SHARP BOUNDS FOR VARIANCE OF TREATMENT EFFECT ESTIMATORS IN THE PRESENCE OF COVARIATES

Ruoyu Wang[1], Qihua Wang[*1,2], Wang Miao[2] and Xiaohua Zhou[3]

[1] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

[2] *University of Chinese Academy of Sciences*

[3] *Peking University*

Supplementary Material

The supplementary material is organized as follows. In Section S1, we provide the proof of Theorem 1. We prove Theorem 2 in Section S2. The proofs of Theorem 3 , 4, 5 and 6 are provided in Section S3, S4, S5 and S6, respectively. Proof of Theorem 7 is similar to that of Theorem 3 and hence is omitted. Section S7 contains further simulation results on the performance of the confidence intervals when the proposed lower bounds are attained.

∗ Corresponding author

S1 Proof of Theorem 1

Proof. Note that

$$\phi^2(\tau) = \sum_{i=1}^{K} \pi_k \frac{1}{N_k} \sum_{w_i=\xi_k} (y_{1i} - y_{0i})^2 - \mu^2(\tau),$$

where $N_k = N\pi_k$. Letting $a_k(s) = F^{-1}_{1|k}(s/N_k)$ and $b_k(s) = F^{-1}_{0|k}(s/N_k)$ for $k = 1, \ldots, K$ and $s = 1, \ldots, N_k$, then we have

$$\frac{1}{N_k} \sum_{w_i=\xi_k} (y_{1i} - y_{0i})^2 = \frac{1}{N_k} \sum_{i=1}^{N_k} a_k(i)^2 + \frac{1}{N_k} \sum_{i=1}^{N_k} b_k(i)^2 - \frac{2}{N_k} \sum_{i=1}^{N_k} a_k(i) b_k(\Pi_k(i))$$

where Π_k is a permutation on $\{1, \ldots, N_k\}$. By the rearrangement inequality, we have

$$\sum_{i=1}^{N_k} a_k(i) b_k(N_k - k + 1) \leq \sum_{i=1}^{N_k} a_k(i) b_k(\Pi_k(i)) \leq \sum_{i=1}^{N_k} a_k(i) b_k(i).$$

Thus

$$\int_0^1 (F^{-1}_{1|k}(u) - F^{-1}_{0|k}(u))^2 du = \frac{1}{N_k} \sum_{i=1}^{N_k} (a_k(i) - b_k(i))^2 \leq \frac{1}{N_k} \sum_{w_i=\xi_k} (y_{1i} - y_{0i})^2$$
$$\leq \frac{1}{N_k} \sum_{i=1}^{N_k} (a_k(i) - b_k(N_k + 1 - i))^2 = \int_0^1 (F^{-1}_{1|k}(u) - F^{-1}_{0|k}(1 - u))^2 du$$

This proves the bound in Theorem 1.

Next, we prove the sharpness of the bound. Let $\mathcal{U} = \{\mathbf{U}^* = (y_1^*, y_0^*, w^*) : P(w^* = \xi_k) = \pi_k$ and $P(y_t^* \leq y \mid w^* = \xi_k) = F_{t|k}(y)$ for $t = 0, 1$ and $k = 1, \ldots, K\}$ be the set of all populations whose covariate distribution and distributions of each potential outcome conditional on the covariate are all

identical to those of $\mathbf{U}$. We first prove that the established bound can be attained by some population among $\mathcal{U}$.

For $k = 1, \ldots, K$, let $i_{k(1)} < \cdots < i_{k(N_k)}$ be the indices in $\mathcal{I}_k = \{i : w_i = \xi_k\}$ in increasing order. Define the population $\mathbf{U}^{\mathrm{L}}$ consisting of N units with two potential outcomes y^{L}_{1i} and y^{L}_{0i} and a vector of covariates w^{L}_i associated with unit i for $i = 1, \ldots, N$. Let $y^{\mathrm{L}}_{1i_{k(j)}} = a_k(j)$, $y^{\mathrm{L}}_{0i_{k(j)}} = b_k(j)$ and $w^{\mathrm{L}}_{i_{k(j)}} = \xi_k$ for $k = 1 \ldots, K$ and $j = 1, \ldots, N_k$. Let $\tau^{\mathrm{L}}_i = y^{\mathrm{L}}_{1i} - y^{\mathrm{L}}_{0i}$ for $i = 1, \ldots, N$. Then $P(y^{\mathrm{L}}_t \leq y \mid w^{\mathrm{L}} = \xi_k) = F_{t|k}(y)$ and $P(w^{\mathrm{L}} = \xi_k) = \pi_k$ for $t = 0, 1$, $k = 1, \ldots, K$ and $y \in \mathbf{R}$ and $1/N_k \sum_{i \in \mathcal{I}_k} (y^{\mathrm{L}}_{1i} - y^{\mathrm{L}}_{0i})^2 = 1/N_k \sum_{j=1}^{N_k} (a_k(j) - b_k(j))^2 = \int_0^1 (F^{-1}_{1|k}(u) - F^{-1}_{0|k}(u))^2 du$ for $k = 1, \ldots, K$. Thus $\phi^2(\tau^{\mathrm{L}}) = \phi^2_{\mathrm{L}}$, which attains the lower bound for $\phi^2(\tau)$.

Define $\mathbf{U}^{\mathrm{H}}$ similarly with $y^{\mathrm{H}}_{1i_{k(j)}} = a_k(j)$, $y^{\mathrm{H}}_{1i_{k(j)}} = b_k(N_k + 1 - j)$ and $w^{\mathrm{H}}_{i_{k(j)}} = \xi_k$ for $k = 1 \ldots, K$ and $j = 1, \ldots, N_k$. Let $\tau^{\mathrm{H}}_i = y^{\mathrm{H}}_{1i} - y^{\mathrm{H}}_{0i}$, then $P(y^{\mathrm{H}}_t \leq y \mid w^{\mathrm{H}} = \xi_k) = F_{t|k}(y)$ and $P(w^{\mathrm{H}} = \xi_k) = \pi_k$ for $t = 0, 1$, $k = 1, \ldots, K$ and $y \in \mathbf{R}$. Moreover, $1/N_k \sum_{i \in \mathcal{I}_k} (y^{\mathrm{H}}_{1i} - y^{\mathrm{H}}_{0i})^2 = 1/N_k \sum_{j=1}^{N_k} (a_k(j) - b_k(N_k + 1 - j))^2 = \int_0^1 (F^{-1}_{1|k}(u) - F^{-1}_{0|k}(1 - u))^2 du$ for $k = 1, \ldots, K$. Thus $\phi^2(\tau^{\mathrm{H}}) = \phi^2_{\mathrm{H}}$, which attains the upper bound for $\phi^2(\tau)$.

Next, we show the sharpness of the bound. We show the result for the lower bound, i.e., ϕ^2_{L} is no smaller than any bound in $\mathcal{B}_{\mathrm{L}}$, and the result for the upper bound follows similarly. For any bound b_{L} in $\mathcal{B}_{\mathrm{L}}$, we

have $b_{\mathrm{L}} = f(\pi_1, \ldots, \pi_K, F_{0|1}, \ldots, F_{0|K}, F_{1|1}, \ldots, F_{1|K})$ for some functional f according to the definition of $\mathcal{B}_{\mathrm{L}}$. For any $\mathbf{U}^*$ in $\mathcal{U}$, define $\pi_k^* = P(w^* = \xi_k)$ and $F_{t|k}^*(y) = P(y_t^* \leq y \mid w = \xi_k)$ for $t = 0, 1$ and $k = 1, \ldots, K$. Let $\tau^* = (y_{11}^* - y_{01}^*, \ldots, y_{1N}^* - y_{0N}^*)$. By the applying the bound to the population $\mathbf{U}^*$, we have

$$f(\pi_1^*, \ldots, \pi_K^*, F_{0|1}^*, \ldots, F_{0|K}^*, F_{1|1}^*, \ldots, F_{1|K}^*) \leq \phi^2(\tau^*). \tag{S1.1}$$

Because $\mathbf{U}^* \in \mathcal{U}$, it holds that $\pi_k^* = \pi_k$ and $F_{t|k}^*(y) = F_{t|k}(y)$ for $t = 0, 1$. Thus (S1.1) implies that

$$\begin{aligned} b_{\mathrm{L}} &= f(\pi_1, \ldots, \pi_K, F_{0|1}, \ldots, F_{0|K}, F_{1|1}, \ldots, F_{1|K}) \\ &= f(\pi_1^*, \ldots, \pi_K^*, F_{0|1}^*, \ldots, F_{0|K}^*, F_{1|1}^*, \ldots, F_{1|K}^*) \leq \phi^2(\tau^*) \end{aligned} \tag{S1.2}$$

for any $\mathbf{U}^* \in \mathcal{U}$. We have shown that there is some $\mathbf{U}^* \in \mathcal{U}$ such that $\phi_{\mathrm{L}}^2 = \phi^2(\tau^*)$. Hence (S1.2) implies $b_{\mathrm{L}} \leq \phi_{\mathrm{L}}^2$ for any $b_{\mathrm{L}} \in \mathcal{B}_{\mathrm{L}}$, which completes the proof. □

S2 Proof of Theorem 2

Proof. Throughout this and the following proofs, for any real number a and b, we let $a \wedge b = \min\{a, b\}$, $a \vee b = \min\{a, b\}$, $a_+ = a \wedge 0$ and $a_- = -(a \wedge 0)$.

For any function H and any constant C, we let

$$H_C(y) = \begin{cases} 0 & y < -C \\ H(y) & -C \le y < C \\ 1 & y \ge C \end{cases}$$

and $H_C^{-1}(y) = (H_C)^{-1}(y)$. We prove consistency of the lower bound estimator only, and the consistency for the upper bound estimator follows similarly. For any $0 < \epsilon < 1$, it is easy to verify that $|\hat{\theta} - \theta| \le \min\{1/(4\theta), 1/\sqrt{2}\}\epsilon$ implies $|\hat{\theta}^2 - \theta^2| \le \epsilon$. Thus $\mathbb{P}(|\hat{\theta}^2 - \theta^2| > \epsilon) \le \mathbb{P}(|\hat{\theta} - \theta| > \min\{1/(4\theta), 1/\sqrt{2}\}\epsilon) \le \mathrm{var}(\hat{\theta}) \max\{16\theta^2, 2\}/\epsilon$ by Chebyshev's inequality. Under Condition 1, we have $\theta \le (2C_{\mathrm{M}})^{1/4}$ and $\sigma^2 \le C_{\mathrm{M}}^{1/2}(1/n_1 + 1/n_0)N/(N-1)$ by Jensen's inequality. Thus for any $0 < \epsilon, \delta < 1$, we have

$$\mathbb{P}(|\hat{\theta}^2 - \theta^2| \le \epsilon) \ge 1 - \delta \tag{S2.3}$$

for sufficiently large N. Let $\hat{\pi}_{tk} = \sum_{T_i = t} 1\{w = \xi_k\}/n_t$ for $t = 0, 1$ and $k = 1, \ldots, K$. Then for any $C > 0$,

$$\begin{aligned}
&|\sum_{k=1}^{K} \hat{\pi}_k \int_0^1 (\hat{F}_{1|k}^{-1}(u) - \hat{F}_{0|k}^{-1}(u))^2 du - \sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k}^{-1}(u) - F_{0|k}^{-1}(u))^2 du| \\
&\le |\sum_{k=1}^{K} \hat{\pi}_k \int_0^1 (\hat{F}_{1|k}^{-1}(u) - \hat{F}_{0|k}^{-1}(u))^2 du - \sum_{k=1}^{K} \hat{\pi}_k \int_0^1 (\hat{F}_{1|k,C}^{-1}(u) - \hat{F}_{0|k,C}^{-1}(u))^2 du| \\
&+ |\sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k}^{-1}(u) - F_{0|k}^{-1}(u))^2 du - \sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k,C}^{-1}(u) - F_{0|k,C}^{-1}(u))^2 du| \\
&+ |\sum_{k=1}^{K} \hat{\pi}_k \int_0^1 (\hat{F}_{1|k,C}^{-1}(u) - \hat{F}_{0|k,C}^{-1}(u))^2 du - \sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k,C}^{-1}(u) - F_{0|k,C}^{-1}(u))^2 du|
\end{aligned}$$

$=: I_1 + I_2 + I_3.$

Hence to prove Theorem 2, it suffices to show $I_1 + I_2 + I_3 \xrightarrow{P} 0$. Note that

$$
\begin{aligned}
I_2 \leq & \,|\sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k}^{-1}(u) - F_{1|k,C}^{-1}(u))(F_{1|k}^{-1}(u) + F_{1|k,C}^{-1}(u) - F_{0|k}^{-1}(u) - F_{0|k,C}^{-1}(u))du| \\
& + |\sum_{k=1}^{K} \pi_k \int_0^1 (F_{0|k}^{-1}(u) - F_{0|k,C}^{-1}(u))(F_{1|k}^{-1}(u) + F_{1|k,C}^{-1}(u) - F_{0|k}^{-1}(u) - F_{0|k,C}^{-1}(u))du| \\
& \leq \Big(\sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k}^{-1}(u) - F_{1|k,C}^{-1}(u))^2 du\Big)^{\frac{1}{2}} \\
& \times \Big(\sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k}^{-1}(u) + F_{1|k,C}^{-1}(u) - F_{0|k}^{-1}(u) - F_{0|k,C}^{-1}(u))^2 du\Big)^{\frac{1}{2}} \\
& + \Big(\sum_{k=1}^{K} \pi_k \int_0^1 (F_{0|k}^{-1}(u) - F_{0|k,C}^{-1}(u))^2 du\Big)^{\frac{1}{2}} \\
& \times \Big(\sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k}^{-1}(u) + F_{1|k,C}^{-1}(u) - F_{0|k}^{-1}(u) - F_{0|k,C}^{-1}(u))^2 du\Big)^{\frac{1}{2}},
\end{aligned}
$$

where the second inequality follows from Cauchy-Schwartz inequality.

Under Condition 1, we have $I_2 \leq 8C_M^{3/4}/C$ because

$$
\sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k}^{-1}(u) - F_{1|k,C}^{-1}(u))^2 du \leq \frac{1}{N} \sum_{|y_{ti}| \geq C} y_{ti}^2 \leq \frac{C_M}{C^2}
$$

for $t = 0, 1$ and

$$
\begin{aligned}
& \sum_{k=1}^{K} \pi_k \int_0^1 (F_{1|k}^{-1}(u) + F_{1|k,C}^{-1}(u) - F_{0|k}^{-1}(u) - F_{0|k,C}^{-1}(u))^2 du \\
& \leq 4 \sum_{k=1}^{K} \pi_k \int_0^1 [(F_{1|k}^{-1}(u))^2 + (F_{1|k,C}^{-1}(u))^2 + (F_{0|k}^{-1}(u))^2 + (F_{0|k,C}^{-1}(u))^2] du \\
& \leq \frac{8}{N} \sum_{i=1}^{N} y_{1i}^2 + \frac{8}{N} \sum_{i=1}^{N} y_{0i}^2 \leq 16\sqrt{C_M}.
\end{aligned}
$$

By similar arguments, we have

$$
\begin{aligned}
I_1 \le & \Big(\sum_{k=1}^{K}\hat{\pi}_k\int_0^1(\hat{F}_{1|k}^{-1}(u)-\hat{F}_{1|k,C}^{-1}(u))^2du\Big)^{\frac{1}{2}} \\
& \times\Big(4\sum_{k=1}^{K}\hat{\pi}_k\int_0^1[(\hat{F}_{1|k}^{-1}(u))^2+(\hat{F}_{1|k,C}^{-1}(u))^2+(\hat{F}_{0|k}^{-1}(u))^2+(\hat{F}_{0|k,C}^{-1}(u))^2]du\Big)^{\frac{1}{2}} \\
& +\Big(\sum_{k=1}^{K}\hat{\pi}_k\int_0^1(\hat{F}_{0|k}^{-1}(u)-\hat{F}_{0|k,C}^{-1}(u))^2du\Big)^{\frac{1}{2}} \\
& \times\Big(4\sum_{k=1}^{K}\hat{\pi}_k\int_0^1[(\hat{F}_{1|k}^{-1}(u))^2+(\hat{F}_{1|k,C}^{-1}(u))^2+(\hat{F}_{0|k}^{-1}(u))^2+(\hat{F}_{0|k,C}^{-1}(u))^2]du\Big)^{\frac{1}{2}}.
\end{aligned}
$$

Because for $t=0,1$,

$$
\sum_{k=1}^{K}\hat{\pi}_k\int_0^1(\hat{F}_{t|k}^{-1}(u)-\hat{F}_{t|k,C}^{-1}(u))^2du \le \max_k\frac{\hat{\pi}_k}{\hat{\pi}_{tk}}\frac{N}{n_t}\frac{1}{N}\sum_{|y_{ti}|\ge C}y_{ti}^2 \le \frac{N}{n_t}\max_k\frac{\hat{\pi}_k}{\hat{\pi}_{tk}}\frac{C_M}{C^2},
$$

and

$$
\begin{aligned}
& \sum_{k=1}^{K}\hat{\pi}_k\int_0^1[(\hat{F}_{1|k}^{-1}(u))^2+(\hat{F}_{1|k,C}^{-1}(u))^2+(\hat{F}_{0|k}^{-1}(u))^2+(\hat{F}_{0|k,C}^{-1}(u))^2]du \\
& \le \frac{N}{n_1\wedge n_0}\max_{t,k}\frac{\hat{\pi}_k}{\hat{\pi}_{tk}}\Big(\frac{2}{N}\sum_{i=1}^{N}y_{1i}^2+\frac{2}{N}\sum_{i=1}^{N}y_{0i}^2\Big) \\
& \le \frac{4N}{n_1\wedge n_0}\max_{t,k}\frac{\hat{\pi}_k}{\hat{\pi}_{tk}}\sqrt{C_M},
\end{aligned}
$$

we have

$$
I_1 \le \frac{8N}{n_1\wedge n_0}\max_{t,k}\frac{\hat{\pi}_k}{\hat{\pi}_{tk}}\frac{C_M^{\frac{3}{4}}}{C}.
$$

For the last term I_3, we have

$$
I_3 = \left|\sum_{k=1}^{K}\hat{\pi}_k\int_0^1(\hat{F}_{1|k,C}^{-1}(u)-\hat{F}_{0|k,C}^{-1}(u))^2du-\sum_{k=1}^{K}\pi_k\int_0^1(F_{1|k,C}^{-1}(u)-F_{0|k,C}^{-1}(u))^2du\right|
$$

$$
\begin{aligned}
&\leq \left|\sum_{k=1}^{K}\hat{\pi}_k\Big(\int_0^1(\hat{F}^{-1}_{1|k,C}(u)-\hat{F}^{-1}_{0|k,C}(u))^2du-\int_0^1(F^{-1}_{1|k,C}(u)-F^{-1}_{0|k,C}(u))^2du\Big)\right| \\
&+\left|\sum_{k=1}^{K}(\hat{\pi}_k-\pi_k)\int_0^1(F^{-1}_{1|k,C}(u)-F^{-1}_{0|k,C}(u))^2du\right| \\
&\leq \max_k\left|\int_0^1(\hat{F}^{-1}_{1|k,C}(u)-\hat{F}^{-1}_{0|k,C}(u))^2du-\int_0^1(F^{-1}_{1|k,C}(u)-F^{-1}_{0|k,C}(u))^2du\right|\sum_{k=1}^{K}\hat{\pi}_k \\
&+\left|\sum_{k=1}^{K}(\frac{\hat{\pi}_k}{\pi_k}-1)\pi_k\int_0^1(F^{-1}_{1|k,C}(u)-F^{-1}_{0|k,C}(u))^2du\right| \\
&\leq \max_k\left|\int_0^1(\hat{F}^{-1}_{1|k,C}(u)-\hat{F}^{-1}_{0|k,C}(u))^2du-\int_0^1(F^{-1}_{1|k,C}(u)-F^{-1}_{0|k,C}(u))^2du\right| \\
&+\max_k\left|\frac{\hat{\pi}_k}{\pi_k}-1\right|\left|\sum_{k=1}^{K}\pi_k\int_0^1(F^{-1}_{1|k,C}(u)-F^{-1}_{0|k,C}(u))^2du\right| \\
&=: I_{31}+I_{32}.
\end{aligned}
$$

By Condition 1, we have

$$
\begin{aligned}
&\sum_{k=1}^{K}\pi_k\int_0^1(F^{-1}_{1|k,C}(u)-F^{-1}_{0|k,C}(u))^2du \\
&\leq 2\sum_{k-1}^{K}\pi_k\int_0^1(F^{-1}_{1|k}(u))^2du+2\sum_{k-1}^{K}\pi_k\int_0^1(F^{-1}_{0|k}(u))^2du \\
&=2\Big(\frac{1}{N}\sum_{i=1}^{N}y_{1i}^2+\frac{1}{N}\sum_{i=1}^{N}y_{0i}^2\Big) \\
&\leq 2\sqrt{C_M}.
\end{aligned}
$$

Thus

$$
I_{32}\leq 2\max_k\left|\frac{\hat{\pi}_k}{\pi_k}-1\right|\sqrt{C_M}.
$$

For $k=1,\ldots,K$, $\int_0^1(F^{-1}_{1|k,C}(u)-F^{-1}_{0|k,C}(u))^2du$ is the square of the

Wasserstein distance induced by L_2 norm between $F_{1|k,C}$ and $F_{0|k,C}$. By the representation theorem (Bobkov and Ledoux, 2019)[Theorem 2.11],

$$\int_0^1 (\hat{F}^{-1}_{1|k,C}(u) - \hat{F}^{-1}_{0|k,C}(u))^2 du$$
$$= 2 \iint_{v \le w} [(\hat{F}_{1|k,C}(v) - \hat{F}_{0|k,C}(w))_+ + (\hat{F}_{0|k,C}(v) - \hat{F}_{1|k,C}(w))_+] dv dw.$$

Because $\hat{F}^{-1}_{1|k,C}(v) - \hat{F}^{-1}_{0|k,C}(w) \le 0$ and $\hat{F}_{0|k,C}(v) - \hat{F}_{1|k,C}(w) \le 0$ when $v < -C$ or $w \ge C$, the integral domain can be restricted to $\{-C \le v \le w < C\}$ without changing the integral.

Similarly, we have

$$\int_0^1 (F^{-1}_{1|k,C}(u) - F^{-1}_{0|k,C}(u))^2 du$$
$$= 2 \iint_{-C \le v \le w < C} [(F_{1|k,C}(v) - F_{0|k,C}(w))_+ + (F_{0|k,C}(v) - F_{1|k,C}(w))_+] dv dw.$$

Because $|(u_1)_+ - (u_2)_+| \le |u_1 - u_2|$ for any u_1, u_2,

$$I_{31} \le 2 \max_k \iint_{-C \le v \le w < C} [|\hat{F}_{1|k,C}(v) - F_{1|k,C}(v)| + |\hat{F}_{0|k,C}(v) - F_{0|k,C}(v)|$$
$$+ |\hat{F}_{1|k,C}(w) - F_{1|k,C}(w)| + |\hat{F}_{0|k,C}(w) - F_{0|k,C}(w)|] dv dw$$
$$\le 2C^2 \max_k \{\sup_v |\hat{F}_{1|k,C}(v) - F_{1|k,C}(v)| + \sup_v |\hat{F}_{0|k,C}(v) - F_{0|k,C}(v)|\}$$
$$\le 4C^2 \max_{t,k} \sup_v |\hat{F}_{t|k,C}(v) - F_{t|k,C}(v)|.$$

For any given M and $t = 0, 1$, let $s_{tk,j} = F^{-1}_{t|k,C}(j/M)$. By the standard

technique in the proof of Glivenko-Cantelli theorem (van der Vaart, 2000),

$$\sup_v |\hat{F}_{t|k,C}(v) - F_{t|k,C}(v)| \leq \max_j |\hat{F}_{t|k,C}(s_{tk,j}) - F_{t|k,C}(s_{tk,j})| + \frac{1}{M}.$$

For $t = 0, 1$ and $i = 1, \ldots, N$, let $y^*_{ti} = y_{ti}1\{-C \leq y_{ti} < C\} + C1\{y_{ti} \geq C\} - C1\{y_{ti} < -C\}$. Then

$$\begin{aligned}
&\hat{F}_{tk,C}(s_{tk,j}) - F_{tk,C}(s_{tk,j}) \\
&= \frac{1}{\hat{\pi}_{tk}} \frac{1}{n_t} \sum_{T_i=t} 1\{y^*_{ti} \leq s_{tk,j}, x = \xi_k\} - \frac{1}{\pi_k} \frac{1}{N} \sum_{i=1}^{N} 1\{y^*_{ti} \leq s_{tk,j}, x = \xi_k\} \\
&= \frac{1}{\pi_k} \left(\frac{1}{n_t} \sum_{T_i=t} 1\{y^*_{ti} \leq s_{tk,j}, x = \xi_k\} - \frac{1}{N} \sum_{i=1}^{N} 1\{y^*_{ti} \leq s_{tk,j}, x = \xi_k\} \right) \\
&- \frac{1}{\pi_k \hat{\pi}_{tk}} (\hat{\pi}_{tk} - \pi_k) \frac{1}{n_t} \sum_{T_i=t} 1\{y^*_{ti} \leq s_{tk,j}, x = \xi_k\}.
\end{aligned}$$

Because

$$\left| \frac{1}{\hat{\pi}_{tk} n_t} \sum_{T_i=t} 1\{y^*_{ti} \leq s_{tk,j}, x = \xi_k\} \right| \leq 1,$$

we have

$$\begin{aligned}
I_{31} \leq 4C^2 \max_{t,k,j} \Big\{ \Big| \frac{1}{\pi_k} \Big(\frac{1}{n_t} \sum_{T_i=t} 1\{y^*_{ti} \leq s_{tk,j}, x = \xi_k\} \\
- \frac{1}{N} \sum_{i=1}^{N} 1\{y^*_{ti} \leq s_{tk,j}, x = \xi_k\} \Big) \Big| + \frac{1}{\pi_k} |\hat{\pi}_{tk} - \pi_k| \Big\} + \frac{1}{M}.
\end{aligned}$$

For any small positive number ϵ, one can choose C and M such that $8C_M^{3/4}/C \leq \epsilon$ and $1/M \leq \epsilon$. Then by Hoeffding inequality for sample without replacement (Bardenet and Maillard, 2015) and the Bonferroni

inequality, we have

$$\mathbb{P}\left(4C^2 \max_{t,k,j}\left\{\frac{1}{\pi_k}\left|\frac{1}{n_t}\sum_{T_i=t} 1\{y_{ti}^* \le s_{tk,j}, x=\xi_k\} - \frac{1}{N}\sum_{i=1}^{N} 1\{y_{ti}^* \le s_{tk,j}, x=\xi_k\}\right|\right\} \ge \epsilon\right)$$
$$\le 4MK\exp\left(-\frac{n_1 \wedge n_0 \epsilon^2 \min_k \pi_k^2}{8C^4}\right),$$

and

$$\mathbb{P}\left(\max_{t,k}\frac{1}{\pi_k}|\hat{\pi}_{tk} - \pi_k| \ge \epsilon\right) \le 2K\exp\left(-2n_1 \wedge n_0 \epsilon^2 \min_k \pi_k^2\right).$$

By Conditions 2 and 3 the right hand side of these two inequalities converge to zero because

$$n_1 \wedge n_0 \min_k \pi_k^2 - C^* \log K \ge C_\pi (n_1 \wedge n_0/N)(N/K^2) - C^* \log K \to \infty,$$

with $C^* = 8C^4 \vee (1/2)$.

Hence, for any $\delta > 0$ and sufficiently large N, $I_{31} \le 3\epsilon$ and

$$\max_{t,k}\left|\frac{\hat{\pi}_k}{\hat{\pi}_{tk}}\right| \le \frac{1+\epsilon}{1-\epsilon}$$

with probability greater than $1-\delta$. Without loss of generality, we let $\epsilon \le 1/3$, then $(1+\epsilon)/(1-\epsilon) \le 2$ and for sufficiently large N we have

$$\left|\sum_{k=1}^{K}\hat{\pi}_k \int_0^1 (\hat{F}_{1|k}^{-1}(u) - \hat{F}_{0|k}^{-1}(u))^2 du - \sum_{k=1}^{K}\int_0^1 (F_{1|k}^{-1}(u) - F_{0|k}^{-1}(u))^2 du\right|$$
$$\le \left(\frac{4}{\rho_1 \wedge \rho_0} + 2\sqrt{C_M} + 4\right)\epsilon$$

with probability greater than $1-\delta$. Combining this with (S2.3) completes the proof of Theorem 2. □

S3 Proof of Theorem 3

In this proof, we use ϵ, δ to denote small positive numbers whose values may change from place to place. Recall the definition of $\mathcal{P}_N$ in (2.6) in the main text. We assume without loss of generality that $\mathcal{P}_N$ is non-empty. Under Condition 3, according to the proof of Theorem 2, for $\mathbf{U}_N \in \mathcal{P}_N$ and any $\epsilon, \delta > 0$, we have $\mathbb{P}(|\hat{\phi}_{\mathrm{L}}^2 - \phi_{\mathrm{L}}^2| \leq \epsilon) \geq 1 - \delta$ for any N larger than a threshold T_1. Note that in the proof of Theorem 2, the threshold T_1 can be chosen to be uniform in $\mathbf{U}_N \in \mathcal{P}_N$. That is, for any small positive numbers ϵ, δ, there is some T_1 such that for $N > T_1$ we have

$$\inf_{\mathbf{U}_N \in \mathcal{P}_N} \mathbb{P}(|\hat{\phi}_{\mathrm{L}}^2 - \phi_{\mathrm{L}}^2| \leq \epsilon) \geq 1 - \delta. \tag{S3.4}$$

Recall the definition of $\hat{\sigma}^2$ in Section 2.3 and the definition of $\hat{\phi}_1^2$, $\hat{\phi}_0^2$ in Section 2.1 in the main text. According to Theorem 1,

$$\begin{aligned}\mathbb{P}(\hat{\sigma}^2 \geq \sigma^2 - \epsilon) \geq 1 - \mathbb{P}\left(\frac{N}{n_1}|\phi^2(y_1) - \hat{\phi}_1^2| \geq \frac{\epsilon}{3}\right) - \mathbb{P}\left(\frac{N}{n_0}|\phi^2(y_1) - \hat{\phi}_0^2| \geq \frac{\epsilon}{3}\right) \\ - \mathbb{P}\left(|\hat{\phi}_{\mathrm{L}}^2 - \phi_{\mathrm{L}}^2| \geq \frac{\epsilon}{3}\right).\end{aligned} \tag{S3.5}$$

Thus for any $\epsilon, \delta > 0$, there is some threshold T_2, for $N > T_2$, we have

$$\inf_{\mathbf{U}_N \in \mathcal{P}_N} \mathbb{P}(\hat{\sigma}^2 \geq \sigma^2 - \epsilon) \geq 1 - \delta \tag{S3.6}$$

by (S3.4), (S3.5), Chebyshev's inequality and straightforward calculation of mean and variance of $\hat{\phi}_1^2$, $\hat{\phi}_0^2$. According to the Cauchy-Schwartz inequality,

we have

$$\begin{aligned}\sigma^2 &= \frac{N}{n_1}\phi^2(y_1) + \frac{N}{n_0}\phi^2(y_0) - \phi^2(\tau) \\ &\geq \frac{N}{n_1}\phi^2(y_1) + \frac{N}{n_0}\phi^2(y_0) - \phi^2(y_1) - \phi^2(y_0) \\ &\geq L_2\frac{n_0}{n_1} + L_2\frac{n_1}{n_0} \\ &> L_2\end{aligned} \tag{S3.7}$$

if $\mathbf{U}_N \in \mathcal{P}_N$. Then (S3.6) and (S3.7) imply

$$\inf_{\mathbf{U}_N \in \mathcal{P}_N} \mathbb{P}(\hat{\sigma} \geq \sigma - \epsilon) \geq 1 - \delta \tag{S3.8}$$

for any $N > T_3$, where T_3 is some threshold that depends on ϵ, δ.

By the finite population central limit theorem (Freedman, 2008)[Theorem 1], for any sequence of finite populations $\{\mathbf{U}_N\}_{N=2}^{\infty}$ such that $\mathbf{U}_N \in \mathcal{P}_N$, we have

$$\sqrt{N}\sigma^{-1}(\hat{\theta} - \theta) \xrightarrow{d} N(0,1)$$

as $N \to \infty$. This and (S3.7) imply for any $0 < \delta < \alpha < 1$ and any sequence of finite populations $\{\mathbf{U}_N\}_{N=2}^{\infty}$ such that $\mathbf{U}_N \in \mathcal{P}_N$, there is some $0 < \epsilon < \sqrt{L_2}$ such that

$$\begin{aligned}&\liminf_{N\to\infty} \mathbb{P}\left(\theta \in \left[\hat{\theta} - q_{\frac{\alpha}{2}}(\sigma - \epsilon)N^{-\frac{1}{2}}, \hat{\theta} + q_{\frac{\alpha}{2}}(\sigma - \epsilon)N^{-\frac{1}{2}}\right]\right) \\ &= \liminf_{N\to\infty} \mathbb{P}\left(\sqrt{N}\sigma^{-1}(\hat{\theta} - \theta) \in \left[-q_{\frac{\alpha}{2}}(1 - \sigma^{-1}\epsilon), q_{\frac{\alpha}{2}}(1 - \sigma^{-1}\epsilon)\right]\right) \\ &\geq \alpha - \delta,\end{aligned} \tag{S3.9}$$

where $q_{\alpha/2}$ is the upper $\alpha/2$ quantile of a standard normal distribution.

Next we show that

$$\liminf_{N\to\infty} \inf_{\mathbf{U}_N\in\mathcal{P}_N} \mathbb{P}\left(\theta \in \left[\hat{\theta} - q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}, \hat{\theta} + q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}\right]\right) \geq \alpha - \delta. \tag{S3.10}$$

We prove this by contradiction. If (S3.10) does not hold, then there is some $c_* > 0$, for any $N_0 > 0$, there is some $N_1 > N_0$ and $\mathbf{U}'_{N_1}$ such that

$$\mathbb{P}\left(\theta \in \left[\hat{\theta} - q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}, \hat{\theta} + q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}\right]\right) \leq \alpha - \delta - c_*$$

under $\mathbf{U}'_{N_1}$. Following the same procedure, we can extract a subsequence $\{\mathbf{U}'_{N_m}\}_{m=1}^{\infty}$ with $N_1 < N_2 < \dots$ and $\mathbf{U}'_{N_m} \in \mathcal{P}_{N_m}$ such that

$$\mathbb{P}\left(\theta \in \left[\hat{\theta} - q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}, \hat{\theta} + q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}\right]\right) \leq \alpha - \delta - c_*$$

under $\mathbf{U}'_{N_m}$ for $m = 1, 2, \dots$ Then it follows that for any sequence of finite populations $\{\mathbf{U}_N\}_{N=2}^{\infty}$ such that $\{\mathbf{U}_N\}_{N=2}^{\infty}$ contains $\{\mathbf{U}'_{N_m}\}_{m=1}^{\infty}$ as a subsequence, we have

$$\lim_{N\to\infty} \mathbb{P}\left(\theta \in \left[\hat{\theta} - q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}, \hat{\theta} + q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}\right]\right) \leq \alpha - \delta - c_*,$$

which is contradict with (S3.9). This proves (S3.10). According to (S3.10), there is some threshold T_3 such that

$$\inf_{\mathcal{P}_N} \mathbb{P}\left(\theta \in \left[\hat{\theta} - q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}, \hat{\theta} + q_{\frac{\alpha}{2}}(\sigma-\epsilon)N^{-\frac{1}{2}}\right]\right) \geq \alpha - 2\delta$$

for $N > T_3$. Combining this with (S3.6), we have

$$\inf_{\mathcal{P}_N} \mathbb{P}\left(\theta \in \left[\hat{\theta} - q_{\frac{\alpha}{2}}\hat{\sigma}N^{-\frac{1}{2}}, \hat{\theta} + q_{\frac{\alpha}{2}}\hat{\sigma}N^{-\frac{1}{2}}\right]\right) \geq \alpha - 3\delta$$

for $N > \max\{T_2, T_3\}$. This proves the theorem since δ is an arbitrary small positive number.

S4 Proof of Theorem 4

Proof. Under Conditions 1 and 4, the conditions of the finite population central limit theorem (Freedman, 2008)[Theorem 1] is satisfied and hence

$$\sqrt{N}V_N^{-1}\left(\frac{1}{n_1}\sum_{T_i=1} y_{1i} - \mu(y_1), \frac{1}{n_1}\sum_{T_i=1} d_{1i} - \mu(d_1), \frac{1}{n_0}\sum_{T_i=0} y_{0i} - \mu(y_0), \frac{1}{n_0}\sum_{T_i=0} d_{0i} - \mu(d_0)\right)^{\mathrm{T}}$$

converges to a multivariate normal distribution, where V_N is defined in Condition 4. Note that we have the decomposition

$$\hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}} = \hat{\pi}_{\mathrm{c}}^{-1}(\hat{\theta} - \hat{\pi}_{\mathrm{c}}\theta_{\mathrm{c}}) = \left(\hat{\pi}_{\mathrm{c}}^{-1} - \pi_{\mathrm{c}}^{-1}\right)\left(\hat{\theta} - \hat{\pi}_{\mathrm{c}}\theta_{\mathrm{c}}\right) + \pi_{\mathrm{c}}^{-1}(\hat{\theta} - \hat{\pi}_{\mathrm{c}}\theta).$$

Under Condition 1, it is easy to show that the trace and hence the spectral norm of V_N is bounded. By the strong instrument assumption, the fact that V_N has bounded spectral norm and the asymptotic normality invoked before, we have

$$\hat{\pi}_{\mathrm{c}}^{-1} - \pi_{\mathrm{c}}^{-1} = O_p\left(\frac{1}{\sqrt{N}}\right)$$

and

$$\hat{\theta} - \hat{\pi}_{\mathrm{c}}\theta_{\mathrm{c}} = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Hence

$$\hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}} = \pi_{\mathrm{c}}^{-1}(\hat{\theta} - \hat{\pi}_{\mathrm{c}}\theta_{\mathrm{c}}) + o_p\left(\frac{1}{\sqrt{N}}\right).$$

Straightforward calculation can show that $(\hat{\theta} - \hat{\pi}_{\mathrm{c}}\theta_{\mathrm{c}})/\pi_{\mathrm{c}}$ has mean zero and variance

$$\frac{1}{\pi_{\mathrm{c}}^2(N-1)}\left(\frac{N}{n_1}\phi^2(\tilde{y}_1) + \frac{N}{n_0}\phi^2(\tilde{y}_0) - \phi^2(\tilde{\tau})\right) = \frac{\sigma_{\mathrm{c}}^2}{N-1}.$$

Again by the finite population central limit theorem we have

$$\sqrt{N}\sigma_{\mathrm{c}}^{-1}(\hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}}) = \sqrt{N}\sigma_{\mathrm{c}}^{-1}\pi_{\mathrm{c}}^{-1}(\hat{\mu}(\tau) - \hat{\pi}_{\mathrm{c}}\theta_{\mathrm{c}}) + o_p(1) \xrightarrow{d} N(0,1).$$

□

S5 Proof of Theorem 5

Proof. According to exclusion restriction, if $g_i = \mathrm{a}$ or n, we have

$$\tilde{\tau}_i = y_{1i} - y_{0i} - \theta_{\mathrm{c}}(d_{1i} - d_{0i}) = 0.$$

By the definition of θ_{c}, we have $\mu(\tilde{\tau}) = 0$. Thus

$$\phi^2(\tilde{\tau}) = \frac{1}{N}\sum_{i=1}^{N}\tilde{\tau}_i^2 = \pi_{\mathrm{c}}\frac{1}{N_{\mathrm{c}}}\sum_{g_i=\mathrm{c}}\tilde{\tau}_i^2$$

where N_{c}. Then the bound can be proved following the same arguments as in Theorem 1. Letting $\mathcal{U}_{\mathrm{c}} = \{\mathbf{U}_{\mathrm{c}}^* = (y_1^*, y_0^*, w^*, g^*) : \mathbf{U}_{\mathrm{c}}^*$ satisfies Assumption 1; $P(g^* = h) = \pi_h,\ P(w^* = \xi_k \mid g^* = h) = \pi_{k|h}$ and $P(y_t^* \leq y \mid w^* = \xi_k, g^* = h) = F_{t|(k,h)}(y)$ for $t = 0, 1$ and $h = \mathrm{a}, \mathrm{c}, \mathrm{n}\}$, the sharpness of the bound can be proved similarly as in Theorem 1. □

S6 Proof of Theorem 6

Proof. Firstly, we provide some relationship that is useful in the proof. According to the monotonicity and exclusion restriction, we have

$$1\{g_i = \mathrm{c}\} = d_{1i} - d_{0i},$$

$$(1 - d_{1i})1\{\tilde{y}_{0i} \leq y\} = (1 - d_{1i})1\{\tilde{y}_{1i} \leq y\}$$

and

$$d_{0i}1\{\tilde{y}_{1i} \leq y\} = d_{0i}1\{\tilde{y}_{0i} \leq y\}.$$

Thus

$$\begin{aligned}
&\pi_{k|\mathrm{c}} \\
&= \frac{\sum_{i=1}^{N} 1\{g_i = \mathrm{c}\}1\{w_i = \xi_k\}}{\sum_{i=1}^{N} 1\{g_i = \mathrm{c}\}} \\
&= \pi_{\mathrm{c}}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} d_{1i}1\{w_i = \xi_k\} - \frac{1}{N}\sum_{i=1}^{N} d_{0i}1\{w_i = \xi_k\}\right) \\
&= \pi_{\mathrm{c}}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}(1 - d_{0i})1\{w_i = \xi_k\} - \frac{1}{N}\sum_{i=1}^{N}(1 - d_{1i})1\{w_i = \xi_k\}\right),
\end{aligned}$$

$$\begin{aligned}
&\tilde{F}_{1|k}(y) \\
&= \frac{\sum_{i=1}^{N}(d_{1i} - d_{0i})1\{\tilde{y}_{1i} \leq y\}1\{w_i = \xi_k\}}{\sum_{i=1}^{N}(d_{1i} - d_{0i})1\{w_i = \xi_k\}} \\
&= \pi_{\mathrm{c}}^{-1}\pi_{k|\mathrm{c}}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} d_{1i}1\{\tilde{y}_{1i} \leq y\}1\{w_i = \xi_k\} - \frac{1}{N}\sum_{i=1}^{N} d_{0i}1\{\tilde{y}_{1i} \leq y\}1\{w_i = \xi_k\}\right) \\
&= \pi_{\mathrm{c}}^{-1}\pi_{k|\mathrm{c}}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} d_{1i}1\{\tilde{y}_{1i} \leq y\}1\{w_i = \xi_k\} - \frac{1}{N}\sum_{i=1}^{N} d_{0i}1\{\tilde{y}_{0i} \leq y\}1\{w_i = \xi_k\}\right),
\end{aligned}$$

and

$$
\begin{aligned}
&\tilde{F}_{0|k}(y) \\
&= \frac{\sum_{i=1}^{N}(d_{1i}-d_{0i})1\{\tilde{y}_{0i}\le y\}1\{w_i=\xi_k\}}{\sum_{i=1}^{N}(d_{1i}-d_{0i})1\{w_i=\xi_k\}} \\
&= \pi_{\mathrm{c}}^{-1}\pi_{k|\mathrm{c}}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}(1-d_{0i})1\{\tilde{y}_{0i}\le y\}1\{w_i=\xi_k\}-\frac{1}{N}\sum_{i=1}^{N}(1-d_{1i})1\{\tilde{y}_{0i}\le y\}1\{w_i=\xi_k\}\right) \\
&= \pi_{\mathrm{c}}^{-1}\pi_{k|\mathrm{c}}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}(1-d_{0i})1\{\tilde{y}_{0i}\le y\}1\{w_i=\xi_k\}-\frac{1}{N}\sum_{i=1}^{N}(1-d_{1i})1\{\tilde{y}_{1i}\le y\}1\{w_i=\xi_k\}\right).
\end{aligned}
$$

Since the estimators in Theorem 6 have the similar structure as those in Theorem 2, we intend to prove the consistency in a similar way. However, there are two extra difficulties. One is that $\hat{y}_{ti}\neq\tilde{y}_{ti}$ and we need to control the error introduced by using $\hat{y}_{ti}$ in place of $\tilde{y}_{ti}$ in the estimators. The other is that the estimators $\check{F}_{t|k}(y)$ for $t=0,1$ and $k=1,\ldots,K$ are not distribution functions and thus the representation theorem (Bobkov and Ledoux, 2019)[Theorem 2.11] can not be used directly. To solve this problem, we define

$$\check{F}^{*}_{t|k}(y)=\sup_{v\le y}\check{F}_{t|k}(v),$$

for $t=0,1$ and $k=1,\ldots,K$. Then by defination, $\check{F}^{*}_{t|k}(y)$ is a distribution function and, for $u\in(0,1)$, $\check{F}^{*-1}_{t|k}(u)=\check{F}^{-1}_{t|k}(u)$. Hence we can use $\check{F}^{*}_{t|k}(y)$ instead of $\check{F}_{t|k}(y)$ in the representation theorem.

We prove only for the lower bound, and the consistency result for the upper bound follows similarly. Let $\lambda_k=\pi_{\mathrm{c}}\pi_{k|\mathrm{c}}$. Because $\hat{\pi}_{\mathrm{c}}\hat{\pi}_{k|\mathrm{c}}=\hat{\lambda}_{1k}$, to

prove $\check{\phi}_{\mathrm{L}}^2 - \tilde{\phi}_{\mathrm{L}}^2 \xrightarrow{P} 0$, we only need to prove

$$\sum_{k=1}^{K} \hat{\lambda}_{1k} \int_0^1 (\check{F}_{1|k}^{-1}(u) - \check{F}_{0|k}^{-1}(u))^2 du - \sum_{k=1}^{K} \lambda_k \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du \xrightarrow{P} 0.$$

Note that

$$\begin{aligned}
&\left| \sum_{k=1}^{K} \hat{\lambda}_{1k} \int_0^1 (\check{F}_{1|k}^{-1}(u) - \check{F}_{0|k}^{-1}(u))^2 du - \sum_{k=1}^{K} \lambda_k \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du \right| \\
&\leq \left| \sum_{k=1}^{K} \hat{\lambda}_{1k} \Big(\int_0^1 (\check{F}_{1|k}^{-1}(u) - \check{F}_{0|k}^{-1}(u))^2 du - \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du \Big) \right| \\
&+ | \sum_{k=1}^{K} (\hat{\lambda}_{1k} - \lambda_k) \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du | \\
&\leq \left| \sum_{k=1}^{K} \Big(\frac{1}{n_1} \sum_{T_i=1} d_{1i} 1\{w_i = \xi_k\} \Big) \Big(\int_0^1 (\check{F}_{1|k}^{-1}(u) - \check{F}_{0|k}^{-1}(u))^2 du - \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du \Big) \right| \\
&+ \left| \sum_{k=1}^{K} \Big(\frac{1}{n_0} \sum_{T_i=0} d_{0i} 1\{w_i = \xi_k\} \Big) \Big(\int_0^1 (\check{F}_{1|k}^{-1}(u) - \check{F}_{0|k}^{-1}(u))^2 du - \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du \Big) \right| \\
&+ \left| \sum_{k=1}^{K} \left(\frac{\hat{\lambda}_{1k}}{\lambda_k} - 1 \right) \lambda_{1k} \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du \right| \\
&\leq 2 \max_k \left| \int_0^1 (\check{F}_{1|k}^{-1}(u) - \check{F}_{0|k}^{-1}(u))^2 du - \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du \right| \\
&+ \max_k \left| \frac{\hat{\lambda}_{1k}}{\lambda_k} - 1 \right| \sum_{k=1}^{K} \lambda_{1k} \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du \\
&=: I_{1,\mathrm{c}} + I_{2,\mathrm{c}}.
\end{aligned}$$

By Condition 1 and Assumption 1(iii), we have

$$\theta \leq \frac{1}{N} \sum_{i=1}^{N} |y_{1i}| + \frac{1}{N} \sum_{i=1}^{N} |y_{0i}| \leq 2C_M^{\frac{1}{4}},$$

and

$$|\theta_{\mathrm{c}}| \leq C_0^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}|y_{1i}| + \frac{1}{N}\sum_{i=1}^{N}|y_{0i}|\right) \leq 2C_0^{-1}C_M^{\frac{1}{4}}$$

due to Jensen's inequality. Note that $|\tilde{y}_{ti}| \leq y_{ti} + |\theta_{\mathrm{c}}|$. By Condition 1 we have

$$\begin{aligned}\sum_{k=1}^{K}\lambda_{1k}\int_0^1(\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2du &\leq \frac{2}{N}\sum_{i=1}^{N}\tilde{y}_{1i}^2 + \frac{2}{N}\sum_{i=1}^{N}\tilde{y}_{0i}^2 \\ &\leq \frac{4}{N}\sum_{i=1}^{N}y_{1i}^2 + \frac{4}{N}\sum_{i=1}^{N}y_{0i}^2 + 8|\theta_{\mathrm{c}}| \\ &\leq 8\sqrt{C_M}(1+2C_0^{-1}).\end{aligned}$$

For any small positive ϵ, according to Hoeffding inequality for sampling without replacement (Bardenet and Maillard, 2015) and the Bonferroni inequality we have

$$\begin{aligned}&\mathbb{P}\left(8\sqrt{C_M}(1+2C_0^{-1})\max_{t,k}\frac{1}{\lambda_k}|\hat{\lambda}_{tk} - \lambda_k| \geq \epsilon\right) \\ &\leq 2K\exp\left(-\frac{1}{32C_M(1+2C_0^{-1})^2}n_1 \wedge n_0\epsilon^2\min_k\lambda_k^2\right) \to 0\end{aligned}$$

by Conditions 5 and 6. Without loss of generality, we assume $\epsilon \leq 1/2$ in the proof. Thus for any $\delta > 0$ and sufficiently large N

$$I_{2,\mathrm{c}} \leq \epsilon$$

with probability at least $1 - \delta/3$.

Define $B_N = (C_N + C_B) \vee 1$ where $C_B = 10C_0^{-2}C_M^{1/4}$. For $\epsilon \leq B_N(C_0 \wedge$

$C_M^{1/4}$), it is easy to verify that

$$\{B_N|\hat{\pi}_{\mathrm{c}} - \pi_{\mathrm{c}}| < \frac{\epsilon}{2}\} \cap \{B_N|\hat{\theta} - \theta| < \frac{\epsilon}{2}\} \subset \{B_N|\hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}}| < \epsilon\}.$$

Thus

$$\begin{aligned}\mathbb{P}\left(B_N|\hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}}| \geq C_B\epsilon\right) &\leq \mathbb{P}\left(B_N|\hat{\pi}_{\mathrm{c}} - \pi_{\mathrm{c}}| \geq \frac{\epsilon}{2}\right) + \mathbb{P}\left(B_N|\hat{\theta} - \theta| \geq \frac{\epsilon}{2}\right) \\ &\leq 4\exp\left(-\frac{2n_1 \wedge n_0\epsilon^2}{B_N^2}\right) + 4\exp\left(-\frac{n_1 \wedge n_0\epsilon^2}{2B_N^2C_N^2}\right)\end{aligned} \tag{S6.11}$$

where the last inequality follows from the Hoeffding inequality. By Condition 6,

$$4\exp\left(-\frac{2n_1 \wedge n_0\epsilon^2}{B_N^2}\right) + 4\exp\left(-\frac{n_1 \wedge n_0\epsilon^2}{2B_N^2C_N^2}\right) \to 0.$$

Thus for sufficiently large N with probability greater than $1 - \delta/3$ we have $B_N|\hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}}| \leq C_B\epsilon$. Note that $C_0 \leq 1$, $|\tilde{y}_{ti}| \leq y_{ti} + |\theta_{\mathrm{c}}| \leq C_N + 2C_0^{-1}C_M^{1/4} \leq B_N$ and $|\hat{y}_{ti}| \leq |y_{ti}| + |\hat{\theta}_{\mathrm{c}}| \leq C_N + 2C_0^{-1}C_M^{1/4} + C_B\epsilon \leq B_N$ when the event $\{B_N|\hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}}| < C_B\epsilon\}$ holds. On the event $\{B_N|\hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}}| < C_B\epsilon\}$, by the representation theorem (Bobkov and Ledoux, 2019)[Theorem 2.11], using the similar arguments as those in the proof of Theorem 2, we can show that

$$\begin{aligned}&\max_k \left|\int_0^1 (\check{F}_{1|k}^{-1}(u) - \check{F}_{0|k}^{-1}(u))^2 du - \int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du\right| \\ &\leq 2\max_k \iint_{-B_N \leq v \leq w \leq B_N} [|\check{F}_{1|k}^*(v) - \tilde{F}_{1|k}(v)| + |\check{F}_{0|k}^*(v) - \tilde{F}_{0|k}(v)| \\ &\qquad + |\check{F}_{1|k}^*(w) - \tilde{F}_{1|k}(w)| + |\check{F}_{0|k}^*(w) - \tilde{F}_{0|k}(w)|]dvdw\end{aligned}$$

$$\leq 4B_N \max_k \int_{-B_N}^{B_N} |\check{F}^*_{1|k}(v) - \tilde{F}_{1|k}(v)|dv + 4B_N \max_k \int_{-B_N}^{B_N} |\check{F}^*_{0|k}(v) - \tilde{F}_{0|k}(v)|dv. \tag{S6.12}$$

Here we only analyze the first term $B_N \max_k \int_{-B_N}^{B_N} |\check{F}^*_{1|k}(v) - F_{1|k}(v)|dv$ and the same result can be proved similarly for the second term. Define

$$\bar{F}_{11|k}(y) = \frac{\hat{\lambda}_{1k}^{-1}}{n_1} \sum_{T_i=1} d_{1i} 1\{\tilde{y}_{1i} \leq y\} 1\{w_i = \xi_k\},$$

$$\bar{F}_{01|k}(y) = \frac{\hat{\lambda}_{1k}^{-1}}{n_0} \sum_{T_i=0} d_{0i} 1\{\tilde{y}_{0i} \leq y\} 1\{w_i = \xi_k\},$$

$$\tilde{F}_{11|k}(y) = \frac{\lambda_k^{-1}}{N} \sum_{i=1}^{N} d_{1i} 1\{\tilde{y}_{1i} \leq y\} 1\{w_i = \xi_k\},$$

$$\tilde{F}_{01|k}(y) = \frac{\lambda_k^{-1}}{N} \sum_{i=1}^{N} d_{0i} 1\{\tilde{y}_{0i} \leq y\} 1\{w_i = \xi_k\}.$$

for $k = 1, \ldots, K$. Let $\bar{F}_{1|k}(y) = \bar{F}_{11|k}(y) - \bar{F}_{01|k}(y)$, then the relationship

$$\check{F}_{1|k}(y) = \bar{F}_{1|k}(y + \hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}})$$

holds. Because $|\sup_{v \leq y} \bar{F}_{1|k}(v) - \sup_{v \leq y} \tilde{F}_{1|k}(v)| \leq \sup_v |\bar{F}_{1|k}(v) - \tilde{F}_{1|k}(v)|$ and $\sup_{v \leq y} \tilde{F}_{1|k}(v) = \tilde{F}_{1|k}(y)$, we have

$$\begin{aligned}
&|\sup_{v \leq y} \bar{F}_{1|k}(v) - \bar{F}_{1|k}(y)| \\
&\leq |\sup_{v \leq y} \bar{F}_{1|k}(v) - \sup_{v \leq y} \tilde{F}_{1|k}(v)| + |\bar{F}_{1|k}(y) - \tilde{F}_{1|k}(y)| \\
&\leq 2 \sup_v |\bar{F}_{1|k}(v) - \tilde{F}_{1|k}(v)|.
\end{aligned}$$

Hence

$$
\begin{aligned}
&|\check{F}^*_{1|k}(y) - \tilde{F}_{1|k}(y)| \\
&\le |\sup_{v \le y+\hat{\theta}_c-\theta_c} \bar{F}_{1|k}(v) - \bar{F}_{1|k}(y+\hat{\theta}_c-\theta_c)| + |\bar{F}_{1|k}(y+\hat{\theta}_c-\theta_c) - \tilde{F}_{1|k}(y)| \\
&\le 2\sup_v |\bar{F}_{1|k}(v) - \tilde{F}_{1|k}(v)| + |\bar{F}_{1|k}(y+\hat{\theta}_c-\theta_c) - \tilde{F}_{1|k}(y+\hat{\theta}_c-\theta_c)| \\
&+ |\tilde{F}_{1|k}(y+\hat{\theta}_c-\theta_c) - \tilde{F}_{1|k}(y)| \\
&\le 3\sup_v |\bar{F}_{1|k}(v) - \tilde{F}_{1|k}(v)| + |\tilde{F}_{1|k}(y+\hat{\theta}_c-\theta_c) - \tilde{F}_{1|k}(y)|.
\end{aligned}
$$

Moreover, because

$$
\begin{aligned}
&|\tilde{F}_{1|k}(y+\hat{\theta}_c-\theta_c) - \tilde{F}_{1|k}(y)| \\
&\le |\tilde{F}_{11|k}(y+\hat{\theta}_c-\theta_c) - \tilde{F}_{11|k}(y)| + |\tilde{F}_{01|k}(y+\hat{\theta}_c-\theta_c) - \tilde{F}_{01|k}(y)| \\
&\le \lambda_k^{-1}\Big(\frac{1}{N}\sum_{i=1}^N d_{1i}1\{|\tilde{y}_{1i}-y| \le |\hat{\theta}_c-\theta_c|\}1\{w_i=\xi_k\} \\
&\qquad + \frac{1}{N}\sum_{i=1}^N d_{0i}1\{|\tilde{y}_{0i}-y| \le |\hat{\theta}_c-\theta_c|\}1\{w_i=\xi_k\}\Big),
\end{aligned}
$$

we have

$$
B_N \max_k \int_{-B_N}^{B_N} |\check{F}^*_{1|k}(v) - F_{1|k}(v)|dv \le 3B_N^2 \max_k \sup_v |\bar{F}_{1|k}(v) - \tilde{F}_{1|k}(v)| + 2B_N|\hat{\theta}_c - \theta_c|.
$$

By inequality (S6.11), we have $B_N|\hat{\theta}_c - \theta_c| \le C_B\epsilon$ with probability at least $1-\delta/3$ for sufficiently large N. Using the similar arguments as those used to analyze I_{31} in the proof of Theorem 2, we can show

$$
B_N^2 \max_k \sup_v |\bar{F}_{1|k}(v) - \tilde{F}_{1|k}(v)| \le \epsilon
$$

with probability at least $1 - \delta/6$ for sufficiently large N. By applying the similar arguments to the second term of expression (S6.12), we have $I_{1,\mathrm{c}} \leq (48 + 32C_B)\epsilon$ with probability at least $1 - 2\delta/3$ for sufficiently large N. Thus we have proved that for any small positive numbers ϵ and δ, we have $|\check{\phi}^2_{\mathrm{L}} - \tilde{\phi}^2_{\mathrm{L}}| \leq (49 + 32C_B)\epsilon$ with probability at least $1 - \delta$ for sufficiently large N and this implies the consistency of the estimator. □

S7 Further simulation results

To explore the reliability of the proposed confidence intervals (CIs), we consider the case where $\phi^2(\tau)$ attains the lower bound ϕ^2_{L}. In this case, the asymptotic variance of $\sqrt{N}(\hat{\theta} - \theta)$ achieves its upper bound. According to the proof of the sharpness in Theorem 1, we can modify the finite populations considered in Section 4.1 to make $\phi^2(\tau)$ equal to ϕ^2_{L} without changing π_k, $F_{1|k}(y)$ or $F_{0|k}(y)$ $(k = 1, \ldots, K)$. Then we conducted the simulation in the same way as in Section 4.1 in the main text under the modified finite populations. The average width (AW) and coverage rate (CR) of 95% CIs based on the naive lower bound zero (Neyman, 1990), the estimator of ϕ^2_{AL} (Aronow et al., 2014), the estimator of ϕ^2_{DL} (Ding et al., 2019) and the estimator of ϕ^2_{L} are summarized in the following table.

Comparing Table S1 with Table 3 in the main text, we find that the

Table S1: Average widths (AWs) and coverage rates (CRs) of 95% CIs based on the naive bound, ϕ^2_{AL}, ϕ^2_{DL} and ϕ^2_{L} under different population sizes when $\phi^2(\tau)$ attains the lower bound ϕ^2_{L} ($n_1 = n_0 = N/2$)

Method	naive		ϕ^2_{AL}		ϕ^2_{DL}		ϕ^2_{L}	
	AW	CR	AW	CR	AW	CR	AW	CR
$N = 400$	1.511	96.4%	1.495	96.3%	1.493	96.3%	1.380	94.7%
$N = 800$	1.033	96.6%	1.025	96.4%	1.025	96.4%	0.944	95.2%
$N = 2000$	0.674	97.0%	0.669	96.9%	0.669	96.9%	0.619	95.7%

AWs are similar while the CRs are smaller in Table S1. This is because the bounds are all the same under the finite populations considered in Table S1 and Table 3, but the variance of $\hat{\theta}$ is larger here. The CI based on ϕ^2_{L} is still quite reliable in this case because its CR is close to 95%.

We then investigate the AW and CR of the CIs in randomized experiments with noncompliance similarly. The following table summarizes the average width (AW) and coverage rate (CR) of 95% CIs based on the naive lower bound zero, and the estimator of lower bounds in Theorem 5 using and without using the covariate.

It can be seen that the CI based on LC is still quite reliable even when the asymptotic variance of $\hat{\theta}_{\mathrm{c}}$ achieves its upper bound.

Table S2: Average widths (AWs) and coverage rates (CRs) of 95% CIs based on the naive bound, HNL and HL under different population sizes when $\phi^2(\tilde{\tau})$ attains the lower bound $\tilde{\phi}^2_{\mathrm{L}}$. LNC: lower bound without covariate; HNC: upper bound without covariate; LC: lower bound with covariate; HC: upper bound with covariate ($n_1 = n_0 = N/2$)

Method	naive		LNC		LC	
	AW	CR	AW	CR	AW	CR
$N = 400$	2.186	96.7%	2.166	96.7%	1.983	94.1%
$N = 800$	1.553	97.0%	1.542	96.8%	1.397	94.5%
$N = 2000$	0.980	96.2%	0.973	96.0%	0.893	94.8%

Bibliography

Aronow, P. M., D. P. Green, and D. K. K. Lee (2014). Sharp bounds on the variance in randomized experiments. *The Annals of Statistics 42*, 850–871.

Bardenet, R. and O.-A. Maillard (2015). Concentration inequalities for sampling without replacement. *Bernoulli 21*, 1361–1385.

Bobkov, S. and M. Ledoux (2019). *One-dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances*, Volume

261. American Mathematical Society.

Ding, P., A. Feller, and L. Miratrix (2019). Decomposing treatment effect variation. *Journal of the American Statistical Association 114*, 304–317.

Freedman, D. A. (2008). On regression adjustments in experiments with several treatments. *The Annals of Applied Statistics 2*, 176–196.

Neyman, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science 5*, 465–472. Reprint of the original 1923 paper.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, New York.