^{-1}}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{w_iA_i\mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_1}t^{\gamma_1}\right)} - \frac{w_i(1-A_i)\mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0}t^{\gamma_0}\right)} - w_iA_iS_h^{(1)}(t) + w_i(1-A_i)S_h^{(0)}(t)\right\}\\
&+ E_h^{-1}\boldsymbol{H}_{\theta_1}^{(1)T}\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) - E_h^{-1}\boldsymbol{H}_{\theta_0}^{(0)T}\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)\\
&+ E_h^{-1}H_{\gamma_1}^{(1)}\sqrt{n}(\hat{\gamma}_1 - \gamma_1) - E_h^{-1}H_{\gamma_0}^{(0)}\sqrt{n}(\hat{\gamma}_0 - \gamma_0)\\
&+ E_h^{-1}\left\{\boldsymbol{H}_{\beta}^{(1)} - \boldsymbol{H}_{\beta}^{(0)}\right\}^T\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1)\\
=&\frac{E_h^{-1}}{\sqrt{n}}\sum_{i=1}^{n}\Bigg\{\underbrace{\frac{w_iA_i\mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_1}t^{\gamma_1}\right)} - \frac{w_i(1-A_i)\mathbb{I}(U_i \geq t)}{\exp\left(-e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0}t^{\gamma_0}\right)} - w_iA_i\hat{S}_h^{(1)}(t) + w_i(1-A_i)\hat{S}_h^{(0)}(t)}_{\text{denoted by } I_{\Delta,i}^{II}}\\
&+ \underbrace{\boldsymbol{H}_{\theta_1}^{(1)^T}\boldsymbol{E}_{\theta_1\theta_1}^{(1)^{-1}}\left\{A_i\boldsymbol{X}_i\left((1-\delta_i) - U_i^{\gamma_1}e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_1}\right)\right\} - \boldsymbol{H}_{\theta_0}^{(0)^T}\boldsymbol{E}_{\theta_0\theta_0}^{(0)^{-1}}\left\{(1-A_i)\boldsymbol{X}_i\left((1-\delta_i) - U_i^{\gamma_0}e^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0}\right)\right\}}_{\text{denoted by } I_{\boldsymbol{\theta},i}^{II}}\\
&+ \underbrace{H_{\gamma_1}^{(1)}E_{\gamma_1\gamma_1}^{(1)^{-1}}A_i\left\{(1-\delta_i)\left(\frac{1}{\gamma_1} + \log U_i\right) - U_i^{\gamma_1}\log U_ie^{\boldsymbol{X}_i^T\boldsymbol{\theta}_1}\right\} - H_{\gamma_0}^{(0)}E_{\gamma_0\gamma_0}^{(0)^{-1}}(1-A_i)\left\{(1-\delta_i)\left(\frac{1}{\gamma_0} + \log U_i\right) - U_i^{\gamma_0}\log U_ie^{\boldsymbol{X}_i^T\boldsymbol{\theta}_0}\right\}}_{\text{denoted by } I_{\boldsymbol{\gamma},i}^{II}}\\
&+ \underbrace{(\boldsymbol{H}_{\beta}^{(1)} - \boldsymbol{H}_{\beta}^{(0)})^T\boldsymbol{E}_{\beta\beta}^{-1}\boldsymbol{X}_i\left(A_i - e(\boldsymbol{X}_i;\boldsymbol{\beta})\right)}_{\text{denoted by } I_{\boldsymbol{\beta},i}^{II}}\Bigg\} + o_p(1)\\
=&\frac{E_h^{-1}}{\sqrt{n}}\sum_{i=1}^{n}(I_{\Delta,i}^{II} + I_{\boldsymbol{\theta},i}^{II} + I_{\boldsymbol{\gamma},i}^{II} + I_{\boldsymbol{\beta},i}^{II}) + o_p(1).
\end{aligned}
$$

The empirical sandwich method uses the sample variance of the influence function

$$
\frac{1}{n^2E_h^2}\sum_{i=1}^{n}\hat{I}_i^{II^2} = \frac{1}{nE_h^2}\sum_{i=1}^{n}(\hat{I}_{\Delta,i}^{II} + \hat{I}_{\boldsymbol{\theta},i}^{II} + \hat{I}_{\boldsymbol{\gamma},i}^{II} + \hat{I}_{\boldsymbol{\beta},i}^{II})^2
$$

to estimate $\mathrm{Var}(\hat{\Delta}_h^{II}(t))$, where $E_h = \frac{1}{n}\sum_{i=1}^{n}h(\boldsymbol{X}_i)$, and $\hat{I}_{\Delta,i}^{II}$, $\hat{I}_{\boldsymbol{\theta},i}^{II}$, $\hat{I}_{\boldsymbol{\gamma},i}^{II}$, and $\hat{I}_{\boldsymbol{\beta},i}^{II}$ are $I_{\Delta,i}^{II}$, $I_{\boldsymbol{\theta},i}^{II}$, $I_{\boldsymbol{\gamma},i}^{II}$, and $I_{\boldsymbol{\beta},i}^{II}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, $\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}_1$, $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_0$, $\gamma_1 = \hat{\gamma}_1$, and $\gamma_0 = \hat{\gamma}_0$, respectively.

Web Appendix E: R Tutorial

1. Aim

In this Appendix, we provide a step-by-step guide for implementation of the proposed propensity score weighting approaches to estimate treatment effects on survival functions. We shall demostrate our proposed methodologies by using a simulated dataset, available at https://github.com/chaochengstat/OW_Survival. The example is written in `R` software.

2. Dataset

The `surv.csv` dataset available at https://github.com/chaochengstat/OW_Survival will be used to demostrate the proposed methods. The treatment variable `z` is binary, which takes values from one of `1` and `0` representing, respectively, treated and control in this example. The outcome `Time` is patient's survival time in months, which is defined as the difference between date of death and the study admission date and then devided it by 30. However, the outcome is subject to right censoring such that we only observed date of first occurance of last follow-up and death. The censoring indicator is `Event`, which takes `1` or `0` to denote the patient is alive or death on the previously given date. There are also 6 pre-treatment covariates `x1-x6`. Now we load the dataset and idenitify those variables.

```
# 1. Load Data
data=read.csv("https://raw.githubusercontent.com/chaochengstat/OW_Survival/main/surv.csv")[,-1]
# 2. Idenity column names of the treatment, survival outcome, and censoring indicator
Treatment = "z"
SurvTime =  "Time"
Status = "Event"
#3. Idenity column names of the pre-treatment covariates
Covariates=c("x1","x2","x3","x4","x5","x6")
# summary of those variables
summary(data[,c(Treatment,SurvTime,Status,Covariates)])
```

```
##        z               Time             Event             x1
##  Min.   :0.0000   Min.   :  0.010   Min.   :0.000   Min.   :-3.35000
##  1st Qu.:0.0000   1st Qu.:  1.240   1st Qu.:0.000   1st Qu.:-0.70000
##  Median :0.0000   Median :  3.130   Median :1.000   Median :-0.01500
##  Mean   :0.4955   Mean   :  4.987   Mean   :0.732   Mean   :-0.04249
##  3rd Qu.:1.0000   3rd Qu.:  6.532   3rd Qu.:1.000   3rd Qu.: 0.63000
##  Max.   :1.0000   Max.   :114.080   Max.   :1.000   Max.   : 3.11000
##        x2                  x3                 x4               x5
##  Min.   :-4.470000   Min.   :-3.88000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:-0.670000   1st Qu.:-0.68000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :-0.020000   Median :-0.04000   Median :0.0000   Median :0.000
##  Mean   : 0.000075   Mean   :-0.03687   Mean   :0.4885   Mean   :0.494
##  3rd Qu.: 0.680000   3rd Qu.: 0.60000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   : 3.470000   Max.   : 3.89000   Max.   :1.0000   Max.   :1.000
##        x6
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :1.0000
##  Mean   :0.5035
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

3. Propsensity Score Modeling

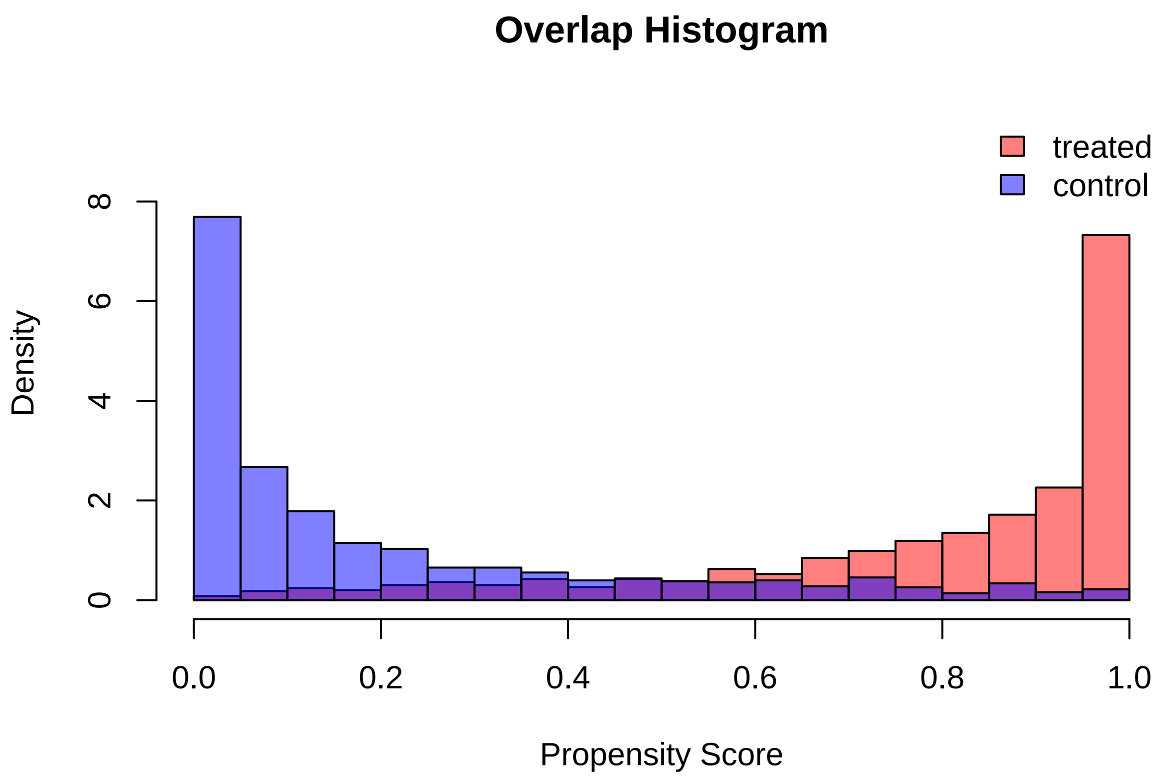
The logistic regressions will be used to estimate the propensity score. We will consider including all of the six pre-treatment covariates (`x1-x6`) into analysis:

```
# 2. Construct the logistic regression formula
PS.formula=as.formula(paste(Treatment,"~",paste(Covariates,collapse="+"),sep=""))
# the PS.formula is shown as below:
```

```
# z ~ x1 + x2 + x3 + x4 + x5 + x6
# 3. run the logistic regression
PS.model=glm(PS.formula,data=data,family=binomial(link="logit"))
# 4. obtain the propensity score
PS = 1/(1+exp(-c(PS.model$linear.predictors)))
```

The distributions of the estimated propensity scores in the treated and untreated group are visualized as below:

```
hist(PS[data[,Treatment]==1], breaks=20,col=rgb(1,0,0,0.5),xlim=c(0,1),
     ylim=c(0,9.5),main="Overlap Histogram",freq=F,xlab="Propensity Score")
hist(PS[data[,Treatment]==0], breaks=20,col=rgb(0,0,1,0.5), add=T,freq=F)
legend("topright", c("treated", "control"), bty = "n",
       fill=c(rgb(1,0,0,0.5),rgb(0,0,1,0.5)))
```



4. Censoring Score Modeling

Here, we will use the Weibull regression model to describe the censoring process. See *PS Weighting* section in the manuscript to learn more details about the parametric Weibull regression. We will treat all of the six pre-treatment covariates as independent variables in our Weibull regression. See the code below

```
# 1. library survival package
library("survival")
# 2. Construct the Weibull model formula
Censor.formula=as.formula(paste("Surv(",SurvTime,",I(1-",Status,"))","~",
                                paste(Covariates,collapse="+"),sep=""))
# Censor.formula is shown below
# Surv(Time, I(1 - Event)) ~ x1 + x2 + x3 + x4 + x5 + x6
# 3. Weibull Model for the treated group
data.trt=subset(data,data[,Treatment]==1)
Censor.trt.model = survreg(Censor.formula,data=data.trt,dist='weibull',score=T)
# 4. Weibull Model for the control group
data.con=subset(data,data[,Treatment]==0)
Censor.con.model = survreg(Censor.formula,data=data.con,dist='weibull',score=T)
```

5. Overlap Weighting

In what follows, we calculate the treatment effect on 6-month survival probability with overlap weighting, i.e., $\Delta_{OW}(6)$. Treatment effect with other balancing weights (IPW and symmetric and asymmetric triming) can be simularly obtained and the code will be briefly introduced in next part. The following code demostrates how to obtain Type I overlap weighting estimator ($\hat{\Delta}^{I}_{OW}(6)$):

```
# 1. Define a function to calculate the predicted censoring score
### Input: i. WeiModel: A Weibull model object;
###        ii. Dataset
###        ii. TimeVec: A vector of time (say, t).
### Output: Probability of P(C_i>t_i|X_i) for i=1,...,n.
CensorScoreFun=function(WeiModel,Dataset,TimeVec) {
  # extract estimatedregression parameters
  theta.est = -WeiModel$coefficients/WeiModel$scale
  gamma.est = 1/WeiModel$scale
  # calculate censoring score
  X=model.matrix(Censor.formula,data=Dataset)
  CensorScore= exp(-exp(c(X %*% theta.est)) * TimeVec^gamma.est)
  return(CensorScore)
}
# 2. Calculate the censoring scores
# censoring scores in the treated group
Kc.trt=CensorScoreFun(WeiModel=Censor.trt.model,Dataset=data.trt,TimeVec=data.trt[,SurvTime])
# censoring scores in the untreated group
Kc.con=CensorScoreFun(WeiModel=Censor.con.model,Dataset=data.con,TimeVec=data.con[,SurvTime])
# 3. Calculate the counterfactual survival probability in the treated group
w.trt = 1-PS[which(data[,Treatment]==1)] # balancing weight in the treated group
S1=sum(w.trt*data.trt[,Status]*(data.trt[,SurvTime]>=6)/Kc.trt)/sum(w.trt*data.trt[,Status]/Kc.trt)
# 4. Calculate the counterfactual survival probability in the untreated group
w.con = PS[which(data[,Treatment]==0)] # balancing weight in the untreated group
S0=sum(w.con*data.con[,Status]*(data.con[,SurvTime]>=6)/Kc.con)/sum(w.con*data.con[,Status]/Kc.con)
# 5. Finally, obtain the treatment effect
Delta.OW=S1-S0
cat(paste("The estimated treatment effect is",round(Delta.OW,4)))
```

```
## The estimated treatment effect is 0.1777
```

The following code demostrates how to obtain Type II overlap weighting estimator ($\hat{\Delta}^{II}_{OW}(6)$):

```
# 1. Calculate the censoring scores
# censoring scores in the treated group
Kc.trt=CensorScoreFun(WeiModel=Censor.trt.model,Dataset=data.trt,
                      TimeVec=rep(6,length(Censor.trt.model$y)))
# censoring scores in the untreated group
Kc.con=CensorScoreFun(WeiModel=Censor.con.model,Dataset=data.con,
                      TimeVec=rep(6,length(Censor.con.model$y)))
# 2. Calculate the counterfactual survival probability in the treated group
w.trt = 1-PS[which(data[,Treatment]==1)] # balancing weight in the treated group
S1=sum(w.trt*as.numeric(data.trt[,SurvTime]>=6)/Kc.trt)/sum(w.trt)
# 3. Calculate the counterfactual survival probability in the untreated group
w.con = PS[which(data[,Treatment]==0)] # balancing weight in the untreated group
S0=sum(w.con*as.numeric(data.con[,SurvTime]>=6)/Kc.con)/sum(w.con)
# 4. Finally, obtain the treatment effect
Delta.OW=S1-S0
cat(paste("The estimated treatment effect is",round(Delta.OW,4)))
```

```
## The estimated treatment effect is 0.1612
```

6. Other Balancing Weights and Confidence Interval Construction

Analogous to the previous part, we can estimate the treatment effect based on the IPW, symmetric weight or asymmetric weight. The R code for implementation of all the balancing weights are summarized in a unified function `SurvEffectWeibull`, available at https://github.com/chaochengstat/OW_Survival. Usage of this function is demostrated as follows

```
SurvEffectWeibull(Data,t,Treatment,SurvTime,Status,PS.formula,Censor.formula,Type,
                  Method,alpha,q)
```

Arguments are

- `Data`: a data frame
- `t`: a time point for evaluation of the treatment effect (i.e., t in $\Delta_h(t)$).
- `Treatment`: treatment variable.
- `SurvTime`: observed survival time.
- `Status`: censoring indicator.
- `PS.formula`: regression formula for the propensity score; see the *Propensity Score Modeling* part for more details.
- `Censor.formula`: regression formula for the Cox model for describing censoring process; see the *Censoring Score Modeling* part for more details.
- `Type`: 1 for estimator I (i.e., $\hat{\Delta}_h^I(t)$) and 2 for estimator II (i.e., $\hat{\Delta}_h^{II}(t)$)
- `Method`: balancing weights; `IPW` for IPW, `OW` for overlap weighting, `Symmetric` for symmetric weighting, and `Asymmetric` for asymmetric weighting.
- `alpha`: the triming threshold for symmetric weighting, i.e., α.
- `q`: the triming threshold for asymmetric weighting, i.e., q.

Output of this function include the point estimate, and the standard error and 95% normality-based confidence interval given by the robust sandwich variance approach.

We now calculate Types I and II estimators with IPW to illustrate usage of this function:

```r
# 1. Load SurvEffectWithCox function
source("https://raw.githubusercontent.com/chaochengstat/OW_Survival/main/functions_Weibull_v3.R")
# 2. define PS and Cox model formulas
PS.formula=as.formula(paste(Treatment,"~",paste(Covariates,collapse="+"),sep=""))
Censor.formula=as.formula(paste("Surv(",SurvTime,",I(1-",Status,"))","~",
                                paste(Covariates,collapse="+"),sep=""))
# 3. Type I IPW estimator
Delta.IPW1=SurvEffectWeibull(Data=data,t=6,Treatment="z",SurvTime="Time",
                             Status="Event",PS.formula=PS.formula,
                             Censor.formula=Censor.formula,Type=1,Method="IPW")
# 4. Type II IPW estimator
Delta.IPW2=SurvEffectWeibull(Data=data,t=6,Treatment="z",SurvTime="Time",
                             Status="Event",PS.formula=PS.formula,
                             Censor.formula=Censor.formula,Type=2,Method="IPW")
cat("Type I estimator: \n");round(Delta.IPW1,3);cat("Type II estimator: \n");round(Delta.IPW2,3)
```

```
## Type I estimator:

## Estimate       SE CI.lower CI.upper
##    0.179    0.078    0.026    0.332

## Type II estimator:

## Estimate       SE CI.lower CI.upper
##    0.151    0.072    0.011    0.292
```

Next, we calculate Types I and II estimators by symmetric triming with triming threshold $\alpha = 0.1$:

```r
# 1. Type I symmetric trimming estimator
Delta.SW1=SurvEffectWeibull(Data=data,t=6,Treatment="z",SurvTime="Time",
                             Status="Event",PS.formula=PS.formula,
                             Censor.formula=Censor.formula,Type=1,Method="Symmetric",alpha=0.1)
# 2. Type II symmetric trimming estimator
```

```r
Delta.SW2=SurvEffectWeibull(Data=data,t=6,Treatment="z",SurvTime="Time",
                            Status="Event",PS.formula=PS.formula,
                            Censor.formula=Censor.formula,Type=2,Method="Symmetric",alpha=0.1)
cat("Type I estimator: \n");round(Delta.SW1,3);cat("Type II estimator: \n");round(Delta.SW2,3)
```

```
## Type I estimator:

## Estimate       SE CI.lower CI.upper
##    0.151    0.045    0.063    0.238

## Type II estimator:

## Estimate       SE CI.lower CI.upper
##    0.154    0.040    0.076    0.232
```

Then, we calculate Types I and II estimators by asymmetric triming with triming threshold $q = 0.01$:

```r
# 1. Type I asymmetric trimming estimator
Delta.AW1=SurvEffectWeibull(Data=data,t=6,Treatment="z",SurvTime="Time",
                            Status="Event",PS.formula=PS.formula,
                            Censor.formula=Censor.formula,Type=1,Method="Asymmetric",q=0.01)
# 2. Type II asymmetric trimming estimator
Delta.AW2=SurvEffectWeibull(Data=data,t=6,Treatment="z",SurvTime="Time",
                            Status="Event",PS.formula=PS.formula,
                            Censor.formula=Censor.formula,Type=2,Method="Asymmetric",q=0.01)
cat("Type I estimator: \n");round(Delta.AW1,3);cat("Type II estimator: \n");round(Delta.AW2,3)
```

```
## Type I estimator:

## Estimate       SE CI.lower CI.upper
##    0.171    0.037    0.098    0.244

## Type II estimator:

## Estimate       SE CI.lower CI.upper
##    0.167    0.037    0.094    0.240
```

Finally, we repeat part 6 to caluate the Types I and II estimators with overlap weighting.

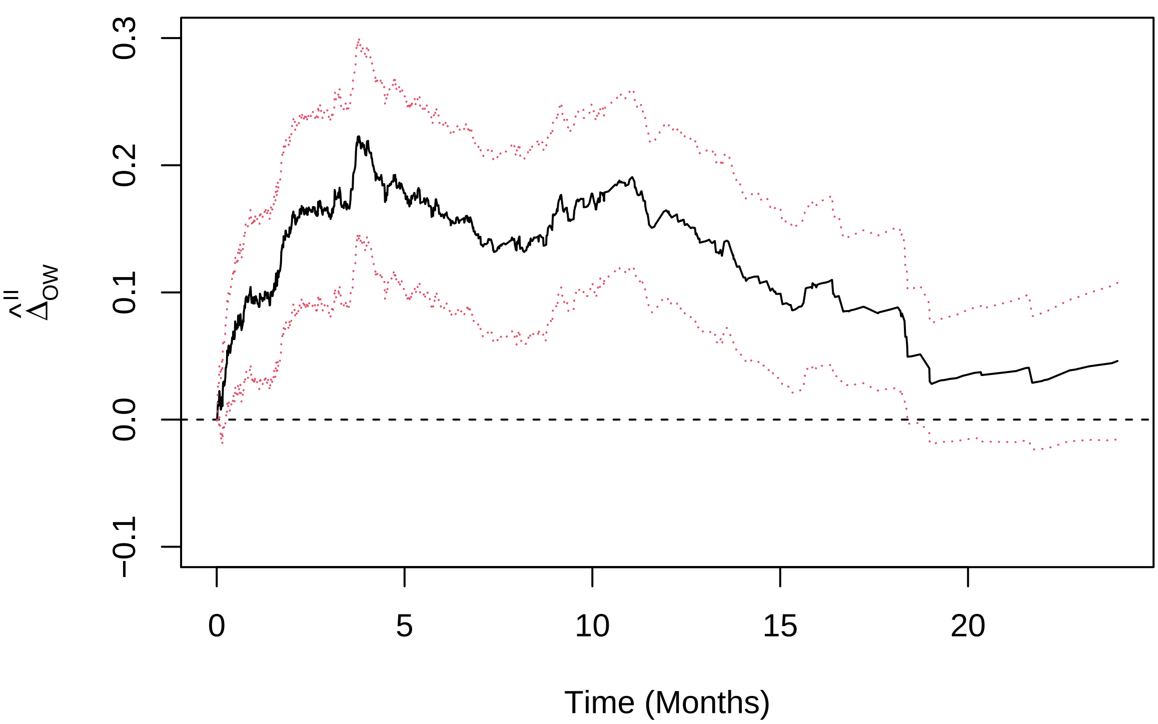
```r
# 1. Type I asymmetric trimming estimator
Delta.OW1=SurvEffectWeibull(Data=data,t=6,Treatment="z",SurvTime="Time",
                            Status="Event",PS.formula=PS.formula,
                            Censor.formula=Censor.formula,Type=1,Method="OW")
# 2. Type II asymmetric trimming estimator
Delta.OW2=SurvEffectWeibull(Data=data,t=6,Treatment="z",SurvTime="Time",
                            Status="Event",PS.formula=PS.formula,
                            Censor.formula=Censor.formula,Type=2,Method="OW")
cat("Type I estimator: \n");round(Delta.OW1,3);cat("Type II estimator: \n");round(Delta.OW2,3)
```

```
## Type I estimator:

## Estimate       SE CI.lower CI.upper
##    0.178    0.047    0.085    0.270

## Type II estimator:

## Estimate       SE CI.lower CI.upper
##    0.161    0.037    0.090    0.233
```

7. Treatment Effect Curves

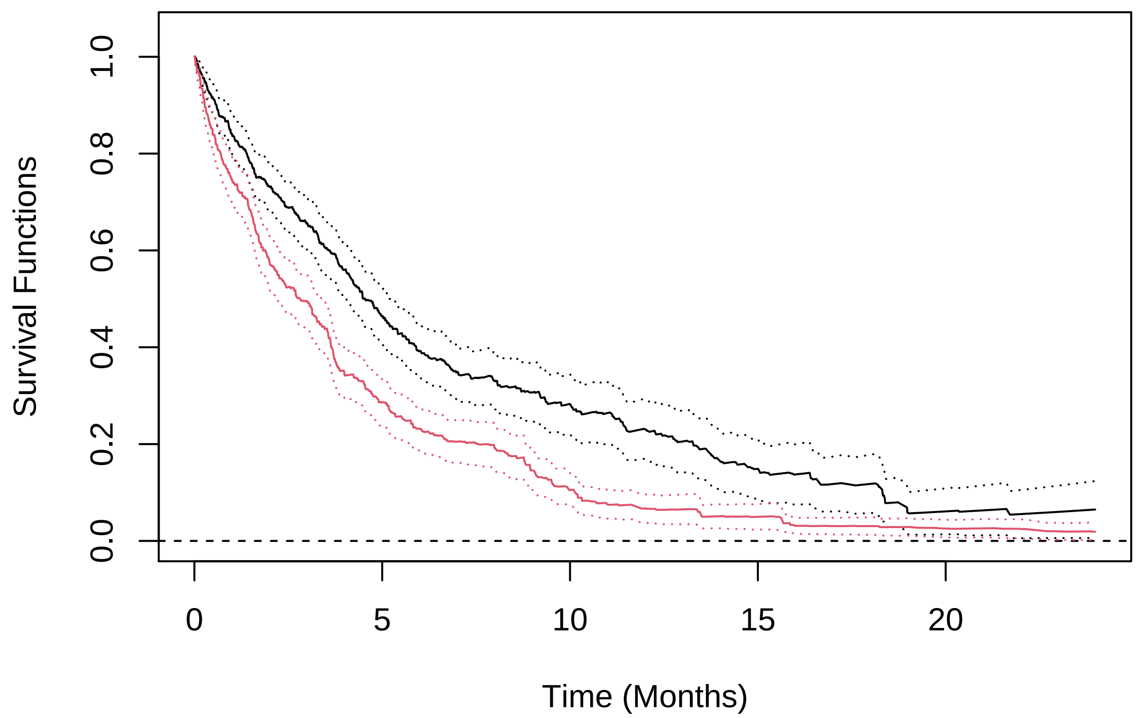
We can intuitively demostrate the treatment effect versus time by drawing a curve of $\hat{\Delta}_h(t)$-by-t with accompanying 95% pointwise confidence intervals. Noticing that the weighted estimator only changes at observed survival times, we can select t as the unique observed survival time in the dataset. In what follows, we use Type II OW as an example to explore the pointwise treatment effect trend by time.

```
# 1. unique observed survival time and <= 24 months
UTime=sort(unique(data$Time[data$Time<=24]))
# 2. A warpped function to calcuate Type II OW estimator on UTime
TypeIIOW=function(time=UTime) {
  n=length(time)
  out=matrix(NA,ncol=4,nrow=n);out[,1]=time
  for (i in 1:n) {
    out[i,2:4] = SurvEffectWeibull(Data=data,t=time[i],Treatment="z",SurvTime="Time",
                              Status="Event",PS.formula=PS.formula,
                              Censor.formula=Censor.formula,Type=2,Method="OW")[c(1,3,4)]
  }
  colnames(out)=c("Time","Estimate","CI.lower","CI.upper")
  out
}
# 3. Obtain Delta_OW(t) for t=UTime
res=TypeIIOW(time=UTime)
# head of `res`
#      Time Estimate CI.lower CI.upper
# [1,] 0.01     0.00     0.00     0.00
# [2,] 0.02     0.00     0.00     0.01
# [3,] 0.03     0.01     0.00     0.03
# [4,] 0.04     0.01     0.00     0.03
# [5,] 0.05     0.01    -0.01     0.03
# [6,] 0.06     0.01     0.00     0.03
# 4. Plot
{
par(mar = c(4.1, 5.1, 4.1, 2.1))
plot(res[,"Time"],res[,"Estimate"],type="l",xlab="Time (Months)",
     ylab=expression(hat(Delta)[OW]^II),ylim=c(-0.1,0.3))
abline(h=0,col=1,lty=2)
lines(res[,"Time"],res[,"CI.lower"],col=2,lty=3)
lines(res[,"Time"],res[,"CI.upper"],col=2,lty=3)
}
```



Beside the treatment effect curve shown above, we may be also interested in investigating the counterfactural survival functions (i.e., $S_h^{(1)}(t)$ and $S_h^{(0)}(t)$). Below, we will illustrate how to draw the estimated counterfactural survival functions under the overlap population with estimator II (i.e., $\hat{S}_{OW}^{II(1)}(t)$ and $\hat{S}_{OW}^{II(0)}(t)$). One can explore other weighting schemes.

```
# 1. identify unique observed survival time and <= 24 months
UTime=sort(unique(data$Time[data$Time<=24]))
# 2 load the R function `SurvFun`. This function is to calcualte the counterfactural
#   survival function under IPW1, IPW2, OW1, or OW2.
source("https://raw.githubusercontent.com/chaochengstat/OW_Survival/main/SurvivalFun.R")
## below is example of how to use the `SurvFun` to calculate S_h(6) based on OW2
## SurvFun(Data=data,t=6,Treatment="z",SurvTime="Time",Status="Event",PS.formula=PS.formula,
##         Censor.formula=Censor.formula,Type=2,Method="OW")
## S1           SE.S1       CI.lower.S1 CI.upper.S1   S0        SE.S0      CI.lower.S0 CI.upper.S0
##  0.39239711  0.02808057  0.33735918  0.44743504  0.23117906  0.02290205  0.18629104  0.27606707
# 3. A warpped function to calcuate Type II OW Survival Function on UTime
TypeIIOW.S=function(time=UTime) {
  n=length(time)
  out=matrix(NA,ncol=7,nrow=n);out[,1]=time
  for (i in 1:n) {
    out[i,2:7] = SurvFun(Data=data,t=time[i],Treatment="z",SurvTime="Time",
                         Status="Event",PS.formula=PS.formula,
                         Censor.formula=Censor.formula,Type=2,Method="OW")[c(1,3,4,5,7,8)]
  }
  colnames(out)=c("Time","S1","CI.lower.S1","CI.upper.S1","S0","CI.lower.S0","CI.upper.S0")
  #out[out[,"CI.upper.S1"]>1,"CI.upper.S1"]=1
  #out[out[,"CI.upper.S0"]>1,"CI.upper.S0"]=1
  #out[out[,"CI.lower.S1"]<0,"CI.lower.S1"]=0
  #out[out[,"CI.lower.S0"]<0,"CI.lower.S0"]=0
  #out[out[,"S1"]>1,"S1"]=1
  #out[out[,"S0"]>1,"S0"]=1
  out
}
# 3. Obtain S_OW^0(t) and S_OW^1(t) for t=UTime
res=TypeIIOW.S(time=UTime)
# head of `res`
#      Time   S1 CI.lower.S1 CI.upper.S1   S0 CI.lower.S0 CI.upper.S0
# [1,] 0.01 1.00        1.00           1 1.00        1.00        1.00
# [2,] 0.02 1.00        1.00           1 0.99        0.99        1.00
# [3,] 0.03 1.00        0.99           1 0.99        0.97        1.00
# [4,] 0.04 1.00        0.99           1 0.98        0.97        1.00
# [5,] 0.05 0.99        0.99           1 0.98        0.97        1.00
# [6,] 0.06 0.99        0.98           1 0.98        0.96        0.99
# 4. Plot
{
par(mar = c(4.1, 5.1, 4.1, 2.1))
plot(res[,"Time"],res[,"S1"],type="l",xlab="Time (Months)",
     ylab="Survival Functions",ylim=c(0.0,1.05))
abline(h=0,col=1,lty=2)
lines(res[,"Time"],res[,"CI.lower.S1"],col=1,lty=3)
lines(res[,"Time"],res[,"CI.upper.S1"],col=1,lty=3)
lines(res[,"Time"],res[,"S0"],col=2)
lines(res[,"Time"],res[,"CI.lower.S0"],col=2,lty=3)
lines(res[,"Time"],res[,"CI.upper.S0"],col=2,lty=3)
}
```



In the above figure, the black lines are $\hat{S}^{II(1)}_{OW}$ with its 95% confidence interval amd the red lines are $\hat{S}^{II(0)}_{OW}$ with its 95% confidence interval.

8. Using Cox Model to Describe The Censoring Process

Previously, we use parametric Weibull regression model to describe the censoring process. Alternatively, we can use semi- or non-parameteric survival models to describe the censoring process, such as Cox proportional hazard model and additive risk model. Here, we will demostrate how to use Cox model to estimate the censoring scores. Specifically, according to the Cox model, we have that

$$P(C^{(z)} \geq t|\mathbf{X}) = \exp\left\{\Lambda_z(t)e^{\theta_z^T\mathbf{X}}\right\},$$

where $C^{(z)}$ is the censoring time with treatment z ($z = 1, 0$ for treated and untreated groups, respectively), Λ_z is the treatment-specific baseline cumulative hazard, and θ_z is treatment-specific coefficients corresponding to pre-treatment covariates $\mathbf{X}$. The partial likelihood approach will be used to estimate coefficients θ_z and the baseline cumulative hazard $\Lambda_z(t)$ will be calculated through the Breslow approach. Because all the parameters are treatment-specific, we will implement two Cox models using the treated and untreated group samples, separately. See the code below.

```
# 1. Construct the Cox model formula
Censor.formula=as.formula(paste("Surv(",SurvTime,",I(1-",Status,"))","~",
                                paste(Covariates,collapse="+"),sep=""))
# Censor.formula is shown below
# Surv(Time, I(1 - Event)) ~ x1 + x2 + x3 + x4 + x5 + x6
# 2. Cox Model for the treated group
data.trt=subset(data,data[,Treatment]==1)
Censor.trt.model = coxph(Censor.formula,data=data.trt)
# 3. Cox Model for the untreated group
data.con=subset(data,data[,Treatment]==0)
Censor.con.model = coxph(Censor.formula,data=data.con)
```

Next, we show example codes to calculate $\hat{\Delta}^{II}_{OW}(6)$ when using Cox model to calculate the censoring scores:

```
# 1. Define a function to calculate the predicted censoring score
### Input: i. CoxModel: A cox model object; ii. TimeVec: A vector of time (say, t).
### Output: Probability of P(C_i>t_i|X_i) for i=1,...,n.
CensorScoreWithCox=function(CoxModel,TimeVec) {
```

```
  LinearPredictor=CoxModel$linear.predictors
  BaselineHazardForm=basehaz(CoxModel,centered=T) # baseline hazard form
  BaselineHazard=sapply(TimeVec, function(x) { # obtain the baseline hazard for Time Vec
    BaselineHazardForm[which.min(abs(BaselineHazardForm$time-x))[1],"hazard"]
  })
  CensorScore=  exp(-BaselineHazard*exp(LinearPredictor)) # obtain censoring probability
  return(CensorScore)
}
# 2. Calculate the censoring scores
# censoring scores in the treated group
Kc.trt=CensorScoreWithCox(CoxModel=Censor.trt.model,TimeVec=rep(6,length(Censor.trt.model$y)))
# censoring scores in the untreated group
Kc.con=CensorScoreWithCox(CoxModel=Censor.con.model,TimeVec=rep(6,length(Censor.con.model$y)))
# 3. Calculate the counterfactual survival probability in the treated group
w.trt = 1-PS[which(data[,Treatment]==1)] # balancing weight in the treated group
S1=sum(w.trt*as.numeric(data.trt[,SurvTime]>=6)/Kc.trt)/sum(w.trt)
# 4. Calculate the counterfactual survival probability in the untreated group
w.con = PS[which(data[,Treatment]==0)] # balancing weight in the untreated group
S0=sum(w.con*as.numeric(data.con[,SurvTime]>=6)/Kc.con)/sum(w.con)
# 5. Finally, obtain the treatment effect
Delta.OW=S1-S0
cat(paste("The estimated treatment effect is",round(Delta.OW,3)))
```

```
## The estimated treatment effect is 0.171
```

Here, we also provide an unified R function `SurvEffectWithCox` to calculate the treatment effects based on the four weighting schemes introduced in manuscript. R code for this function is available at https://github.com/chaochengstat/OW_Survival. Usage of this function is analogous to the function `SurvEffectWeibull` in Part 6 *Other Balancing Weights and Confidence Interval Construction*. Specifically, we can implement this function by

```
SurvEffectWithCox(Data,t,Treatment,SurvTime,Status,PS.formula,Censor.formula,Type,
                  Method,alpha,q)
```

where arguments are same with those in `SurvEffectWeibull`. The output of `SurvEffectWithCox` is the point estimate corresponding to the weighted estimators specified by `Type=` and `Method=`. If we want to calculate Type II estimate with overlap weight, we can specify `Type=2` and `Method="OW"` as below

```
# 1. Load SurvEffectWithCox function
source("https://raw.githubusercontent.com/chaochengstat/OW_Survival/main/functions_Cox_v2.R")
# 2. define PS and Cox model formulas
PS.formula=as.formula(paste(Treatment,"~",paste(Covariates,collapse="+"),sep=""))
Censor.formula=as.formula(paste("Surv(",SurvTime,",I(1-",Status,"))","~",
                                paste(Covariates,collapse="+"),sep=""))
# 3. Type II OW estimator
Delta.OW=SurvEffectWithCox(Data=data,t=6,Treatment="z",SurvTime="Time",
                           Status="Event",PS.formula=PS.formula,
                           Censor.formula=Censor.formula,Type=2,Method="OW")
cat(paste("Type II OW estimator: ",round(Delta.OW,3)))
```

```
## Type II OW estimator:  0.171
```

One can try other weighting schemes with `SurvEffectWithCox` by setting `Method="IPW"`, `"Symmetric"`, or `"Asymmetric"`.

Because estimation of the baseline cumulative hazard function ($\Lambda_z(t)$) is fully non-parametric, it is not straightforward to derive the asymptotic distribution of estimated censoring scores. As a result, derivation of the asymptotic distribution of the weighted estimator is cumbersome when we choose to use Cox model. As an alternative, we can always use nonparametric bootstrap to construct the 95% confidence interval. Specifically, we first resample the original dataset for B times with replacement. Next, for each bootstrap dataset, we calculate $\hat{\Delta}_{h,b}(t)$ for $b = 1, \ldots, B$. Then, the lower and upper bounds of the 95% confidence interval can be obtained by setting 2.5% and 97.5% percentiles of the empirical distribution of $\left\{\hat{\Delta}_{h,b}(t)\right\}_{b=1}^{B}$, respectively. Below we present example codes to calulate the 95% confidence interval.

The `boot` package will be utlized for bootstrapping. To simple the bootstrapping process, we first define several wapper functions that summarizes all eight approaches (four balancing weights (IPW, OW, symmetric trimming with $\alpha = 0.1$, and asymmetric trimming with $q = 0.01$) multiply two types of estimator (Type I and Type II):

```
# 1. load boot package
library("boot")
# 2. Define a wapper function that summarizes all kinds of estimators
AllSurvEffect=function(Data=rhc,t=6,Treatment="z",SurvTime="Time",Status="Event",
                       PS.formula,Censor.formula,alpha,q) {
  Delta1=SurvEffectWithCox(Data=Data,t=t,Treatment=Treatment,SurvTime=SurvTime,Status=Status,
                           PS.formula=PS.formula,Censor.formula=Censor.formula,Type=1,Method="IPW")
  Delta2=SurvEffectWithCox(Data=Data,t=t,Treatment=Treatment,SurvTime=SurvTime,Status=Status,
                           PS.formula=PS.formula,Censor.formula=Censor.formula,Type=1,Method="OW")
  Delta3=SurvEffectWithCox(Data=Data,t=t,Treatment=Treatment,SurvTime=SurvTime,Status=Status,
                           PS.formula=PS.formula,Censor.formula=Censor.formula,Type=1,
                           Method="Symmetric",alpha=alpha)
  Delta4=SurvEffectWithCox(Data=Data,t=t,Treatment=Treatment,SurvTime=SurvTime,Status=Status,
                           PS.formula=PS.formula,Censor.formula=Censor.formula,Type=1,
                           Method="Asymmetric",q=q)
  Delta5=SurvEffectWithCox(Data=Data,t=t,Treatment=Treatment,SurvTime=SurvTime,Status=Status,
                           PS.formula=PS.formula,Censor.formula=Censor.formula,Type=2,Method="IPW")
  Delta6=SurvEffectWithCox(Data=Data,t=t,Treatment=Treatment,SurvTime=SurvTime,Status=Status,
                           PS.formula=PS.formula,Censor.formula=Censor.formula,Type=2,Method="OW")
  Delta7=SurvEffectWithCox(Data=Data,t=t,Treatment=Treatment,SurvTime=SurvTime,Status=Status,
                           PS.formula=PS.formula,Censor.formula=Censor.formula,Type=2,
                           Method="Symmetric",alpha=alpha)
  Delta8=SurvEffectWithCox(Data=Data,t=t,Treatment=Treatment,SurvTime=SurvTime,Status=Status,
                           PS.formula=PS.formula,Censor.formula=Censor.formula,Type=2,
                           Method="Asymmetric",q=q)
  output=c(Delta1,Delta2,Delta3,Delta4,Delta5,Delta6,Delta7,Delta8)
  names(output)=paste(rep(c("IPW","OW","Symmetric","Asymmetric"),2),rep(c(1,2),each=4),sep="")
  output
}
# 3. Define a function for each bootstrap step
ConstructBootFun=function(d,i) {
  out=AllSurvEffect(Data=d[i,],t=6,Treatment="z",SurvTime="Time",Status="Event",
                    PS.formula=PS.formula,Censor.formula=Censor.formula,alpha=0.1,q=0.01)
  out
}
# 4. Define a function to summarize bootstrapping
# where R is  number of bootstrap replicates
GetBootCI=function(R=200) {
  myboot=boot(data, ConstructBootFun, R = R, stype = "i")
  out <- as.data.frame(matrix(NA,ncol=3,nrow=8))
  out[,1]=myboot$t0
  rownames(out)=names(myboot$t0)
  for (j in (1:8)) {
    out[j,2:3]=boot.ci(myboot,type="perc",index=j)$percent[4:5]
  }
  colnames(out) <- c("Estimate","CI.lower","CI.upper")
  out
}
```

Now, we run the bootstrap.

```
set.seed(2021)
BootSummary=GetBootCI(R=200)
```

```
BootSummary
```

```
##                Estimate   CI.lower  CI.upper
## IPW1        0.1677221 0.02687072 0.2523178
## OW1         0.1568592 0.08726048 0.2367292
## Symmetric1  0.1436646 0.06521123 0.2301729
## Asymmetric1 0.1609344 0.08269305 0.2767431
## IPW2        0.1597199 0.03290690 0.2455347
## OW2         0.1711008 0.10536792 0.2446221
## Symmetric2  0.1575849 0.07906868 0.2510091
## Asymmetric2 0.1602268 0.08802597 0.2814505
```

The first four rows displays the point, standard error, and 95% confidence interval estimators of Type I estimators. The second four rows presents the estimators of Type II estimators.

Web Figures and Tables

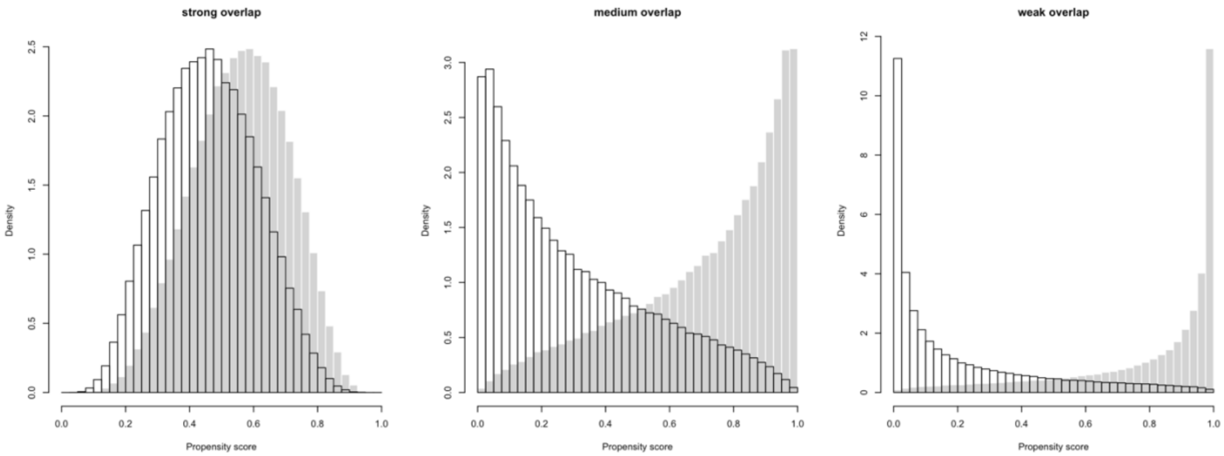


Figure 1. Distributions of true propensity scores with three levels of covariate overlap, where the shaded bars denote the treated group and the unshaded bars represent the untreated group.

Table 1. Percent bias of the estimators in the presence of increasingly strong tails in the propensity score distribution (censoring rate = 50%).

Estimator	Overlap	$\widehat{\Delta}_h^I(t)$				$\widehat{\Delta}_h^{II}(t)$			
		t_1	t_2	t_3	t_4	t_1	t_2	t_3	t_4
Overlap Weighting	$\psi = 1$	-4.6	-5.8	-7.0	-10.0	0.0	-0.6	0.0	0.1
	$\psi = 3$	-5.2	-6.2	-7.8	-10.1	-0.2	-0.5	-0.4	0.1
	$\psi = 5$	-6.3	-6.6	-7.9	-10.0	-1.2	-0.8	-0.6	0.3
Inverse Propensity Weighting (IPW)									
No Trimming	$\psi = 1$	-4.5	-5.8	-7.2	-10.4	-0.1	-0.6	-0.1	0.1
	$\psi = 3$	-5.4	-7.2	-9.0	-12.0	-0.2	-0.9	-0.7	0.0
	$\psi = 5$	-12.8	-11.2	-12.2	-14.3	-7.4	-5.2	-4.0	-1.3
Symmetric Trimming									
$\alpha = 0.05$	$\psi = 1$	-4.5	-5.8	-7.2	-10.4	-0.1	-0.6	-0.1	0.1
	$\psi = 3$	-5.5	-6.6	-8.3	-10.9	-0.2	-0.5	-0.4	0.1
	$\psi = 5$	-7.4	-7.6	-9.0	-11.2	-1.6	-0.8	-0.6	0.5
$\alpha = 0.1$	$\psi = 1$	-4.7	-5.9	-7.2	-10.4	-0.2	-0.6	-0.1	0.1
	$\psi = 3$	-5.9	-6.7	-8.4	-10.9	-0.4	-0.5	-0.4	0.2
	$\psi = 5$	-7.3	-7.8	-9.3	-12.0	-0.9	-0.5	-0.2	0.5
$\alpha = 0.15$	$\psi = 1$	-4.8	-6.0	-7.2	-10.3	-0.2	-0.7	0.0	0.1
	$\psi = 3$	-6.1	-7.2	-8.9	-11.3	-0.3	-0.6	-0.4	0.2
	$\psi = 5$	-7.7	-8.4	-10.0	-12.9	-0.4	-0.4	-0.1	0.6
Asymmetric Trimming									
$q = 0$	$\psi = 1$	-4.8	-5.9	-7.1	-10.2	-0.3	-0.7	0.0	0.3
	$\psi = 3$	-3.5	-5.0	-6.5	-9.2	1.4	1.1	1.6	2.3
	$\psi = 5$	-6.0	-4.1	-5.1	-7.5	-0.8	1.5	2.9	5.1
$q = 0.01$	$\psi = 1$	-8.7	-7.6	-6.9	-8.3	-4.2	-2.5	0.1	1.9
	$\psi = 3$	-12.8	-10.0	-7.9	-7.1	-7.8	-4.1	0.1	4.6
	$\psi = 5$	-11.9	-9.8	-8.8	-8.8	-6.2	-2.7	-0.1	3.6
$q = 0.05$	$\psi = 1$	-14.1	-10.6	-7.4	-6.1	-9.5	-5.4	-0.1	4.5
	$\psi = 3$	-12.4	-10.2	-8.6	-8.0	-6.5	-3.6	0.2	4.5
	$\psi = 5$	-11.0	-10.8	-11.0	-13.0	-3.2	-1.9	0.3	3.3

Table 2. Relative efficiency of the estimators relative to the Original Approach IPW estimator in the presence of increasingly strong tails in the propensity score distribution (censoring rate = 50%).

Estimator	Overlap	$\widehat{\Delta}_h^I(t)$				$\widehat{\Delta}_h^{II}(t)$			
		t_1	t_2	t_3	t_4	t_1	t_2	t_3	t_4
Overlap Weighting	$\psi = 1$	1.07	1.03	1.01	1.01	1.34	1.94	2.89	4.55
	$\psi = 3$	3.37	2.1	1.49	1.2	3.65	3.29	3.51	4.42
	$\psi = 5$	9.66	5.08	2.84	1.66	9.89	7.27	6.18	5.45
Inverse Propensity Weighting (IPW)									
No Trimming	$\psi = 1$	1.00	1.00	1.00	1.00	1.24	1.92	2.96	4.66
	$\psi = 3$	1.00	1.00	1.00	1.00	0.92	1.14	1.61	2.4
	$\psi = 5$	1.00	1.00	1.00	1.00	0.90	0.86	0.81	1.27
Symmetric Trimming									
$\alpha = 0.05$	$\psi = 1$	1.00	1.00	1.00	1.00	1.24	1.92	2.96	4.66
	$\psi = 3$	2.48	1.75	1.37	1.15	2.54	2.52	2.91	3.64
	$\psi = 5$	7.27	4.26	2.59	1.63	7.18	5.44	4.72	4.32
$\alpha = 0.1$	$\psi = 1$	1.00	1.00	1.00	1.00	1.24	1.91	2.93	4.63
	$\psi = 3$	2.88	1.86	1.38	1.13	3.12	2.84	3.00	3.75
	$\psi = 5$	7.95	4.49	2.63	1.58	8.15	5.88	5.16	4.44
$\alpha = 0.15$	$\psi = 1$	1.02	1.01	1.01	1.01	1.26	1.89	2.88	4.59
	$\psi = 3$	2.85	1.81	1.33	1.08	2.98	2.68	2.77	3.48
	$\psi = 5$	7.69	4.19	2.45	1.47	8.18	6.00	4.80	4.18
Asymmetric Trimming									
$q = 0$	$\psi = 1$	1.00	1.00	1.00	1.00	1.24	1.90	2.91	4.62
	$\psi = 3$	1.04	1.00	0.98	0.99	0.96	1.12	1.52	2.25
	$\psi = 5$	0.92	0.92	0.91	0.92	0.82	0.78	0.73	1.07
$q = 0.01$	$\psi = 1$	1.02	0.98	0.98	0.98	1.27	1.81	2.71	4.26
	$\psi = 3$	2.62	1.76	1.33	1.13	2.84	2.59	2.84	3.56
	$\psi = 5$	7.46	4.22	2.52	1.59	7.54	5.42	4.79	4.26
$q = 0.05$	$\psi = 1$	0.98	0.91	0.92	0.93	1.19	1.58	2.36	3.64
	$\psi = 3$	2.53	1.59	1.17	0.95	2.72	2.38	2.44	3.04
	$\psi = 5$	6.12	3.54	2.05	1.31	6.44	4.70	3.50	3.18

Table 3. Coverage rate of the 95% confidence intervals for the estimators in the presence of increasingly strong tails in the propensity score distribution (censoring rate = 50%).

Estimator	Overlap	$\widehat{\Delta}_h^I(t)$				$\widehat{\Delta}_h^{II}(t)$			
		t_1	t_2	t_3	t_4	t_1	t_2	t_3	t_4
Overlap Weighting	$\psi = 1$	93.4	91.7	89.2	83.1	96.2	96.7	97.0	96.9
	$\psi = 3$	93.8	90.8	88.2	84.0	96.3	95.6	96.3	96.7
	$\psi = 5$	93.2	91.2	88.2	84.5	95.5	96.0	95.7	96.0
Inverse Propensity Weighting (IPW)									
No Trimming	$\psi = 1$	93.3	91.8	89.0	81.7	95.9	97.1	97.1	97.2
	$\psi = 3$	91.7	90.8	89.6	82.3	95.0	94.4	95.2	96.1
	$\psi = 5$	85.2	87.7	85.8	80.5	92.0	92.4	93.0	93.0
Symmetric Trimming									
$\alpha = 0.05$	$\psi = 1$	93.3	91.9	89.0	81.7	96.0	97.0	97.1	97.2
	$\psi = 3$	93.4	91.3	89.6	84.5	95.1	95.1	96.4	96.1
	$\psi = 5$	93.1	91.7	89.3	85.1	95.3	95.7	96.3	96.5
$\alpha = 0.1$	$\psi = 1$	93.1	92.0	89.0	81.9	95.9	96.9	97.0	97.0
	$\psi = 3$	93.8	91.8	89.3	84.8	95.5	96.1	96.4	96.7
	$\psi = 5$	94.0	92.7	89.2	85.1	96.4	96.2	97.0	96.9
$\alpha = 0.15$	$\psi = 1$	92.9	91.8	89.4	81.9	96.1	96.6	96.9	97.0
	$\psi = 3$	93.6	92.0	89.1	84.9	95.7	96.1	96.5	97.2
	$\psi = 5$	93.7	93.2	89.2	84.9	96.8	96.8	97.1	97.5
Asymmetric Trimming									
$q = 0$	$\psi = 1$	93.0	92.1	88.8	82.5	96.0	96.9	97.4	97.2
	$\psi = 3$	93.0	92.7	91.3	86.1	95.5	94.1	94.7	94.9
	$\psi = 5$	92.3	93.1	91.4	88.5	95.5	94.2	92.9	93.9
$q = 0.01$	$\psi = 1$	91.3	89.0	89.7	85.8	95.8	96.4	96.9	96.6
	$\psi = 3$	89.2	88.8	89.1	88.5	94.3	95.3	96.0	95.0
	$\psi = 5$	91.8	90.7	89.9	87.1	95.4	95.8	96.2	96.3
$q = 0.05$	$\psi = 1$	86.5	87.2	89.2	88.8	93.9	94.6	97.3	95.9
	$\psi = 3$	91.3	90.5	89.3	87.8	95.7	96.1	96.4	96.4
	$\psi = 5$	92.9	91.6	90.8	86.9	96.5	96.4	96.1	97.0