A SEQUENTIAL SIGNIFICANCE TEST FOR TREATMENT BY COVARIATE INTERACTIONS

Min Qian[1], Bibhas Chakraborty[2,3], Raju Maiti[2] and Ying-Kuen Cheung[1]

[1]*Columbia University,* [2]*National University of Singapore and* [3]*Duke University*

Supplementary Materials

Section S1 contains proofs of Theorem 3. The required Assumptions are also included for completeness. Theorems 1 and 2 follow from Theorem 3 with $J = \emptyset$ and $\widetilde{\boldsymbol{X}}_J = 1$. In this case, $U_k = X_k - EW^2X_k/EW^2$. And the proofs are omitted.

Section S2 contains proofs of Theorem 4. The required Assumptions are also included for completeness.

Section S3 contains discussion of the uniqueness condition on k_0 and an extension of Theorem 1 to the case of non-unique k_0.

Section S4 contains discussion of the doubly robust method when used in randomized trial.

Section S5 contains details of simulations.

S1 Assumptions and proofs of Theorem 3

Assumptions:

(C1) The error term ϵ in model (2.1) has mean zero, finite variance and is uncorrelated with W and $W\boldsymbol{X}$.

(C2) $EX_k^4 < \infty$ for $k = 1, \ldots, p$.

(C3) $\hat{\phi}_n(\boldsymbol{X})$ is estimated from a P-Donsker class of measurable functions, and there exists some fourth-moment integrable function $\tilde{\phi}(\boldsymbol{X})$ such that $E[\hat{\phi}_n(\boldsymbol{X}) - \tilde{\phi}(\boldsymbol{X})]^4 \xrightarrow{P} 0$ as $n \to \infty$.

(C4) $k_0' \triangleq \arg\max_{k:k \in J^C} \left|\mathrm{Corr}(WU_k, W\boldsymbol{U}^\mathsf{T}\boldsymbol{\beta}_{0,J^C})\right|$ is unique if $\boldsymbol{\beta}_{0,J^C} \neq \mathbf{0}$.

Proof for part i) of Theorem 3.

For $k \in J^C$, let $(\alpha_k', \theta_k') = \arg\min_{(\alpha,\theta)} E[Y - E(Y|\boldsymbol{X}) - (\alpha + \theta U_k)W]^2$. By first order conditions, we have

$$\begin{aligned}
\begin{pmatrix} \alpha_k' \\ \theta_k' \end{pmatrix} &= \begin{pmatrix} EW^2 & EW^2U_k \\ EW^2U_k & E(WU_k)^2 \end{pmatrix}^{-1} \begin{pmatrix} E[W(Y - E(Y|\boldsymbol{X}))] \\ E[WU_k(Y - E(Y|\boldsymbol{X}))] \end{pmatrix} \\
&= \begin{pmatrix} E[W(Y - E(Y|\boldsymbol{X}))]/EW^2 \\ E[WU_k(Y - E(Y|\boldsymbol{X}))]/E(WU_k)^2 \end{pmatrix},
\end{aligned}$$

where the second equality follows from the fact that $E(W^2U_k) = 0$.

Under Assumption (C1), it is easy to verify that the new error term ϵ' in model (3.2) has mean zero, and is uncorrelated with W and WU_k for $k \in J^C$. Replacing Y by the right hand side of (3.2) yields $\alpha_k' = \alpha_0'$ and

$\theta_k' = \quad \mathrm{Cov}(WU_k, W\boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C})/E(WU_k)^2$. In addition, note that

$$
\begin{aligned}
&E[Y - E(Y|\boldsymbol{X}) - (\alpha_k' + \theta_k' U_k)W]^2\\
=&E[h_0'(\boldsymbol{X}) + (\alpha_0' + \boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C})q_0(\boldsymbol{X}) + \epsilon' - E(Y|\boldsymbol{X}) + (\alpha_0' + \boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C} - \alpha_k' - \theta_k' U_k)W]^2\\
=&E[h_0'(\boldsymbol{X}) + (\alpha_0' + \boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C})q_0(\boldsymbol{X}) + \epsilon' - E(Y|\boldsymbol{X})]^2 + E[(\alpha_0' + \boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C} - \alpha_k' - \theta_k' U_k)W]^2\\
=&E[h_0'(\boldsymbol{X}) + (\alpha_0' + \boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C})q_0(\boldsymbol{X}) + \epsilon' - E(Y|\boldsymbol{X})]^2 + E(\boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C})^2\\
&- [\mathrm{Corr}(WU_k, W\boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C})]^2 E(W\boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C}).
\end{aligned}
$$

Thus choosing k to maximize $\mathrm{Corr}(WU_k, W\boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C})$ is equivalent to minimizing $E[Y - E(Y|\boldsymbol{X}) - (\alpha_k' + \theta_k' U_k)W]^2$. So $\theta_0' = \theta_{k_0'}'$, where

$$
k_0' = \arg\max_{k\in J^C} |\mathrm{Corr}(WU_k, W\boldsymbol{U}^\top\boldsymbol{\beta}_{0,J^C})| = \arg\min_{k\in J^C} E[Y - E(Y|\boldsymbol{X}) - (\alpha_k' + \theta_k' U_k)W]^2.
$$

Similarly, we can verify that $\hat{\theta}_n' = \hat{\theta}_{\hat{k}_n'}'$, where

$$
\begin{aligned}
(\hat{\alpha}_k', \hat{\theta}_k') &= \arg\min_{(\alpha,\theta)\in\mathbb{R}^2} \mathbb{P}_n[Y - \hat{\phi}_n(\boldsymbol{X}) - (\alpha + \theta\hat{U}_k)W]^2\\
&= \left(\frac{\mathbb{P}_n[W(Y - \hat{\phi}_n(\boldsymbol{X}))]}{\mathbb{P}_n W^2}, \frac{\mathbb{P}_n[W\hat{U}_k(Y - \hat{\phi}_n(\boldsymbol{X}))]}{\mathbb{P}_n(W\hat{U}_k)^2}\right),\\
\text{and } \hat{k}_n' &= \arg\min_{k\in J^C} \mathbb{P}_n[Y - \hat{\phi}_n(\boldsymbol{X}) - (\hat{\alpha}_k' + \hat{\theta}_k'\hat{U}_k)W]^2 = \arg\max_{k\in J^C} \frac{(\mathbb{P}_n[W\hat{U}_k(Y - \hat{\phi}_n(\boldsymbol{X}))])^2}{\mathbb{P}_n(W\hat{U}_k)^2}
\end{aligned}
$$

with $\hat{U}_k = X_k - \widetilde{\boldsymbol{X}}_J^\top\hat{\boldsymbol{\gamma}}_k$ and $\hat{\boldsymbol{\gamma}}_k = \arg\min_{\boldsymbol{\gamma}} \mathbb{P}_n[W(X_k - \widetilde{\boldsymbol{X}}_J^\top\boldsymbol{\gamma})]^2$.

Note that $\mathbb{P}_n W^2 \hat{U}_k = 0.$, Again using first order conditions, we have

$$\begin{aligned}
\mathbb{P}_n\big(W\hat{U}_k\big)^2 n^{1/2}(\hat{\theta}'_k - \theta'_k) =&\ n^{1/2}\mathbb{P}_n\left[W\hat{U}_k\left(Y - \hat{\phi}_n(\boldsymbol{X}) - \theta'_k W\hat{U}_k\right)\right] \\
=&\,\mathbb{G}_n\left[WU_k\left(Y - \tilde{\phi}(\boldsymbol{X}) - \theta'_k WU_k\right)\right] \\
&+ n^{1/2}\mathbb{P}_n\left[W(\hat{U}_k - U_k)\left(Y - \tilde{\phi}(\boldsymbol{X}) - \theta'_k WU_k\right)\right] \\
&+ n^{1/2}\mathbb{P}_n\Big[W\hat{U}_k\Big(\tilde{\phi}(\boldsymbol{X}) - \hat{\phi}_n(\boldsymbol{X}) + \theta'_k W(U_k - \hat{U}_k)\Big)\Big], \qquad \text{(S1.1)}
\end{aligned}$$

where the second equality follows from the fact that

$$\begin{aligned}
&E\left[WU_k\left(Y - \tilde{\phi}(\boldsymbol{X}) - \theta'_k WU_k\right)\right] \\
=&E\left[WU_k\left(Y - E(Y|\boldsymbol{X}) - (\alpha'_k + \theta'_k U_k)W\right) + WU_k\left(E(Y|\boldsymbol{X}) + \alpha'_k W - \tilde{\phi}(\boldsymbol{X})\right)\right] \\
=&0.
\end{aligned}$$

By definition of U_k and $\hat{U}_k$, it is easy to verify that $\mathbb{P}_n W^2 \hat{U}_k \widetilde{\boldsymbol{X}}_J = \boldsymbol{0}$ and $n^{1/2}(\hat{U}_k - U_k) = -\widetilde{\boldsymbol{X}}_J^\intercal\left(\mathbb{P}_n W^2 \widetilde{\boldsymbol{X}}_J \widetilde{\boldsymbol{X}}_J^\intercal\right)^{-1}\mathbb{G}_n(W^2\widetilde{\boldsymbol{X}}_J U_k)$. Thus the third term in (S1.1) equals $\mathbb{G}_n\big[W\hat{U}_k\big(\tilde{\phi}(\boldsymbol{X}) - \hat{\phi}_n(\boldsymbol{X})\big)\big]$. Note that $E\big[W\hat{U}_k\big(\tilde{\phi}(\boldsymbol{X}) - \hat{\phi}_n(\boldsymbol{X})\big)\big]^2 \le \big[E(W\hat{U}_k)^4 E\big(\tilde{\phi}(\boldsymbol{X}) - \hat{\phi}_n(\boldsymbol{X})\big)^4\big]^{1/2} \xrightarrow{P} 0$ under Assumptions (C2) and (C3). by Lemma 19.24 of van der Vaart (1998), we have $\mathbb{G}_n\big[W\hat{U}_k\big(\tilde{\phi}(\boldsymbol{X}) - \hat{\phi}_n(\boldsymbol{X})\big)\big] = o_P(1)$. The second term in (S1.1) equals

$$-\mathbb{P}_n\left[W\left(Y - \tilde{\phi}(\boldsymbol{X}) - \theta'_k WU_k\right)\widetilde{\boldsymbol{X}}_J^\intercal\right]\left(\mathbb{P}_n W^2 \widetilde{\boldsymbol{X}}_J \widetilde{\boldsymbol{X}}_J^\intercal\right)^{-1}\mathbb{G}_n(W^2\widetilde{\boldsymbol{X}}_J U_k). \qquad \text{(S1.2)}$$

Plugging in (S1.2) into (S1.1) and using LLN yields

$$\begin{aligned}
&\mathbb{P}_n\big(W\hat{U}_k\big)^2 n^{1/2}(\hat{\theta}'_k - \theta'_k)\\
=&\,\mathbb{G}_n\Bigg\{WU_k\Big[Y - \tilde{\phi}(\boldsymbol{X}) - \theta'_k WU_k - W\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big(PW^2\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big)^{-1}\\
&\qquad\qquad E\Big(W\widetilde{\boldsymbol{X}}_J\big(Y - \tilde{\phi}(\boldsymbol{X}) - \theta'_k WU_k\big)\Big)\Big]\Bigg\} + o_P(1)\\
=&\,\mathbb{G}_n[WU_k(Y - \tilde{\phi}(\boldsymbol{X}) - M_k)] + o_P(1) \qquad \text{(S1.3)}
\end{aligned}$$

where

$$M_k = \theta'_k WU_k + W\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big(PW^2\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big)^{-1} E\Big[W\widetilde{\boldsymbol{X}}_J\Big(Y - \tilde{\phi}(\boldsymbol{X})\Big)\Big]. \qquad \text{(S1.4)}$$

Case 1. $\boldsymbol{\beta}_{0,J^C} \neq \mathbf{0}$. In this case,

$$\begin{aligned}
\frac{\big\{\mathbb{P}_n[W\hat{U}_k(Y - \hat{\phi}_n(\boldsymbol{X}))]\big\}^2}{\mathbb{P}_n(W\hat{U}_k)^2} \xrightarrow{P}& \frac{\{E[WU_k(Y - \tilde{\phi}(\boldsymbol{X}))]\}^2}{E(WU_k)^2}\\
=&\frac{\{E[W^2U_k(\boldsymbol{U}^{\mathsf{T}}\boldsymbol{\beta}_{0,J^C})]\}^2}{E(WU_k)^2}\\
=&\mathrm{Var}(W\boldsymbol{U}^{\mathsf{T}}\boldsymbol{\beta}_{0,J^C})[\mathrm{Corr}(WU_k, W\boldsymbol{U}^{\mathsf{T}})\boldsymbol{\beta}_{0,J^C}]^2,
\end{aligned}$$

which is maximized at unique k'_0 when $\boldsymbol{\beta}_{0,J^C} \neq \mathbf{0}$. Since $\hat{k}'_n$ maximizes the left hand side of the above display, it follows immediately that $\hat{k}'_n \xrightarrow{P} k'_0$ as $n \to \infty$. Hence

$$\begin{aligned}
n^{1/2}(\hat{\theta}'_n - \theta'_0) =& n^{1/2}(\hat{\theta}'_{k'_0} - \theta'_{k'_0}) + o_P(1)\\
=&\frac{\mathbb{G}_n[WU_{k'_0}(Y - \tilde{\phi}(\boldsymbol{X}) - M_{k'_0})]}{P(WU_{k'_0})^2} + o_P(1).
\end{aligned}$$

The proof is completed by using Slutsky's lemma and the CLT and noticing that $\theta'_k = E[WU_k(Y - E(Y|\boldsymbol{X}))]/E(WU_k)^2 = E[WU_k(Y - \tilde{\phi}(\boldsymbol{X}))]/E(WU_k)^2$.

Case 2. $\boldsymbol{\beta}_{0,J^C} = \mathbf{0}$**.** In this case $\theta'_k = 0$ for all $k \in J^C$. Thus

$$\mathbb{P}_n\big(W\hat{U}_k\big)^2 n^{1/2}\hat{\theta}'_k = n^{1/2}\mathbb{P}_n[W\hat{U}_k(Y - \hat{\phi}_n(\boldsymbol{X}))] = \mathbb{G}_n[WU_k(Y - \tilde{\phi}(\boldsymbol{X}) - M_k)] + o_P(1).$$

This implies that

$$\begin{aligned}
&\begin{pmatrix} n^{1/2}\hat{\theta}'_k \\ n(\mathbb{P}_n[W\hat{U}_k(Y - \hat{\phi}_n(\boldsymbol{X}))])^2/\mathbb{P}_n(W\hat{U}_k)^2 \end{pmatrix}_{k \in J^C} \\
&= \left(\frac{1}{\mathbb{P}_n(W\hat{U}_k)^2} \begin{pmatrix} \mathbb{G}_n[WU_k(Y - \tilde{\phi}(\boldsymbol{X}) - M_k)] \\ (\mathbb{G}_n[WU_k(Y - \tilde{\phi}(\boldsymbol{X}) - M_k)])^2 \end{pmatrix} \right)_{k \in J^C} + o_P(1) \\
&\xrightarrow{d} \begin{pmatrix} Z_k/E(WU_k)^2 \\ Z_k^2/E(WU_k)^2 \end{pmatrix}_{k \in J^C}, \qquad \text{(S1.5)}
\end{aligned}$$

where $\{Z_k : k \in J^C\}$ is a normal random vector with covariance matrix given by that of the random vector with components $\{WU_k(Y - \tilde{\phi}(\boldsymbol{X}) - M_k), k \in J^C\}$.

For any $\mathbf{t} \in \mathbb{R}^{|J^C|}$, let $h(\mathbf{t})$ be a J^C-dimensional vector of zeros, except a 1 at the maximal element of $\mathbf{t}$. We can re-write $n^{1/2}\hat{\theta}'_n$ as

$$n^{1/2}\hat{\theta}'_n = n^{1/2}(\{\hat{\theta}_k\}_{k \in J^C})^\mathsf{T} h\big(\{n(\mathbb{P}_n[W\hat{U}_k(Y - \hat{\phi}_n(\boldsymbol{X}))])^2/\mathbb{P}_n(W\hat{U}_k)^2\}_{k \in J^C}\big).$$

Under Assumption (C4), we have $|\mathrm{Corr}(WU_k, WU_j)| < 1$ for $k \neq j$, and thus $|\mathrm{Corr}(Z_k, Z_j)| < 1$. Since $\{Z_k : k \in J^C\}$ is a normal random vector, we have

$$\frac{Z_j^2}{E(WU_j)^2} \neq \frac{Z_k^2}{E(WU_k)^2} \text{ for any } j \neq k \ \text{ a.s.} \tag{S1.6}$$

So K is unique a.s. Thus h is continuous at $\left(\{Z_k^2/E(WU_k)^2\}_{k\in J^C}\right)$ a.s. And the result follows by applying the continuous mapping theorem to (S1.5).

Proof for part ii) of Theorem 3.

We use $\mathbb{P}_m^*$ to denote average over the bootstrap sample of size m, and $\mathbb{G}_m^* = \sqrt{m}(\mathbb{P}_m^* - \mathbb{P}_n)$. In the case of $\mathbb{Z}_{n,k}$ in which ϵ is not observed, we also replace ϵ by $\hat{\epsilon}_n \equiv Y - \hat{\alpha}_n - \hat{\theta}_n\hat{X}$, resulting in $\mathbb{Z}_{n,k}^* = \mathbb{G}_n^*[\hat{\epsilon}_n(X_k - \mathbb{P}_n^* X_k)]$ where $\mathbb{G}^* = \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$.

Let E^M denote expectation conditional on the data, and let P^M be the corresponding probability measure. The bootstrap analog of $\hat{\theta}_n'$ is $\hat{\theta}_m'^* = \hat{\theta}_{\hat{k}_m'}'^*$, where

$$\begin{aligned}
\hat{k}_m'^* &= \arg\min_{k\in J^C} \mathbb{P}_m^*[Y - \hat{\phi}_n(\boldsymbol{X}) - (\hat{\alpha}_k'^* + \hat{\theta}_k'^*\hat{U}_k^*)W]^2 \\
&= \arg\max_{k\in J^C} \frac{\left\{\mathbb{P}_m^*[W\hat{U}_k^*(Y - \hat{\phi}_n(\boldsymbol{X}))]\right\}^2}{\mathbb{P}_m^*(W\hat{U}_k^*)^2}, \\
(\hat{\alpha}_k'^*, \hat{\theta}_k'^*) &= \arg\min_{(\alpha,\theta)} \mathbb{P}_m^*[Y - \hat{\phi}_n(\boldsymbol{X}) - (\alpha + \theta\hat{U}_k^*)W]^2, \\
&\text{and } \hat{U}_k^* = X_k - \widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\hat{\boldsymbol{\gamma}}_k^* \text{ with } \hat{\boldsymbol{\gamma}}_k^* = \arg\min_{\boldsymbol{\gamma}} \mathbb{P}_m^*[W(X_k - \widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\boldsymbol{\gamma})]^2.
\end{aligned}$$

By first order conditions, $\mathbb{P}_m^*\{W(1, \hat{U}_k^*)^{\mathsf{T}}[Y - \hat{\phi}_n(\boldsymbol{X}) - (\hat{\alpha}_k'^* + \hat{\theta}_k'^*\hat{U}_k^*)W]\} = 0.$

In addition, by the definition of $\hat{U}_k^*$, $\mathbb{P}_m^* W^2 \hat{U}_k^* = 0$. This implies that

$$
\begin{aligned}
&\mathbb{P}_m^*\big(W\hat{U}_k^*\big)^2 m^{1/2}(\hat{\theta}_k'^* - \hat{\theta}_k') \\
=& m^{1/2}\mathbb{P}_m^*\left[W\hat{U}_k^*\left(Y - \hat{\phi}_n(\boldsymbol{X}) - \theta_k' W\hat{U}_k^*\right)\right] + \mathbb{P}_m^*\big(W\hat{U}_k^*\big)^2 m^{1/2}(\theta_k' - \hat{\theta}_k') \\
=& \mathbb{G}_m^*\left[WU_k\left(Y - \tilde{\phi}(\boldsymbol{X}) - \theta_k' WU_k\right)\right] \\
&+ m^{1/2}\mathbb{P}_m^*\left[W(\hat{U}_k^* - U_k)\left(Y - \tilde{\phi}(\boldsymbol{X}) - \theta_k' WU_k\right)\right] \\
&+ m^{1/2}\mathbb{P}_m^*\left[W\hat{U}_k^*\left(\tilde{\phi}(\boldsymbol{X}) - \hat{\phi}_n(\boldsymbol{X}) + \theta_k' W(U_k - \hat{U}_k^*)\right)\right] \\
&+ (m/n)^{1/2}\mathbb{G}_n\left[WU_k\left(Y - \tilde{\phi}(\boldsymbol{X}) - \theta_k' WU_k\right)\right] \\
&+ \mathbb{P}_m^*\big(W\hat{U}_k^*\big)^2 m^{1/2}(\theta_k' - \hat{\theta}_k') \qquad \text{(S1.7)}
\end{aligned}
$$

By the definition of $\hat{U}_k^*$ and U_k, It is easy to verify that $\mathbb{P}_m^* W^2 \hat{U}_k^* \widetilde{\boldsymbol{X}}_J = \mathbf{0}$ and $EW^2 U_k \widetilde{\boldsymbol{X}}_J = \mathbf{0}$. Thus

$$
\begin{aligned}
&m^{1/2}(\hat{U}_k^* - U_k) \\
=& - m^{1/2}\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\left[\mathbb{P}_m^*\left(W^2\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right)\right]^{-1}\mathbb{P}_m^*\left(W^2 U_k\widetilde{\boldsymbol{X}}_J\right) \\
=& - \widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\left[\mathbb{P}_m^*\left(W^2\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right)\right]^{-1}\left[\mathbb{G}_m^*\left(W^2 U_k\widetilde{\boldsymbol{X}}_J\right) + (m/n)^{1/2}\mathbb{G}_n\left(W^2 U_k\widetilde{\boldsymbol{X}}_J\right)\right].
\end{aligned}
$$

So the second term in (S1.7) equals

$$
\begin{aligned}
&-\mathbb{P}_m^*\left[W\left(Y-\tilde{\phi}(\boldsymbol{X})-\theta_k' W U_k\right)\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right]\left[\mathbb{P}_m^*\left(W^2\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right)\right]^{-1}\\
&\qquad\qquad\times\left[\mathbb{G}_m^*\left(W^2U_k\widetilde{\boldsymbol{X}}_J\right)+(m/n)^{1/2}\mathbb{G}_n\left(W^2U_k\widetilde{\boldsymbol{X}}_J\right)\right]\\
=&-E\left[W\left(Y-\tilde{\phi}(\boldsymbol{X})\right)\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right]\left[E\left(W^2\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right)\right]^{-1}\\
&\qquad\qquad\times\left[\mathbb{G}_m^*\left(W^2U_k\widetilde{\boldsymbol{X}}_J\right)+(m/n)^{1/2}\mathbb{G}_n\left(W^2U_k\widetilde{\boldsymbol{X}}_J\right)\right]+o_{PM}(1),
\end{aligned}
\tag{S1.8}
$$

conditionally on the data, in probability. And the third term in (S1.7) equals

$$
\begin{aligned}
&m^{1/2}\mathbb{P}_m^*\left[W\hat{U}_k^*\left(\tilde{\phi}(\boldsymbol{X})-\hat{\phi}_n(\boldsymbol{X})\right)\right]\\
=&m^{1/2}\mathbb{P}_m^*\left[WU_k\left(\tilde{\phi}(\boldsymbol{X})-\hat{\phi}_n(\boldsymbol{X})\right)\right]-\mathbb{P}_m^*\left[W\left(\tilde{\phi}(\boldsymbol{X})-\hat{\phi}_n(\boldsymbol{X})\right)\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right]\\
&\qquad\times\left[\mathbb{P}_m^*\left(W^2\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right)\right]^{-1}\left[\mathbb{G}_m^*\left(W^2U_k\widetilde{\boldsymbol{X}}_J\right)+(m/n)^{1/2}\mathbb{G}_n\left(W^2U_k\widetilde{\boldsymbol{X}}_J\right)\right]\\
=&\mathbb{G}_m^*\left[WU_k\left(\tilde{\phi}(\boldsymbol{X})-\hat{\phi}_n(\boldsymbol{X})\right)\right]+(m/n)^{1/2}\mathbb{G}_n\left[WU_k\left(\tilde{\phi}(\boldsymbol{X})-\hat{\phi}_n(\boldsymbol{X})\right)\right]\\
&-\mathbb{P}_m^*\left[W\left(\tilde{\phi}(\boldsymbol{X})-\hat{\phi}_n(\boldsymbol{X})\right)\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right]\left[\mathbb{P}_m^*\left(W^2\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right)\right]^{-1}\\
&\qquad\qquad\times\left[\mathbb{G}_m^*\left(W^2U_k\widetilde{\boldsymbol{X}}_J\right)+(m/n)^{1/2}\mathbb{G}_n\left(W^2U_k\widetilde{\boldsymbol{X}}_J\right)\right],
\end{aligned}
$$

which converges to zero, conditionally on the data, in probability, under Assumptions (C2) and (C3). The last term in (S1.7) equals

$$
\begin{aligned}
&\mathbb{P}_n\left(W\hat{U}_k\right)^2m^{1/2}(\theta_k'-\hat{\theta}_k')+\left[\mathbb{P}_m^*\left(W\hat{U}_k^*\right)^2-\mathbb{P}_n\left(W\hat{U}_k\right)^2\right]m^{1/2}(\theta_k'-\hat{\theta}_k')\\
=&-(m/n)^{1/2}\mathbb{G}_n[WU_k(Y-\tilde{\phi}(\boldsymbol{X})-M_k)]+o_{PM}(1),
\end{aligned}
\tag{S1.9}
$$

where $M_k = \theta'_k W U_k + W\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\left(PW^2\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right)^{-1} P\left[W\widetilde{\boldsymbol{X}}_J\left(Y - \tilde{\phi}(\boldsymbol{X})\right)\right]$ is defined in (S1.4). Plugging in (S1.8) and (S1.9) into (S1.7) yields

$$\mathbb{P}_m^*\big(W\hat{U}_k^*\big)^2 m^{1/2}(\hat{\theta}_k'^* - \hat{\theta}_k') = \mathbb{G}_m^*\Big[WU_k\big(Y - \tilde{\phi}(\boldsymbol{X}) - M_k\big)\Big] + o_{P^M}(1), \tag{S1.10}$$

conditionally on the data, in probability.

When $\boldsymbol{\beta}_{0,C} \neq 0$, it is easy to verify that

$$\begin{aligned}
&\frac{\big\{\mathbb{P}_m^*[W\hat{U}_k^*(Y - \hat{\phi}_n(\boldsymbol{X}))]\big\}^2}{\mathbb{P}_m^*(W\hat{U}_k^*)^2}\\
=&\frac{\Big\{\mathbb{P}_m^*[WU_k(Y - \tilde{\phi}(\boldsymbol{X}))] + \mathbb{P}_m^*\big[W(\hat{U}_k^* - U_k)(Y - \hat{\phi}_n(\boldsymbol{X})) + WU_k(\tilde{\phi}(\boldsymbol{X}) - \hat{\phi}_n(\boldsymbol{X}))\big]\Big\}^2}{\mathbb{P}_m^*(WU_k)^2 + \mathbb{P}_m^*(W\hat{U}_k^* - WU_k)^2 + 2\mathbb{P}_m^*[W^2U_k(\hat{U}_k^* - U_k)]}\\
\xrightarrow{P^M}&\frac{\{E[WU_k(Y - \tilde{\phi}(\boldsymbol{X}))]\}^2}{E(WU_k)^2} = \mathrm{Var}(W\boldsymbol{U}^{\mathsf{T}}\boldsymbol{\beta}_{0,J^C})[\mathrm{Corr}(WU_k, W\boldsymbol{U}^{\mathsf{T}})\boldsymbol{\beta}_{0,J^C}]^2
\end{aligned}$$

conditionally on the data, a.s. for $k \in J^C$. This implies that

$$\begin{aligned}
&P^M(\hat{k}_m'^* \neq k_0')\\
=&P^M\left(\bigcup_{k:k\neq k_0'}\left\{\frac{\big\{\mathbb{P}_m^*[W\hat{U}_{k_0'}^*(Y - \hat{\phi}_n(\boldsymbol{X}))]\big\}^2}{\mathbb{P}_m^*(W\hat{U}_{k_0'}^*)^2} \leq \frac{\big\{\mathbb{P}_m^*[W\hat{U}_k^*(Y - \hat{\phi}_n(\boldsymbol{X}))]\big\}^2}{\mathbb{P}_m^*(W\hat{U}_k^*)^2}\right\}\right)\\
\leq&\sum_{k:k\neq k_0} P^M\left(\frac{\big\{\mathbb{P}_m^*[W\hat{U}_{k_0'}^*(Y - \hat{\phi}_n(\boldsymbol{X}))]\big\}^2}{\mathbb{P}_m^*(W\hat{U}_{k_0'}^*)^2} \leq \frac{\big\{\mathbb{P}_m^*[W\hat{U}_k^*(Y - \hat{\phi}_n(\boldsymbol{X}))]\big\}^2}{\mathbb{P}_m^*(W\hat{U}_k^*)^2}\right)\\
\to& 0 \quad \text{a.s.},
\end{aligned}$$

where the convergence follows from the condition that, when $\boldsymbol{\beta}_{0,J^C} \neq \boldsymbol{0}$, k_0' uniquely maximizes $|\mathrm{Corr}(WU_k, W\boldsymbol{U}^{\mathsf{T}})\boldsymbol{\beta}_{0,J^C}|$. This, together with (S1.10) and $\hat{k}'_n \xrightarrow{P} k'_0$, yields

$$\begin{aligned} m^{1/2}(\hat{\theta}'^*_m - \hat{\theta}'_n)\mathbb{P}^*_m(W\hat{U}^*_{\hat{k}'^*_m})^2 =& m^{1/2}(\hat{\theta}'^*_{k'_0} - \hat{\theta}_{k'_0})\mathbb{P}^*_m(W\hat{U}^*_{k'_0})^2 + o_{PM}(1) \\ =& \mathbb{G}^*_m\left[WU_{k'_0}\big(Y - \tilde{\phi}(\boldsymbol{X}) - M_{k'_0}\big)\right] + o_{PM}(1) \end{aligned} \tag{S1.11}$$

conditionally in probability. The result follows from bootstrap CLT, continuous mapping theorem, and Slutsky's lemma.

When $\boldsymbol{\beta}_{0,J^C} = \mathbf{0}$, $\theta'_k = 0$ for $k \in J^C$ and $\theta'_0 = 0$. In this case, we need $m/n = o(1)$. Thus

$$m^{1/2}(\hat{\theta}'^*_m - \hat{\theta}'_n) = m^{1/2}\hat{\theta}'^*_m - m^{1/2}(\hat{\theta}'_n - \theta'_0) = m^{1/2}\hat{\theta}'^*_m + o_P(1),$$

$$\begin{aligned} \text{and} & \begin{pmatrix} m^{1/2}\hat{\theta}'^*_k \\ \left\{m^{1/2}\mathbb{P}^*_m[W\hat{U}^*_k(Y - \hat{\phi}_n(\boldsymbol{X}))]\right\}^2/\mathbb{P}^*_m(W\hat{U}^*_k)^2 \end{pmatrix}_{k\in J^C} \\ =& \left(\frac{1}{\mathbb{P}^*_m(W\hat{U}^*_k)^2}\begin{pmatrix} m^{1/2}\mathbb{P}^*_m[W\hat{U}^*_k(Y - \hat{\phi}_n(\boldsymbol{X}) - \theta'_k W\hat{U}^*_k)] \\ \left\{m^{1/2}\mathbb{P}^*_m[W\hat{U}^*_k(Y - \hat{\phi}_n(\boldsymbol{X}) - \theta'_k W\hat{U}^*_k)]\right\}^2 \end{pmatrix}\right)_{k\in J^C} \\ =& \left(\frac{1}{\mathbb{P}^*_m(W\hat{U}^*_k)^2}\begin{pmatrix} \mathbb{G}^*_m[WU_k(Y - \tilde{\phi}(\boldsymbol{X}) - M_k)] \\ (\mathbb{G}^*_m[WU_k(Y - \tilde{\phi}(\boldsymbol{X}) - M_k)])^2 \end{pmatrix}\right)_{k\in J^C} + o_{PM}(1) \\ \xrightarrow{d}& \begin{pmatrix} Z_k/E(WU_k)^2 \\ Z_k^2/E(WU_k)^2 \end{pmatrix}_{k\in J^C}. \end{aligned}$$

conditionally on the data, in probability. The result follows by using similar arguments are those at the end of the proof of part i).

S2 Assumptions and proofs of Theorem 4

Assumptions.

(A1) $EX_k^4 < \infty$ for $k = 1, \dots, p$.

(A2) There exist functions $\tilde{h}(\boldsymbol{X})$ and $\tilde{q}(\boldsymbol{X})$ such that $n^{1/2}[\hat{h}_n(\boldsymbol{x}) - \tilde{h}(\boldsymbol{x})] = \Delta_h(\boldsymbol{x})\hat{S}_h + o_P(1)$ and $n^{1/2}[\hat{q}_n(\boldsymbol{x}) - \tilde{q}(\boldsymbol{x})] = \Delta_q(\boldsymbol{x})\hat{S}_q + o_P(1)$, where $\Delta_h(\boldsymbol{x})$ and $\Delta_q(\boldsymbol{x})$ are vector-valued deterministic functions of $\boldsymbol{x}$, and $\hat{S}_h$ and $\hat{S}_q$ are data dependent random vectors satisfying

i). $\Delta_h(\boldsymbol{X})$ and $\Delta_q(\boldsymbol{X})$ are square integrable random vectors; and

ii)
$$\left(\left\{\mathbb{G}_n\left[\widetilde{W}L_k\left(Y - \tilde{h}(\boldsymbol{X}) - \psi_k A L_k - E\left[\widetilde{W}\left(Y - \tilde{h}(\boldsymbol{X})\right)\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right]\left[E\left(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\right)\right]^{-1} \times A\widetilde{\boldsymbol{X}}_J\right)\right]\right\}_{k\in J^C}, \hat{S}_h,\, \hat{S}_q\right)^{\mathsf{T}} \xrightarrow{d} (\{Z_k^o : k \in J^C\}, S_h, S_q)^{\mathsf{T}} \sim N(0, \Sigma^o)$$

for some variance-covariance matrix Σ^o assumed to exist.

(A3) The error term ϵ in model (2.1) has mean zero, finite variance, and is uncorrelated with $(\widetilde{W}, \widetilde{W}\boldsymbol{X})$, where $\widetilde{W} = A - \tilde{q}(\boldsymbol{X})$.

(A4) k_0^o is unique when $\boldsymbol{\beta}_{0,J^C} \neq 0$.

(A5) $\tilde{q}(\boldsymbol{X}) = q_0(\boldsymbol{X})$ or $\tilde{h}(\boldsymbol{X}) = h_0(\boldsymbol{X})$ a.s.

(A6) $m^{1/2}[\hat{h}_m^*(\boldsymbol{x}) - \hat{h}_n(\boldsymbol{x})] = \Delta_h(\boldsymbol{x})\hat{S}_h^* + o_{P_M}(1)$ and $m^{1/2}[\hat{q}_m^*(\boldsymbol{x}) - \hat{q}_n(\boldsymbol{x})] = \Delta_q(\boldsymbol{x})\hat{S}_q^* + o_{P_M}(1)$ conditionally on the data (in probability), where

$\Delta_h(\boldsymbol{x})$ and $\Delta_q(\boldsymbol{x})$ are defined in Assumption (A2), and $\hat{S}_h^*$ and $\hat{S}_q^*$ are bootstrap sample dependent random vectors satisfying

$$\left(\left\{\mathbb{G}_m^*\Big[\widetilde{W}L_k\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k - E\Big[\widetilde{W}\Big(Y-\tilde{h}(\boldsymbol{X})\Big)\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big]\Big[E\Big(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big)\Big]^{-1} A\widetilde{\boldsymbol{X}}_J\Big)\Big]\right\}_{k\in J^C}, \hat{S}_h^*, \hat{S}_q^*\right)^{\mathsf{T}} \xrightarrow{d} (\{Z_k^o : k\in J^C\}, S_h, S_q)^{\mathsf{T}} \sim N(0,\Sigma^o)$$

conditional on the data, in probability.

Proof for part i) of Theorem 4.

First note that for $k \in J^C$,

$$\begin{aligned}
\mathbb{P}_n[A\widehat{W}\hat{L}_k^2]n^{1/2}(\hat{\psi}_k-\psi_k) &= n^{1/2}\mathbb{P}_n\Big[\widehat{W}\hat{L}_k\Big(Y-\hat{h}_n(\boldsymbol{X})-\psi_k A\hat{L}_k\Big)\Big] \\
&= \mathbb{G}_n\Big[\widetilde{W}L_k\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k\Big)\Big] + n^{1/2}\mathbb{P}_n\Big[\widehat{W}(\hat{L}_k-L_k)\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k\Big)\Big] \\
&\quad + n^{1/2}\mathbb{P}_n\Big[(\widehat{W}-\widetilde{W})L_k\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k\Big)\Big] \\
&\quad + n^{1/2}\mathbb{P}_n\Big[\widehat{W}\hat{L}_k\Big(\tilde{h}(\boldsymbol{X})-\hat{h}_n(\boldsymbol{X})+\psi_k A(L_k-\hat{L}_k)\Big)\Big]. \qquad \text{(S2.12)}
\end{aligned}$$

Recall that $L_k = X_k - \widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\boldsymbol{\eta}_k$ and $\hat{L}_k = X_k - \widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\hat{\boldsymbol{\eta}}_k$, where $\boldsymbol{\eta}_k = \arg\min_{\boldsymbol{\eta}} E\big[A\widetilde{W}(X_k - \widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\boldsymbol{\eta})^2\big]$ and $\hat{\boldsymbol{\eta}}_k = \arg\min_{\boldsymbol{\eta}} \mathbb{P}_n\big[A\widehat{W}(X_k - \widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\boldsymbol{\eta})^2\big]$, respectively. First order conditions implies that $E(A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J)=0$, $\mathbb{P}_n(A\widehat{W}\hat{L}_k\widetilde{\boldsymbol{X}}_J)=0$, and

$$\begin{aligned}
n^{1/2}(\hat{\boldsymbol{\eta}}_k-\boldsymbol{\eta}_k) &= [\mathbb{P}_n(A\widehat{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}})]^{-1}\Big(\mathbb{G}_n[A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J] + n^{1/2}\mathbb{P}_n[A(\widehat{W}-\widetilde{W})L_k\widetilde{\boldsymbol{X}}_J]\Big) \\
&= [E(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}})]^{-1}\Big(\mathbb{G}_n[A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J] - E\big[AL_k\widetilde{\boldsymbol{X}}_J\Delta_q(\boldsymbol{X})\big]\hat{S}_q\Big) + o_P(1),
\end{aligned}$$

where the second equality follows from Assumptions (A1) and (A2). Thus the second term of (S2.12) equals

$$-E\Big[\widetilde{W}\Big(Y-\tilde{h}(\boldsymbol{X})\Big)\widetilde{\boldsymbol{X}}_J^{\intercal}\Big]\Big[E\Big(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\intercal}\Big)\Big]^{-1}$$
$$\Big(\mathbb{G}_n[A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J]-E\big[AL_k\widetilde{\boldsymbol{X}}_J\Delta_q(\boldsymbol{X})\big]\hat{S}_q\Big)+o_P(1),$$

the third term of (S2.12) equals $-E\Big[L_k\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k\Big)\Delta_q(\boldsymbol{X})\Big]\hat{S}_q+o_P(1)$, and the fourth term equals $-E\left[\widetilde{W}L_k\Delta_h(\boldsymbol{X})\right]\hat{S}_h+o_P(1)$. Plugging these into (S2.12), we have

$$\begin{aligned}
&\mathbb{P}_n[A\widehat{W}\hat{L}_k^2]n^{1/2}(\hat{\psi}_k-\psi_k)\\
=&\mathbb{G}_n\Big[\widetilde{W}L_k\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k-E\Big[\widetilde{W}\Big(Y-\tilde{h}(\boldsymbol{X})\Big)\widetilde{\boldsymbol{X}}_J^{\intercal}\Big]\Big[E\Big(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\intercal}\Big)\Big]^{-1}A\widetilde{\boldsymbol{X}}_J\Big)\Big]\\
&-E\Bigg\{L_k\left(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k-E\Big[\widetilde{W}\Big(Y-\tilde{h}(\boldsymbol{X})\Big)\widetilde{\boldsymbol{X}}_J^{\intercal}\Big]\Big[E\Big(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\intercal}\Big)\Big]^{-1}A\widetilde{\boldsymbol{X}}_J\right)\\
&\qquad\Delta_q(\boldsymbol{X})\Bigg\}\hat{S}_q-E\Big[\widetilde{W}L_k\Delta_h(\boldsymbol{X})\Big]\hat{S}_h+o_P(1)
\end{aligned}\tag{S2.13}$$

Case 1. $\boldsymbol{\beta}_{0,J^C}\neq\mathbf{0}$. In this case, (S2.13) implies that $\hat{\psi}_k\xrightarrow{P}\psi_k$. In addition, it is easy to verify that $\mathbb{P}_n(A\widehat{W}\hat{L}_k^2)\xrightarrow{p}E(A\widetilde{W}L_k^2)$. By Slutsky's Lemma,

$$\hat{\psi}_k^2\mathbb{P}_n(A\widehat{W}L_k^2)\xrightarrow{p}\psi_k^2E(A\widetilde{W}L_k^2),$$

which is maximized at unique k_0^o when $\boldsymbol{\beta}_{0,J^C}\neq\mathbf{0}$ by Assumption (A4). Since $\hat{k}_n^o=\arg\max_{kinJ^C}\left[\hat{\psi}_k^2\mathbb{P}_n(A\widehat{W}L_k^2)\right]$, it follows immediately that $\hat{k}_n^o\xrightarrow{P}k_0^o$. Hence

$$n^{1/2}(\hat{\psi}_n-\psi_0)=n^{1/2}(\hat{\psi}_{k_0^o}-\psi_{k_0^o})+o_P(1)$$

The result follows from Assumption (A2')

Case 2. $\boldsymbol{\beta}_{0,J^C} = \mathbf{0}$**.** In this case $\psi_k = 0$ for $k \in J^C$ under Assumptions (A3) and (A5). Thus $n\hat{\psi}_k^2\mathbb{P}_n(A\widehat{W}\hat{L}_k^2) = [n^{1/2}(\hat{\psi}_k - \psi_k)]^2\mathbb{P}_n(A\widehat{W}\hat{L}_k^2)$. The result follows using similar techniques as those in the proof of part i) of Theorem 3.

Proof for part ii) of Theorem 4.

For $k \in J^C$, let $(\hat{\boldsymbol{\delta}}_k^*, \hat{\psi}_k^*)$ be the solution to

$$\mathbb{P}_m^*\Big[(\widetilde{\boldsymbol{X}}_J^\top, X_k)^\top\widehat{W}^*\big(Y - \hat{h}_m^*(\boldsymbol{X}) - (\widetilde{\boldsymbol{X}}_J^\top\boldsymbol{\delta} + X_k\psi)A\big)\Big] = 0,$$

where $\widehat{W}^* = A - \hat{q}_m^*(\boldsymbol{X})$. Then

$$\hat{\psi}_k^* = \mathbb{P}_m^*\Big[\widehat{W}^*(Y - \hat{h}_m^*(\boldsymbol{X}))\hat{L}_k^*\Big]\Big/\mathbb{P}_m^*\Big[A\widehat{W}^*(\hat{L}_k^*)^2\Big],$$

where $\hat{L}_k^* = X_k - \widetilde{\boldsymbol{X}}_J^\top\hat{\boldsymbol{\eta}}_k^*$ and $\hat{\boldsymbol{\eta}}_k^* = \arg\min_{\boldsymbol{\eta}}\mathbb{P}_m^*[A\widehat{W}^*(X_k - \widetilde{\boldsymbol{X}}_J^\top\boldsymbol{\eta})^2]$. The m-out-of-n bootstrap analog of $\hat{\psi}_n$ is

$$\hat{\psi}_m^* = \hat{\psi}_{\hat{k}_m^{o*}}^*, \text{ where } \hat{k}_n^{o*} = \arg\max_{k\in J^C}\big\{(\hat{\psi}_k^*)^2\mathbb{P}_m^*[A\widehat{W}^*(\hat{L}_k^*)^2]\big\}.$$

For $k \in J^C$, note that $E\Big[\widetilde{W}L_k\Big(Y - \tilde{h}(\boldsymbol{X}) - \psi_k AL_k\Big)\Big] = 0$. Thus

$$
\begin{aligned}
&\mathbb{P}_m^*[A\widehat{W}^*(\hat{L}_k^*)^2]m^{1/2}(\hat{\psi}_k^* - \hat{\psi}_k)\\
=&\, m^{1/2}\mathbb{P}_m^*\Big[\widehat{W}^*\hat{L}_k^*\Big(Y - \hat{h}_m^*(\boldsymbol{X}) - \psi_k A\hat{L}_k^*\Big)\Big] + m^{1/2}\mathbb{P}_m^*\Big[A\widehat{W}^*(\hat{L}_k^*)^2(\psi_k - \hat{\psi}_k)\Big]\\
=&\, \mathbb{G}_m^*\Big[\widetilde{W}L_k\Big(Y - \tilde{h}(\boldsymbol{X}) - \psi_k AL_k\Big)\Big] + m^{1/2}\mathbb{P}_m^*\Big[\widehat{W}^*(\hat{L}_k^* - L_k)\Big(Y - \tilde{h}(\boldsymbol{X}) - \psi_k AL_k\Big)\Big]\\
&+ m^{1/2}\mathbb{P}_m^*\Big[(\widehat{W}^* - \widetilde{W})L_k\Big(Y - \tilde{h}(\boldsymbol{X}) - \psi_k AL_k\Big)\Big]\\
&+ m^{1/2}\mathbb{P}_m^*\Big[\widehat{W}^*\hat{L}_k^*\Big(\tilde{h}(\boldsymbol{X}) - \hat{h}_m^*(\boldsymbol{X}) + A\psi_k(L_k - \hat{L}_k^*)\Big)\Big]\\
&+ (m/n)^{1/2}\mathbb{G}_n\Big[\widetilde{W}L_k\Big(Y - \tilde{h}(\boldsymbol{X}) - \psi_k AL_k\Big)\Big] + m^{1/2}\mathbb{P}_m^*\Big[A\widehat{W}^*(\hat{L}_k^*)^2(\psi_k - \hat{\psi}_k)\Big]
\end{aligned} \tag{S2.14}
$$

By definition of $\hat{\boldsymbol{\eta}}_k^*$ and $\boldsymbol{\eta}_k$, we have

$$
\begin{aligned}
&m^{1/2}(\hat{\boldsymbol{\eta}}_k^* - \boldsymbol{\eta}_k)\\
=&\Big[\mathbb{P}_m^*\Big(A\widehat{W}^*\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big)\Big]^{-1}\Big[\mathbb{G}_m^*(A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J) + (m/n)^{1/2}\mathbb{G}_n(A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J)\\
&\qquad + m^{1/2}\mathbb{P}_m^*(A(\widehat{W}^* - \widetilde{W})L_k\widetilde{\boldsymbol{X}}_J)\Big]\\
=&\Big[\mathbb{P}_m^*\Big(A\widehat{W}^*\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big)\Big]^{-1}\Big[\mathbb{G}_m^*(A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J) + (m/n)^{1/2}\mathbb{G}_n(A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J)\\
&\qquad - \mathbb{P}_m^*[AL_k\widetilde{\boldsymbol{X}}_J\Delta_q(\boldsymbol{X})](\hat{S}_q^* + (m/n)^{1/2}\hat{S}_q) + o_{PM}(1)\Big]\\
=&\Big[E\Big(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big)\Big]^{-1}\Big[\mathbb{G}_m^*(A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J) + (m/n)^{1/2}\mathbb{G}_n(A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J)\\
&\qquad - E[AL_k\widetilde{\boldsymbol{X}}_J\Delta_q(\boldsymbol{X})](\hat{S}_q^* + (m/n)^{1/2}\hat{S}_q)\Big] + o_{PM}(1)
\end{aligned}
$$

conditional on the data, in probability. Thus the second term of (S2.14) equals

$$
\begin{aligned}
& -m^{1/2}\mathbb{P}_m^*\Big[\widehat{W}^*\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k\Big)\Big](\hat{\boldsymbol{\eta}}_k^*-\boldsymbol{\eta}_k) \\
=& -E\Big[\widetilde{W}\Big(Y-\tilde{h}(\boldsymbol{X})\Big)\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big]\Big[E\Big(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big)\Big]^{-1}\Big[\mathbb{G}_m^*(A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J) \\
& + (m/n)^{1/2}\mathbb{G}_n(A\widetilde{W}L_k\widetilde{\boldsymbol{X}}_J) - E[AL_k\widetilde{\boldsymbol{X}}_J\Delta_q(\boldsymbol{X})](\hat{S}_q^*+(m/n)^{1/2}\hat{S}_q)\Big] + o_{PM}(1).
\end{aligned}
\tag{S2.15}
$$

conditionally in probability. Similarly, the third term of (S2.14) equals

$$
-E\Big[L_k\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k\Big)\Delta_q(\boldsymbol{X})\Big]\big[\hat{S}_q^*+(m/n)^{1/2}\hat{S}_q\big]+o_{PM}(1)
\tag{S2.16}
$$

conditionally in probability, and the fourth term of (S2.14) equals

$$
-E\big[\widetilde{W}L_k\Delta_h(\boldsymbol{X})\big][\hat{S}_h^*+(m/n)^{1/2}\hat{S}_h]+o_{PM}(1) \tag{S2.17}
$$

conditionally in probability. Note that the last two terms in (S2.14) are negligible. Plugging (S2.13) and (S2.15)-(S2.17) into (S2.14), we have

$$
\begin{aligned}
& \mathbb{P}_m^*[A\widehat{W}^*(\hat{L}_k^*)^2]m^{1/2}(\hat{\psi}_k^*-\hat{\psi}_k) \\
=& \mathbb{G}_m^*\Big[\widetilde{W}L_k\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k - E\Big[\widetilde{W}\Big(Y-\tilde{h}(\boldsymbol{X})\Big)\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big]\Big[E\Big(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big)\Big]^{-1}A\widetilde{\boldsymbol{X}}_J\Big)\Big] \\
& - E\Big\{L_k\Big(Y-\tilde{h}(\boldsymbol{X})-\psi_k AL_k - E\Big[\widetilde{W}\Big(Y-\tilde{h}(\boldsymbol{X})\Big)\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big]\Big[E\Big(A\widetilde{W}\widetilde{\boldsymbol{X}}_J\widetilde{\boldsymbol{X}}_J^{\mathsf{T}}\Big)\Big]^{-1}A\widetilde{\boldsymbol{X}}_J\Big) \\
& \qquad \Delta_q(\boldsymbol{X})\Big\}\hat{S}_q^* - E\big[\widetilde{W}L_k\Delta_h(\boldsymbol{X})\big]\hat{S}_h^* + o_{P_M}(1)
\end{aligned}
$$

conditionally in probability. The result follows using similar arguments as those in the proof of Theorem 4.

S3 Non-Uniqueness of the Most Informative Predictor

Theorems 1-4 requires that the most informative predictor of $T(\boldsymbol{X})$ to be unique under H_a so that the parameter in the hypotheses are well defined. In fact, this condition can be removed with a slight modification of the parameter and test statistic. In this section, we demonstrate the extension of Theorem 1 to the case of non-unique k_0. Extension to other theorems can be derived in a similar fashion.

Denote the set of maximizers by $\mathcal{K}_0 := \arg\max_{k\in\{1,\ldots,p\}} |\mathrm{Corr}(WX_k', W\boldsymbol{X}^\top)\boldsymbol{\beta}_0|$. Note that

$$\mathrm{Corr}(WX_k', W\boldsymbol{X}^\top\boldsymbol{\beta}_0) = \left[\frac{E(WX_k')^2}{\mathrm{Var}(W\boldsymbol{X}^\top\boldsymbol{\beta}_0)}\right]^{1/2}\theta_k.$$

Since $\mathrm{Var}(W\boldsymbol{X}^\top\boldsymbol{\beta}_0)$ does not depend on k, we have $\mathcal{K}_0 = \arg\max_{k\in\{1,\ldots,p\}} |\left[E(WX_k')^2\right]^{1/2}\theta_k|$, and Hypothesis (2.4) in Section 2 is equivalent to

$$H_0 : \tau_0 = 0 \text{ vs. } H_a : \tau_0 > 0$$

where $\tau_0 = \max_{k\in\{1,\ldots,p\}} \left|\left[E(WX_k')^2\right]^{1/2}\theta_k\right|$, which can be estimated by $\hat{\tau} = \max_{k\in\{1,\ldots,p\}} \left|\left[\mathbb{P}_n(W\hat{X}_k')^2\right]^{1/2}\hat{\theta}_k\right|$.

Theorem S1. *Assume conditions (C1) - (C3) in Section S1 hold. Then* *under model (2.1),*

$$n^{1/2}\left(\hat{\tau}-\tau_0\right) \xrightarrow{d} \begin{cases} \max_{k\in\mathcal{K}_0}\left[2(1_{\theta_k>0}-1/2)Z_k'\right] & \textit{if } \tau_0>0 \\ \max_{k\in\{1,\ldots,p\}}|Z_k'| & \textit{if } \tau_0=0, \end{cases}$$

where $(Z_1',\ldots,Z_p')^{\mathsf{T}}\in\mathbb{R}^p$ *is a mean zero normal random vector. with covariance matrix given by that of the random vector with components*

$$\frac{WX_k'}{[E(WX_k')^2]^{1/2}}\left\{Y-\tilde{\phi}(\boldsymbol{X})-\frac{E[W(Y-\tilde{\phi}(\boldsymbol{X}))]}{EW^2}W-\frac{E[WX_k'(Y-\tilde{\phi}(\boldsymbol{X}))]}{2E(WX_k')^2}WX_k'\right\}.$$

Proof. Denote $\mathbb{Z}_{n,k}=n^{1/2}\left\{\left[\mathbb{P}_n(W\hat{X}_k')^2\right]^{1/2}\hat{\theta}_k-[E(WX_k')^2]^{1/2}\theta_k\right\}$.

First, consider under H_0. In this case, $[E(WX_k')^2]^{1/2}\theta_k=0$ for all k. Thus

$$\begin{aligned} & n^{1/2}(\hat{\tau}-\tau_0) \\ =& n^{1/2}\left(\max_{k\in\{1,\ldots,p\}}\left|\left[\mathbb{P}_n(W\hat{X}_k')^2\right]^{1/2}\hat{\theta}_k-\left[E(WX_k')^2\right]^{1/2}\theta_k\right|\right) \\ =& \max_{k\in\{1,\ldots,p\}}|\mathbb{Z}_{n,k}|. \end{aligned}$$

Second, consider under H_a. When $\mathcal{K}_0=\{1,\ldots,p\}$, $\left|[E(WX_k')^2]^{1/2}\theta_k\right|$ is positive and takes the same value for all k. We have

$$\begin{aligned} & n^{1/2}\left(\hat{\tau}-\tau_0\right) \\ =& \max_{k\in\{1,\ldots,p\}}\left(\left|\mathbb{Z}_{n,k}+n^{1/2}\left[E(WX_k')^2\right]^{1/2}\theta_k\right|-n^{1/2}\left|\left[E(WX_k')^2\right]^{1/2}\theta_k\right|\right) \\ =& \max_{k\in\{1,\ldots,p\}}\left[2(1_{\theta_k>0}-1/2)\mathbb{Z}_{n,k}\right]+o_P(1), \end{aligned}$$

where the second equality follows since $n^{1/2}\big|\left[E(WX_k')^2\right]^{1/2}\theta_k\big| \to \infty$ and $\mathbb{Z}_{n,k} = O_p(1)$ as $n \to \infty$.

When $\mathcal{K}_0 \neq \{1, \ldots, p\}$, denote

$$\delta_n := 1_{\left\{\max_{k\in\mathcal{K}_0^C}\left|\left[E(WX_k')^2\right]^{1/2}\theta_k\right| \geq \max_{k\in\mathcal{K}_0}\left|\left[E(WX_k')^2\right]^{1/2}\theta_k\right|\right\}}.$$

It is easy to see that

$$\begin{aligned}
&P\Big(\max_{k\in\mathcal{K}_0^C}\big|\left[E(WX_k')^2\right]^{1/2}\theta_k\big| \geq \max_{k\in\mathcal{K}_0}\big|\left[E(WX_k')^2\right]^{1/2}\theta_k\big|\Big)\\
\leq &P\left(\bigcup_{k\in\mathcal{K}_0^C}\bigcap_{j\in\mathcal{K}_0}\left\{\big|\left[E(WX_k')^2\right]^{1/2}\theta_k\big| \geq \big|\left[E(WX_j')^2\right]^{1/2}\theta_j\big|\right\}\right)\\
\leq &\sum_{k\in\mathcal{K}_0^C} P\left(\bigcap_{j\in\mathcal{K}_0}\left\{\big|\left[E(WX_k')^2\right]^{1/2}\theta_k\big| \geq \big|\left[E(WX_j')^2\right]^{1/2}\theta_j\big|\right\}\right) \to 0.
\end{aligned}$$

Thus $\delta_n = o_P(1)$. Therefore,

$$\begin{aligned}
&n^{1/2}\left(\hat{\tau} - \tau_0\right)\\
=&n^{1/2}\Big(\max_{k\in\mathcal{K}_0}\Big|\left[\mathbb{P}_n(W\hat{X}_k')^2\right]^{1/2}\hat{\theta}_k\Big|(1-\delta_n) + \max_{k\in\mathcal{K}_0^C}\Big|\left[\mathbb{P}_n(W\hat{X}_k')^2\right]^{1/2}\hat{\theta}_k\Big|\delta_n\\
&\qquad\qquad - \max_{k\in\mathcal{K}_0}\Big|\left[E(WX_k')^2\right]^{1/2}\theta_k\Big|\Big)\\
=&n^{1/2}\left(\max_{k\in\mathcal{K}_0}\Big|\left[\mathbb{P}_n(W\hat{X}_k')^2\right]^{1/2}\hat{\theta}_k\Big| - \max_{k\in\mathcal{K}_0}\Big|\left[E(WX_k')^2\right]^{1/2}\theta_k\Big|\right)\\
&+ n^{1/2}\left(\max_{k\in\mathcal{K}_0^C}\Big|\left[\mathbb{P}_n(W\hat{X}_k')^2\right]^{1/2}\hat{\theta}_k\Big| - \max_{k\in\mathcal{K}_0}\Big|\left[\mathbb{P}_n(W\hat{X}_k')^2\right]^{1/2}\hat{\theta}_k\Big|\right)\delta_n\\
=&\max_{k\in\mathcal{K}_0}\left[2(1_{\theta_k>0} - 1/2)\mathbb{Z}_{n,k}\right] + o_P(1)
\end{aligned}$$

where the last equality follows since

$$\begin{aligned}
0 \leq & n^{1/2} \left(\max_{k \in \mathcal{K}_0^C} \left| \left[\mathbb{P}_n (W \hat{X}_k')^2 \right]^{1/2} \hat{\theta}_k \right| - \max_{k \in \mathcal{K}_0} \left| \left[\mathbb{P}_n (W \hat{X}_k')^2 \right]^{1/2} \hat{\theta}_k \right| \right) \delta_n \\
\leq & n^{1/2} \max_{k \in \mathcal{K}_0^C} \left(\left| \left[\mathbb{P}_n (W \hat{X}_k')^2 \right]^{1/2} \hat{\theta}_k \right| - \left| \left[E (W X_k')^2 \right]^{1/2} \theta_k \right| \right) \delta_n \\
& - n^{1/2} \max_{k \in \mathcal{K}_0} \left(\left| \left[\mathbb{P}_n (W \hat{X}_k')^2 \right]^{1/2} \hat{\theta}_k \right| - \left| \left[E (W X_k')^2 \right]^{1/2} \theta_k \right| \right) \delta_n \\
\leq & 2 \max_{k \in \{1, \ldots, p\}} |\mathbb{Z}_{n,k}| \delta_n = o_P(1).
\end{aligned}$$

The result follows by showing that $(\mathbb{Z}_{n,1}, \ldots, \mathbb{Z}_{n,p})^{\mathsf{T}}$ converges in distribution to $(Z_1', \ldots, Z_p')^{\mathsf{T}}$, using arguments similar to that in Section S1.

□

S4 More on the Doubly Robust Method

Note that the doubly robust method presented in Section 4 can also be used for randomized trials. However, as compared to the method presented in Sections 2 and 3, this approach may cause dispersion in variance of the estimate. Below we illustrate this point at the initial step (i.e. when $J = \emptyset$).

Consider a trial where two treatments are randomized with equal probability $q_0(\boldsymbol{X}) = 1/2$. We estimate the propensity score by the sample proportion of patients who were assigned to treatment $A = 1$. In this case, $\tilde{q}(\boldsymbol{X}) = q_0(\boldsymbol{X}) = 1/2$ and $\widetilde{W} = W = 1_{A=1} - 1/2$. We can further verify that $X_k' = L_k = X - EX_k$, $\theta_k = \psi_k$, and $k_0 = k_0^o$, where θ_k and k_0 are defined in

Section 2 below equation (2.3), and ψ_k and k_0^o are defined in Section 4 by setting $J = \emptyset$. Thus the parameters in the hypotheses (2.4) and (4.2) are the same (i.e. $\theta_0 = \psi_0$).

To ensure a fair comparison, we further assume that the main effect can be consistently estimated by both methods, namely, $\tilde{\phi}(\boldsymbol{X}) = E(Y|\boldsymbol{X})$ and $\tilde{h}(\boldsymbol{X}) = E(Y|A = 0)$.

Denote $g_k(\boldsymbol{X}) := X_k'[(\boldsymbol{X} - E\boldsymbol{X})^{\mathsf{T}}\boldsymbol{\beta}_0 - \theta_k X_k']$. Based on Remark 2 of Theorem 1, it is easy to see that the asymptotic variance of $\hat{\theta}_k$ is

$$\mathrm{Var}\left(\frac{Z_k}{E[WX_k']^2}\right) = \frac{\mathrm{Var}\,(WX_k'\epsilon) + \mathrm{Var}\,[W^2 g_k(\boldsymbol{X})]}{\{E[WX_k']^2\}^2} = \frac{4\mathrm{Var}\,(WX_k'\epsilon) + \mathrm{Var}\,[g_k(\boldsymbol{X})]}{\{E[X_k']^2\}^2}.$$

Similarly, by Theorem 4, the asymptotic variance of $\hat{\psi}_k$ is

$$\begin{aligned}\mathrm{Var}\left(\frac{\tilde{Z}_k}{E(AWX_k'^2)}\right) =& \frac{\mathrm{Var}(WX_k'\epsilon) + \mathrm{Var}(W[Ag_k(X) - EAg_k(X)])}{[E(AWX_k'^2)]^2} \\ =& \frac{4\mathrm{Var}\,(WX_k'\epsilon) + 2\mathrm{Var}\,[g_k(\boldsymbol{X})]}{\{E[X_k']^2\}^2}.\end{aligned}$$

That is, $\hat{\psi}_k$ has larger asymptotic variance than $\hat{\theta}_k$. So we expect the doubly robust method to be more conservative than the method in Section 2 when applied to randomized trials.

S5 Details of simulations and tables

S5.1 Simulations for randomized trials

In the randomized trial setting, we compare the sampling from null (NULL), m-out-of-n bootstrap for known propensity score ($\hat{m}$-boot), and the doubly robust method ($\hat{m}$-boot-DR) procedures with the following competing methods.

Likelihood ratio test (LRT). This test is based on assuming a full linear model (3.1) of the interaction terms. At each step, under the null hypothesis, $\boldsymbol{\beta}_{0,J^C} = \mathbf{0}$, the reduction in the residual sum of squares is compared to the residual sum of squares for the full model using an F-ratio.

Multiple testing with Bonferroni correction (BONF). At each step, marginal regression models are used. A t-test with Bonferroni correction is then carried out to detect whether each regression coefficient θ'_k is non-zero. The intersection of the $|J^C|$ null hypotheses coincides with our null in each step.

n-out-of-n bootstrap (n-boot). This procedure is similar to the proposed m-out-of-n approach, except that the usual n-out-of-n bootstrap is used at each step.

m-out-of-n bootstrap with m chosen by Bickel and Sakov's method ($\hat{m}^{BS}$-boot). This procedure is similar to the proposed m-out-of-n approach, except that m is chosen via Bickel and Sakov's method at each step.

We consider three examples for the data generating model: **i)** $Y = \epsilon$, **ii)** $Y = 0.6X_1(A - 0.5) + \epsilon$, and **iii)** $Y = 0.6(X_1 + X_2)(A - 0.5) + \epsilon$. In all examples treatment $A \sim Bernoulli(0.5)$, and $\mathbf{X}$ is generated from a mean zero p-dimensional normal distribution with an exchangeable variance-covariance structure $\mathrm{Var}(X_k) = 1$ and $\mathrm{Cov}(X_j, X_k) = \rho$ for $j \neq k$, where ρ takes values 0 and 0.6, and the noise $\epsilon \sim N(0, 1)$ is independent of $\mathbf{X}$.

In the first model, there is no active interaction term. We perform one step screening test to evaluate the type I error rate of the proposed test. In the second model, there is one active interaction term. Thus we perform sequential tests in two steps. The first step evaluates the power of the test and the second step evaluates the type I error rate. Similarly, in the third example, we conduct the test in three steps, the first two steps for power and the last step for type I error rate control.

We consider $n = 200$, and $p = 10, 50, 100$. A nominal 5% significance level is used throughout. The number of bootstrap resamples is taken as

$1,000$. Empirical rejection rates based on 500 Monte Carlo replications are reported in Tables S1 and S2. The two proposed methods (NULL and $\hat{m}$-boot) provide good control of type I error rate and good power in all cases. $\hat{m}$-boot-DR, $\hat{m}^{BS}$-boot and LRT are less powerful as compared to the proposed methods. n-boot fails to control the type I error rate. In the case of independent $\boldsymbol{X}$, BONF is as good as our proposed methods in terms of type I error rate control and power (Table S1). However, when the components of $\boldsymbol{X}$ are highly correlated, BONF is less powerful for large p (Table S2).

S5.2 Simulations for observational studies

In the observational study setting, we compare the proposed $\hat{m}$-boot-DR method with $\hat{m}^{BS}$-boot and n-boot methods. We consider four data generating models:

i') $\text{logit}P(A = 1|X) = (X_1 + X_2)/2 + (X_4 - X_3)/4$, $Y = (X_1 + X_2 + X_3)^2/4 + \epsilon$;

ii') $\text{logit}P(A = 1|X) = (X_1 + X_2)/2 + (X_4 - X_3)/4$, $Y = (X_1 + X_2 + X_3)^2/4 + (1 + X_2)A + \epsilon$;

iii') $\text{logit}P(A = 1|X) = (X_1 + X_2)^2/2 - (X_3 + X_4)^2/2$, $Y = (X_1 + X_2 + X_3)/2 + \epsilon$;

Table S1: Rejection rate (%) over 500 Monte Carlo replications for independent $\boldsymbol{X}$ for randomized trials ($n = 200$).

Model	Step	p	NULL	$\hat{m}$-boot	$\hat{m}$-boot-DR	$\hat{m}^{BS}$-boot	n-boot	LRT	BONF
i)	step 1	10	4.0	6.6	3.8	5.8	36.2	2.4	4.2
	(type I error	50	4.0	7.2	1.0	4.2	70.4	4.0	4.4
	rate)	100	3.6	6.8	1.8	2.4	83.6	3.0	4.8
ii)	step 1	10	90.4	88.6	81.0	82.6	98.6	79.0	89.4
	(power)	50	79.8	83.2	52.0	65.4	99.0	32.4	81.6
		100	71.2	76.6	38.2	52.2	98.6	16.6	73.2
	step 2	10	4.0	5.2	3.8	5.2	34.6	2.4	3.8
	(type I error	50	4.0	6.0	1.0	4.0	68.0	4.2	4.0
	rate)	100	3.0	3.4	1.0	2.0	82.2	2.6	3.2
iii)	step 1	10	97.8	97.0	93.8	91.4	99.8	98.6	97.6
	(power)	50	93.2	95.6	73.0	84.4	99.8	73.0	94.0
		100	91.6	94.6	54.0	77.4	100	44.0	92.6
	step 2	10	82.0	75.4	68.0	68.4	97.2	65.4	82.4
	(power)	50	63.2	63.2	27.6	40.4	97.8	22.2	64.8
		100	50.8	49.8	13.4	27.4	97.0	11.4	53.8
	step 3	10	3.8	4.2	4.4	4.2	30.2	2.8	3.8
	(type I error	50	4.0	3.8	1.6	2.6	68.0	4.2	3.8
	rate)	100	3.0	1.2	0.8	1.2	81.4	2.8	3.2

Table S2: Rejection rate (%) over 500 Monte Carlo replications for $\boldsymbol{X}$ with pairwise correlation of 0.6 for randomized trials ($n = 200$).

Model	Step	p	NULL	$\hat{m}$-boot	$\hat{m}$-boot-DR	$\hat{m}^{BS}$-boot	n-boot	LRT	BONF
i)	step 1	10	3.8	5.0	4.4	5.0	22.0	2.4	1.8
	(type I error	50	4.4	6.6	3.2	5.6	35.0	4.0	2.4
	rate)	100	4.0	6.8	3.6	6.0	36.6	3.0	1.6
ii)	step 1	10	94.4	91.2	86.2	82.6	99.2	79.6	92.4
	(power)	50	90.8	91.8	69.2	75.6	98.0	33.4	86.6
		100	90.2	89.4	72.4	72.2	99.0	15.8	84.6
	step 2	10	2.4	3.6	2.8	3.6	20	1.4	2.2
	(type I error	50	2.6	4.8	2.8	4.2	37.2	2.6	2.4
	rate)	100	2.2	5.4	2.2	5.2	43.6	1.6	1.6
iii)	step 1	10	100	100	99.8	99.2	100	100	100
	(power)	50	100	100	99.4	99.0	100	93.0	100
		100	100	100	99.6	97.4	100	73.4	100
	step 2	10	50.4	43.8	40.4	42.8	82.6	26.2	47.8
	(power)	50	26.2	30.4	13.4	28.4	84.2	9.4	24.0
		100	19.6	27.6	15.4	26.6	87.0	5.2	16.2
	step 3	10	3.4	5.0	4.4	5.0	15.8	1.0	3.2
	(type I error	50	2.4	4.4	2.6	4.2	39.0	2.8	2.2
	rate)	100	1.6	5.0	2.2	5.0	47.2	1.6	1.2

iv') $\text{logit}P(A = 1|X) = (X_1 + X_2)^2/2 - (X_3 + X_4)^2/2$, $Y = (X_1 + X_2 + X_3)/2 + (1 + X_2)A + \epsilon$.

In all examples $\mathbf{X}$ is generated from a mean zero p-dimensional normal distribution with an exchangeable variance-covariance structure $\text{Var}(X_k) = 1$ and $\text{Cov}(X_j, X_k) = \rho$ for $j \neq k$, where ρ takes values 0 and 0.5, and the noise $\epsilon \sim N(0, 1)$ is independent of $\mathbf{X}$.

In the analysis, linear logistic regression model with adaptive lasso is used to estimate the propensity score model $q_0(\boldsymbol{X})$, and linear regression with adaptive lasso is used to estimate the main effect $h_0(\boldsymbol{X})$. So in models i') and ii'), $q_0(\boldsymbol{X})$ is correctly specified, while $h_0(\boldsymbol{X})$ is misspecified; in models iii') and iv'), $q_0(\boldsymbol{X})$ is misspecified, while $h_0(\boldsymbol{X})$ is correctly specified.

Similar to the randomized trial setting, there is no active interaction term in models i') and iii'). We perform one step screening test to evaluate the type I error rate of the proposed test. In models ii') and iv'), there is one active interaction term. Thus we perform sequential tests in two steps. The first step evaluates the power of the test and the second step evaluates the type I error rate.

We consider $n = 200$, and $p = 10, 50, 100$. A nominal 5% significance level is used throughout. The number of bootstrap resamples is taken as

1,000. Empirical rejection rates based on 500 Monte Carlo replications are reported in Table S3. The proposed $\hat{m}$-boot-DR method provides good control of type I error rate and good power in all cases. $\hat{m}^{BS}$-boot lacks power as compared to $\hat{m}$-boot, and n-boot fails to control the type I error rate.

S5.3 Simulation for Test of Global Null

In this section, we report simulation results for testing the global null hypothesis

$$\begin{aligned} H_0 &: \text{ there is no treatment by covariate interaction} \\ \text{vs. } H_a &: \text{there is treatment by covariate interaction} \end{aligned}$$

This corresponds to the first step of our method. We compare out method with two competing methods:

Kernel Machine based Score test. This test is proposed by Shen and Cai (2016) to identify whether a set of covariates are predictive of treatment difference in the setting of randomized trials. They consider three kernels (linear, quadratic, and Gaussian) and an Omnibus test to choose the best kernel. Since our models are linear, we consider the linear kernel and denote the method by $\mathbf{KM}_l$.

Gene Environment Set Association Test (GESAT): This is a variance

Table S3: Rejection rate (%) over 500 Monte Carlo replications for observational studies ($n = 200$).

			$\mathrm{Corr}(X_j, X_k) = 0$			$\mathrm{Corr}(X_j, X_k) = 0.5$		
Model	Step	p	$\hat{m}$-boot-DR	$\hat{m}^{BS}$-boot	n-boot	$\hat{m}$-boot-DR	$\hat{m}^{BS}$-boot	n-boot
i')	step 1	10	4.6	3.4	28.4	6.6	4.6	24.4
	(type I error	50	1.0	1.0	21.6	4.2	4.0	15.8
	rate)	100	1.4	1.4	27.8	6.6	6.6	20.2
ii')	step 1	10	80.4	65.4	94.4	43.2	24.2	64.8
	(power)	50	68.0	52.0	91.6	30.0	18.6	48.2
		100	60.2	55.8	87.4	29.6	27.8	40.8
	step 2	10	3.2	2.6	25.8	2.4	2.0	16.2
	(type I error	50	1.0	1.0	21.8	0.8	0.8	10.8
	rate)	100	1.8	1.8	25.6	1.4	1.4	13.4
iii')	step 1	10	4.0	4.0	26.4	6.4	6.0	16.6
	(type I error	50	2.4	2.0	23.6	3.6	3.4	12.8
	rate)	100	4.4	3.8	22.8	4.4	4.4	11.2
iv')	step 1	10	99.8	93.0	100	100	91.2	100
	(power)	50	97.6	87.2	100	96.4	76.8	99.8
		100	96.0	91.4	99.8	94.4	77.8	98.0
	step 2	10	4.2	4.2	22.4	5.8	5.8	24.6
	(type I error	50	1.6	1.6	20.4	2.8	2.8	14.2
	rate)	100	3.0	2.6	21.2	4.6	4.6	16.4

component test proposed by Lin et al. (2013). It relies on the correct specification of the full linear model. We use it as an example of set-based test for gene by environmental interactions.

For the randomized trial setting, we consider models **i)** and **ii)** in Section S5.1, and a new model **iv)** $Y = (\sum_{k=1}^{p} X_k)(A - 0.5)/10$. Model **i)** represents the null hypothesis, model **ii)** represents the presence of a strong sparse signal (since there is only one large nonzero treatment-by-covariate interaction term), and model **iv)** represents the case of weak dense signals (since all treatment-by-covariate interaction terms are nonzero and small). For observational studies, we consider the two null models **i')** and **iii')** in Section S5.2 to check the validity of the two competing methods.

Simulation results are presented in Tables S4 and S5. In randomized trials, the $\hat{m}$-boot method is slightly anti-conservative when $p = 50$ in the case of independent covariates. (note, with 500 Monte-Carlo replications, a rate of 7% or more is considered as significantly bigger than the nominal 5% rate.) Otherwise, all methods provide good control of type I error rates. In the case of strong sparse signal (Model **ii**), our methods have significant larger power as compared to two competing methods. In the weak dense signal case (Model **iv**), all methods are comparable and have increasing power as p increases (i.e. more signals), while KM_l performs slightly better when p is large. This is expected as our methods are based on the test of selected covariate, while the two competing methods incorporate all covariates in the test. If those extra covariates bring in signals to the model, then the power of the test would increase; otherwise, those covariates would just increase noise and thus decrease the power. All methods perform better when covariates are correlated. From Table S5, we see that KM_l fails to control type I error rate as expected since the method is designed for randomized trial data, while GESAT is only valid in Model **iii')** where the main effect model is correctly specified.

S5.4 Tuning parameter selection

In this section, we present simulation results of the m-out-of-n bootstrap method for $p = 10$ when different tuning parameter d is used. It can be seen from Tables S6 and S7 that the result is pretty robust to the choice of $d \in [0.7, 0.9]$.

Table S4: Rejection rate (%) over 500 Monte Carlo replications for the test of global null for randomized trials ($n = 200$).

		Corr$(X_j, X_k) = 0$				Corr$(X_j, X_k) = 0.6$			
Model	p	NULL	$\hat{m}$-boot	KM$_l$	GESAT	NULL	$\hat{m}$-boot	KM$_l$	GESAT
i)	10	4.0	6.6	4.0	4.2	3.8	5.0	4.4	3.4
	50	4.0	7.2	3.8	2.2	4.4	6.6	5.2	6.0
	100	3.6	6.8	2.8	3.0	4.0	6.8	5.8	5.2
ii)	10	90.4	88.6	81.4	78.6	94.4	91.2	92.0	90.6
	50	79.8	83.2	41.6	24.2	90.8	91.8	89.2	77.0
	100	71.2	76.6	29.4	7.6	90.2	89.4	89.2	69.4
iv)	10	9.2	14.2	13.2	11.6	49.0	46.8	79.8	77.0
	50	9.2	21.4	20.4	14.6	92.6	94.2	100	100
	100	11.2	25.0	30.2	11.8	97.6	99.2	100	100

Bibliography

Lin, X., S. Lee, D. C. Christiani, and X. Lin (2013). A test for the interaction between a genetic marker set and environment in generalized linear models. *Biostatistics 14*, 667–681.

Shen, Y. and T. Cai (2016). Identifying predictive markers for personalized treatment selection. *Biometrics 72*, 1017–1025.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge university

Table S5: Rejection rate (%) over 500 Monte Carlo replications for the global test of treatment by covariates interactions for models i') and iii') (null models) in Section S5.2 ($n = 200$).

		$\mathrm{Corr}(X_j, X_k) = 0$			$\mathrm{Corr}(X_j, X_k) = 0.5$		
Model	p	$\hat{m}$-DR	KM_l	GESAT	$\hat{m}$-DR	KM_l	GESAT
i')	10	4.6	28.8	39.6	6.6	88.6	97.0
	50	1.0	12.2	12.6	4.2	84.2	95.2
	100	1.4	7.8	6.6	6.6	87.4	93.4
iii')	10	4.0	97.8	3.8	6.4	98.4	6.6
	50	2.4	97.6	4.0	3.6	74.4	3.0
	100	4.4	97.2	0.8	4.4	64.8	5.8

press.

Table S6: Rejection rate (%) of $\hat{m}$-boot method over 500 Monte Carlo replications for randomized trial examples with tuning parameter $d = 0.9, 9.85, 0.8, 0.75$ and 0.7 ($n = 200, p = 10$).

		$\mathrm{Corr}(X_j, X_k) = 0$					$\mathrm{Corr}(X_j, X_k) = 0.6$				
Model	Step	0.9	0.85	0.8	0.75	0.7	0.9	0.85	0.8	0.75	0.7
i)	1	7.8	5.4	6.6	3.6	5.0	4.6	4.8	5.0	4.4	6.0
ii)	1	87.6	89.4	88.6	87.2	89.0	89.4	91.6	91.2	91.2	94.6
	2	4.4	5.0	5.2	3.6	5.8	3.4	3.0	3.6	3.2	4.2
iii)	1	96.6	96.2	97.0	97.2	97.8	100	100	100	100	100
	2	76.2	74.0	75.4	70.8	76.4	33.6	37.0	43.8	46.0	49.6
	3	3.8	3.8	4.2	4.2	2.8	3.6	5.2	5.0	4.4	5.0

Table S7: Rejection rate (%) of $\hat{m}$-boot-DR method over 500 Monte Carlo replications for observational study examples with tuning parameter $d = 0.9, 9.85, 0.8, 0.75$ and 0.7 ($n = 200, p = 10$).

		$\mathrm{Corr}(X_j, X_k) = 0$					$\mathrm{Corr}(X_j, X_k) = 0.5$				
Model	Step	0.9	0.85	0.8	0.75	0.7	0.9	0.85	0.8	0.75	0.7
i')	1	3.4	4.0	4.6	4.2	4.0	7.0	6.6	6.6	5.6	6.0
ii')	1	81.4	80.0	80.4	80.0	79.8	44.6	44.6	43.2	43.2	44.4
	2	3.2	2.8	3.2	2.2	3.0	3.4	3.0	2.4	2.6	3.6
iii')	1	2.6	3.2	4.0	3.2	3.6	4.2	4.6	6.4	4.4	6.4
iv')	1	99.6	99.4	99.8	100	99.8	99.4	99.8	100	99.6	99.6
	2	2.8	3.8	4.2	2.6	3.4	4.4	5.4	5.8	6.4	6.6