CROSS-VALIDATION MODEL AVERAGING UNDER COVARIATE-ADAPTIVE RANDOMIZATION

Fuyi Tu[1], Jiahui Xin[2] and Wei Ma[2]

[1] *Chongqing University of Posts and Telecommunications*

and [2] *Renmin University of China*

Supplementary Material

This document serves as the online supplementary material for the paper entitled “Cross-validation model averaging under covariate-adaptive randomization”. This supplement contains the proofs of all theorems stated in the main text. Sections S1 and S2 establish the optimality of the cross-validated weights; Section S3 proves the asymptotic normality of the proposed model averaging estimator; and Section S4 collects several technical lemmas used throughout the proofs.

S1 Proof of Theorem 1

Proof. The following result is Lemma 1 in Gao et al. (2019) and forms the cornerstone of the proof of Theorem 1.

Lemma S1 (Gao et al. (2019))**.** *Let $\mathcal{W}$ be a set of weight vectors. Define*

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}\in\mathcal{W}}{\operatorname{argmin}} \left(R(\boldsymbol{w}) + a_n(\boldsymbol{w})\right).$$

If

$$\sup_{\boldsymbol{w}\in\mathcal{W}} \frac{|a_n(\boldsymbol{w})|}{R^*(\boldsymbol{w})} \xrightarrow{P} 0, \ \sup_{\boldsymbol{w}\in\mathcal{W}} \left| \frac{R(\boldsymbol{w})}{R^*(\boldsymbol{w})} - 1 \right| \xrightarrow{P} 0,$$

and there exists a constant κ_3 such that

$$\inf_{\boldsymbol{w}\in\mathcal{W}} R^*(\boldsymbol{w}) \geq \kappa_3 > 0,$$

then

$$\frac{R\left(\hat{\boldsymbol{w}}\right)}{\inf_{\boldsymbol{w}\in\mathcal{W}} R(\boldsymbol{w})} \xrightarrow{P} 1.$$

For simplicity, we assume $n = KJ$, i.e., each fold has the same number of units J. For $i \in \mathcal{J}_{[s]}$, denote the expectation-adjusted transformed outcome as

$$r_i^{\dagger}(a) = Y_i(a) - \Big[(1-\pi)E\{Y_i(1)|\boldsymbol{X}_i, B_i\} + \pi E\{Y_i(0)|\boldsymbol{X}_i, B_i\}\Big].$$

Let $CV_K^{\dagger}(\boldsymbol{w}) = CV_K(\boldsymbol{w}) - \varsigma^2_{r^{\dagger}}$, where the second term is unrelated to $\boldsymbol{w}$. Then

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}\in\mathcal{W}} CV_K(\boldsymbol{w}) = \arg\min_{\boldsymbol{w}\in\mathcal{W}} CV_K^{\dagger}(\boldsymbol{w}).$$

Then, by Lemma S1, Theorem 1 holds if

$$\sup_{\boldsymbol{w}\in\mathcal{W}} \frac{|R(\boldsymbol{w}) - R^*(\boldsymbol{w})|}{R^*(\boldsymbol{w})} = o(1), \tag{S1.1}$$

and

$$\sup_{\boldsymbol{w}\in\mathcal{W}} \frac{|CV_K^{\dagger}(\boldsymbol{w}) - R^*(\boldsymbol{w})|}{R^*(\boldsymbol{w})} = o_P(1). \tag{S1.2}$$

Denote

$$\mu_{is} = \Big[(1-\pi)E\{Y_i(1)|\boldsymbol{X}_i, B_i = s\} + \pi E\{Y_i(0)|\boldsymbol{X}_i, B_i = s\}\Big],$$

$$\hat{\Delta}_{m[s]a}^{(-k)} = \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, a) - h_{m[s]}^*(\boldsymbol{X}_i, a),$$

$$\hat{\Delta}_{m[s]a} = \sum_{k=1}^{K} \mathbf{1}_{\{i\in I_k\}} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, a) - h_{m[s]}^*(\boldsymbol{X}_i, a).$$

For Equation (S1.1), we observe that for any fixed treatment a, fold k, weight $\boldsymbol{w}$ and stratum s,

$$\begin{aligned}
& E\Big\{\Big(\sum_{m=1}^{M} w_{(m)a}\hat{\Delta}_{m[s]}^{(-k)}(\boldsymbol{X}_i, a)\Big)^2 \mid B_i = s, \{Y_j, \boldsymbol{X}_j, A_j, B_j\}_{j\in I_k^c}\Big\} \\
= \;& \sum_{m=1}^{M}\sum_{m'=1}^{M} w_{(m)a}w_{(m')a} E\big(\hat{\Delta}_{m[s]a}^{(-k)}\hat{\Delta}_{m'[s]a}^{(-k)} \mid B_i = s, \{Y_j, \boldsymbol{X}_j, A_j, B_j\}_{j\in I_k^c}\big) \\
\le \;& \sum_{m=1}^{M}\sum_{m'=1}^{M} w_{(m)a}w_{(m')a} \cdot \\
& \sqrt{E\Big[\big\{\hat{\Delta}_{m[s]a}^{(-k)}\big\}^2 \mid B_i = s, \{Y_j, \boldsymbol{X}_j, A_j, B_j\}_{j\in I_k^c}\Big] \cdot E\Big[\big\{\hat{\Delta}_{m'[s]a}^{(-k)}\big\}^2 \mid B_i = s, \{Y_j, \boldsymbol{X}_j, A_j, B_j\}_{j\in I_k^c}\Big]} \\
\le \;& \sup_{m,s} E\Big[\big\{\hat{\Delta}_{m[s]a}^{(-k)}\big\}^2 \mid B_i = s, \{Y_j, \boldsymbol{X}_j, A_j, B_j\}_{j\in I_k^c}\Big] \cdot \sum_{m=1}^{M}\sum_{m'=1}^{M} w_{(m)a}w_{(m')a} \\
= \;& \sup_{m,s} E\Big[\big\{\hat{\Delta}_{m[s]a}^{(-k)}\big\}^2 \mid B_i = s, \{Y_j, \boldsymbol{X}_j, A_j, B_j\}_{j\in I_k^c}\Big] \cdot \Big(\sum_{m=1}^{M} w_{(m)a}\Big)^2 \\
= \;& \sup_{m,s} E\Big[\big\{\hat{\Delta}_{m[s]a}^{(-k)}\big\}^2 \mid B_i = s, \{Y_j, \boldsymbol{X}_j, A_j, B_j\}_{j\in I_k^c}\Big] \\
= \;& o_P(1),
\end{aligned}$$

which does not depend on $\boldsymbol{w}$. Let $\mathcal{D}_{-k} = \{Y_j, \boldsymbol{X}_j, A_j, B_j : j \in I_k^c\}$ and

$$Z_{n,k,s,a}(\boldsymbol{w}) = E\left\{\left(\sum_{m=1}^{M} w_{(m)a}\hat{\Delta}_{m[s]a}^{(-k)}\right)^2 \mid B_i = s, \mathcal{D}_{-k}\right\}.$$

Because the weights are nonnegative and sum to one,

$$0 \le \sup_{\boldsymbol{w} \in \mathcal{W}} Z_{n,k,s,a}(\boldsymbol{w}) \le \max_{1 \le m \le M} E\left[\{\hat{\Delta}^{(-k)}_{m[s]a}\}^2 \mid B_i = s, \mathcal{D}_{-k}\right].$$

Assumption 5 gives the right-hand side $o_P(1)$, and the added uniform integrability condition plus the fixed number of candidate models gives uniform integrability of the finite maximum. Hence, by Vitali's theorem,

$$E\left\{\sup_{\boldsymbol{w} \in \mathcal{W}} Z_{n,k,s,a}(\boldsymbol{w}) \mid B_i = s\right\} = o(1).$$

The tower property then yields

$$\sup_{\boldsymbol{w} \in \mathcal{W}} E\left(\left|\sum_{m=1}^{M} w_{(m)a} \hat{\Delta}^{(-k)}_{m[s]a}\right|^2 \mid B_i = s\right) = o(1).$$

Since the fold number K is fixed, it further holds that

$$\sup_{\boldsymbol{w} \in \mathcal{W}} E\left(\left|\sum_{m=1}^{M} w_{(m)a} \hat{\Delta}_{m[s]a}\right|^2 \mid B_i = s\right) = o(1).$$

So we have

$$\begin{aligned}
& \sup_{w \in \mathcal{W}} |R(\boldsymbol{w}, s) - R^*(\boldsymbol{w}, s)| \\
\le \quad & \sup_{\boldsymbol{w} \in \mathcal{W}} \Bigg| E\Bigg(\Bigg[\Bigg\{(1-\pi)\sum_{m=1}^{M} w_{(m)1} \sum_{k=1}^{K} \mathbb{1}_{[i \in I_k]} \hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i, 1) + \pi \sum_{m=1}^{M} w_{(m)0} \sum_{k=1}^{K} \mathbb{1}_{[i \in I_k]} \hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i, 0)\Bigg\} - \mu_{is}\Bigg]^2 \\
& \quad - \Bigg[\Bigg\{(1-\pi)\sum_{m=1}^{M} w_{(m)1} h^*_{m[s]}(\boldsymbol{X}_i, 1) + \pi \sum_{m=1}^{M} w_{(m)0} h^*_{m[s]}(\boldsymbol{X}_i, 0)\Bigg\} - \mu_{is}\Bigg]^2 \mid B_i = s\Bigg)\Bigg| \\
+ \quad & \sup_{\boldsymbol{w} \in \mathcal{W}} \Bigg| - E^2\Bigg[\Bigg\{(1-\pi)\sum_{m=1}^{M} w_{(m)1} \sum_{k=1}^{K} \mathbb{1}_{[i \in I_k]} \hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i, 1) + \pi \sum_{m=1}^{M} w_{(m)0} \sum_{k=1}^{K} \mathbb{1}_{[i \in I_k]} \hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i, 0)\Bigg\} - \mu_{is} \mid B_i = s\Bigg] \\
& \quad + E^2\Bigg[\Bigg\{(1-\pi)\sum_{m=1}^{M} w_{(m)1} h^*_{m[s]}(\boldsymbol{X}_i, 1) + \pi \sum_{m=1}^{M} w_{(m)0} h^*_{m[s]}(\boldsymbol{X}_i, 0)\Bigg\} - \mu_{is} \mid B_i = s\Bigg]\Bigg| \\
\le \quad & \sup_{\boldsymbol{w} \in \mathcal{W}} \Bigg| E\Bigg[\Bigg\{(1-\pi)\sum_{m=1}^{M} w_{(m)1} \hat{\Delta}_{m[s]1} + \pi \sum_{m=1}^{M} w_{(m)0} \hat{\Delta}_{m[s]0}\Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&\times\Bigg\{(1-\pi)\sum_{m=1}^{M} w_{(m)1}\hat{\Delta}_{m[s]1} + \pi\sum_{m=1}^{M} w_{(m)0}\hat{\Delta}_{m[s]0} + 2(1-\pi)\sum_{m=1}^{M} w_{(m)1}h^{*}_{m[s]}(\boldsymbol{X}_i,1)\\
&\qquad\qquad +2\pi\sum_{m=1}^{M} w_{(m)0}h^{*}_{m[s]}(\boldsymbol{X}_i,0) - 2\mu_{is}\Bigg\}\ \Bigg|\ B_i = s\Bigg]\Bigg|\\
&+\sup_{\boldsymbol{w}\in\mathcal{W}}\Bigg|E\Big\{(1-\pi)\sum_{m=1}^{M} w_{(m)1}\hat{\Delta}_{m[s]1} + \pi\sum_{m=1}^{M} w_{(m)0}\hat{\Delta}_{m[s]0} \mid B_i = s\Big\}\\
&\times E\Bigg\{(1-\pi)\sum_{m=1}^{M} w_{(m)1}\hat{\Delta}_{m[s]1} + \pi\sum_{m=1}^{M} w_{(m)0}\hat{\Delta}_{m[s]0} + 2(1-\pi)\sum_{m=1}^{M} w_{(m)1}h^{*}_{m[s]}(\boldsymbol{X}_i,1)\\
&\qquad\qquad +2\pi\sum_{m=1}^{M} w_{(m)0}h^{*}_{m[s]}(\boldsymbol{X}_i,0) - 2\mu_{is}\ \Bigg|\ B_i = s\Bigg\}\Bigg|\\
=\ & o(1).
\end{aligned}
$$

where the last equality follows from the Cauchy–Schwarz inequality: the second moment of $\sum_m w_{(m)a}\hat{\Delta}_{m[s]a}$ vanishes uniformly, whereas the second moments of the oracle adjustment and μ_{is} are uniformly bounded under Assumption 4.

Together with Assumption 6, the fact that $\sum_{s=1}^{S} p_s = 1$ implies that Equation (S1.1) holds.

For Equation (S1.2), define the limiting transformed outcome by

$$
r_i^{*}(\boldsymbol{w},a) = Y_i(a) - \Big\{(1-\pi)\sum_{m=1}^{M} w_{(m)1}h^{*}_{m[s]}(\boldsymbol{X}_i,1) + \pi\sum_{m=1}^{M} w_{(m)0}h^{*}_{m[s]}(\boldsymbol{X}_i,0)\Big\}.
$$

Then, in fold k, we have

$$
\begin{aligned}
&\hat{r}_i(\boldsymbol{w},a) - r_i^{*}(\boldsymbol{w},a)\\
=\ & -(1-\pi)\sum_{m=1}^{M} w_{(m)1}\{\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,1) - h^{*}_{m[s]}(\boldsymbol{X}_i,1)\} - \pi\sum_{m=1}^{M} w_{(m)0}\{\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,0) - h^{*}_{m[s]}(\boldsymbol{X}_i,0)\}\\
=\ & -(1-\pi)\sum_{m=1}^{M} w_{(m)1}\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,1) - \pi\sum_{m=1}^{M} w_{(m)0}\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,0).
\end{aligned}
$$

To bound $\sup_{w\in\mathcal{W}}|\hat{\varsigma}^2_{\hat{r}(\boldsymbol{w})k}-\hat{\varsigma}^2_{r^*(\boldsymbol{w})k}|$, we focus on the following term within stratum s: $\frac{1}{n_{[s]ka}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}\mathbf{1}_{\{A_i=a\}}\Big\{\hat{r}_i(\boldsymbol{w},a)-\frac{1}{n_{[s]ka}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}\mathbf{1}_{\{A_j=a\}}\hat{r}_j(\boldsymbol{w},a)\Big\}^2$.

Because of the linearity of expectation, it is sufficient to bound the following term.

$$
\begin{aligned}
&\sup_{w\in\mathcal{W}}E\Bigg|\Big\{\hat{r}_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{r}_j(\boldsymbol{w},1)\Big\}^2-\Big\{r^*_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_jr^*_j(\boldsymbol{w},1)\Big\}^2\mid B_i=s\Bigg|\\
=&\sup_{w\in\mathcal{W}}E\Bigg|\Big[\hat{r}_i(\boldsymbol{w},1)-r^*_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\{\hat{r}_j(\boldsymbol{w},1)-r^*_j(\boldsymbol{w},1)\}\Big]\cdot\\
&\Big[\hat{r}_i(\boldsymbol{w},1)+r^*_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\{\hat{r}_j(\boldsymbol{w},1)+r^*_j(\boldsymbol{w},1)\}\Big]\mid B_i=s\Bigg|\\
=&\sup_{w\in\mathcal{W}}E\Bigg|\Big[\hat{r}_i(\boldsymbol{w},1)-r^*_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\{\hat{r}_j(\boldsymbol{w},1)-r^*_j(\boldsymbol{w},1)\}\Big]\\
&\cdot\Big[\hat{r}_i(\boldsymbol{w},1)-r^*_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\{\hat{r}_j(\boldsymbol{w},1)-r^*_j(\boldsymbol{w},1)\}\\
&+2\{r^*_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_jr^*_j(\boldsymbol{w},1)\}\Big]\mid B_i=s\Bigg|\\
=&\sup_{w\in\mathcal{W}}E\Bigg|\Big[(1-\pi)\sum_{m=1}^{M}w_{(m)1}\big\{\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,1)\big\}\\
&+\pi\sum_{m=1}^{M}w_{(m)0}\big\{\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,0)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,0)\big\}\Big]\\
&\cdot\Big[(1-\pi)\sum_{m=1}^{M}w_{(m)1}\big\{\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,1)\big\}\\
&+\pi\sum_{m=1}^{M}w_{(m)0}\big\{\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,0)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,0)\big\}\\
&-2\{r^*_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_jr^*_j(\boldsymbol{w},1)\}\Big]\mid B_i=s\Bigg|.
\end{aligned}
$$

By simple algebra, we have two main terms to treat:

$$
\sup_{w\in\mathcal{W}}E\Bigg|\Bigg[\sum_{m=1}^{M}w_{(m)a}\big\{\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)-\frac{1}{n_{[s]ka}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}\mathbf{1}_{\{A_j=a\}}\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,a)\big\}\Bigg]^2\mid B_i=s\Bigg|,
$$

and

$$\sup_{w\in\mathcal{W}} E\left|\sum_{m=1}^{M} w_{(m)a}\big\{\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)-\frac{1}{n_{[s]ka}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}\mathbf{1}_{\{A_j=a\}}\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,a)\big\}\right.$$
$$\left.\{r^*_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j r^*_j(\boldsymbol{w},1)\}\mid B_i=s\right|.$$

From the linearity of expectation, it is sufficient to show $E\left[\left\{\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,1)\right\}^2\mid B_i=s\right]=o(1)$ and $\sup_{w\in\mathcal{W}}E\{r^{*2}_i(\boldsymbol{w},1)\mid B_i=s\}=O(1)$. The latter can be checked by simple algebra. We verify the former as follows.

Conditional on $B_i=s$ and the training data outside fold I_k, we have

$$\begin{aligned}
&E\left[\left\{\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,1)\right\}^2\,\middle|\,B_i=s,\{Y_{j_0},\boldsymbol{X}_{j_0},A_{j_0},B_{j_0}\}_{j_0\in I^c_k}\right]\\
&\le 2E\left[\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,1)^2\,\middle|\,B_i=s,\{Y_{j_0},\boldsymbol{X}_{j_0},A_{j_0},B_{j_0}\}_{j_0\in I^c_k}\right]\\
&\quad+2E\left[\left\{\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,1)\right\}^2\,\middle|\,B_i=s,\{Y_{j_0},\boldsymbol{X}_{j_0},A_{j_0},B_{j_0}\}_{j_0\in I^c_k}\right]\\
&\le 2E\left[\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_i,1)^2\,\middle|\,B_i=s,\{Y_{j_0},\boldsymbol{X}_{j_0},A_{j_0},B_{j_0}\}_{j_0\in I^c_k}\right]\\
&\quad+2E\left[\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{\Delta}^{(-k)}_{m[s]}(\boldsymbol{X}_j,1)^2\,\middle|\,B_i=s,\{Y_{j_0},\boldsymbol{X}_{j_0},A_{j_0},B_{j_0}\}_{j_0\in I^c_k}\right]\\
&\le 4\sup_{m,s}E\left[\left\{\hat{h}^{(-k)}_{m[s]}(\tilde{\boldsymbol{X}}_i,1)-h^*_{m[s]}(\tilde{\boldsymbol{X}}_i,1)\right\}^2\,\middle|\,B_i=s,\{Y_{j_0},\boldsymbol{X}_{j_0},A_{j_0},B_{j_0}\}_{j_0\in I^c_k}\right]\\
&=o_P(1).
\end{aligned}$$

The first inequality follows from $(x-y)^2\le 2x^2+2y^2$, and the second follows from the Cauchy–Schwarz inequality. The final bound follows from the tower property, Assumptions 1, 2, and 5, and the independence of the fold partition.

Then, because the relevant functions are uniformly integrable and finitely many integrable functions are involved, we obtain

$$E\left[\left\{\hat{\Delta}_{m[s]}^{(-k)}(\boldsymbol{X}_i,1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{\Delta}_{m[s]}^{(-k)}(\boldsymbol{X}_j,1)\right\}^2\mid B_i=s\right]=o(1).$$

Similarly, we have $E\left\{\hat{\Delta}_{m[s]}^{(-k)}(\boldsymbol{X}_i,0)-\frac{1}{n_{[s]k0}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}(1-A_j)\hat{\Delta}_{m[s]}^{(-k)}(\boldsymbol{X}_j,0)\mid B_i=s,\{Y_{j_0},\boldsymbol{X}_{j_0},A_{j_0},B_{j_0}\}_{j_0\in I_k^c}\right\}^2=o_P(1)$ and

$$E\left[\left\{\hat{\Delta}_{m[s]}^{(-k)}(\boldsymbol{X}_i,0)-\frac{1}{n_{[s]k0}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}(1-A_j)\hat{\Delta}_{m[s]}^{(-k)}(\boldsymbol{X}_j,0)\right\}^2\mid B_i=s\right]=o(1).$$

Hence we obtain

$$\sup_{w\in\mathcal{W}}E\left|\left\{\hat{r}_i(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_j\hat{r}_j(\boldsymbol{w},1)\right\}^2-\left\{r_i^*(\boldsymbol{w},1)-\frac{1}{n_{[s]k1}}\sum_{j\in\mathcal{J}_{[s]}\cap I_k}A_jr_j^*(\boldsymbol{w},1)\right\}^2\mid B_i=s\right|=o(1).$$

By the same argument applied to the transformed outcomes of the control group, we have $\sup_{w\in\mathcal{W}}|\hat{\varsigma}^2_{\hat{r}(\boldsymbol{w})k}-\hat{\varsigma}^2_{r^*(\boldsymbol{w})k}|=o_P(1)$. Here the treatment-selected fold cells are justified as follows. The fold split is independent of the trial data; within each stratum and fold, conditioning on the realized CAR treatment counts lets us use the usual coupling argument as in Bugni et al. (2018) for selected treated and control validation cells. Thus the fold-stratum-treatment empirical means above obey the same LLN bounds needed for the displayed mean-square control.

Next, we show $\sup_{w\in\mathcal{W}}|\hat{\varsigma}^2_{r^*(\boldsymbol{w})k}-\varsigma^2_{r^*(\boldsymbol{w})k}|=o_P(1)$. Notice the variance estimator in the fold k, $\hat{\varsigma}^2_{r^*(\boldsymbol{w})k}$, contains the two potential outcome parts, each of which is a finite sum of sample variances in stratum s,

$$\frac{1}{n_{[s]ka}} \sum_{i \in \mathcal{J}_{[s]} \cap I_k} \mathbf{1}_{\{A_i=a\}} \left\{ r_i^*(\boldsymbol{w}, a) - \frac{1}{n_{[s]ka}} \sum_{j \in \mathcal{J}_{[s]} \cap I_k} \mathbf{1}_{\{A_j=a\}} r_j^*(\boldsymbol{w}, a) \right\}^2.$$

It is sufficient to show its uniform convergence with respect to $\boldsymbol{w}$.

Observe that this sample variance could be expressed as the difference between the sample mean of $\{r_i^*(\boldsymbol{w}, a)\}^2$ and the square of the sample mean of $r_i^*(\boldsymbol{w}, a)$. Conditional on the fold partition, strata, assignments, and realized cell counts, the CAR coupling argument applies to each fold–stratum–arm average, with $n_{[s]ka} \xrightarrow{P} \infty$ and $n/n_{[s]ka} = O_P(1)$. Moreover, because $r_i^*(\boldsymbol{w}, a)$ is affine in the fixed-dimensional vector $\boldsymbol{w}$, each cell variance is a quadratic polynomial in $\boldsymbol{w}$. Its coefficients are finitely many cell averages of oracle products. Lemmas S3 and S4, applied to these coefficients after the CAR coupling, therefore give uniform convergence of $\frac{1}{n_{[s]ka}} \sum_{i \in \mathcal{J}_{[s]} \cap I_k} \mathbf{1}_{\{A_i=a\}} \left\{ r_i^*(\boldsymbol{w}, a) - \frac{1}{n_{[s]ka}} \sum_{j \in \mathcal{J}_{[s]} \cap I_k} \mathbf{1}_{\{A_j=a\}} r_j^*(\boldsymbol{w}, a) \right\}^2$ over the compact set $\mathcal{W}$. Since both the stratum number and fold number are fixed, we obtain

$$\sup_{w \in \mathcal{W}} |\hat{\varsigma}^2_{r^*(\boldsymbol{w})k} - \varsigma^2_{r^*(\boldsymbol{w})k}| = o_P(1).$$

As a consequence,

$$\begin{aligned}
&\sup_{w \in \mathcal{W}} |CV_K^{\dagger}(\boldsymbol{w}) - (\varsigma^2_{r^*(\boldsymbol{w})} - \varsigma^2_{r^{\dagger}})| \\
&\leq \sup_{w \in \mathcal{W}} \frac{1}{K} \sum_{k=1}^{K} |\hat{\varsigma}^2_{r^*(\boldsymbol{w})k} - \varsigma^2_{r^*(\boldsymbol{w})k}|
\end{aligned}$$

$$\leq \frac{1}{K}\sum_{k=1}^{K}\sup_{w\in\mathcal{W}}|\hat{\varsigma}^2_{r^*(\boldsymbol{w})k}-\varsigma^2_{r^*(\boldsymbol{w})k}|$$

$$= o_P(1).$$

Combining the results above gives $\sup_{w\in\mathcal{W}}|CV_K^{\dagger}(\boldsymbol{w}) - R^*(\boldsymbol{w})| = o_P(1)$. Together with Assumption 6, Equation (S1.2) holds. This completes the proof of optimality. □

S2 Proof of Theorem 2

Proof. From the proof of Theorem 1, we have

$$\sup_{\boldsymbol{w}\in\mathcal{W}}|R(\boldsymbol{w}) - R^*(\boldsymbol{w})| = o(1),$$

and

$$\sup_{\boldsymbol{w}\in\mathcal{W}}|CV_K^{\dagger}(\boldsymbol{w}) - R^*(\boldsymbol{w})| = o_P(1).$$

Furthermore, we have

$$\sup_{\boldsymbol{w}\in\mathcal{W}}|CV_K^{\dagger}(\boldsymbol{w}) - R(\boldsymbol{w})| = o_P(1).$$

Since $\inf_{\boldsymbol{w}\in\mathcal{W}} R^*(\boldsymbol{w}) = 0$, we have $\inf_{\boldsymbol{w}\in\mathcal{W}} CV_K^{\dagger}(\boldsymbol{w}) = o_P(1)$.

Recall that $\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}\in\mathcal{W}} CV_K(\boldsymbol{w}) = \arg\min_{\boldsymbol{w}\in\mathcal{W}} CV_K^{\dagger}(\boldsymbol{w})$. We obtain the result

$$CV_K^{\dagger}(\hat{\boldsymbol{w}}) = o_P(1).$$

Combining with the previous result $\sup_{\boldsymbol{w}\in\mathcal{W}} |CV_K^{\dagger}(\boldsymbol{w}) - R(\boldsymbol{w})| = o_P(1)$,

$$R(\hat{\boldsymbol{w}}) = o_P(1).$$

Moreover, let $\mathcal{W}_0 = \{\boldsymbol{w} \in \mathcal{W} : R^*(\boldsymbol{w}) = 0\}$. Since $R^*(\boldsymbol{w})$ is a finite-dimensional quadratic form in $\boldsymbol{w}$, it is continuous on the compact set $\mathcal{W}$, and $\mathcal{W}_0$ is nonempty under the present zero-risk case. For any $\epsilon > 0$,

$$c_\epsilon = \inf_{\{\boldsymbol{w}:\mathrm{dist}(\boldsymbol{w},\mathcal{W}_0)\geq\epsilon\}} R^*(\boldsymbol{w}) > 0.$$

Together with $\sup_{\boldsymbol{w}\in\mathcal{W}} |R(\boldsymbol{w}) - R^*(\boldsymbol{w})| = o(1)$, this gives

$$R^*(\hat{\boldsymbol{w}}) \leq R(\hat{\boldsymbol{w}}) + \sup_{\boldsymbol{w}\in\mathcal{W}} |R(\boldsymbol{w}) - R^*(\boldsymbol{w})| = o_P(1).$$

$$P\{\mathrm{dist}(\hat{\boldsymbol{w}}, \mathcal{W}_0) \geq \epsilon\} \leq P\{R^*(\hat{\boldsymbol{w}}) \geq c_\epsilon\} \to 0.$$

Hence $\mathrm{dist}(\hat{\boldsymbol{w}}, \mathcal{W}_0) = o_P(1)$, which is used in Theorem 3. $\square$

S3 Proof of Theorem 3

Proof. For the i-th observation in fold k and stratum s, let

$$\hat{h}_{[s]}(\boldsymbol{X}_i, \hat{\boldsymbol{w}}) = (1-\pi)\sum_{m=1}^{M} \hat{w}_{(m)1}\hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 1) + \pi\sum_{m=1}^{M} \hat{w}_{(m)0}\hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 0),$$

$$\bar{\hat{h}}_{[s]1} = \frac{1}{n_{[s]1}}\sum_{i\in\mathcal{J}_{[s]}} A_i\hat{h}_{[s]}(\boldsymbol{X}_i, \hat{\boldsymbol{w}}), \text{ and } \bar{\hat{h}}_{[s]0} = \frac{1}{n_{[s]0}}\sum_{i\in\mathcal{J}_{[s]}} (1-A_i)\hat{h}_{[s]}(\boldsymbol{X}_i, \hat{\boldsymbol{w}}).$$

Then

$$\hat{\tau} \;=\; \sum_{s=1}^{S} p_{n[s]}\left[\left\{\bar{Y}_{[s]1} - \frac{1}{n_{[s]1}}\sum_{i\in\mathcal{J}_{[s]}}(A_i - \pi_{n[s]})\sum_{m=1}^{M}\hat{w}_{(m)1}\hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 1)\right\}\right.$$

$$
\begin{aligned}
&\quad -\left\{\bar{Y}_{[s]0} + \frac{1}{n_{[s]0}} \sum_{i\in\mathcal{J}_{[s]}} (A_i - \pi_{n[s]}) \sum_{m=1}^{M} \hat{w}_{(m)0} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 0)\right\}\Bigg]\\
&= \sum_{s=1}^{S} p_{n[s]} \Bigg[\left\{\bar{Y}_{[s]1} - \sum_{i\in\mathcal{J}_{[s]}} \frac{A_i - \pi_{n[s]}}{n_{[s]}\pi_{n[s]}} \sum_{m=1}^{M} \hat{w}_{(m)1} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 1)\right\}\\
&\quad -\left\{\bar{Y}_{[s]0} + \sum_{i\in\mathcal{J}_{[s]}} \frac{A_i - \pi_{n[s]}}{n_{[s]}(1-\pi_{n[s]})} \sum_{m=1}^{M} \hat{w}_{(m)0} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 0)\right\}\Bigg]\\
&= \sum_{s=1}^{S} \frac{p_{n[s]}}{n_{[s]}} \sum_{i\in\mathcal{J}_{[s]}} \Bigg\{\frac{A_i Y_i}{\pi_{n[s]}} - \frac{A_i - \pi_{n[s]}}{\pi_{n[s]}} \sum_{m=1}^{M} \hat{w}_{(m)1} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 1)\\
&\quad - \frac{(1-A_i)Y_i}{1-\pi_{n[s]}} - \frac{A_i - \pi_{n[s]}}{1-\pi_{n[s]}} \sum_{m=1}^{M} \hat{w}_{(m)0} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 0)\Bigg\}\\
&= \sum_{s=1}^{S} \frac{p_{n[s]}}{n_{[s]}} \sum_{i\in\mathcal{J}_{[s]}} \left[\frac{A_i}{\pi_{n[s]}}\{Y_i - \hat{h}_{[s]}(\boldsymbol{X}_i, \hat{\boldsymbol{w}})\} - \frac{1-A_i}{1-\pi_{n[s]}}\{Y_i - \hat{h}_{[s]}(\boldsymbol{X}_i, \hat{\boldsymbol{w}})\}\right]\\
&\quad + \sum_{s=1}^{S} \frac{p_{n[s]}}{n_{[s]}} \sum_{i\in\mathcal{J}_{[s]}} \Bigg[\frac{A_i}{\pi_{n[s]}}(\pi_{n[s]} - \pi)\left\{\sum_{m=1}^{M} \hat{w}_{(m)1} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 1) - \sum_{m=1}^{M} \hat{w}_{(m)0} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 0)\right\}\\
&\quad - \frac{1-A_i}{1-\pi_{n[s]}}(\pi_{n[s]} - \pi)\left\{\sum_{m=1}^{M} \hat{w}_{(m)1} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 1) - \sum_{m=1}^{M} \hat{w}_{(m)0} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 0)\right\}\Bigg]\\
&= \sum_{s=1}^{S} p_{n[s]}(\pi_{n[s]} - \pi)\frac{1}{n_{[s]1}} \sum_{i\in\mathcal{J}_{[s]}} A_i \left\{\sum_{m=1}^{M} \hat{w}_{(m)1} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 1) - \sum_{m=1}^{M} \hat{w}_{(m)0} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 0)\right\}\\
&\quad - \sum_{s=1}^{S} p_{n[s]}(\pi_{n[s]} - \pi)\frac{1}{n_{[s]0}} \sum_{i\in\mathcal{J}_{[s]}} (1-A_i) \left\{\sum_{m=1}^{M} \hat{w}_{(m)1} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 1) - \sum_{m=1}^{M} \hat{w}_{(m)0} \hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, 0)\right\}\\
&\quad + \sum_{s=1}^{S} p_{n[s]}\Big[\bar{Y}_{[s]1} - \bar{\hat{h}}_{[s]1} - \{\bar{Y}_{[s]0} - \bar{\hat{h}}_{[s]0}\}\Big]. \qquad \text{(S3.3)}
\end{aligned}
$$

Let F_i be the fold containing observation i. In the full-stratum sums in (S3.3), the notation $\hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, a)$ is interpreted as $\hat{h}_{m[s]}^{(-F_i)}(\boldsymbol{X}_i, a)$; when the sum is later restricted to $i \in \mathcal{J}_{[s]} \cap I_k$, this reduces to the usual fold-specific notation.

Next, we show that the last term is the leading term and is asymptotically normal in each case, whereas the first two terms are asymptotically negligible.

Let $n_{[s]k*} = \sum_{i\in\mathcal{J}_{[s]}\cap I_k} 1$ and $n_{[s]ka} = \sum_{i\in\mathcal{J}_{[s]}\cap I_k} \mathbb{1}_{[A_i=a]}$ denote the numbers of units in stratum s and fold k, overall and in treatment arm a, respectively. Since K is fixed and the fold partition is independent of the trial data, for $a = 0, 1$,

$$\frac{n_{[s]ka}}{n_{[s]a}} = \frac{1}{K} + O_P(n^{-1/2}), \qquad \frac{n}{n_{[s]ka}} = O_P(1).$$

For $i \in \mathcal{J}_{[s]}$, let

$$\hat{h}_{i[s]} = \sum_{m=1}^{M} \hat{w}_{(m)1}\hat{h}_{m[s]}^{(-F_i)}(\boldsymbol{X}_i, 1) - \sum_{m=1}^{M} \hat{w}_{(m)0}\hat{h}_{m[s]}^{(-F_i)}(\boldsymbol{X}_i, 0).$$

For each fixed m, a, and k, conditional on the training data outside fold I_k, the function $\hat{h}_{m[s]}^{(-k)}(\cdot, a)$ is fixed on the validation fold. Hence Proposition 1 of Liu et al. (2023) gives

$$\frac{1}{n_{[s]k1}} \sum_{i\in\mathcal{J}_{[s]}\cap I_k} A_i\hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, a) - \frac{1}{n_{[s]k0}} \sum_{i\in\mathcal{J}_{[s]}\cap I_k} (1-A_i)\hat{h}_{m[s]}^{(-k)}(\boldsymbol{X}_i, a) = O_P(n^{-1/2}).$$

Because $0 \le \hat{w}_{(m)a} \le 1$ and M is fixed, the same order holds for the corresponding foldwise contrast of $\hat{h}_{i[s]}$.

Splitting the full-stratum averages over the K folds gives

$$\begin{aligned}
&\frac{1}{n_{[s]1}} \sum_{i\in\mathcal{J}_{[s]}} A_i\hat{h}_{i[s]} - \frac{1}{n_{[s]0}} \sum_{i\in\mathcal{J}_{[s]}} (1 - A_i)\hat{h}_{i[s]} \\
&= \sum_{k=1}^{K} \frac{n_{[s]k1}}{n_{[s]1}} \left\{ \frac{1}{n_{[s]k1}} \sum_{i\in\mathcal{J}_{[s]}\cap I_k} A_i\hat{h}_{i[s]} - \frac{1}{n_{[s]k0}} \sum_{i\in\mathcal{J}_{[s]}\cap I_k} (1 - A_i)\hat{h}_{i[s]} \right\}
\end{aligned}$$

$$+ \sum_{k=1}^{K} \left(\frac{n_{[s]k1}}{n_{[s]1}} - \frac{n_{[s]k0}}{n_{[s]0}} \right) \frac{1}{n_{[s]k0}} \sum_{i \in \mathcal{J}_{[s]} \cap I_k} (1 - A_i) \hat{h}_{i[s]}.$$

The first sum is $O_P(n^{-1/2})$. In the second sum,

$$\frac{n_{[s]k1}}{n_{[s]1}} - \frac{n_{[s]k0}}{n_{[s]0}} = O_P(n^{-1/2}),$$

whereas each foldwise mean is $O_P(1)$ under the second-moment conditions. Since K is fixed,

$$\frac{1}{n_{[s]1}} \sum_{i \in \mathcal{J}_{[s]}} A_i \hat{h}_{i[s]} - \frac{1}{n_{[s]0}} \sum_{i \in \mathcal{J}_{[s]}} (1 - A_i) \hat{h}_{i[s]} = O_P(n^{-1/2}).$$

Together with $\pi_{n[s]} - \pi = o_P(1)$, this shows that the first two terms in (S3.3) are $o_P(n^{-1/2})$.

We present a lemma based on Assumption 5 in the main text. The first conclusion gives the root-n replacement that transfers the CLT from the oracle transformed outcome to the plug-in estimator, while the second gives the mean-square replacement that transfers the plug-in variance estimator to the oracle variance estimator.

Lemma S2 (Tu et al., 2024)**.** *Under Assumptions 1–3 and 5 and the independent fold partition, for $s = 1, \ldots, S$, $k = 1, \ldots, K$, $m = 1, \ldots, M$, and*

$a = 0, 1$,

$$\sqrt{n}\Bigg[\frac{1}{n_{[s]k1}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k} A_i\left\{\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - h^*_{m[s]}(\boldsymbol{X}_i,a)\right\}$$
$$- \frac{1}{n_{[s]k0}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}(1-A_i)\left\{\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - h^*_{m[s]}(\boldsymbol{X}_i,a)\right\}\Bigg] = o_P(1),$$
$$\frac{1}{n_{[s]k*}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}\left\{\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - h^*_{m[s]}(\boldsymbol{X}_i,a)\right\}^2 = o_P(1).$$

Here the subscripts $k1$ and $k0$ correspond to the treated and control validation units in fold k and stratum s; the argument a remains the fixed potential-outcome arm being evaluated.

Case 1. ξ is positive.

In this case, we have that $\frac{R(\hat{\boldsymbol{w}})}{\inf_{\boldsymbol{w}\in\mathcal{W}} R(\boldsymbol{w})} \xrightarrow{P} 1$. Recall that $R(\boldsymbol{w}) = \{\pi(1-\pi)\}^{-1}\sum_{s=1}^{S} p_s R(\boldsymbol{w}, s)$, and

$$\begin{aligned}
R^*(\boldsymbol{w}, s) &= \mathrm{Var}\Bigg(\Bigg\{(1-\pi)\sum_{m=1}^{M} w_{(m)1}h^*_{m[s]}(\boldsymbol{X}_i,1) + \pi\sum_{m=1}^{M} w_{(m)0}h^*_{m[s]}(\boldsymbol{X}_i,0)\Bigg\} \\
&\quad - \Big[(1-\pi)E\{Y_i(1)|\boldsymbol{X}_i, B_i = s\} + \pi E\{Y_i(0)|\boldsymbol{X}_i, B_i = s\}\Big]\Bigg) \\
&= \mathrm{Var}\Bigg(\Bigg\{(1-\pi)\sum_{m=1}^{M} w_{(m)1}h^*_{m[s]}(\boldsymbol{X}_i,1) + \pi\sum_{m=1}^{M} w_{(m)0}h^*_{m[s]}(\boldsymbol{X}_i,0)\Bigg\} \\
&\quad - \Big[(1-\pi)\sum_{m=1}^{M} w_{(m)1}E\{Y_i(1)|\boldsymbol{X}_i, B_i = s\} + \pi\sum_{m=1}^{M} w_{(m)0}E\{Y_i(0)|\boldsymbol{X}_i, B_i = s\}\Big]\Bigg) \\
&= \mathrm{Var}\{\boldsymbol{w}^T\boldsymbol{h}^*_{\mathrm{cent}}(s)\} = \boldsymbol{w}^T\mathrm{Var}\{\boldsymbol{h}^*_{\mathrm{cent}}(s)\}\boldsymbol{w},
\end{aligned}$$

where $\boldsymbol{h}^*_{\mathrm{cent}}(s) = \Big((1-\pi)\big[h^*_{1[s]}(X_i,1) - E\{Y_i(1)|\boldsymbol{X}_i, B_i = s\}\big], \ldots, (1-\pi)\big[h^*_{M[s]}(X_i,1) - E\{Y_i(1)|\boldsymbol{X}_i, B_i = s\}\big], \pi\big[h^*_{1[s]}(X_i,0) - E\{Y_i(0)|\boldsymbol{X}_i, B_i =$

$s\}\big], \ldots, \pi\big[h^*_{M[s]}(X_i, 0) - E\{Y_i(0)|\boldsymbol{X}_i, B_i = s\}\big]\Big)^{\mathrm{T}}$.

Let

$$\mathcal{W}_\star = \arg\min_{\boldsymbol{w}\in\mathcal{W}} R^*(\boldsymbol{w}), \qquad R_\star = \inf_{\boldsymbol{w}\in\mathcal{W}} R^*(\boldsymbol{w}).$$

Theorem 1 and the uniform replacement $R(\boldsymbol{w}) - R^*(\boldsymbol{w}) = o(1)$ imply

$$R^*(\hat{\boldsymbol{w}}) - R_\star = o_P(1).$$

Since $R^*(\boldsymbol{w})$ is continuous on the compact set $\mathcal{W}$,

$$\mathrm{dist}(\hat{\boldsymbol{w}}, \mathcal{W}_\star) = o_P(1).$$

Let $\boldsymbol{w}^\circ = \Pi_{\mathcal{W}_\star}(\hat{\boldsymbol{w}})$, so that $\|\hat{\boldsymbol{w}} - \boldsymbol{w}^\circ\| = o_P(1)$. Here $\boldsymbol{w}^\circ$ is the possibly random projection; a fixed reference minimizer is denoted by $\boldsymbol{w}^*$ below.

We have the following contrast-level replacement:

$$\begin{aligned}
&\sqrt{n}\Bigg[\frac{1}{n_{[s]k1}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k} A_i\Big\{\sum_{m=1}^M \hat{w}_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - \sum_{m=1}^M w^\circ_{(m)a}h^*_{m[s]}(\boldsymbol{X}_i,a)\Big\} \\
&\qquad -\frac{1}{n_{[s]k0}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}(1-A_i)\Big\{\sum_{m=1}^M \hat{w}_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - \sum_{m=1}^M w^\circ_{(m)a}h^*_{m[s]}(\boldsymbol{X}_i,a)\Big\}\Bigg] \\
= &\sqrt{n}\Bigg[\frac{1}{n_{[s]k1}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k} A_i\Big\{\sum_{m=1}^M \hat{w}_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - \sum_{m=1}^M w^\circ_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)\Big\} \\
&\qquad -\frac{1}{n_{[s]k0}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}(1-A_i)\Big\{\sum_{m=1}^M \hat{w}_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - \sum_{m=1}^M w^\circ_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)\Big\}\Bigg] \\
&+\sqrt{n}\Bigg[\frac{1}{n_{[s]k1}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k} A_i\Big\{\sum_{m=1}^M w^\circ_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - \sum_{m=1}^M w^\circ_{(m)a}h^*_{m[s]}(\boldsymbol{X}_i,a)\Big\} \\
&\qquad -\frac{1}{n_{[s]k0}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}(1-A_i)\Big\{\sum_{m=1}^M w^\circ_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - \sum_{m=1}^M w^\circ_{(m)a}h^*_{m[s]}(\boldsymbol{X}_i,a)\Big\}\Bigg] \\
= &\sum_{m=1}^M\{\hat{w}_{(m)a} - w^\circ_{(m)a}\}\sqrt{n}\Bigg[\frac{1}{n_{[s]k1}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k} A_i\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a) - \frac{1}{n_{[s]k0}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}(1-A_i)\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{m=1}^{M} w^{\circ}_{(m)a}\sqrt{n}\left[\frac{1}{n_{[s]k1}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k} A_i\{\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)-h^{*}_{m[s]}(\boldsymbol{X}_i,a)\}-\frac{1}{n_{[s]k0}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}(1-A_i)\{\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)-h^{*}_{m[s]}(\boldsymbol{X}_i,a)\}\right]\\
=\;& o_P(1)\cdot O_P(1)+o_P(1)=o_P(1).
\end{aligned}
$$

The last equality follows because $\|\hat{\boldsymbol{w}}-\boldsymbol{w}^{\circ}\|=o_P(1)$, each fixed-component CAR contrast in the first bracket is $O_P(1)$ after multiplication by $\sqrt{n}$, and the second bracketed term is exactly the contrast replacement in Lemma S2.

In addition,

$$
\begin{aligned}
&\frac{1}{n_{[s]k*}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}\left\{\sum_{m=1}^{M}\hat{w}_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)-\sum_{m=1}^{M}w^{\circ}_{(m)a}h^{*}_{m[s]}(\boldsymbol{X}_i,a)\right\}^2\\
\le\;& \frac{2}{n_{[s]k*}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}\left\{\sum_{m=1}^{M}\hat{w}_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)-\sum_{m=1}^{M}w^{\circ}_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)\right\}^2\\
&+\frac{2}{n_{[s]k*}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}\left\{\sum_{m=1}^{M}w^{\circ}_{(m)a}\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)-\sum_{m=1}^{M}w^{\circ}_{(m)a}h^{*}_{m[s]}(\boldsymbol{X}_i,a)\right\}^2\\
=\;& \frac{2}{n_{[s]k*}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}(\hat{\boldsymbol{w}}_a-\boldsymbol{w}^{\circ}_a)^{\mathrm{T}}\hat{\boldsymbol{h}}^{(-k)}_{[s]}(X_i,a)\hat{\boldsymbol{h}}^{(-k)\mathrm{T}}_{[s]}(X_i,a)(\hat{\boldsymbol{w}}_a-\boldsymbol{w}^{\circ}_a)\\
&+\sup_{m\in\{1,\ldots,M\}}\frac{2}{n_{[s]k*}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}\left\{\hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i,a)-h^{*}_{m[s]}(\boldsymbol{X}_i,a)\right\}^2\\
=\;& (\hat{\boldsymbol{w}}_a-\boldsymbol{w}^{\circ}_a)^{\mathrm{T}}\left\{\frac{2}{n_{[s]k*}}\sum_{i\in\mathcal{J}_{[s]}\cap I_k}\hat{\boldsymbol{h}}^{(-k)}_{[s]}(X_i,a)\hat{\boldsymbol{h}}^{(-k)\mathrm{T}}_{[s]}(X_i,a)\right\}(\hat{\boldsymbol{w}}_a-\boldsymbol{w}^{\circ}_a)+o_P(1)\\
\le\;& \|\hat{\boldsymbol{w}}_a-\boldsymbol{w}^{\circ}_a\|^2\cdot O_P(1)+o_P(1)\\
=\;& o_P(1)\cdot O_P(1)+o_P(1)=o_P(1),
\end{aligned}
$$

where the third-to-last equality holds under Assumption 5, the second-to-last inequality holds under Assumption 4, and the last equality holds because $||\hat{\boldsymbol{w}}-\boldsymbol{w}^{\circ}||^2\xrightarrow{P}0$.

It remains to show that the oracle conclusion does not depend on which element of $\mathcal{W}_\star$ is selected. For any $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{W}_\star$, the midpoint $(\boldsymbol{w}_1 + \boldsymbol{w}_2)/2$ also belongs to $\mathcal{W}$. Since $R^*(\boldsymbol{w}, s) = \mathrm{Var}\{\boldsymbol{w}^{\mathrm{T}} \boldsymbol{h}^*_{\mathrm{cent}}(s)\}$, quadratic expansion gives

$$R^* \left(\frac{\boldsymbol{w}_1 + \boldsymbol{w}_2}{2} \right) = \frac{1}{2} R^*(\boldsymbol{w}_1) + \frac{1}{2} R^*(\boldsymbol{w}_2) \\ - \frac{1}{4\pi(1-\pi)} \sum_{s=1}^{S} p_s \mathrm{Var}\big\{ (\boldsymbol{w}_1 - \boldsymbol{w}_2)^{\mathrm{T}} \boldsymbol{h}^*_{\mathrm{cent}}(s) \mid B_i = s \big\} .$$

The first two terms on the right equal $R_\star$, whereas the left-hand side cannot be smaller than $R_\star$. Since $p_s > 0$ and every variance is nonnegative,

$$\mathrm{Var}\big\{ (\boldsymbol{w}_1 - \boldsymbol{w}_2)^{\mathrm{T}} \boldsymbol{h}^*_{\mathrm{cent}}(s) \mid B_i = s \big\} = 0, \qquad s = 1, \ldots, S.$$

The treatment-arm weights and the control-arm weights each sum to one, so their differences each sum to zero.

Consequently, the oracle adjustments induced by $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ differ only by a stratum-specific constant. Such a common constant cancels from the stratified difference-in-means estimator, leaves the within-stratum variances unchanged, and leaves the stratum-specific treatment effects unchanged. Hence all elements of $\mathcal{W}_\star$ induce the same oracle ATE estimator and the same asymptotic variance.

Fix any $\boldsymbol{w}^* \in \mathcal{W}_\star$. The projected weight $\boldsymbol{w}^\circ$ and this fixed reference minimizer therefore differ only through stratum-specific constants in their induced oracle adjustments.

Define the oracle transformed outcomes

$$r_i^*(\boldsymbol{w}^\circ, a) = Y_i(a) - \left\{ (1-\pi) \sum_{m=1}^{M} w_{(m)1}^\circ h_{m[s]}^*(\boldsymbol{X}_i, 1) + \pi \sum_{m=1}^{M} w_{(m)0}^\circ h_{m[s]}^*(\boldsymbol{X}_i, 0) \right\}.$$

Applying the same fold-to-full decomposition to the selected-to-oracle difference, the preceding foldwise contrast and mean-square bounds give

$$\sqrt{n}\{\hat{\tau}(\hat{\boldsymbol{w}}) - \hat{\tau}(r^*(\boldsymbol{w}^\circ))\} = o_P(1).$$

The oracle estimators based on $\boldsymbol{w}^\circ$ and the fixed minimizer $\boldsymbol{w}^*$ are identical because their adjustments differ only by stratum-specific constants. Applying Proposition 1 in the main text, which follows from Theorem 1 in Tu et al. (2024), to the fixed transformed outcomes $r_i^*(\boldsymbol{w}^*, a)$ and Slutsky's theorem therefore yield asymptotic normality with mean τ and variance $\varsigma^2_{r^*(\boldsymbol{w}^*)}(\pi) + \varsigma^2_{Hr^*(\boldsymbol{w}^*)}$.

Then, for each stratum and treatment arm,

$$\left| \widehat{\mathrm{Var}}(x) - \widehat{\mathrm{Var}}(y) \right| \le \widehat{E}\{(x-y)^2\} + 2\widehat{E}\{(x-y)^2\}^{1/2} \widehat{E}(y^2)^{1/2}.$$

Taking $x = \hat{r}_i(\hat{\boldsymbol{w}}, a)$ and $y = r_i^*(\boldsymbol{w}^\circ, a)$, the preceding mean-square replacement and $n_{[s]k}/n_{[s]ka} = O_P(1)$ give

$$\widehat{E}_{[s]a}\left[\{\hat{r}_i(\hat{\boldsymbol{w}}, a) - r_i^*(\boldsymbol{w}^\circ, a)\}^2\right] = o_P(1),$$

and the oracle empirical second moments are $O_P(1)$ under Assumption 4.

Consequently, the displayed inequality controls the within-stratum variances; $|\widehat{E}(x-y)| \leq \widehat{E}\{(x-y)^2\}^{1/2}$ controls the arm means in the between-stratum component. Since the numbers of strata and arms are fixed, the plug-in variance estimator differs by $o_P(1)$ from the variance estimator computed from the oracle transformed outcomes $r_i^*(\boldsymbol{w}^\circ, a)$.

Finally, the oracle adjustments evaluated at $\boldsymbol{w}^\circ$ and at the fixed minimizer $\boldsymbol{w}^*$ differ only by stratum-specific constants. These constants leave the limiting variance unchanged, and the two corresponding empirical variance estimators differ by $o_P(1)$ under CAR balance. Proposition 1 applied to $r_i^*(\boldsymbol{w}^*, a)$ therefore yields

$$\hat{\varsigma}^2_{r(\hat{\boldsymbol{w}})}(\pi) + \hat{\varsigma}^2_{Hr(\hat{\boldsymbol{w}})} \xrightarrow{P} \varsigma^2_{r^*(\boldsymbol{w}^*)}(\pi) + \varsigma^2_{Hr^*(\boldsymbol{w}^*)}.$$

Case 2. There exists a convex combination of candidate models that achieves zero population risk.

In this case, we have $\inf_{\boldsymbol{w}\in\mathcal{W}} R^*(\boldsymbol{w}) = 0$. Define the set $\mathcal{W}_0 = \{\boldsymbol{w} \in \mathcal{W} : R^*(\boldsymbol{w}) = 0\}$. By the proof of Theorem 2, it follows that $R(\hat{\boldsymbol{w}}) = o_P(1)$ and $\mathrm{dist}(\hat{\boldsymbol{w}}, \mathcal{W}_0) = o_P(1)$.

Since $\mathcal{W}$ is compact and $R^*(\boldsymbol{w})$ is continuous, $\mathcal{W}_0$ is nonempty and compact. Let

$$\boldsymbol{w}^\circ = \Pi_{\mathcal{W}_0}(\hat{\boldsymbol{w}}), \qquad \|\hat{\boldsymbol{w}} - \boldsymbol{w}^\circ\| = o_P(1).$$

For this projected oracle weight, $R^*(\boldsymbol{w}^\circ) = 0$ implies

$$R^*(\boldsymbol{w}^\circ, s) = 0, \qquad s = 1, \ldots, S,$$

because $R^*(\boldsymbol{w}^\circ) = \{\pi(1-\pi)\}^{-1} \sum_s p_s R^*(\boldsymbol{w}^\circ, s)$, $p_s > 0$, and every stratum risk is nonnegative. Thus the oracle adjustment $h^*_{[s]}(\boldsymbol{X}_i, \boldsymbol{w}^\circ)$ is optimal in the same L_2/variance sense needed in Case 1. Repeating the two replacement bounds already verified in Case 1, with $\boldsymbol{w}^*$ replaced by $\boldsymbol{w}^\circ$, gives

$$\sqrt{n}\left[\frac{1}{n_{[s]k1}} \sum_{i \in \mathcal{J}_{[s]} \cap I_k} A_i \left\{ \sum_{m=1}^{M} \hat{w}_{(m)a} \hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i, a) - \sum_{m=1}^{M} w^\circ_{(m)a} h^*_{m[s]}(\boldsymbol{X}_i, a) \right\} \right.$$
$$\left. - \frac{1}{n_{[s]k0}} \sum_{i \in \mathcal{J}_{[s]} \cap I_k} (1 - A_i) \left\{ \sum_{m=1}^{M} \hat{w}_{(m)a} \hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i, a) - \sum_{m=1}^{M} w^\circ_{(m)a} h^*_{m[s]}(\boldsymbol{X}_i, a) \right\} \right] = o_P(1),$$

and

$$\frac{1}{n_{[s]k*}} \sum_{i \in \mathcal{J}_{[s]} \cap I_k} \left\{ \sum_{m=1}^{M} \hat{w}_{(m)a} \hat{h}^{(-k)}_{m[s]}(\boldsymbol{X}_i, a) - \sum_{m=1}^{M} w^\circ_{(m)a} h^*_{m[s]}(\boldsymbol{X}_i, a) \right\}^2 = o_P(1).$$

Hence the same fixed-oracle CLT and the same plug-in variance-estimator argument established in Case 1 apply, and Case 2 follows. □

S4 Auxiliary Results

Lemma S3. *Let M be fixed. For $j = 1, \ldots, M$, suppose that $\{X_{n,j} : n \geq 1\}$ is uniformly integrable. Let $w_{n,1}, \ldots, w_{n,M}$ be nonnegative weights, deterministic or random, with $\sum_{j=1}^{M} w_{n,j} = 1$ almost surely. Then*

$$\left\{ \sum_{j=1}^{M} w_{n,j} X_{n,j} : n \geq 1, \ (w_{n,1}, \ldots, w_{n,M}) \in \Delta_M \right\}$$

is uniformly integrable. The fixed-random-variable version follows as the special case in which $X_{n,j}$ does not depend on n.

Proof. Let $S_n = \sum_{j=1}^M w_{n,j} X_{n,j}$ and set $u = t/M$. Since the weights are nonnegative and sum to one,

$$|S_n| \leq \sum_{j=1}^M |X_{n,j}|.$$

Hence

$$|S_n| \mathbf{1}_{\{|S_n|>t\}} \leq \sum_{j=1}^M \sum_{\ell=1}^M |X_{n,j}| \mathbf{1}_{\{|X_{n,\ell}|>u\}}.$$

For nonnegative a, b,

$$a\mathbf{1}_{\{b>u\}} \leq a\mathbf{1}_{\{a>u\}} + u\mathbf{1}_{\{b>u\}},$$

and $uP(|X_{n,\ell}| > u) \leq E\{|X_{n,\ell}| \mathbf{1}_{\{|X_{n,\ell}|>u\}}\}$. Therefore

$$\sup_{n,w} E\{|S_n| \mathbf{1}_{\{|S_n|>t\}}\} \leq 2M \sum_{j=1}^M \sup_n E\{|X_{n,j}| \mathbf{1}_{\{|X_{n,j}|>t/M\}}\}.$$

The right-hand side tends to zero by the assumed uniform integrability of each sequence and fixed M. □

Lemma S4. *For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,k}$ be i.i.d. with mean $\mu_n = \mathbb{E}X_{n,1}$. Assume the family $\{X_{n,1} : n \geq 1\}$ is uniformly integrable, i.e.*

$$\lim_{M\to\infty} \sup_{n\geq 1} \mathbb{E}\big[|X_{n,1}| \mathbf{1}_{\{|X_{n,1}|>M\}}\big] = 0.$$

Then

$$\sup_{n\geq 1} \mathbb{E}\big|\overline{X}_{n,k} - \mu_n\big| \xrightarrow[k\to\infty]{} 0.$$

Consequently, for every $\varepsilon > 0$,

$$\sup_{n\geq 1} \mathbb{P}\big(|\overline{X}_{n,k} - \mu_n| > \varepsilon\big) \xrightarrow[k\to\infty]{} 0,$$

i.e. $\overline{X}_{n,k} \to \mu_n$ *in probability uniformly in* n.

Proof. Fix $M > 0$ and define the truncation $Y_{n,i} := X_{n,i}\mathbf{1}_{\{|X_{n,i}|\leq M\}}$, $\overline{Y}_{n,k} = k^{-1}\sum_{i=1}^{k} Y_{n,i}$ and $\mu_{n,M} = \mathbb{E}Y_{n,1}$. As usual,

$$\mathbb{E}|\overline{X}_{n,k} - \mu_n| \leq \mathbb{E}|\overline{X}_{n,k} - \overline{Y}_{n,k}| + \mathbb{E}|\overline{Y}_{n,k} - \mu_{n,M}| + |\mu_{n,M} - \mu_n|.$$

By identical distribution and the triangle inequality,

$$\mathbb{E}|\overline{X}_{n,k} - \overline{Y}_{n,k}| \leq \mathbb{E}|X_{n,1} - Y_{n,1}| = \mathbb{E}\big[|X_{n,1}|\mathbf{1}_{\{|X_{n,1}|>M\}}\big],$$

and similarly $|\mu_{n,M} - \mu_n| \leq \mathbb{E}[|X_{n,1}|\mathbf{1}_{\{|X_{n,1}|>M\}}]$. For the middle term, $|Y_{n,1}| \leq M$ so

$$\mathbb{E}|\overline{Y}_{n,k} - \mu_{n,M}| \leq \sqrt{\operatorname{Var}(\overline{Y}_{n,k})} = \frac{\sigma_{n,M}}{\sqrt{k}} \leq \frac{M}{\sqrt{k}},$$

where $\sigma_{n,M}^2 = \operatorname{Var}(Y_{n,1}) \leq M^2$. Combine these three bounds to obtain the uniform estimate

$$\sup_{n\geq 1} \mathbb{E}|\overline{X}_{n,k} - \mu_n| \leq 2\sup_{n\geq 1} \mathbb{E}\big[|X_{n,1}|\mathbf{1}_{\{|X_{n,1}|>M\}}\big] + \frac{M}{\sqrt{k}}.$$

Now fix $\varepsilon > 0$. Choose M so large that the first term is $< \varepsilon/2$ (which is possible by uniform integrability), and then choose K so large that $M/\sqrt{K} < \varepsilon/2$. For all $k \geq K$ the right-hand side is $< \varepsilon$. Thus $\sup_n \mathbb{E}|\overline{X}_{n,k} - \mu_n| \to 0$ as $k \to \infty$.

Finally, we apply Markov's inequality:

$$\sup_n \mathbb{P}(|\overline{X}_{n,k} - \mu_n| > \varepsilon) \leq \frac{1}{\varepsilon} \sup_n \mathbb{E}|\overline{X}_{n,k} - \mu_n|,$$

which yields uniform convergence in probability. □

S5 Detailed Simulation Settings and Results

Here, we present the detailed simulation settings and results. For $a \in \{0, 1\}$ and $1 \leq i \leq n$, the potential outcomes are generated according to

$$Y_i(a) = g_a(\boldsymbol{X}_i) + \varepsilon_{a,i},$$

where $g_a(\boldsymbol{X}_i)$ is specified below for $a = 0, 1$. In each model, the vectors $(\boldsymbol{X}_i, \varepsilon_{0,i}, \varepsilon_{1,i}), 1 \leq i \leq n$ are i.i.d. For all models except Model 2, the error terms are generated as $\varepsilon_{0,i} \sim \mathcal{N}(0, 1), \varepsilon_{1,i} \sim \mathcal{N}(0, 9)$. For Model 2, we instead generate heavier-tailed and skewed errors, with $\varepsilon_{0,i}$ following a noncentral t distribution with 3 degrees of freedom and noncentrality parameter 2, and $\varepsilon_{1,i} \sim t(5)$.

Model 1: $\boldsymbol{X}_i$ is a four-dimensional vector,

$$g_a(\boldsymbol{X}_i) = \mu_a + \boldsymbol{\beta}_a^{\mathrm{T}} \boldsymbol{X}_i, a = 0, 1$$

where $\mu_0 = 1, \mu_1 = 4, X_{i1} \sim \text{Beta}(3, 4)$, $X_{i2} \sim \text{Unif}[-2, 2]$, X_{i3} takes values in $\{-1, 1\}$ with equal probability, X_{i4} takes values in $\{3, 5\}$ with probabilities 0.6, 0.4, and they are mutually independent. $\boldsymbol{\beta}_0 = (75, 35, 125, 80)^{\mathrm{T}}, \boldsymbol{\beta}_1 =$

$(100, 80, 60, 40)^{\mathrm{T}}$. The stratum used for randomization is an additional covariate taking values in $\{1, 2, 3, 4\}$ with probabilities $0.2, 0.3, 0.3, 0.2$ and is independent of $\boldsymbol{X}_i$. All the covariates are used in the working models.

Model 2: $\boldsymbol{X}_i$ is a two-dimensional vector,

$$g_0(\boldsymbol{X}_i) = \mu_0 + 60\text{sin}(3X_{i2}) + 80/(1 + \exp(-5(X_{i1} - 2))),$$

$$g_1(\boldsymbol{X}_i) = \mu_1 + 80\text{sin}(3X_{i1}) + 60/(1 + \exp(-5(X_{i2} - 2))),$$

where $\mu_0 = -3, \mu_1 = 0, X_{i1} \sim \text{Beta}(3, 4)$, $X_{i2} \sim \text{Unif}[-2, 2]$, and they are mutually independent. The stratum used for randomization is an additional covariate taking values in $\{1, 2, 3, 4\}$ with probabilities $0.2, 0.3, 0.3, 0.2$ and is independent of $\boldsymbol{X}_i$. All the covariates are used in the working models.

Model 3: $\boldsymbol{X}_i$ is a four-dimensional vector,

$$g_0(\boldsymbol{X}_i) = 6(X_{i1} - 15)^2,$$

$$g_1(\boldsymbol{X}_i) = (10\text{sin}(X_{i1}\pi))\mathbb{1}_{[X_{i2} \le 0]} + 4\exp(X_{i3} + X_{i4})\mathbb{1}_{[X_{i2} > 0]},$$

where $X_{i1} \sim 0.8 \cdot \text{Unif}[10, 20] + 0.13 \cdot \text{Unif}[0, 10] + 0.07 \cdot \text{Unif}[20, 30]$, $X_{i2} \sim \text{Unif}[-1, 3], X_{i3} \sim \text{Unif}[0, 3], X_{i4} \sim \text{Unif}[1, 5]$, and they are mutually independent. The stratum used for randomization is an additional covariate taking values in $\{1, 2\}$ with equal probability and is independent of $\boldsymbol{X}_i$. All the covariates are used in the working models.

Model 4: $\boldsymbol{X}_i$ is a three-dimensional vector,

$$g_0(\boldsymbol{X}_i) = 100\sin(X_{i1}) + 10\exp(X_{i2}),$$

$$g_1(\boldsymbol{X}_i) = 75X_{i1}\sin(X_{i3}) + 30/\big[1 + \exp\{-2(X_{i1} + X_{i3})\}\big],$$

where the latent covariates $U_{i1} \sim Beta(3,4), U_{i2} \sim Unif[-2,2], U_{i3} \sim Gamma(2,3)$. $X_{i1} = U_{i1}, X_{i2} = 3U_{i1}^2 + 0.5U_{i2}$, and $X_{i3} = U_{i1}U_{i3}$. U_{i1}, U_{i2}, and U_{i3} are mutually independent, so the observed covariates $\boldsymbol{X}_i$ are correlated with each other. The stratum used for randomization is an additional covariate taking values in $\{1,2,3,4\}$ with probabilities $0.2, 0.3, 0.3, 0.2$ and is independent of $\boldsymbol{X}_i$. X_{i1} and X_{i2} are used for spline fitting, while X_{i1} and X_{i3} are used for other models.

In addition, the block size used in stratified block randomization is 6, equal weights are used in Pocock and Simon's minimization, and the overall, marginal, and within-stratum weights are set to be 0.3, 0.2 and 0.5, respectively, for Hu and Hu's randomization. The empirical ratio is calculated as SE/SD.

S5.1 Simulation Results with Different Numbers of Folds

Table S1: Monte Carlo results for each estimator under different randomization methods with $K = 2$ and $n = 500$.

Model	Estimator	Complete Rand.					Stratified Block Rand.					Minimization					Hu & Hu's Rand.				
		RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP
1	$\hat{\tau}_{\text{ma}}$	-0.06%	4.13	4.11	1.00	0.95	-0.02%	4.19	4.11	0.98	0.95	0.10%	4.10	4.11	1.00	0.95	-0.09%	4.21	4.11	0.98	0.96
	$\hat{\tau}_{\text{linear}}$	-0.06%	4.14	4.10	0.99	0.95	-0.03%	4.15	4.10	0.99	0.95	0.10%	4.09	4.10	1.00	0.95	-0.08%	4.19	4.10	0.98	0.96
	$\hat{\tau}_{\text{kernel}}$	-0.06%	4.14	4.10	0.99	0.95	-0.03%	4.15	4.10	0.99	0.95	0.10%	4.09	4.10	1.00	0.94	-0.08%	4.19	4.10	0.98	0.96
	$\hat{\tau}_{\text{nspline}}$	0.05%	10.56	10.84	1.03	0.95	0.02%	10.94	10.83	0.99	0.94	0.24%	10.76	10.82	1.01	0.95	0.10%	10.60	10.82	1.02	0.96
	$\hat{\tau}_{\text{rf}}$	0.00%	6.03	6.10	1.01	0.95	0.05%	6.33	6.08	0.96	0.94	0.23%	6.16	6.09	0.99	0.95	0.06%	5.98	6.09	1.02	0.96
	$\hat{\tau}_{\text{ms}}$	-0.06%	4.14	4.10	0.99	0.95	-0.03%	4.15	4.10	0.99	0.95	0.10%	4.09	4.10	1.00	0.94	-0.08%	4.19	4.10	0.98	0.96
2	$\hat{\tau}_{\text{ma}}$	-0.05%	2.10	2.15	1.02	0.96	-0.02%	2.16	2.15	1.00	0.95	0.02%	2.11	2.15	1.02	0.95	0.00%	2.13	2.15	1.01	0.95
	$\hat{\tau}_{\text{linear}}$	-0.12%	2.80	2.83	1.01	0.95	-0.10%	2.87	2.83	0.99	0.94	-0.03%	2.75	2.83	1.03	0.95	0.03%	2.82	2.83	1.00	0.94
	$\hat{\tau}_{\text{kernel}}$	-0.07%	2.08	2.14	1.03	0.96	-0.03%	2.12	2.14	1.01	0.95	0.00%	2.11	2.14	1.02	0.96	0.00%	2.12	2.14	1.01	0.95
	$\hat{\tau}_{\text{nspline}}$	-0.11%	2.12	2.19	1.03	0.96	-0.01%	2.20	2.19	1.00	0.95	-0.01%	2.16	2.19	1.02	0.96	-0.02%	2.15	2.19	1.02	0.96
	$\hat{\tau}_{\text{rf}}$	-0.09%	2.21	2.26	1.02	0.96	-0.05%	2.27	2.26	1.00	0.94	-0.01%	2.19	2.26	1.03	0.95	0.01%	2.24	2.26	1.01	0.95
	$\hat{\tau}_{\text{ms}}$	-0.06%	2.08	2.14	1.03	0.96	-0.03%	2.13	2.14	1.01	0.95	0.00%	2.10	2.14	1.02	0.96	0.00%	2.12	2.14	1.01	0.95
3	$\hat{\tau}_{\text{ma}}$	-0.67%	63.93	65.17	1.02	0.94	-0.08%	67.55	65.07	0.96	0.94	-0.29%	66.77	65.17	0.98	0.93	-0.30%	64.06	65.27	1.02	0.95
	$\hat{\tau}_{\text{linear}}$	-0.33%	72.33	75.37	1.04	0.94	-0.02%	77.98	75.63	0.97	0.93	-0.02%	76.13	75.11	0.99	0.94	0.31%	74.84	75.55	1.01	0.94
	$\hat{\tau}_{\text{kernel}}$	5.48%	501.79	162.52	0.32	0.97	8.55%	798.08	180.95	0.23	0.95	-1.79%	556.01	169.68	0.31	0.95	7.47×10^{12}%	1.23×10^{15}	3.91×10^{13}	0.03	0.97
	$\hat{\tau}_{\text{nspline}}$	-0.63%	81.03	83.41	1.03	0.94	0.16%	85.27	83.70	0.98	0.94	-0.41%	84.50	83.14	0.98	0.93	0.41%	83.65	83.71	1.00	0.95
	$\hat{\tau}_{\text{rf}}$	-0.51%	68.10	70.87	1.04	0.94	0.13%	73.16	71.16	0.97	0.94	-0.23%	71.65	70.62	0.99	0.94	0.39%	70.28	71.04	1.01	0.95
	$\hat{\tau}_{\text{ms}}$	-0.90%	66.13	68.30	1.03	0.94	-0.06%	71.25	68.49	0.96	0.94	-0.32%	69.87	68.10	0.97	0.94	0.13%	68.80	68.55	1.00	0.95
4	$\hat{\tau}_{\text{ma}}$	-0.29%	1.32	1.32	1.00	0.94	-0.26%	1.34	1.32	0.98	0.94	-0.24%	1.34	1.32	0.98	0.95	-0.07%	1.38	1.32	0.96	0.94
	$\hat{\tau}_{\text{linear}}$	-0.04%	1.57	1.55	0.99	0.94	-0.10%	1.53	1.54	1.01	0.94	0.05%	1.54	1.54	1.00	0.95	0.06%	1.59	1.54	0.97	0.94
	$\hat{\tau}_{\text{kernel}}$	22.44%	237.02	10.48	0.04	0.95	-1.65%	7.95	2.55	0.32	0.94	2.71%	29.82	3.40	0.11	0.95	0.21%	3.18	2.21	0.70	0.95
	$\hat{\tau}_{\text{nspline}}$	-0.06%	1.35	1.35	1.01	0.95	0.01%	1.38	1.35	0.98	0.94	-0.10%	1.37	1.35	0.98	0.95	0.09%	1.39	1.35	0.98	0.95
	$\hat{\tau}_{\text{rf}}$	0.01%	1.60	1.56	0.97	0.94	-0.07%	1.54	1.55	1.01	0.94	0.01%	1.55	1.55	1.00	0.94	0.10%	1.60	1.56	0.97	0.94
	$\hat{\tau}_{\text{ms}}$	-0.19%	1.39	1.35	0.98	0.95	-0.10%	1.40	1.35	0.96	0.93	-0.19%	1.38	1.35	0.97	0.94	-0.02%	1.41	1.35	0.96	0.94

Abbreviations: RB, relative bias; SD, standard deviation; SE, standard error; ER, empirical ratio; CP, coverage probability; Rand.: randomization; ma: model averaging; nspline: natural spline; rf: random forest; ms: model selection.

Table S2: Monte Carlo results for each estimator under different randomization methods with $K = 5$ and $n = 500$.

Model	Estimator	Complete Rand.					Stratified Block Rand.					Minimization					Hu & Hu's Rand.				
		RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP
1	$\hat{\tau}_{\text{ma}}$	0.09%	4.40	4.11	0.93	0.94	0.09%	4.09	4.11	1.01	0.96	0.09%	4.03	4.11	1.02	0.96	-0.12%	4.14	4.11	0.99	0.96
	$\hat{\tau}_{\text{linear}}$	0.11%	4.35	4.10	0.94	0.94	0.11%	4.06	4.10	1.01	0.96	0.09%	4.03	4.10	1.02	0.95	-0.13%	4.09	4.10	1.00	0.95
	$\hat{\tau}_{\text{kernel}}$	0.11%	4.35	4.10	0.94	0.93	0.11%	4.06	4.10	1.01	0.96	0.09%	4.04	4.10	1.02	0.95	-0.13%	4.09	4.10	1.00	0.95
	$\hat{\tau}_{\text{nspline}}$	0.13%	10.93	10.75	0.98	0.94	-0.15%	10.42	10.74	1.03	0.95	0.27%	10.66	10.73	1.01	0.96	-0.18%	11.30	10.73	0.95	0.94
	$\hat{\tau}_{\text{rf}}$	0.02%	6.22	5.92	0.95	0.94	-0.05%	6.03	5.92	0.98	0.95	0.23%	6.01	5.92	0.98	0.95	-0.13%	6.06	5.92	0.98	0.95
	$\hat{\tau}_{\text{ms}}$	0.11%	4.36	4.10	0.94	0.93	0.11%	4.06	4.10	1.01	0.96	0.09%	4.04	4.10	1.02	0.95	-0.13%	4.09	4.10	1.00	0.95
2	$\hat{\tau}_{\text{ma}}$	0.07%	2.17	2.15	0.99	0.95	-0.16%	2.18	2.15	0.99	0.95	0.07%	2.08	2.15	1.03	0.96	0.06%	2.13	2.15	1.01	0.95
	$\hat{\tau}_{\text{linear}}$	0.00%	2.86	2.82	0.99	0.94	0.06%	2.87	2.82	0.98	0.95	0.16%	2.73	2.82	1.03	0.96	0.00%	2.83	2.82	1.00	0.95
	$\hat{\tau}_{\text{kernel}}$	0.05%	2.17	2.15	0.99	0.94	-0.18%	2.16	2.14	0.99	0.95	0.04%	2.07	2.14	1.03	0.96	0.05%	2.11	2.14	1.02	0.95
	$\hat{\tau}_{\text{nspline}}$	0.02%	2.20	2.18	0.99	0.94	-0.19%	2.20	2.18	0.99	0.95	0.01%	2.11	2.18	1.04	0.96	0.06%	2.15	2.18	1.01	0.95
	$\hat{\tau}_{\text{rf}}$	0.02%	2.23	2.22	1.00	0.95	-0.13%	2.25	2.22	0.98	0.95	0.07%	2.13	2.21	1.04	0.96	0.03%	2.20	2.22	1.01	0.95
	$\hat{\tau}_{\text{ms}}$	0.03%	2.15	2.14	0.99	0.94	-0.18%	2.16	2.14	0.99	0.95	0.04%	2.07	2.14	1.03	0.96	0.05%	2.11	2.14	1.02	0.95
3	$\hat{\tau}_{\text{ma}}$	-0.45%	65.39	63.84	0.98	0.93	0.07%	64.13	63.72	0.99	0.94	-0.08%	66.00	63.97	0.97	0.94	-0.81%	63.92	63.43	0.99	0.94
	$\hat{\tau}_{\text{linear}}$	0.04%	75.69	75.10	0.99	0.93	0.63%	75.85	75.44	0.99	0.95	0.49%	76.14	75.43	0.99	0.94	-0.56%	72.58	74.72	1.03	0.95
	$\hat{\tau}_{\text{kernel}}$	-2.00%	303.89	110.45	0.36	0.95	-0.91%	322.10	122.08	0.38	0.96	-26.56%	7056.73	383.97	0.05	0.95	5.24%	590.94	134.76	0.23	0.96
	$\hat{\tau}_{\text{nspline}}$	-0.09%	83.10	82.77	1.00	0.94	0.73%	83.76	83.08	0.99	0.94	0.44%	82.80	83.14	1.00	0.95	-0.59%	82.63	82.26	1.00	0.94
	$\hat{\tau}_{\text{rf}}$	0.05%	68.93	68.79	1.00	0.93	0.61%	67.98	69.16	1.02	0.95	0.43%	69.74	69.13	0.99	0.95	-0.44%	68.16	68.50	1.00	0.94
	$\hat{\tau}_{\text{ms}}$	-0.37%	65.21	65.41	1.00	0.94	0.30%	65.19	65.67	1.01	0.94	0.13%	68.25	65.72	0.96	0.94	-0.76%	66.26	65.04	0.98	0.94
4	$\hat{\tau}_{\text{ma}}$	-0.09%	1.42	1.32	0.93	0.93	-0.09%	1.34	1.31	0.98	0.95	-0.17%	1.33	1.32	0.99	0.95	-0.21%	1.32	1.32	1.00	0.95
	$\hat{\tau}_{\text{linear}}$	0.16%	1.63	1.54	0.95	0.93	0.16%	1.54	1.53	0.99	0.94	0.16%	1.53	1.54	1.01	0.95	0.06%	1.53	1.54	1.01	0.96
	$\hat{\tau}_{\text{kernel}}$	-1.41%	11.70	2.85	0.24	0.94	-0.64%	9.44	2.60	0.27	0.95	0.07%	4.58	2.32	0.51	0.96	9.33%	53.37	4.88	0.09	0.96
	$\hat{\tau}_{\text{nspline}}$	0.11%	1.43	1.35	0.95	0.94	0.12%	1.36	1.34	0.99	0.94	0.12%	1.36	1.35	0.99	0.95	0.03%	1.34	1.35	1.00	0.96
	$\hat{\tau}_{\text{rf}}$	0.22%	1.63	1.55	0.95	0.94	0.22%	1.55	1.54	0.99	0.95	0.21%	1.53	1.55	1.01	0.95	0.08%	1.53	1.54	1.01	0.97
	$\hat{\tau}_{\text{ms}}$	-0.01%	1.46	1.35	0.93	0.94	0.02%	1.38	1.34	0.97	0.94	0.02%	1.38	1.35	0.98	0.95	-0.08%	1.36	1.35	0.99	0.95

Abbreviations: RB, relative bias; SD, standard deviation; SE, standard error; ER, empirical ratio; CP, coverage probability; Rand.: randomization; ma: model averaging; nspline: natural spline; rf: random forest; ms: model selection.

Table S3: Monte Carlo results for each estimator under different randomization methods with $K = 10$ and $n = 500$.

Model	Estimator	Complete Rand.					Stratified Block Rand.					Minimization					Hu & Hu's Rand.				
		RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP
1	$\hat{\tau}_{\text{ma}}$	-0.15%	3.96	4.11	1.04	0.96	0.07%	4.35	4.11	0.94	0.94	0.02%	4.07	4.11	1.01	0.95	0.28%	4.36	4.11	0.94	0.94
	$\hat{\tau}_{\text{linear}}$	-0.14%	3.95	4.10	1.04	0.95	0.08%	4.32	4.10	0.95	0.94	0.03%	4.04	4.10	1.01	0.95	0.31%	4.34	4.10	0.94	0.94
	$\hat{\tau}_{\text{kernel}}$	-0.14%	3.95	4.10	1.04	0.95	0.08%	4.32	4.10	0.95	0.94	0.03%	4.04	4.10	1.01	0.95	0.31%	4.34	4.10	0.94	0.94
	$\hat{\tau}_{\text{nspline}}$	-0.43%	10.36	10.72	1.04	0.96	0.31%	11.06	10.72	0.97	0.94	0.10%	10.86	10.70	0.98	0.95	-0.01%	10.69	10.73	1.00	0.95
	$\hat{\tau}_{\text{rf}}$	-0.15%	5.58	5.89	1.06	0.96	0.15%	6.12	5.89	0.96	0.94	0.11%	5.95	5.89	0.99	0.95	0.06%	6.04	5.88	0.97	0.95
	$\hat{\tau}_{\text{ms}}$	-0.14%	3.95	4.10	1.04	0.95	0.08%	4.32	4.10	0.95	0.94	0.03%	4.04	4.10	1.01	0.95	0.31%	4.34	4.10	0.94	0.94
2	$\hat{\tau}_{\text{ma}}$	0.02%	2.16	2.15	1.00	0.96	-0.11%	2.12	2.15	1.01	0.96	0.01%	2.17	2.15	0.99	0.95	-0.16%	2.17	2.15	0.99	0.95
	$\hat{\tau}_{\text{linear}}$	-0.03%	2.90	2.82	0.97	0.95	-0.22%	2.76	2.82	1.02	0.95	0.04%	2.83	2.82	1.00	0.96	-0.30%	2.84	2.82	0.99	0.94
	$\hat{\tau}_{\text{kernel}}$	-0.02%	2.13	2.14	1.00	0.95	-0.13%	2.10	2.14	1.02	0.96	0.00%	2.14	2.14	1.00	0.95	-0.20%	2.15	2.14	0.99	0.94
	$\hat{\tau}_{\text{nspline}}$	-0.01%	2.17	2.18	1.01	0.96	-0.12%	2.14	2.18	1.02	0.96	-0.03%	2.20	2.18	0.99	0.95	-0.19%	2.20	2.18	0.99	0.95
	$\hat{\tau}_{\text{rf}}$	-0.01%	2.23	2.21	0.99	0.95	-0.18%	2.17	2.21	1.02	0.96	0.01%	2.19	2.21	1.01	0.95	-0.23%	2.21	2.21	1.00	0.95
	$\hat{\tau}_{\text{ms}}$	-0.02%	2.13	2.14	1.01	0.95	-0.13%	2.10	2.14	1.02	0.96	0.00%	2.14	2.14	1.00	0.95	-0.20%	2.16	2.14	0.99	0.94
3	$\hat{\tau}_{\text{ma}}$	-0.81%	62.79	63.21	1.01	0.94	-0.61%	63.69	63.31	0.99	0.93	-0.95%	64.41	63.16	0.98	0.93	-0.86%	63.08	63.43	1.01	0.94
	$\hat{\tau}_{\text{linear}}$	-0.19%	75.48	74.86	0.99	0.94	-0.45%	76.25	74.81	0.98	0.94	-0.36%	75.63	75.01	0.99	0.93	-0.24%	73.16	75.24	1.03	0.95
	$\hat{\tau}_{\text{kernel}}$	1.06%	204.79	94.30	0.46	0.96	2.38%	409.78	113.94	0.28	0.95	-2.20%	629.36	118.36	0.19	0.95	-1.10%	241.26	102.56	0.43	0.96
	$\hat{\tau}_{\text{nspline}}$	-0.88%	81.74	82.40	1.01	0.94	-0.30%	84.12	82.30	0.98	0.93	-0.40%	84.42	82.49	0.98	0.93	-0.33%	79.78	82.82	1.04	0.95
	$\hat{\tau}_{\text{rf}}$	-0.33%	67.36	68.25	1.01	0.94	-0.33%	67.97	68.08	1.00	0.93	-0.31%	68.57	68.33	1.00	0.94	-0.27%	66.46	68.45	1.03	0.96
	$\hat{\tau}_{\text{ms}}$	-0.60%	65.46	64.74	0.99	0.94	-0.69%	65.21	64.55	0.99	0.93	-0.89%	66.05	64.52	0.98	0.93	-0.68%	64.20	64.55	1.01	0.95
4	$\hat{\tau}_{\text{ma}}$	-0.08%	1.38	1.32	0.96	0.93	-0.40%	1.31	1.31	1.00	0.95	-0.28%	1.34	1.32	0.98	0.94	-0.10%	1.33	1.32	1.00	0.95
	$\hat{\tau}_{\text{linear}}$	0.06%	1.56	1.54	0.99	0.94	-0.17%	1.54	1.52	0.99	0.94	0.06%	1.50	1.53	1.02	0.95	0.10%	1.55	1.55	1.00	0.95
	$\hat{\tau}_{\text{kernel}}$	-0.04%	12.57	2.85	0.23	0.94	-0.30%	6.82	2.54	0.37	0.94	1.31%	7.45	2.73	0.37	0.94	-0.57%	6.51	2.42	0.37	0.95
	$\hat{\tau}_{\text{nspline}}$	0.12%	1.36	1.35	0.99	0.94	-0.13%	1.33	1.34	1.00	0.95	0.00%	1.33	1.34	1.01	0.95	0.18%	1.37	1.36	0.99	0.95
	$\hat{\tau}_{\text{rf}}$	0.15%	1.58	1.54	0.98	0.95	-0.17%	1.53	1.53	1.00	0.94	-0.03%	1.52	1.54	1.01	0.94	0.14%	1.57	1.55	0.99	0.96
	$\hat{\tau}_{\text{ms}}$	-0.02%	1.40	1.35	0.96	0.94	-0.28%	1.36	1.34	0.98	0.95	-0.12%	1.34	1.34	1.00	0.95	0.04%	1.39	1.35	0.98	0.95

Abbreviations: RB, relative bias; SD, standard deviation; SE, standard error; ER, empirical ratio; CP, coverage probability; Rand.: randomization; ma: model averaging; nspline: natural spline; rf: random forest; ms: model selection.

S5.2 Simulation Results with Different Numbers of Units

Table S4: Monte Carlo results for each estimator under different randomization methods with $K = 5$ and $n = 200$.

Model	Estimator	Complete Rand.					Stratified Block Rand.					Minimization					Hu & Hu's Rand.				
		RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP
1	$\hat{\tau}_{\mathrm{ma}}$	0.10%	6.84	6.47	0.95	0.92	-0.15%	6.36	6.47	1.02	0.95	0.17%	6.70	6.47	0.97	0.94	-0.08%	6.83	6.49	0.95	0.94
	$\hat{\tau}_{\mathrm{linear}}$	0.12%	6.74	6.42	0.95	0.93	-0.13%	6.26	6.43	1.03	0.95	0.15%	6.62	6.44	0.97	0.93	-0.09%	6.71	6.44	0.96	0.94
	$\hat{\tau}_{\mathrm{kernel}}$	0.12%	6.75	6.42	0.95	0.93	-0.13%	6.26	6.43	1.03	0.95	0.14%	6.63	6.44	0.97	0.93	-0.09%	6.71	6.44	0.96	0.94
	$\hat{\tau}_{\mathrm{nspline}}$	0.46%	18.25	17.25	0.95	0.93	0.05%	17.30	17.30	1.00	0.95	0.48%	17.01	17.27	1.02	0.96	0.19%	17.71	17.32	0.98	0.95
	$\hat{\tau}_{\mathrm{rf}}$	0.18%	10.30	9.93	0.96	0.94	0.06%	9.88	9.92	1.00	0.95	0.37%	9.73	9.93	1.02	0.95	0.08%	10.27	9.92	0.97	0.94
	$\hat{\tau}_{\mathrm{ms}}$	0.12%	6.75	6.42	0.95	0.93	-0.14%	6.27	6.42	1.03	0.95	0.14%	6.63	6.43	0.97	0.93	-0.09%	6.71	6.44	0.96	0.94
2	$\hat{\tau}_{\mathrm{ma}}$	0.17%	3.49	3.40	0.97	0.94	-0.33%	3.44	3.40	0.99	0.94	0.01%	3.66	3.40	0.93	0.92	0.10%	3.48	3.40	0.98	0.93
	$\hat{\tau}_{\mathrm{linear}}$	0.14%	4.43	4.44	1.00	0.94	-0.40%	4.41	4.45	1.01	0.96	-0.07%	4.71	4.45	0.95	0.94	0.19%	4.52	4.45	0.99	0.95
	$\hat{\tau}_{\mathrm{kernel}}$	0.16%	3.38	3.37	1.00	0.95	-0.29%	3.37	3.37	1.00	0.94	0.02%	3.57	3.37	0.94	0.93	0.05%	3.37	3.37	1.00	0.94
	$\hat{\tau}_{\mathrm{nspline}}$	0.15%	3.53	3.47	0.98	0.95	-0.37%	3.47	3.47	1.00	0.95	-0.01%	3.70	3.47	0.94	0.92	0.04%	3.45	3.47	1.01	0.94
	$\hat{\tau}_{\mathrm{rf}}$	0.19%	3.64	3.63	1.00	0.94	-0.28%	3.64	3.64	1.00	0.94	-0.08%	3.87	3.64	0.94	0.93	0.14%	3.64	3.64	1.00	0.94
	$\hat{\tau}_{\mathrm{ms}}$	0.16%	3.42	3.37	0.98	0.95	-0.30%	3.39	3.37	0.99	0.94	0.02%	3.59	3.37	0.94	0.93	0.06%	3.40	3.37	0.99	0.94
3	$\hat{\tau}_{\mathrm{ma}}$	-0.42%	112.74	105.41	0.93	0.93	-1.73%	108.25	103.89	0.96	0.93	-1.13%	106.08	103.70	0.98	0.93	-1.12%	110.46	104.21	0.94	0.93
	$\hat{\tau}_{\mathrm{linear}}$	0.65%	122.60	118.92	0.97	0.92	-0.73%	117.93	117.86	1.00	0.93	-0.49%	118.93	117.40	0.99	0.92	-0.38%	120.15	118.27	0.98	0.94
	$\hat{\tau}_{\mathrm{kernel}}$	12.94%	1326.07	266.96	0.20	0.96	34.73%	5522.72	480.84	0.09	0.95	-4.43×10^{15}%	7.33×10^{17}	2.33×10^{16}	0.03	0.96	6.72%	1383.20	308.66	0.22	0.95
	$\hat{\tau}_{\mathrm{nspline}}$	0.79%	140.49	132.70	0.94	0.92	-0.54%	128.93	131.35	1.02	0.94	-1.67%	130.82	130.77	1.00	0.92	0.43%	138.21	132.19	0.96	0.92
	$\hat{\tau}_{\mathrm{rf}}$	0.49%	118.24	114.33	0.97	0.93	-0.45%	114.08	113.29	0.99	0.94	-0.94%	114.17	112.92	0.99	0.93	0.30%	118.07	113.76	0.96	0.92
	$\hat{\tau}_{\mathrm{ms}}$	-0.20%	115.40	109.36	0.95	0.93	-1.13%	111.66	108.79	0.97	0.93	-1.53%	111.38	108.53	0.97	0.92	-0.44%	116.06	109.00	0.94	0.92
4	$\hat{\tau}_{\mathrm{ma}}$	-0.33%	2.17	2.10	0.97	0.93	-0.42%	2.12	2.09	0.99	0.94	-0.36%	2.16	2.06	0.95	0.93	-0.31%	2.10	2.09	0.99	0.94
	$\hat{\tau}_{\mathrm{linear}}$	0.08%	2.47	2.43	0.98	0.93	-0.09%	2.45	2.43	0.99	0.93	0.24%	2.56	2.40	0.94	0.93	0.19%	2.42	2.42	1.00	0.94
	$\hat{\tau}_{\mathrm{kernel}}$	0.43%	10.87	3.60	0.33	0.94	1.12%	19.78	3.86	0.20	0.94	-0.04%	4.71	3.09	0.66	0.94	0.21%	6.55	3.34	0.51	0.96
	$\hat{\tau}_{\mathrm{nspline}}$	0.21%	2.15	2.14	1.00	0.94	0.17%	2.16	2.15	0.99	0.94	0.25%	2.21	2.12	0.96	0.93	0.12%	2.08	2.14	1.03	0.94
	$\hat{\tau}_{\mathrm{rf}}$	0.10%	2.55	2.48	0.97	0.93	-0.02%	2.44	2.48	1.01	0.94	0.10%	2.56	2.45	0.96	0.93	0.26%	2.47	2.47	1.00	0.94
	$\hat{\tau}_{\mathrm{ms}}$	-0.43%	2.21	2.13	0.96	0.92	-0.41%	2.26	2.13	0.94	0.92	-0.21%	2.28	2.11	0.92	0.92	-0.45%	2.16	2.13	0.99	0.93

Abbreviations: RB, relative bias; SD, standard deviation; SE, standard error; ER, empirical ratio; CP, coverage probability; Rand.: randomization; ma: model averaging; nspline: natural spline; rf: random forest; ms: model selection.

Table S5: Monte Carlo results for each estimator under different randomization methods with $K = 5$ and $n = 400$.

Model	Estimator	Complete Rand.					Stratified Block Rand.					Minimization					Hu & Hu's Rand.				
		RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP
1	$\hat{\tau}_{\text{ma}}$	0.04%	4.63	4.60	0.99	0.94	0.11%	4.66	4.60	0.99	0.95	-0.05%	4.45	4.59	1.03	0.96	0.08%	4.76	4.60	0.96	0.94
	$\hat{\tau}_{\text{linear}}$	0.06%	4.61	4.59	0.99	0.95	0.10%	4.62	4.58	0.99	0.95	-0.05%	4.41	4.57	1.04	0.96	0.07%	4.71	4.58	0.97	0.94
	$\hat{\tau}_{\text{kernel}}$	0.06%	4.61	4.59	0.99	0.95	0.11%	4.61	4.58	0.99	0.95	-0.05%	4.41	4.58	1.04	0.96	0.07%	4.71	4.58	0.97	0.94
	$\hat{\tau}_{\text{nspline}}$	0.17%	12.58	12.03	0.96	0.94	0.20%	12.13	12.04	0.99	0.95	0.21%	11.90	12.03	1.01	0.95	0.43%	12.09	12.03	0.99	0.94
	$\hat{\tau}_{\text{rf}}$	-0.05%	6.86	6.70	0.98	0.94	0.02%	6.85	6.69	0.98	0.95	0.04%	6.52	6.68	1.02	0.95	0.25%	6.88	6.69	0.97	0.94
	$\hat{\tau}_{\text{ms}}$	0.06%	4.62	4.59	0.99	0.95	0.10%	4.62	4.58	0.99	0.95	-0.05%	4.41	4.57	1.04	0.96	0.07%	4.71	4.58	0.97	0.94
2	$\hat{\tau}_{\text{ma}}$	-0.07%	2.52	2.41	0.96	0.94	0.16%	2.38	2.40	1.01	0.96	-0.06%	2.49	2.40	0.96	0.94	0.23%	2.44	2.40	0.99	0.94
	$\hat{\tau}_{\text{linear}}$	-0.10%	3.20	3.15	0.98	0.95	0.06%	3.06	3.15	1.03	0.95	-0.03%	3.23	3.15	0.98	0.94	0.19%	3.23	3.15	0.97	0.94
	$\hat{\tau}_{\text{kernel}}$	-0.09%	2.48	2.40	0.97	0.94	0.14%	2.34	2.39	1.02	0.96	-0.09%	2.45	2.39	0.97	0.94	0.20%	2.39	2.39	1.00	0.94
	$\hat{\tau}_{\text{nspline}}$	-0.09%	2.52	2.45	0.97	0.94	0.12%	2.40	2.44	1.02	0.96	-0.09%	2.51	2.44	0.97	0.95	0.19%	2.44	2.44	1.00	0.94
	$\hat{\tau}_{\text{rf}}$	-0.12%	2.59	2.50	0.96	0.95	0.18%	2.46	2.49	1.01	0.96	-0.07%	2.58	2.49	0.97	0.94	0.24%	2.53	2.49	0.98	0.94
	$\hat{\tau}_{\text{ms}}$	-0.08%	2.48	2.40	0.96	0.94	0.14%	2.34	2.39	1.02	0.96	-0.08%	2.47	2.39	0.97	0.94	0.19%	2.39	2.39	1.00	0.94
3	$\hat{\tau}_{\text{ma}}$	-0.43%	76.11	72.31	0.95	0.94	-0.35%	73.76	72.14	0.98	0.94	-0.52%	72.19	72.21	1.00	0.93	-0.57%	70.90	71.76	1.01	0.95
	$\hat{\tau}_{\text{linear}}$	0.18%	84.54	84.37	1.00	0.94	0.18%	83.15	84.04	1.01	0.96	0.27%	82.42	83.88	1.02	0.95	0.13%	82.60	83.97	1.02	0.95
	$\hat{\tau}_{\text{kernel}}$	3.02%	414.09	144.42	0.35	0.94	-22.07%	4346.41	295.96	0.07	0.96	-38.08%	6331.68	333.93	0.05	0.96	-0.91%	285.57	130.53	0.46	0.95
	$\hat{\tau}_{\text{nspline}}$	1.09%	94.40	93.17	0.99	0.94	0.22%	95.00	92.68	0.98	0.94	0.14%	93.30	92.62	0.99	0.93	-0.40%	88.81	92.66	1.04	0.95
	$\hat{\tau}_{\text{rf}}$	0.52%	78.71	78.26	0.99	0.94	0.20%	78.62	77.89	0.99	0.95	0.33%	77.89	77.86	1.00	0.94	-0.11%	75.75	77.92	1.03	0.95
	$\hat{\tau}_{\text{ms}}$	-0.49%	77.25	74.36	0.96	0.93	-0.19%	76.32	74.55	0.98	0.94	-0.10%	76.11	74.46	0.98	0.94	-0.48%	73.80	74.36	1.01	0.94
4	$\hat{\tau}_{\text{ma}}$	-0.44%	1.49	1.47	0.98	0.94	-0.33%	1.57	1.46	0.93	0.93	-0.46%	1.50	1.47	0.98	0.94	-0.28%	1.51	1.48	0.98	0.94
	$\hat{\tau}_{\text{linear}}$	0.01%	1.78	1.70	0.96	0.93	-0.23%	1.81	1.71	0.94	0.93	-0.29%	1.72	1.71	0.99	0.94	0.00%	1.72	1.72	1.00	0.94
	$\hat{\tau}_{\text{kernel}}$	-0.17%	4.73	2.53	0.54	0.94	1.59%	21.21	3.08	0.15	0.94	2.35%	25.51	3.92	0.15	0.94	-0.80%	8.42	2.81	0.33	0.94
	$\hat{\tau}_{\text{nspline}}$	-0.14%	1.49	1.50	1.01	0.95	0.00%	1.59	1.50	0.95	0.93	-0.19%	1.52	1.50	0.99	0.94	0.04%	1.54	1.51	0.98	0.93
	$\hat{\tau}_{\text{rf}}$	0.00%	1.79	1.72	0.96	0.93	-0.30%	1.81	1.71	0.95	0.93	-0.24%	1.75	1.72	0.98	0.94	0.09%	1.73	1.73	1.00	0.94
	$\hat{\tau}_{\text{ms}}$	-0.35%	1.53	1.50	0.98	0.94	-0.18%	1.61	1.50	0.93	0.93	-0.41%	1.55	1.50	0.97	0.93	-0.13%	1.56	1.51	0.96	0.93

Abbreviations: RB, relative bias; SD, standard deviation; SE, standard error; ER, empirical ratio; CP, coverage probability; Rand.: randomization; ma: model averaging; nspline: natural spline; rf: random forest; ms: model selection.

Table S6: Monte Carlo results for each estimator under different randomization methods with $K = 5$ and $n = 600$.

Model	Estimator	Complete Rand.					Stratified Block Rand.					Minimization					Hu & Hu's Rand.				
		RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP
1	$\hat{\tau}_{\mathrm{ma}}$	-0.01%	3.82	3.76	0.98	0.95	-0.04%	3.86	3.76	0.97	0.93	0.03%	3.79	3.75	0.99	0.94	0.04%	3.77	3.76	1.00	0.94
	$\hat{\tau}_{\mathrm{linear}}$	-0.02%	3.80	3.75	0.99	0.95	-0.04%	3.85	3.75	0.97	0.93	0.03%	3.77	3.75	0.99	0.94	0.05%	3.74	3.75	1.00	0.94
	$\hat{\tau}_{\mathrm{kernel}}$	-0.02%	3.80	3.75	0.99	0.95	-0.04%	3.85	3.75	0.97	0.93	0.03%	3.77	3.75	0.99	0.94	0.05%	3.74	3.75	1.00	0.94
	$\hat{\tau}_{\mathrm{nspline}}$	0.17%	9.74	9.78	1.00	0.95	-0.30%	9.87	9.79	0.99	0.95	0.08%	9.65	9.78	1.01	0.95	0.33%	9.93	9.79	0.99	0.95
	$\hat{\tau}_{\mathrm{rf}}$	0.10%	5.40	5.37	0.99	0.95	-0.06%	5.37	5.37	1.00	0.95	0.07%	5.25	5.36	1.02	0.95	0.12%	5.42	5.37	0.99	0.95
	$\hat{\tau}_{\mathrm{ms}}$	-0.02%	3.80	3.75	0.99	0.95	-0.04%	3.85	3.75	0.97	0.93	0.03%	3.77	3.75	0.99	0.94	0.05%	3.74	3.75	1.00	0.94
2	$\hat{\tau}_{\mathrm{ma}}$	0.03%	1.95	1.96	1.01	0.96	-0.08%	1.96	1.96	1.00	0.96	-0.08%	2.05	1.96	0.96	0.93	0.08%	1.95	1.96	1.01	0.95
	$\hat{\tau}_{\mathrm{linear}}$	-0.02%	2.61	2.57	0.99	0.95	-0.01%	2.56	2.58	1.01	0.95	-0.11%	2.70	2.58	0.95	0.93	0.05%	2.55	2.58	1.01	0.95
	$\hat{\tau}_{\mathrm{kernel}}$	0.02%	1.92	1.96	1.02	0.96	-0.08%	1.95	1.96	1.01	0.95	-0.12%	2.03	1.96	0.96	0.93	0.07%	1.93	1.96	1.01	0.96
	$\hat{\tau}_{\mathrm{nspline}}$	0.02%	1.98	1.99	1.01	0.96	-0.10%	1.98	1.99	1.01	0.96	-0.09%	2.09	1.99	0.95	0.93	0.07%	1.97	1.99	1.01	0.95
	$\hat{\tau}_{\mathrm{rf}}$	0.00%	2.01	2.01	1.00	0.96	-0.08%	2.00	2.01	1.01	0.96	-0.11%	2.12	2.01	0.95	0.93	0.05%	2.00	2.01	1.01	0.95
	$\hat{\tau}_{\mathrm{ms}}$	0.02%	1.92	1.96	1.02	0.96	-0.08%	1.94	1.96	1.01	0.95	-0.12%	2.03	1.96	0.96	0.93	0.07%	1.94	1.96	1.01	0.96
3	$\hat{\tau}_{\mathrm{ma}}$	-0.27%	56.67	57.84	1.02	0.95	-0.62%	58.01	57.59	0.99	0.94	-0.75%	58.63	57.76	0.99	0.93	-0.67%	56.26	57.55	1.02	0.96
	$\hat{\tau}_{\mathrm{linear}}$	0.29%	67.23	68.96	1.03	0.95	-0.07%	69.42	68.53	0.99	0.94	0.01%	69.94	68.60	0.98	0.93	-0.45%	67.51	68.47	1.01	0.95
	$\hat{\tau}_{\mathrm{kernel}}$	1.66%	229.70	87.09	0.38	0.96	1.53%	317.79	98.31	0.31	0.94	-13.80%	1854.03	161.71	0.09	0.94	-4.01%	525.68	111.47	0.21	0.96
	$\hat{\tau}_{\mathrm{nspline}}$	0.36%	74.61	75.90	1.02	0.94	0.17%	76.88	75.45	0.98	0.94	-0.27%	77.37	75.43	0.97	0.94	-0.66%	73.87	75.28	1.02	0.94
	$\hat{\tau}_{\mathrm{rf}}$	0.13%	60.92	62.51	1.03	0.95	-0.05%	62.48	62.10	0.99	0.94	-0.18%	63.02	62.14	0.99	0.94	-0.43%	60.72	62.15	1.02	0.95
	$\hat{\tau}_{\mathrm{ms}}$	-0.26%	57.95	59.14	1.02	0.95	-0.51%	59.26	58.72	0.99	0.93	-0.57%	60.51	58.82	0.97	0.93	-0.65%	58.21	58.81	1.01	0.95
4	$\hat{\tau}_{\mathrm{ma}}$	-0.19%	1.23	1.20	0.98	0.94	-0.23%	1.21	1.20	0.99	0.94	-0.24%	1.24	1.20	0.97	0.94	-0.13%	1.19	1.20	1.01	0.94
	$\hat{\tau}_{\mathrm{linear}}$	-0.02%	1.38	1.41	1.02	0.96	0.03%	1.42	1.40	0.99	0.93	-0.04%	1.47	1.41	0.96	0.94	0.09%	1.37	1.40	1.03	0.95
	$\hat{\tau}_{\mathrm{kernel}}$	7.43%	75.01	4.45	0.06	0.95	0.57%	4.31	2.19	0.51	0.95	1.00%	7.84	2.37	0.30	0.96	-0.39%	5.55	2.31	0.42	0.95
	$\hat{\tau}_{\mathrm{nspline}}$	0.03%	1.24	1.23	0.99	0.94	-0.04%	1.23	1.23	1.00	0.94	0.00%	1.26	1.23	0.98	0.94	0.12%	1.23	1.23	1.00	0.95
	$\hat{\tau}_{\mathrm{rf}}$	0.05%	1.40	1.41	1.01	0.95	-0.07%	1.42	1.40	0.99	0.94	-0.03%	1.46	1.41	0.96	0.93	0.15%	1.38	1.41	1.02	0.94
	$\hat{\tau}_{\mathrm{ms}}$	-0.03%	1.26	1.23	0.98	0.94	-0.11%	1.24	1.23	0.99	0.94	-0.06%	1.27	1.23	0.96	0.94	0.08%	1.23	1.23	1.00	0.95

Abbreviations: RB, relative bias; SD, standard deviation; SE, standard error; ER, empirical ratio; CP, coverage probability; Rand.: randomization; ma: model averaging; nspline: natural spline; rf: random forest; ms: model selection.

Table S7: Monte Carlo results for each estimator under different randomization methods with $K = 5$ and $n = 800$.

Model	Estimator	Complete Rand.					Stratified Block Rand.					Minimization					Hu & Hu's Rand.				
		RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP
1	$\hat{\tau}_{\text{ma}}$	0.00%	3.25	3.25	1.00	0.95	0.06%	3.25	3.26	1.00	0.95	0.03%	3.20	3.26	1.02	0.95	0.10%	3.36	3.25	0.97	0.95
	$\hat{\tau}_{\text{linear}}$	0.00%	3.23	3.25	1.00	0.95	0.06%	3.25	3.25	1.00	0.95	0.04%	3.19	3.25	1.02	0.96	0.10%	3.35	3.25	0.97	0.95
	$\hat{\tau}_{\text{kernel}}$	0.00%	3.23	3.25	1.00	0.95	0.06%	3.25	3.25	1.00	0.95	0.04%	3.19	3.25	1.02	0.96	0.10%	3.35	3.25	0.97	0.95
	$\hat{\tau}_{\text{nspline}}$	-0.02%	8.46	8.46	1.00	0.94	0.27%	8.23	8.45	1.03	0.96	-0.30%	8.43	8.46	1.00	0.95	0.16%	8.42	8.45	1.00	0.95
	$\hat{\tau}_{\text{rf}}$	0.04%	4.61	4.60	1.00	0.96	0.14%	4.44	4.60	1.04	0.96	-0.15%	4.58	4.60	1.00	0.95	0.13%	4.73	4.60	0.97	0.95
	$\hat{\tau}_{\text{ms}}$	0.00%	3.24	3.25	1.00	0.95	0.06%	3.25	3.25	1.00	0.95	0.04%	3.19	3.25	1.02	0.96	0.10%	3.35	3.25	0.97	0.95
2	$\hat{\tau}_{\text{ma}}$	-0.03%	1.72	1.70	0.99	0.95	-0.17%	1.76	1.70	0.96	0.94	-0.06%	1.71	1.70	1.00	0.95	0.13%	1.74	1.70	0.98	0.94
	$\hat{\tau}_{\text{linear}}$	0.03%	2.20	2.23	1.01	0.95	-0.23%	2.31	2.23	0.97	0.94	-0.06%	2.27	2.23	0.98	0.95	0.18%	2.25	2.23	0.99	0.95
	$\hat{\tau}_{\text{kernel}}$	-0.05%	1.70	1.70	0.99	0.95	-0.16%	1.75	1.70	0.97	0.95	-0.07%	1.69	1.69	1.00	0.95	0.11%	1.72	1.70	0.99	0.94
	$\hat{\tau}_{\text{nspline}}$	-0.03%	1.74	1.72	0.99	0.95	-0.18%	1.78	1.73	0.97	0.95	-0.07%	1.70	1.72	1.01	0.96	0.11%	1.74	1.72	0.99	0.95
	$\hat{\tau}_{\text{rf}}$	-0.02%	1.75	1.73	0.99	0.95	-0.19%	1.79	1.73	0.97	0.95	-0.07%	1.73	1.73	1.00	0.96	0.09%	1.76	1.73	0.99	0.94
	$\hat{\tau}_{\text{ms}}$	-0.05%	1.70	1.69	0.99	0.95	-0.17%	1.75	1.70	0.97	0.95	-0.07%	1.69	1.69	1.00	0.95	0.12%	1.72	1.70	0.99	0.94
3	$\hat{\tau}_{\text{ma}}$	-0.57%	49.94	49.51	0.99	0.95	0.13%	50.07	49.75	0.99	0.94	-0.87%	50.01	49.25	0.98	0.94	-0.48%	51.08	49.44	0.97	0.94
	$\hat{\tau}_{\text{linear}}$	-0.22%	58.64	59.44	1.01	0.95	0.34%	59.84	59.78	1.00	0.95	-0.41%	59.09	59.21	1.00	0.94	-0.10%	60.36	59.46	0.99	0.94
	$\hat{\tau}_{\text{kernel}}$	0.30%	235.98	67.97	0.29	0.95	11.81%	1691.81	121.82	0.07	0.95	7.13%	1344.34	119.67	0.09	0.94	0.02%	128.33	61.39	0.48	0.95
	$\hat{\tau}_{\text{nspline}}$	-0.15%	65.43	65.26	1.00	0.94	0.35%	65.82	65.71	1.00	0.94	-0.45%	65.48	65.03	0.99	0.94	0.11%	66.12	65.35	0.99	0.95
	$\hat{\tau}_{\text{rf}}$	-0.14%	53.25	53.02	1.00	0.95	0.50%	53.15	53.33	1.00	0.95	-0.55%	53.61	52.78	0.98	0.94	-0.01%	54.04	53.05	0.98	0.94
	$\hat{\tau}_{\text{ms}}$	-0.43%	50.99	49.99	0.98	0.94	0.24%	50.41	50.27	1.00	0.95	-0.73%	50.41	49.78	0.99	0.94	-0.39%	51.88	50.01	0.96	0.94
4	$\hat{\tau}_{\text{ma}}$	-0.27%	1.09	1.04	0.96	0.93	-0.10%	1.07	1.04	0.97	0.95	-0.02%	1.03	1.04	1.01	0.95	-0.24%	1.05	1.03	0.98	0.94
	$\hat{\tau}_{\text{linear}}$	-0.14%	1.26	1.22	0.97	0.93	0.01%	1.24	1.21	0.98	0.94	0.11%	1.18	1.22	1.03	0.96	-0.05%	1.19	1.21	1.02	0.95
	$\hat{\tau}_{\text{kernel}}$	-0.18%	3.03	1.92	0.63	0.94	0.63%	5.59	2.06	0.37	0.95	1.30%	7.83	2.38	0.30	0.96	-0.32%	3.38	1.86	0.55	0.94
	$\hat{\tau}_{\text{nspline}}$	-0.12%	1.07	1.07	0.99	0.95	0.08%	1.07	1.07	1.00	0.95	0.21%	1.05	1.07	1.02	0.95	-0.07%	1.08	1.06	0.99	0.95
	$\hat{\tau}_{\text{rf}}$	-0.17%	1.26	1.21	0.97	0.94	0.01%	1.24	1.21	0.98	0.94	0.14%	1.19	1.22	1.03	0.95	-0.04%	1.20	1.21	1.01	0.95
	$\hat{\tau}_{\text{ms}}$	-0.13%	1.07	1.07	0.99	0.95	0.07%	1.07	1.07	1.00	0.95	0.18%	1.06	1.07	1.01	0.95	-0.13%	1.09	1.06	0.98	0.94

Abbreviations: RB, relative bias; SD, standard deviation; SE, standard error; ER, empirical ratio; CP, coverage probability; Rand.: randomization; ma: model averaging; nspline: natural spline; rf: random forest; ms: model selection.

Table S8: Monte Carlo results for each estimator under different randomization methods with $K = 5$ and $n = 1000$.

Model	Estimator	Complete Rand.					Stratified Block Rand.					Minimization					Hu & Hu's Rand.				
		RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP	RB	SD	SE	ER	CP
1	$\hat{\tau}_{\text{ma}}$	-0.04%	3.03	2.91	0.96	0.93	0.07%	2.82	2.91	1.03	0.96	0.02%	2.92	2.91	0.99	0.94	0.03%	2.78	2.91	1.05	0.96
	$\hat{\tau}_{\text{linear}}$	-0.04%	3.04	2.90	0.96	0.93	0.08%	2.82	2.91	1.03	0.96	0.01%	2.92	2.91	0.99	0.95	0.04%	2.77	2.91	1.05	0.96
	$\hat{\tau}_{\text{kernel}}$	-0.04%	3.03	2.90	0.96	0.93	0.08%	2.82	2.91	1.03	0.96	0.01%	2.92	2.91	0.99	0.95	0.04%	2.77	2.91	1.05	0.96
	$\hat{\tau}_{\text{nspline}}$	-0.05%	7.78	7.54	0.97	0.94	-0.06%	7.46	7.55	1.01	0.96	0.16%	7.51	7.55	1.01	0.95	0.04%	7.52	7.55	1.00	0.95
	$\hat{\tau}_{\text{rf}}$	0.00%	4.18	4.09	0.98	0.94	0.03%	4.07	4.09	1.00	0.95	0.04%	4.06	4.09	1.01	0.95	0.05%	3.94	4.09	1.04	0.95
	$\hat{\tau}_{\text{ms}}$	-0.04%	3.04	2.90	0.96	0.93	0.08%	2.82	2.91	1.03	0.96	0.01%	2.92	2.91	0.99	0.95	0.04%	2.77	2.91	1.05	0.96
2	$\hat{\tau}_{\text{ma}}$	0.01%	1.55	1.52	0.98	0.94	0.01%	1.54	1.52	0.99	0.95	-0.21%	1.54	1.52	0.99	0.94	-0.08%	1.53	1.52	0.99	0.94
	$\hat{\tau}_{\text{linear}}$	-0.08%	1.97	2.00	1.02	0.95	-0.01%	2.03	2.00	0.98	0.94	-0.18%	1.98	1.99	1.01	0.95	-0.12%	2.02	2.00	0.99	0.96
	$\hat{\tau}_{\text{kernel}}$	0.00%	1.53	1.52	0.99	0.94	0.01%	1.53	1.52	0.99	0.95	-0.22%	1.53	1.52	0.99	0.95	-0.08%	1.52	1.52	1.00	0.94
	$\hat{\tau}_{\text{nspline}}$	0.00%	1.58	1.54	0.98	0.95	-0.02%	1.57	1.54	0.98	0.95	-0.21%	1.56	1.54	0.99	0.94	-0.07%	1.53	1.54	1.01	0.95
	$\hat{\tau}_{\text{rf}}$	0.00%	1.56	1.55	0.99	0.95	0.00%	1.57	1.54	0.99	0.95	-0.20%	1.57	1.54	0.99	0.95	-0.10%	1.56	1.54	0.99	0.94
	$\hat{\tau}_{\text{ms}}$	0.00%	1.53	1.52	0.99	0.94	0.01%	1.53	1.52	0.99	0.95	-0.22%	1.53	1.52	0.99	0.95	-0.08%	1.52	1.52	1.00	0.94
3	$\hat{\tau}_{\text{ma}}$	-0.33%	43.59	43.78	1.00	0.95	-0.70%	43.14	43.84	1.02	0.95	-0.33%	44.86	44.00	0.98	0.95	0.11%	44.39	43.98	0.99	0.94
	$\hat{\tau}_{\text{linear}}$	0.33%	53.89	53.21	0.99	0.95	-0.16%	52.77	53.10	1.01	0.95	-0.13%	53.57	53.24	0.99	0.95	0.58%	53.33	53.36	1.00	0.94
	$\hat{\tau}_{\text{kernel}}$	0.53%	97.23	53.06	0.55	0.95	0.60%	132.77	52.90	0.40	0.96	-0.40%	90.93	50.67	0.56	0.95	0.39%	48.52	46.80	0.96	0.95
	$\hat{\tau}_{\text{nspline}}$	0.26%	58.25	58.45	1.00	0.95	-0.46%	59.26	58.29	0.98	0.94	-0.18%	58.48	58.43	1.00	0.95	0.44%	59.41	58.62	0.99	0.95
	$\hat{\tau}_{\text{rf}}$	0.13%	45.93	46.81	1.02	0.95	-0.34%	45.99	46.80	1.02	0.95	-0.06%	47.31	46.89	0.99	0.96	0.46%	47.04	46.97	1.00	0.95
	$\hat{\tau}_{\text{ms}}$	-0.19%	43.68	44.13	1.01	0.95	-0.63%	43.60	44.17	1.01	0.95	-0.16%	45.19	44.19	0.98	0.95	0.17%	45.39	44.25	0.97	0.94
4	$\hat{\tau}_{\text{ma}}$	-0.16%	0.93	0.93	1.00	0.95	-0.13%	0.92	0.94	1.02	0.96	-0.05%	0.94	0.93	0.98	0.95	-0.19%	0.96	0.93	0.97	0.94
	$\hat{\tau}_{\text{linear}}$	0.03%	1.07	1.09	1.02	0.95	0.02%	1.07	1.09	1.02	0.95	0.13%	1.12	1.09	0.97	0.94	-0.05%	1.12	1.09	0.97	0.94
	$\hat{\tau}_{\text{kernel}}$	-2.14%	23.23	2.72	0.12	0.94	0.87%	5.69	2.11	0.37	0.96	-0.13%	4.14	1.95	0.47	0.94	0.89%	5.65	1.88	0.33	0.94
	$\hat{\tau}_{\text{nspline}}$	-0.02%	0.95	0.96	1.01	0.95	0.02%	0.92	0.95	1.04	0.96	0.13%	0.96	0.95	1.00	0.96	-0.03%	0.99	0.95	0.96	0.94
	$\hat{\tau}_{\text{rf}}$	0.00%	1.07	1.09	1.02	0.95	0.04%	1.06	1.09	1.02	0.94	0.18%	1.12	1.09	0.97	0.94	-0.06%	1.11	1.08	0.98	0.94
	$\hat{\tau}_{\text{ms}}$	-0.03%	0.95	0.96	1.00	0.95	0.00%	0.92	0.95	1.04	0.96	0.11%	0.97	0.95	0.99	0.96	-0.04%	0.99	0.95	0.96	0.94

Abbreviations: RB, relative bias; SD, standard deviation; SE, standard error; ER, empirical ratio; CP, coverage probability; Rand.: randomization; ma: model averaging; nspline: natural spline; rf: random forest; ms: model selection.

S5.3 Summaries of the selected weights

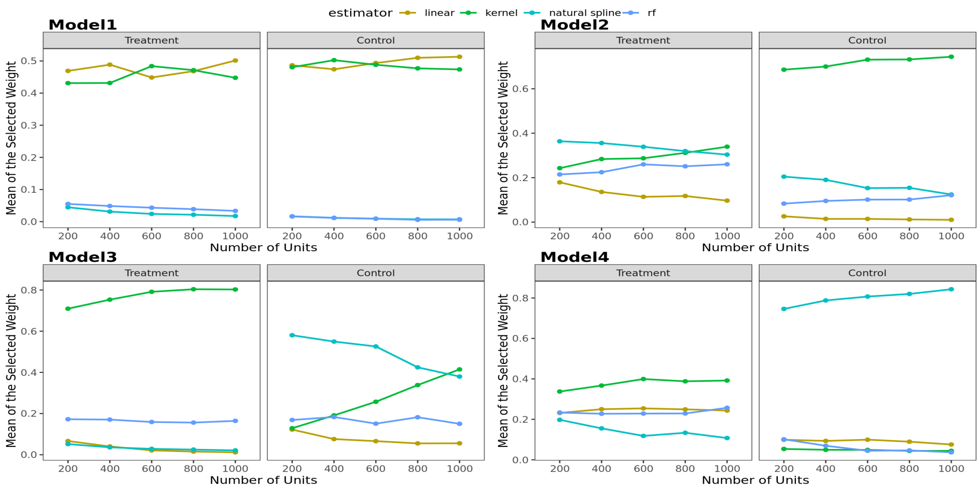

Figure S1: Mean of the selected weights for the adjustment methods under different models.

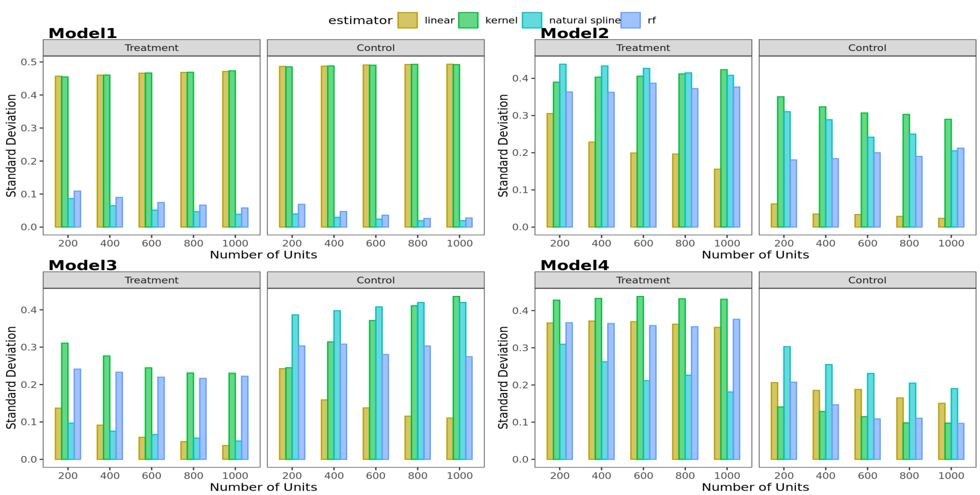

Figure S2: Standard deviation of the selected weights for the adjustment methods under different models.

Figures S1 and S2 summarize the mean selected weights and their standard deviations across simulation scenarios. The results show that the proposed cross-validation procedure assigns weights adaptively according to the underlying outcome structure, and the selected weights are consistent with the patterns in the dominance counts reported in the main text.

The standard deviations of the selected weights provide additional evidence on the stability of the weighting procedure. In settings where one or two methods clearly dominate, such as the control group of Model 4, the corresponding weights tend to be relatively stable. In contrast, when several candidate models have comparable predictive performance, the selected weights show larger variability, as seen in the treatment group of Model 2 and the control group of Model 3. Overall, these patterns indicate that the model averaging procedure does not simply select a fixed adjustment method across all scenarios, but instead adapts to treatment-arm-specific and model-specific covariate-outcome relationships.

Table S9: Summaries of the selected weights under Model 1 with 1000 replications.

Number of Units	Group	Linear		Kernel		Spline		RF	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
200	treatment	0.47	0.46	0.43	0.45	0.04	0.09	0.06	0.11
	control	0.49	0.49	0.48	0.49	0.02	0.04	0.02	0.07
400	treatment	0.49	0.46	0.43	0.46	0.03	0.06	0.05	0.09
	control	0.47	0.49	0.50	0.49	0.01	0.03	0.01	0.05
600	treatment	0.45	0.47	0.48	0.47	0.02	0.05	0.04	0.07
	control	0.49	0.49	0.49	0.49	0.01	0.02	0.01	0.04
800	treatment	0.47	0.47	0.47	0.47	0.02	0.05	0.04	0.07
	control	0.51	0.49	0.48	0.49	0.01	0.02	0.01	0.03
1000	treatment	0.50	0.47	0.45	0.47	0.02	0.04	0.03	0.06
	control	0.51	0.49	0.47	0.49	0.01	0.02	0.01	0.03

Table S10: Summaries of the selected weights under Model 2 with 1000 replications.

Number of Units	Group	Linear		Kernel		Spline		RF	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
200	treatment	0.18	0.31	0.24	0.39	0.36	0.44	0.21	0.36
	control	0.03	0.06	0.69	0.35	0.20	0.31	0.08	0.18
400	treatment	0.14	0.23	0.28	0.40	0.36	0.43	0.22	0.36
	control	0.01	0.04	0.70	0.32	0.19	0.29	0.10	0.18
600	treatment	0.11	0.20	0.29	0.41	0.34	0.43	0.26	0.39
	control	0.01	0.03	0.73	0.31	0.15	0.24	0.10	0.20
800	treatment	0.12	0.20	0.31	0.41	0.32	0.41	0.25	0.37
	control	0.01	0.03	0.73	0.30	0.15	0.25	0.10	0.19
1000	treatment	0.10	0.16	0.34	0.42	0.30	0.41	0.26	0.38
	control	0.01	0.02	0.74	0.29	0.12	0.21	0.12	0.21

Table S11: Summaries of the selected weights under Model 3 with 1000 replications.

Number of Units	Group	Linear		Kernel		Spline		RF	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
200	treatment	0.07	0.14	0.71	0.31	0.05	0.10	0.17	0.24
	control	0.12	0.24	0.13	0.25	0.58	0.39	0.17	0.30
400	treatment	0.04	0.09	0.75	0.28	0.04	0.08	0.17	0.23
	control	0.08	0.16	0.19	0.31	0.55	0.40	0.18	0.31
600	treatment	0.02	0.06	0.79	0.24	0.03	0.07	0.16	0.22
	control	0.07	0.14	0.26	0.37	0.53	0.41	0.15	0.28
800	treatment	0.02	0.05	0.80	0.23	0.02	0.06	0.16	0.22
	control	0.06	0.12	0.34	0.41	0.42	0.42	0.18	0.30
1000	treatment	0.01	0.04	0.80	0.23	0.02	0.05	0.16	0.22
	control	0.06	0.11	0.41	0.44	0.38	0.42	0.15	0.27

Table S12: Summaries of the selected weights under Model 4 with 1000 replications.

Number of Units	Group	Linear		Kernel		Spline		RF	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
200	treatment	0.23	0.37	0.34	0.43	0.20	0.31	0.23	0.37
	control	0.10	0.21	0.05	0.14	0.75	0.30	0.10	0.21
400	treatment	0.25	0.37	0.37	0.43	0.16	0.26	0.23	0.37
	control	0.09	0.19	0.05	0.13	0.79	0.25	0.07	0.15
600	treatment	0.25	0.37	0.40	0.44	0.12	0.21	0.23	0.36
	control	0.10	0.19	0.05	0.11	0.81	0.23	0.04	0.11
800	treatment	0.25	0.36	0.39	0.43	0.13	0.23	0.23	0.36
	control	0.09	0.17	0.04	0.10	0.82	0.21	0.05	0.11
1000	treatment	0.24	0.35	0.39	0.43	0.11	0.18	0.26	0.38
	control	0.08	0.15	0.04	0.10	0.84	0.19	0.04	0.10

S5.4 Dominance counts for the adjustment methods

Table S13: Dominance counts for each method under Model 1 with 1000 replications.

Number of Units	Group	Linear	Kernel	Spline	RF
200	treatment	515	476	0	9
	control	499	495	0	6
400	treatment	532	466	0	2
	control	485	515	0	0
600	treatment	483	517	0	0
	control	502	498	0	0
800	treatment	498	502	0	0
	control	518	482	0	0
1000	treatment	528	472	0	0
	control	518	482	0	0

Table S14: Dominance counts for each method under Model 2 with 1000 replications.

Number of Units	Group	Linear	Kernel	Spline	RF
200	treatment	168	243	378	211
	control	0	734	202	64
400	treatment	97	296	371	236
	control	0	749	181	70
600	treatment	64	305	351	280
	control	0	795	125	80
800	treatment	65	326	344	265
	control	0	793	135	72
1000	treatment	32	354	335	279
	control	0	822	83	95

Table S15: Dominance counts for each method under Model 3 with 1000 replications.

Number of Units	Group	Linear	Kernel	Spline	RF
200	treatment	44	794	8	154
	control	106	98	624	172
400	treatment	15	831	5	149
	control	41	175	600	184
600	treatment	4	883	1	112
	control	33	259	570	138
800	treatment	1	893	0	106
	control	19	339	457	185
1000	treatment	0	887	0	113
	control	10	432	412	146

Table S16: Dominance counts of each method under Model 4 with 1000 replications.

Number of Units	Group	Linear	Kernel	Spline	RF
200	treatment	233	340	186	241
	control	81	36	793	90
400	treatment	262	384	125	229
	control	78	22	859	41
600	treatment	272	416	78	234
	control	70	15	892	23
800	treatment	257	402	92	249
	control	44	5	935	16
1000	treatment	263	414	57	266
	control	40	10	940	10

S6 Detailed Description of Clinical Trial Example

Table S17: Description of selected covariates in the Nefazodone CBASP trial.

Variable	Description
Mstatus	Marital status, 1 if married or living with someone and 0 otherwise
TreatPD	Previous treatment for depression, 1 yes and 0 no
AGE	Age of patients in years
HAMA_SOMATI	HAMA Somatic Anxiety Score
HAMD_COGNID	HAMD cognitive disturbance score
HAMD17	Total HAMD-17 score
HAMD24	Total HAMD-24 score
NDE	Number of depressive episodes

Abbreviation: HAMA, Hamilton Anxiety Scale; HAMD, Hamilton Rating Scale for Depression.

Table S18: Covariates used for treatment effect adjustment in each model.

Adjusting Model	Used Covariates
Linear	Mstatus, TreatPD, AGE, NDE
Kernel	Mstatus, AGE
Natural Spline	Mstatus, AGE, TreatPD
Random Forest	AGE, NDE

Bibliography

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association 113*(524), 1784–1796.

Gao, Y., X. Zhang, S. Wang, T. T. L. Chong, and G. Zou (2019). Frequentist model averaging for threshold models. *Annals of the Institute of Statistical Mathematics 71*(2), 275–306.

Liu, H., F. Tu, and W. Ma (2023). Lasso-adjusted treatment effect estimation under covariate-adaptive randomization. *Biometrika 110*(2), 431–447.

Tu, F., W. Ma, and H. Liu (2024). A unified framework for covariate adjustment under stratified randomisation. *Stat 13*(4), e70016.