Supplement to "A Sparse Random Projection-based Test for Overall Qualitative Treatment Effects"

Chengchun Shi, Wenbin Lu and Rui Song

This supplementary article is organized as follows. In Section C, we present some examples on the estimation of the contrast function to better understand Assumption (C2), (A4), (A5) and (A6). Section D contains proofs of Lemma 3.1, Theorem 3.1, Theorem 3.2, Theorem 3.4 and additional simulation results. We omit the proof of Theorem 3.3 since it is very similar to that of Theorem 3.1. Additional simulation results are given in Section E.

C More on the technical conditions

In this section, we present some examples on the estimation of the contrast function to better understand Assumption (C2), (A4), (A5), (A6) and (A7).

C.1 Detailed discussion on (C2)

We consider the case where the structure parameter γ in (C1) equals 1. As commented in Section B, this assumption holds when $\tau(X)$ has a bounded density near 0.

Example 1 (B-spline methods) *Assume $p = 1$ and X has a bounded probability density function on a closed interval. For $j = 0, 1$, let $\hat{h}_{j,\mathcal{I}}(\cdot)$ be the B-spline regression estimators of the conditional mean functions $h_j(\cdot)$ based on the sub-dataset $\{(X_i, Y_i)\}_{i\in\mathcal{I},A_i=j}$. Define the estimated contrast function $\hat{\tau}_{\mathcal{I}}(\cdot) = \hat{h}_{1,\mathcal{I}}(\cdot) - \hat{h}_{0,\mathcal{I}}(\cdot)$. Then similar to Equation (8) in Zhou et al. (1998), we can show $E|\hat{\tau}_{\mathcal{I}}(X) - \tau(X)|^2 = O(|\mathcal{I}|^{-4/5})$ when $\tau(\cdot)$ is twice continuously differentiable and the number of interior knots K satisfies $K = C|\mathcal{I}|^{1/5}$ for some $C > 0$. Condition (C2) is thus satisfied.*

Example 2 (Kernel ridge regression) *For $j = 0, 1$, let $\hat{h}_{j,\mathcal{I}}(\cdot)$ be the kernel ridge regression estimator of $h_j(\cdot)$ by minimizing,*

$$\hat{h}_{j,\mathcal{I}} = \arg\min_{h \in \mathcal{H}} \left\{ \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} I(A_i = j)\{Y_i - h(X_i)\}^2 + \lambda \|h\|_{\mathcal{H}}^2 \right\},$$

where $\lambda > 0$ is a regularization parameter, $\mathcal{H}$ is a reproducing kernel Hilbert space with a reproducing kernel $K(\cdot, \cdot)$ and $\|\cdot\|_{\mathcal{H}}$ is the corresponding Hilbert norm. It follows from Mercer's theorem that

$$K(x, x') = \sum_{j=1}^{\infty} \mu_j \phi_j(x) \phi_j(x'),$$

for some nonnegative and nonincreasing sequence $\{\mu_j\}_{j=1}^{\infty}$, and some orthogonal basis functions $\{\phi_j(\cdot)\}_{j=1}^{\infty}$. Assume $h_0, h_1 \in \mathcal{H}$, $\sup_x E[\{Y^*(a) - h_a(X)\}^2 | X = x] = O(1)$ for $a = 0, 1$. Assume for some $k \geq 2$, there exists a constant $\rho < \infty$ such that $E\{\phi_j^{2k}(X)\} \leq \rho^{2k}$ for any integer $j \geq 1$. Further assume one of the following three conditions is satisfied:
(i) $\mu_j = 0, \forall j \geq r$ for some integer r that satisfies $r = o(|\mathcal{I}|^{1/4})$, and $\lambda = r/|\mathcal{I}|$;
(ii) $\mu_j = O(j^{-2\nu}), \forall j \geq 1$ for some $\nu > 3/2$, and $\lambda = |\mathcal{I}|^{-2\nu/(1+2\nu)}$;
(iii) $\mu_j = O(\exp(-\bar{c} j^2)), \forall j \geq 1$ for some constant $\bar{c} > 0$, and $\lambda = |\mathcal{I}|^{-1}$.
Then, we have $E|\hat{h}_{j,\mathcal{I}}(X) - h_j(X)|^2 = o(|\mathcal{I}|^{-3/4})$ (see Corollaries 2-4 in Zhang et al., 2013). Set $\hat{\tau}_{\mathcal{I}}(x) = \hat{h}_{\mathcal{I},1}(x) - \hat{h}_{\mathcal{I},0}(x)$. It follows from Cauchy-Schwarz inequality that $E|\hat{\tau}_{\mathcal{I}}(X) - \tau(X)|^2 = o(|\mathcal{I}|^{-3/4})$. Notice that the convergence rates are independent of the dimension p. Condition (C2) therefore holds.

C.2 Discussion on (A4)-(A6)

We assume covariates follow an elliptical distribution with mean zero and covariance matrix Σ. Further assume there exists some constant $c_0 \geq 1$ such that $c_0^{-1} \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq c_0$. Then, X has the following stochastic representation:

$$X \stackrel{d}{=} \Sigma^{1/2} U, \tag{S.1}$$

where U is spherically distributed. For $1 \leq j \leq p$, we denoted by $U^{(j)}$ the j-th element of U. Assume $U^{(1)}$ has a positive probability density function that is bounded away from 0 and ∞ on $(-u_{\max}, u_{\max})$ for some $u_{\max} < \infty$ where $(-u_{\max}, u_{\max})$ is the support of $U^{(1)}$.

Assume $\tau(X) = a_0 S^* X + b_0$ for some sketching matrix $S^* \in \mathbb{R}^{1\times p}$ with $\|S^*\|_0 \leq s$ and $\|S^*\|_2 = 1$ and some $a_0, b_0 \in \mathbb{R}$ with $a_0 \neq 0$. Then the optimal choice of the projected dimension q would be 1. In the following, we focus on the case where $q = 1$ and show (A4)-(A6) hold.

More generally, one may consider the following model:

$$\tau(X) = \psi(S^{*(1)}X, S^{*(2)}X, \dots, S^{*(q_0)}X),$$

for some sketching matrices $S^{*(1)}, S^{*(2)}, \dots, S^{*(q_0)} \in \mathbb{R}^{1\times p}$, some integer $q_0 > 1$ and some function $\psi(\cdot)$. The choice of q then involves a trade-off. For $q < q_0$, the contrast function might not be well approximated. However, the convergence rate of the estimated contrast function slows as q increases.

C.2.1 Validity of (A4)

The random vector U (see (S.1)) has a spherical distribution. For any sketching matrix $S \in \mathbb{R}^{1\times p}$, it follows from Theorem 2.4 of Fang et al. (1990) that

$$S\Sigma^{1/2}U \overset{d}{=} \|S\Sigma^{1/2}\|_2 U^{(1)}.$$

This together with (S.1) implies that

$$SX \overset{d}{=} \|S\Sigma^{1/2}\|_2 U^{(1)}. \tag{S.2}$$

For any $S \in \mathcal{S}$, we have $\|S\|_2 = 1$. Hence, $c_0^{-1/2} \leq \|S\Sigma^{1/2}\| \leq c_0^{1/2}$. It follows from (S.2) that there exists some constant $c_* \geq 1$ such that the probability density function $q_S(\cdot)$ of SX satisfies

$$c_*^{-1} \leq q_S(u) \leq c_*, \quad \forall u \in (-\|S\Sigma^{1/2}\|_2 u_{\max}, \|S\Sigma^{1/2}\|_2 u_{\max}) \text{ and } S \in \mathcal{S}. \tag{S.3}$$

We focus on the cubic B-spline methods discussed in Section 3.3.2. Let $\mathbb{Q}^S$ denote the matrix

$$\begin{pmatrix} \sum_{i\in\mathcal{I}} N_1^{(S)}(SX_i)N_1^{(S)}(SX_i) & \sum_{i\in\mathcal{I}} N_1^{(S)}(SX_i)N_2^{(S)}(SX_i) & \cdots & \sum_{i\in\mathcal{I}} N_1^{(S)}(SX_i)N_{K+4}^{(S)}(SX_i) \\ \sum_{i\in\mathcal{I}} N_1^{(S)}(SX_i)N_2^{(S)}(SX_i) & \sum_{i\in\mathcal{I}} N_2^{(S)}(SX_i)N_2^{(S)}(SX_i) & \cdots & \sum_{i\in\mathcal{I}} N_2^{(S)}(SX_i)N_{K+4}^{(S)}(SX_i) \\ \vdots & \vdots & & \vdots \\ \sum_{i\in\mathcal{I}} N_1^{(S)}(SX_i)N_{K+4}^{(S)}(SX_i) & \sum_{i\in\mathcal{I}} N_2^{(S)}(SX_i)N_{K+4}^{(S)}(SX_i) & \cdots & \sum_{i\in\mathcal{I}} N_{K+4}^{(S)}(SX_i)N_{K+4}^{(S)}(SX_i) \end{pmatrix}.$$

Assume the interior knots are placed at equally-spaced quantiles of SX. Since $|\mathcal{I}| \geq n/2$, similar to Equation (13) in Zhou (2009) and Theorem 2 in de Boor (1973), we can show there exists some constant $c_{**} > 1$ such that

$$(2c_{**}K)^{-1}n \leq c_{**}^{-1}|\mathcal{I}|K^{-1} \leq \lambda_{\min}(\mathrm{E}\mathbb{Q}^S) \leq \lambda_{\max}(\mathrm{E}\mathbb{Q}^S) \leq c_{**}|\mathcal{I}|K^{-1} \leq 2c_{**}K^{-1}n, \forall S \in \mathcal{S}, \quad \text{(S.4)}$$

under the condition in (S.3). The B-spline bases are uniformly bounded. Therefore, we have

$$\begin{aligned} \max_{1\leq k_1,k_2\leq K+4} \sup_{S\in\mathcal{S}} \mathrm{Var}\{N_{k_1}^{(S)}(SX)N_{k_2}^{(S)}(SX)\} &\leq \max_{1\leq k_1,k_2\leq K+4} \sup_{S\in\mathcal{S}} \mathrm{E}\{N_{k_1}^{(S)}(SX)N_{k_2}^{(S)}(SX)\}^2 \\ &= O\left(\max_{1\leq k\leq K} \sup_{S\in\mathcal{S}} \{\mathrm{E}N_k^{(S)}(SX)\}^2 \right) = O(K^{-1}). \quad \text{(S.5)} \end{aligned}$$

Assume $K = Cn^{1/5}$ for some constant $C > 0$, and $B = O(n^{\kappa_B})$ for some $\kappa_B > 0$. It follows from Bernstein's inequality (see Lemma 2.2.9, van der Vaart and Wellner, 1996) that the following event occurs with probability tending to 1,

$$\max_{\substack{b\in\{1,\dots,B\} \\ 1\leq k_1,k_2\leq K+4}} \left| \sum_{i\in\mathcal{I}} \left(N_{k_1}^{(S_b)}(S_bX_i)N_{k_2}^{(S_b)}(S_bX_i) - \mathrm{E}N_{k_1}^{(S_b)}(S_bX)N_{k_2}^{(S_b)}(S_bX) \right) \right| = O(n^{2/5}\sqrt{\log n}). \quad \text{(S.6)}$$

Notice that $N_{k_1}^{(S_b)}(S_bX)N_{k_2}^{(S_b)}(S_bX) = 0$ when $|k_1 - k_2| > 4$, $\mathbb{Q}^{S_b} - \mathrm{E}\mathbb{Q}^{S_b}$ is a band matrix. Under the event defined in (S.6), we have $\|\mathbb{Q}^{S_b} - \mathrm{E}\mathbb{Q}^{S_b}\|_1 = O(n^{2/5}\sqrt{\log n})$ and $\|\mathbb{Q}^{S_b} - \mathrm{E}\mathbb{Q}^{S_b}\|_\infty = O(n^{2/5}\sqrt{\log n})$. As a result, we have

$$\max_{b\in\{1,\dots,B\}} \|\mathbb{Q}^{S_b} - \mathrm{E}\mathbb{Q}^{S_b}\|_2 \leq \max_{b\in\{1,\dots,B\}} \sqrt{\|\mathbb{Q}^{S_b} - \mathrm{E}\mathbb{Q}^{S_b}\|_1 \|\mathbb{Q}^{S_b} - \mathrm{E}\mathbb{Q}^{S_b}\|_\infty} = O(n^{2/5}\sqrt{\log n}).$$

with probability tending to 1. This together with (S.4) yields

$$\min_{b\in\{1,\dots,B\}} \lambda_{\min}(\mathbb{Q}^{S_b}) \geq \bar{c}_* n^{4/5}, \tag{S.7}$$

for some constant $\bar{c}_* > 0$. By (S.4), (S.6) and (S.7), we have

$$\begin{aligned} & \max_{b\in\{1,\dots,B\}} \|(\mathbb{Q}^{S_b})^{-1} - (\mathrm{E}\mathbb{Q}^{S_b})^{-1}\|_2 \\ \leq & \max_{b\in\{1,\dots,B\}} \|(\mathbb{Q}^{S_b})^{-1}\|_2 \|\mathbb{Q}^{S_b} - \mathrm{E}\mathbb{Q}^{S_b}\|_2 \|(\mathrm{E}\mathbb{Q}^{S_b})^{-1}\|_2 = O(n^{-6/5}\sqrt{\log n}). \end{aligned} \tag{S.8}$$

Consider the pseudo-outcome defined in (12). The estimated propensity score function $\hat{\pi}^{\mathcal{I}}(\cdot)$ and conditional mean functions $\hat{h}_0^{\mathcal{I}}, \hat{h}_1^{\mathcal{I}}$ will converge to some $\pi^*(\cdot), h_0^*(\cdot)$ and $h_1^*(\cdot)$ respectively. Under certain regularity conditions, we can show the following event holds with probability tending to 1,

$$\max_{i\in\mathcal{I}} \left|\hat{\tau}_i^{\mathcal{I}} - \tau_i^*\right| = O(n^{-2/5}\sqrt{\log n}), \tag{S.9}$$

where

$$\tau_i^* = \left(\frac{A_i}{\pi^*(X_i)} - \frac{1-A_i}{1-\pi^*(X_i)}\right) Y_i + \left(\frac{A_i}{\pi^*(X_i)} - 1\right) h_1^*(X_i) - \left(\frac{1-A_i}{1-\pi^*(X_i)} - 1\right) h_0^*(X_i).$$

Similar to (S.6), we have by Bernstein's inequality that

$$\max_{\substack{b\in\{1,\dots,B\}\\ 1\leq k\leq K+4}} \left|\sum_{i\in\mathcal{I}} \{N_k^{(S_b)}(S_b X_i)\tau_i^* - \mathrm{E}N_k^{(S_b)}(S_b X_i)\tau_i^*\}\right| = O(n^{2/5}\sqrt{\log n}),$$

with probability tending to 1. By the doubly-robustness property, we have $\mathrm{E}(\tau_i^*|X_i) = \tau(X_i)$ and hence $\mathrm{E}N_k^{(S)}(SX_i)\tau_i^* = \mathrm{E}N_k^{(S)}(SX)\tau(X)$. Therefore,

$$\Pr\left(\max_{\substack{b\in\{1,\dots,B\}\\ 1\leq k\leq K+4}} \left|\sum_{i\in\mathcal{I}} \{N_k^{(S_b)}(S_b X_i)\tau_i^* - \mathrm{E}N_k^{(S_b)}(S_b X_i)\tau(X_i)\}\right| = O(n^{2/5}\sqrt{\log n})\right) \to 1. \tag{S.10}$$

For any $1 \leq k \leq K+4$, $1 \leq b \leq B$, $N_k^{(S_b)}(\cdot)$ has bounded support of length $O(K^{-1}) = O(n^{-1/5})$. It follows from (S.3) that

$$\sup_{S\in\mathcal{S}} \max_{k\in\{1,\dots,K\}} \mathrm{E}|N_k^{(S)}(SX)| = O(n^{-1/5}). \tag{S.11}$$

Similar to (S.6), we have by Bernstein's inequality that

$$\max_{\substack{b\in\{1,\dots,B\}\\1\le k\le K+4}}\left|\sum_{i\in\mathcal{I}}\{|N_k^{(S_b)}(S_bX_i)|-\mathrm{E}|N_k^{(S_b)}(S_bX_i)|\}\right|=O(n^{2/5}\sqrt{\log n}),$$

with probability tending to 1. This together with (S.11) yields

$$\max_{\substack{b\in\{1,\dots,B\}\\1\le k\le K+4}}\sum_{i\in\mathcal{I}}|N_k^{(S_b)}(S_bX_i)|=O(n^{4/5}),$$

with probability tending to 1. By (S.9) and (S.10), we have

$$\max_{\substack{b\in\{1,\dots,B\}\\1\le k\le K+4}}\left|\sum_{i\in\mathcal{I}}\{N_k^{(S_b)}(S_bX_i)\hat{\tau}_i^{\mathcal{I}}-\mathrm{E}N_k^{(S_b)}(S_bX_i)\tau(X_i)\}\right|\le\max_{\substack{b\in\{1,\dots,B\}\\1\le k\le K+4}}\left(\sum_{i\in\mathcal{I}}|N_k^{(S_b)}(S_bX_i)|\right)$$
$$\times\max_{i\in\mathcal{I}}\left|\hat{\tau}_i^{\mathcal{I}}-\tau_i^*\right|+\max_{\substack{b\in\{1,\dots,B\}\\1\le k\le K+4}}\left|\sum_{i\in\mathcal{I}}\{N_k^{(S_b)}(S_bX_i)\tau_i^*-\mathrm{E}N_k^{(S_b)}(S_bX_i)\tau(X_i)\}\right|=O(n^{2/5}\sqrt{\log n}),$$

with probability tending to 1. As a result, we have

$$\max_{b\in\{1,\dots,B\}}\sqrt{\sum_{k=1}^{K+4}\left(\sum_{i\in\mathcal{I}}\{N_k^{(S_b)}(S_bX_i)\hat{\tau}_i^{\mathcal{I}}-\mathrm{E}N_k^{(S_b)}(S_bX)\tau(X)\}\right)^2}=O\left(\sqrt{n\log n}\right),\quad\text{(S.12)}$$

with probability tending to 1. In view of (S.2), since $\|S^*\|_2=1,\lambda_{\max}(\Sigma)\le c_0$ and $\|U^{(1)}\|_\infty\le u_{\max}$, the contrast function $\tau(\cdot)$ is uniformly bounded. Similar to (S.11), we can show $\sup_{S\in\mathcal{S}}\max_{k\in\{1,\dots,K\}}|\mathrm{E}N_k(SX)\tau(X)|=O(n^{-1/5})$. Therefore, we have

$$\|\{\mathrm{E}N_1(SX)\tau(X),\cdots,\mathrm{E}N_{K+4}(SX)\tau(X)\}\|_2=O(\sqrt{K}n^{-1/5})=O(n^{-1/10}),$$

where the big-O term is uniform in $S\in\mathcal{S}$. This together with (S.7), (S.8) and (S.12) yields

$$\begin{aligned}&\max_{b\in\{1,\dots,B\}}\|\hat{\xi}^{\mathcal{I},S_b}-\xi^{S_b}\|_2 &\text{(S.13)}\\ \le\;&\max_{b\in\{1,\dots,B\}}\|(\mathbb{Q}^{S_b})^{-1}-(\mathrm{E}(\mathbb{Q}^{S_b})^{-1})\|_2|\mathcal{I}|\|\{\mathrm{E}N_1(SX)\tau(X),\cdots,\mathrm{E}N_{K+4}(SX)\tau(X)\}\|_2\\ +\;&\max_{b\in\{1,\dots,B\}}\|(\mathbb{Q}^{S_b})^{-1}\|_2\sqrt{\sum_{k=1}^{K+4}\left(\sum_{i\in\mathcal{I}}\{N_k(S_bX_i)\hat{\tau}_i^{\mathcal{I}}-\mathrm{E}N_k(S_bX)\tau(X)\}\right)^2}=O(n^{-3/10}\sqrt{\log n}),\end{aligned}$$

with probability tending to 1, where

$$
\begin{aligned}
\hat{\xi}^{\mathcal{I},S_b} &= (\mathbb{Q}^{S_b})^{-1}\left(\sum_{i\in\mathcal{I}} N_1(S_bX_i)\hat{\tau}_i^{\mathcal{I}},\cdots,\sum_{i\in\mathcal{I}} N_{K+4}(S_bX_i)\hat{\tau}_i^{\mathcal{I}}\right)^T,\\
\xi^{S_b} &= (\mathrm{E}\mathbb{Q}^{S_b})^{-1}|\mathcal{I}|\{\mathrm{E}N_1(S_bX)\tau(X),\cdots,\mathrm{E}N_{K+4}\tau(X)\}.
\end{aligned}
$$

By Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
& \max_{1\le b\le B} \mathrm{E}|\hat{\tau}_{\mathcal{I}}^{S_b}(S_bX)-\tau^{S_b}(S_bX)|^2\\
= & \max_{1\le b\le B} \mathrm{E}\left|\tau^{S_b}(S_bX)-\sum_{k=1}^{K+4} N_k^{(S_b)}(S_bX)\xi_k^{S_b}+\sum_{k=1}^{K+4} N_k^{(S_b)}(S_bX)\xi_k^{S_b}-\sum_{k=1}^{K+4} N_k^{(S_b)}(S_bX)\hat{\xi}_k^{\mathcal{I},S_b}\right|^2\\
\le & \underbrace{2\max_{1\le b\le B} \mathrm{E}\left|\tau^{S_b}(S_bX)-\sum_{k=1}^{K+4} N_k^{(S_b)}(S_bX)\xi_k^{S_b}\right|^2}_{\mathbb{I}_1}+\underbrace{2\max_{1\le b\le B} \mathrm{E}\left|\sum_{k=1}^{K+4} N_k^{(S_b)}(S_bX)(\hat{\xi}_k^{\mathcal{I},S_b}-\xi_k^{S_b})\right|^2}_{\mathbb{I}_2}.
\end{aligned}
$$

It follows from (S.4) and (S.13) that

$$
\begin{aligned}
\mathbb{I}_2 &\le 2\max_{1\le b\le B}(\hat{\xi}^{\mathcal{I},S_b}-\xi^{S_b})^T\{\mathrm{E}(\mathbb{Q}^{S_b}/|\mathcal{I}|)\}(\hat{\xi}^{\mathcal{I},S_b}-\xi^{S_b})\\
&\le 2\max_{1\le b\le B}\lambda_{\max}\{\mathrm{E}(\mathbb{Q}^{S_b}/|\mathcal{I}|)\}\|\hat{\xi}^{\mathcal{I},S_b}-\xi^{S_b}\|_2^2 = O(n^{-4/5}\log n).
\end{aligned}
$$

Besides, since X is elliptically distributed, we have

$$
\mathrm{E}(S^*X|SX) = (S\Sigma S^T)^{-1}(S^*\Sigma S^T)(SX), \tag{S.14}
$$

for any sketching matrix $S\in\mathbb{R}^{1\times p}$. Thus,

$$
\mathrm{E}\{\tau(X)|SX\} = a_0(S\Sigma S^T)^{-1}(S^*\Sigma S^T)(SX)+b_0. \tag{S.15}
$$

This implies that $\tau^{S_1},\cdots,\tau^{S_B}$ have uniformly bounded second-order derivatives. As a result, we have $\mathbb{I}_1 = O(n^{-4/5})$ (see the proof of Equation (5) in Zhou, 2009). Therefore, (A4) holds with $r_0 = 4/5$. The condition $r_0 > (2+\gamma)/(2+2\gamma)$ in Theorem 3.4 holds as long as $\gamma > 2/3$.

C.2.2 Validity of (A5)

Below, we show Assumption (9) holds. In view of (S.15), we have for any $S \in \mathbb{R}^{1\times p}$ that

$$\begin{aligned} &\mathrm{E}|\tau(X) - \mathrm{E}\{\tau(X)|SX\}|^2 = a_0^2\mathrm{E}|S^*X - (S\Sigma S^T)^{-1}(S^*\Sigma S^T)(SX)|^2 \qquad \text{(S.16)} \\ = \; & a_0^2\left(S^*\Sigma(S^*)^T - 2\frac{(S^*\Sigma S^T)^2}{S\Sigma S^T} + \frac{(S^*\Sigma S^T)^2}{S\Sigma S^T}\right) = a_0^2\left(S^*\Sigma(S^*)^T - \frac{(S^*\Sigma S^T)^2}{S\Sigma S^T}\right). \end{aligned}$$

By Cauchy-Schwarz inequality, we have

$$\frac{(S^*\Sigma S^T)^2}{S\Sigma S^T} + S\Sigma S^T \geq 2S^*\Sigma S^T.$$

This together with (S.16) yields that

$$\mathrm{E}|\tau(X) - \mathrm{E}\{\tau(X)|SX\}|^2 \leq a_0^2(S^* - S)\Sigma(S^* - S)^T \leq a_0\lambda_{\max}(\Sigma)\|S^* - S\|_2^2 \leq a_0c_0\|S^* - S\|_2^2.$$

Assumption (9) is thus satisfied.

C.2.3 Validity of (A6)

In view of (S.15), we have

$$\begin{aligned} V(d_{S^*}^{opt}) - V(d_S^{opt}) &= \mathrm{E}(a_0S^*X + b_0)[I(a_0S^*X > -b_0) - I\{a_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX) > -b_0\}] \\ &= -\mathrm{E}\{a_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX) + b_0\}I\{a_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX) > -b_0\} \\ &- a_0\mathrm{E}\{S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX)\}I\{a_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX) > -b_0\} \\ &+ \mathrm{E}(a_0S^*X + b_0)[I(a_0S^*X > -b_0) = \mathrm{E}(a_0S^*X + b_0)[I(a_0S^*X > -b_0) \\ &- \mathrm{E}\{a_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX) + b_0\}I\{a_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX) > -b_0\} \\ &= |a_0|(S^*\Sigma S^{*T})^{1/2}\underbrace{\{\mathrm{E}(U^{(1)} - r_0)I(U^{(1)} > r_0) - \mathrm{E}(\rho U^{(1)} - r_0)I(\rho U^{(1)} > r_0)\}}_{\Psi(\rho)}, \end{aligned}$$

where $U^{(1)}$ is the first element of U defined in (S.1), $\rho = \sqrt{(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(S\Sigma S^{*T})/(S^*\Sigma S^{*T})}$ and $r_0 = -b_0/\{|a_0|(S^*\Sigma S^{*T})^{1/2}\}$. The third equality is due to (S.15), and the last equality follows from Theorem 2.5 (i) of Fang et al. (1990).

Case 1: $r_0 = 0$. Then we have $\Psi(\rho) = (1-\rho)\mathrm{E}\max(U^{(1)}, 0)$. Since $\mathrm{E}\max(U^{(1)}, 0) \geq \bar{c}$

for some constant $\bar{c} > 0$, for any $\varepsilon_0 > 0$ that satisfies $V(d_{S^*}^{opt}) - V(d_S^{opt}) \leq \varepsilon_0$, we have

$$1 - \rho \leq \frac{\varepsilon_0}{\bar{c}|a_0|(S^*\Sigma S^{*T})^{1/2}} \leq \frac{\varepsilon_0}{\bar{c}|a_0|c_0^{-1}}.$$

By (S.14), we have $\mathrm{E}\{\tau(X)|SX\} \overset{d}{=} \bar{C}U^{(1)}$ for some $\bar{C}$ that is uniformly bounded away from 0. Since $U^{(1)}$ has a bounded probability density function near 0, (A6) thus holds with $\gamma = 1$.

Case 2: $r_0 > 0$, $Pr(U^{(1)} \geq r_0) = 0$. It follows that $\Pr(\rho U^{(1)} - r_0 < 0) = 1, \forall \rho \leq 1$. Thus (A6) holds for any $\gamma > 0$.

Case 3: $r_0 < 0$, $Pr(U^{(1)} \leq r_0) = 0$. It follows that $\Pr(\rho U^{(1)} - r_0 > 0) = 1, \forall \rho \leq 1$. Thus (A6) holds for any $\gamma > 0$.

Case 4: $r_0 \neq 0$, $Pr(|U^{(1)}| > |r_0|) > 0$. Let $f_U^{(1)}(\cdot)$ denote the probability density of $U^{(1)}$. With some calculations, we have

$$\frac{d\Psi(\rho)}{d\rho} = -\int_{\frac{r_0}{\rho}}^{+\infty} u f_{U^{(1)}}(u) du.$$

The derivation of $\Psi(\cdot)$ is nonpositive for all $0 < \rho \leq 1$. Since $\Pr(|U^{(1)}| > |r_0|) > 0$, there exists some $\rho_0 > 0$ such that $|d\Phi(\rho)/d\rho|_{\rho=\rho_0}| > 0$ and As a result, we have $\Psi(\rho) \geq \Psi(\rho_0)$ for all $0 \leq \rho \leq \rho_0$. In addition, $|d\Psi(\rho)/d\rho|$ is monotonically increasing as a function of ρ. For $\rho_0 \leq \rho \leq 1$, it follows from Taylor's theorem that

$$\Psi(\rho) \geq \left| \left. \frac{d\Phi(\rho)}{d\rho} \right|_{\rho=\rho_0} \right| (1 - \rho).$$

Therefore, for any sufficiently small ε_0 such that $V(d_{S^*}^{opt}) - V(d_S^{opt}) \leq \varepsilon_0$, we have

$$1 - \rho \leq \frac{\varepsilon_0}{|a_0|(S^*\Sigma S^{*T})^{1/2}|d\Phi(\rho)/d\rho|_{\rho=\rho_0}|}.$$

As a result, for any sufficiently small $\varepsilon_0 > 0$ and any sketching matrix S such that $V(d_{S^*}^{opt}) - V(d_S^{opt}) \leq \varepsilon_0$, we have $1 - \rho \leq \bar{c}_* \varepsilon_0 / \{|a_0|(S^*\Sigma S^{*T})^{1/2}\}$ for some constant $\bar{c}_* > 0$. By (S.14), it is immediate to see that (A6) holds with $\gamma = 1$ since $U^{(1)}$ has a bounded probability density function.

C.3 Additional discussion on (A5)

In Section C.2.2, we show (9) in (A5) is satisfied when X follows an elliptical distribution and the contrast function is linear. More generally, (A5) holds as long as $\tau(X) = \psi(S^*X)$ for some Lipschitz continuous function $\psi(\cdot)$, some $S^* \in \mathbb{S}^*$, and $\Sigma = \mathrm{E}XX^T$ satisfies $\lambda_{\max}(\Sigma) = O(1)$. Notice that

$$
\begin{aligned}
&\mathrm{E}|\tau(X) - \mathrm{E}\{\tau(X)|SX\}|^2 = \mathrm{E}|\psi(S^*X) - \psi(SX) + \psi(SX) - \mathrm{E}\{\psi(S^*X)|SX\}|^2 \\
\leq\ & 2\mathrm{E}|\psi(S^*X) - \psi(SX)|^2 + 2\mathrm{E}|\mathrm{E}[\{\psi(S^*X) - \psi(SX)\}|SX]|^2 \leq 4\mathrm{E}|\psi(S^*X) - \psi(SX)|^2 \\
=\ & O\left(\mathrm{E}\|S^*X - SX\|_2^2\right) = O\left(\sum_{j=1}^{q}(S^{(j)} - S^{*(j)})^T(\mathrm{E}XX^T)(S^{(j)} - S^{*(j)})\right),
\end{aligned}
$$

where the first inequality is due to Cauchy-Schwarz inequality, the second inequality is due to Jensen's inequality and the second equality follows by the Lipschitz continuity of $\psi(\cdot)$. Since $\lambda_{\max}(\mathrm{E}XX^T) = O(1)$, it is immediate to see that (9) holds.

C.4 Additional discussion on (A6)

In Section C.2.3, we assume X follows an elliptical distribution and the contrast function is linear. If we are willing to assume $X \sim N(0, \Sigma)$ for some Σ that satisfies $c_0^{-1} \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq c_0$ for some $c_0 \geq 1$, then we can show (A6) holds when $\tau(\cdot)$ takes some other nonlinear forms as well.

Example 3 (Quadratic contrast function) *Assume $\tau(x) = a_0(S^*x)^2 - b_0$ for some $S^* \in \mathbb{R}^{1\times p}$ such that $\|S^*\|_2 = 1$ and $a_0, b_0 > 0$. Notice that for any sketching matrix $S \in \mathcal{S} \subseteq \mathbb{R}^{q\times p}$, SX is independent of $S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX$. Hence,*

$$
\begin{aligned}
&E\{\tau(X)|SX\} = a_0E\{S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX + (S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX|SX\}^2 - b_0 \\
=\ & a_0E\{S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX\}^2 + a_0|(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX)|^2 - b_0 \\
=\ & a_0\{S^*\Sigma S^{*T} - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}(S\Sigma S^{*T})\} + a_0|(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX)|^2 - b_0. \quad \text{(S.17)}
\end{aligned}
$$

Notice that $S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX)$ *is independent of* SX. *Thus, we have*

$$\begin{aligned}
& V(d_S^{opt}) = \{a_0 E(S^*X)^2 - b_0\} I[E\{\tau(X)|SX\} > 0] \\
= \;& [a_0 E\{(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX)\}^2 - b_0] I[E\{\tau(X)|SX\} > 0] \\
+ \;& a_0 E\{S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX)\}^2 I[E\{\tau(X)|SX\} > 0] \\
+ \;& 2a_0 E\{S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX)\}\{(S^*\Sigma S^T)(S\Sigma S^T)^{-1}(SX)\} I[E\{\tau(X)|SX\} > 0] \\
= \;& a_0 E[E\{\tau(X)|SX\} I\{E(\tau(X)|SX) > 0\}].
\end{aligned}$$

Let $r_0 = \sqrt{b_0/(a_0 S^*\Sigma S^{*T})}$ *and* $\rho = \sqrt{(S^*\Sigma S)(S\Sigma S^T)^{-1}(S\Sigma S^{*T})/(S^*\Sigma S^{*T})}$. *Then we have*

$$\begin{aligned}
& V(d_{S^*}^{opt}) - V(d_S^{opt}) = a_0 E\tau(X) I\{\tau(X) > 0\} - a_0 E[E\{\tau(X)|SX\} I\{E(\tau(X)|SX) > 0\}] \\
= \;& a_0(S^*\Sigma S^{*T}) \underbrace{\{E(\mathbb{Z}^2 - r_0^2) I(|\mathbb{Z}| > r_0) - E(1 - \rho^2 + \rho^2\mathbb{Z}^2 - r_0^2) I(1 - \rho^2 + \rho^2\mathbb{Z}^2 > r_0^2)\}}_{\Psi(\rho)}.
\end{aligned}$$

Case 1: $r_0 < 1$. *For any* ρ *such that* $1 - \rho^2 \geq r_0^2$, *we have*

$$\Psi(\rho) = E(\mathbb{Z}^2 - r_0^2) I(|\mathbb{Z}| > r_0) - (1 - r_0^2) > E(\mathbb{Z}^2 - r_0^2) - (1 - r_0^2) > 0.$$

For any ρ *such that* $1 - \rho^2 < r_0^2$, *we have*

$$\frac{d\Psi(\rho)}{d\rho} = 2\rho \int_{\sqrt{r_0^2+\rho^2-1}/\rho}^{+\infty} (1 - x^2) \exp\left(-\frac{x^2}{2}\right) dx = 2\rho \int_0^{\sqrt{r_0^2+\rho^2-1}/\rho} (x^2 - 1) \exp\left(-\frac{x^2}{2}\right) dx. \tag{S.18}$$

Since $r_0 < 1$, *we have* $\sqrt{r_0^2 + \rho^2 - 1} < \rho$ *and hence* $d\Psi(\rho)/d\rho$ *is negative. Therefore, for any* ρ *such that* $1 - \rho^2 \geq r_0^2/2$, *we have* $\Psi(\rho) \geq \Psi(\sqrt{1 - r_0^2/2})$. *For any* ρ *such that* $1 - \rho^2 < r_0^2/2$, *we have*

$$\frac{d\Psi(\rho)}{d\rho} \leq -2\sqrt{1 - r_0^2/2} \int_0^{r_0/\sqrt{2-r_0^2}} (1 - x^2) \exp\left(-\frac{x^2}{2}\right) dx.$$

Therefore, for any sufficiently small $\varepsilon_0 > 0$ *such that* $V(d_{S^*}^{opt}) - V(d_S^{opt}) \leq \varepsilon_0$, *we have*

$$1 - \rho \leq \frac{\varepsilon_0}{a_0(S^*\Sigma S^{*T})} \left\{2\sqrt{1 - r_0^2/2} \int_0^{r_0/\sqrt{2-r_0^2}} (1 - x^2) \exp\left(-\frac{x^2}{2}\right) dx\right\}^{-1}.$$

Case 2: $r_0 \geq 1$. *Since* $r_0 \geq 1$, *we have* $\sqrt{r_0^2+\rho^2-1}/\rho \geq 1$. *Hence, the integral*

$$\int_0^{\sqrt{r_0^2+\rho^2-1}/\rho} (1-x^2)\exp\left(-\frac{x^2}{2}\right) dx \tag{S.19}$$

decreases as $\sqrt{r_0^2+\rho^2-1}/\rho$ *increases for any* $\sqrt{r_0^2+\rho^2-1}/\rho \geq 1$. *Notice that* (S.19) *is equal to* 0 *when setting* $\sqrt{r_0^2+\rho^2-1}/\rho = \infty$. *In view of* (S.18), $d\Psi(\rho)/d\rho$ *is negative for any* $0 < \rho \leq 1$. *As a result, we have* $\Psi(\rho) \geq \Psi(1/2)$ *for any* ρ *such that* $0 < \rho \leq 1/2$. *For any* ρ *such that* $1/2 < \rho \leq 1$, *we have*

$$\frac{d\Psi(\rho)}{d\rho} \leq -\int_0^{\sqrt{r_0^2+\rho^2-1}/\rho} (1-x^2)\exp\left(-\frac{x^2}{2}\right) dx \leq -\int_0^{\sqrt{r_0^2-3/4}} (1-x^2)\exp\left(-\frac{x^2}{2}\right) dx.$$

Therefore, for any sufficiently small $\varepsilon_0 > 0$ *such that* $V(d_{S^*}^{opt}) - V(d_S^{opt}) \leq \varepsilon_0$, *we have*

$$1-\rho \leq \frac{\varepsilon_0}{a_0(S^*\Sigma S^{*T})}\left\{\int_0^{\sqrt{r_0^2-3/4}} (1-x^2)\exp\left(-\frac{x^2}{2}\right) dx\right\}^{-1}.$$

As a result, for any $r_0 > 0$ *and sufficiently small* $\varepsilon_0 > 0$ *such that* $V(d_{S^*}^{opt}) - V(d_S^{opt}) \leq \varepsilon_0$, *we have* $1-\rho \leq \bar{c}_*\varepsilon_0$ *for some constant* $\bar{c}_* > 0$. *By* (S.17), $E\{\tau(X)|SX\} \stackrel{d}{=} C_0\rho^2\mathbb{Z}^2 + C_1(S)$ *for some constant* $C_0 > 0$, *some function* $C_1(\cdot)$, *and* $\mathbb{Z} \sim N(0,1)$. *For any* ρ *that satisfies* $1-\rho \leq \bar{c}_*\varepsilon_0$ *for some sufficiently small* $\varepsilon_0 > 0$, *the density function of* $E\{\tau(X)|SX\}$ *is uniformly bounded. (A6) thus holds.*

Example 4 (Trigonometric contrast function) *Assume* $\tau(x) = a_0\sin(b_0S^*x)$ *for some* $S^* \in \mathbb{R}^{1\times p}$ *and* $a_0, b_0 \in \mathbb{R}$ *such that* $\|S^*\|_2 = 1$, $a_0 \neq 0$, $b_0 > 0$. *For any sketching matrix* $S \in \boldsymbol{\mathcal{S}} \subseteq \mathbb{R}^{q\times p}$, SX *is independent of* $S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX$. *Since*

$$\begin{aligned}
\sin(b_0S^*X) &= \sin[b_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX + b_0\{S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX\}] \\
&= \sin\{b_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX\}\cos[b_0\{S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX\}] \\
&+ \cos\{b_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX\}\sin[b_0\{S^*X - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX\}],
\end{aligned}$$

we have $E\{\sin(b_0S^*X)|SX\} = \sin\{b_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX\}E\cos(b_0\kappa\mathbb{Z})$ *where* $\kappa^2 = S^*\Sigma S^{*T} - (S^*\Sigma S^T)(S\Sigma S^T)^{-1}(S\Sigma S^{*T})$ *and* $\mathbb{Z}$ *follows a standard normal distribution. Using integra-*

tion by parts, we have

$$\frac{\partial E\cos(b_0\kappa\mathbb{Z})}{\partial\kappa} = -b_0 E\mathbb{Z}\sin(b_0\kappa\mathbb{Z}) = -\frac{b_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x\exp\left(-\frac{x^2}{2}\right)\sin(b_0\kappa x)dx$$

$$= \frac{b_0}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\sin(b_0\kappa x)\Big|_{-\infty}^{\infty} - \frac{b_0^2\kappa}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2}{2}\right)\cos(b_0\kappa x) = -b_0^2\kappa E\cos(b_0\kappa\mathbb{Z}).$$

Therefore, $E\cos(b_0\kappa\mathbb{Z}) = C\exp(-b_0^2\kappa^2/2)$ *for some constant* $C > 0$. *Since* $E\cos(b_0\kappa\mathbb{Z}) = 1$ *when* $\kappa = 0$, *we have* $E\cos(b_0\kappa\mathbb{Z}) = \exp(-b_0^2\kappa^2/2)$.

Let $\rho = |S^*\Sigma S^T|(S\Sigma S^T)^{-1/2}(S^*\Sigma S^{*T})^{-1/2}$. *To summarize, we've shown*

$$E\{\tau(X)|SX\} = a_0\exp\{-b_0^2 S^*\Sigma S^{*T}(1-\rho^2)\}\sin\{b_0(S^*\Sigma S^T)(S\Sigma S^T)^{-1}SX\}. \tag{S.20}$$

Since $V(d_{S^*}^{opt}) \overset{d}{=} |a_0|\max(sgn(a_0)\sin(b_0\sqrt{S^*\Sigma S^{*T}}\mathbb{Z}, 0)$ *where* $\mathbb{Z}\sim N(0,1)$, *we have* $V(d_{S^*}^{opt}) \geq 2\varepsilon_*$ *for some* $\varepsilon_* > 0$. *Set* $\varepsilon_0 = \varepsilon_*$. *For any* S *such that* $V(d_{S^*}^{opt}) - V(d_S^{opt}) \leq \varepsilon_0$, *we have* $V(d_S^{opt}) \geq \varepsilon_0$. *It follows from* (S.20) *that*

$$V(d_S^{opt}) = |a_0|\exp\{-b_0^2 S^*\Sigma S^{*T}(1-\rho^2)\}\max(sgn(a_0)\sin\{b_0(S^*\Sigma S^{*T})^{1/2}\rho\mathbb{Z}\}, 0).$$

For any such S, *we have* $|a_0|\max(sgn(a_0)\sin\{b_0(S^*\Sigma S^{*T})^{1/2}\rho\mathbb{Z}\}, 0) \geq \varepsilon_0$ *and hence* ρ *satisfies* $\rho \geq \bar{\varepsilon}$ *for some* $\bar{\varepsilon} > 0$. *Notice that for any such* S, *we have* $E\{\tau(X)|SX\} \overset{d}{=} \kappa_1\sin(\kappa_2\mathbb{Z})$ *with* κ_1 *and* κ_2 *uniformly bounded away from* 0. *For any sufficiently small* $t > 0$, *we have*

$$Pr(|E\{\tau(X)|SX\}| \leq t) \leq \sum_{k=0,\pm1,\pm2\ldots,} Pr(|\mathbb{Z} - k\pi| \leq \arcsin(t/\kappa_1)/\kappa_2) = O(t).$$

Assumption (A6) is thus satisfied.

D Proofs

D.1 Proof of Lemma 3.1

We first show (i)⇒(ii). Assume (i) holds. We have $\Pr(\tau(X)\geq 0) = 1$ or $\Pr(\tau(X)\leq 0) = 1$. When $\Pr(\tau(X)\geq 0) = 1$, it follows from Conditions (A1)-(A3) that

$$V(d^{opt}) - V(1) = \mathrm{E}\tau(X)I(\tau(X) > 0) - \mathrm{E}\tau(X) = \mathrm{E}\{-\tau(X)I(\tau(X) < 0)\} = 0. \tag{S.1}$$

Similarly, when $\Pr(\tau(X) \leq 0) = 1$, we have $V(d^{opt}) = V(0)$. This verifies that $V(d^{opt}) \leq \max(V(0), V(1))$. Besides, since d^{opt} maximizes V, we have $V(d^{opt}) \geq \max(V(0), V(1))$. (ii) therefore follows.

We now show (ii)$\Rightarrow$(i). Assume $V(d^{opt}) = V(1)$. Then it follows from (S.1) that

$$\mathrm{E}\{-\tau(X)I(\tau(X) < 0)\} = 0.$$

Observe that $-\tau(X)I(\tau(X) < 0)$ is nonnegative and integrable. Therefore, we have $\tau(X)I(\tau(X) < 0) = 0$, almost surely. This implies that $\Pr(\tau(X) \leq 0) = 1$. Similarly, when $V(d^{opt}) = V(0)$, we can show $\Pr(\tau(X) \geq 0) = 1$. This completes the proof.

D.2 Proof of Theorem 3.1

It follows from Bonferroni's inequality that

$$\begin{aligned}
\Pr(\widehat{T}_{CV} > z_{\alpha/2}) &\leq \Pr\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2})}{\max(\hat{\sigma}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2}), \delta_n)} > z_{\alpha/2}\right) + \Pr\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_2}(\hat{d}_{\mathcal{I}_1})}{\max(\hat{\sigma}_{\mathcal{I}_2}(\hat{d}_{\mathcal{I}_1}), \delta_n)} > z_{\alpha/2}\right) \\
&\leq 2\Pr\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2}\right),
\end{aligned}$$

where $\mathcal{I}$ is a random subset of $\{1, \ldots, N\}$ of size n. Hence, it suffices to show

$$\Pr\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2}\right) \leq \frac{\alpha}{2}.$$

Under H_0, we have $V(d^{opt}) = V(1)$. This implies for any treatment regime d, we have $V(d) \leq V(1)$ and hence $\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) \leq 0$. Therefore, it suffices to show

$$\Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2}\right) \leq \frac{\alpha}{2}. \tag{S.2}$$

Given $\mathcal{I}$ and $\{O_i\}_{i\in\mathcal{I}^c}$, the estimated treatment regime $\hat{d}_{\mathcal{I}^c}$ is fixed. Let

$$\sigma_0^2(\hat{d}_{\mathcal{I}^c}) = \mathrm{Var}\left\{\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right)Y\{1 - \hat{d}_{\mathcal{I}^c}(X)\} \mid \mathcal{I}, \{O_i\}_{i\in\mathcal{I}^c}\right\}.$$

For a given $\varepsilon > 0$, define the event $\mathcal{A} = \{\sigma_0(\hat{d}_{\mathcal{I}^c}) \geq \varepsilon\delta_n\}$. In the following, we show

$$\limsup_n \Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2} \mid \mathcal{A}\right) \leq \frac{\alpha}{2}, \tag{S.3}$$

$$\limsup_n \Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2} \mid \mathcal{A}^c\right) \leq \frac{\varepsilon^2}{z^2_{\alpha/2}}. \tag{S.4}$$

Combining (S.3) together with (S.4), we obtain

$$\limsup_n \Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2}\right) \leq \max\left(\frac{\alpha}{2}, \frac{\varepsilon^2}{z^2_{\alpha/2}}\right).$$

Since ε can be arbitrarily small, this proves (S.2).

Proof of (S.3). Define

$$\xi_i = \left(\frac{1-A_i}{1-\pi(X_i)} - \frac{A_i}{\pi(X_i)}\right) Y_i\{1-\hat{d}_{\mathcal{I}^c}(X_i)\} - \mathrm{E}\left\{\left(\frac{1-A_i}{1-\pi(X_i)} - \frac{A_i}{\pi(X_i)}\right) Y_i\{1-\hat{d}_{\mathcal{I}^c}(X_i)\}\right\}.$$

Notice that

$$\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} \leq \frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})} = \frac{\sqrt{n}\bar{\xi}_{\mathcal{I}}}{\sqrt{(\sum_{i\in\mathcal{I}}\xi_i^2 - n\bar{\xi}^2_{\mathcal{I}})/(n-1)}},$$

where $\bar{\xi}_{\mathcal{I}} = \sum_{i\in\mathcal{I}}\xi_i/|\mathcal{I}|$. Therefore, we have

$$\begin{aligned}
& \Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2} \big| \mathcal{I}, \{O_i\}_{i\in\mathcal{I}^c}\right) \\
\leq\ & \Pr\left(\frac{\sqrt{n}\bar{\xi}_{\mathcal{I}}}{\sqrt{(\sum_{i\in\mathcal{I}}\xi_i^2 - n\bar{\xi}^2_{\mathcal{I}})/(n-1)}} > z_{\alpha/2} \big| \mathcal{I}, \{O_i\}_{i\in\mathcal{I}^c}\right) \\
\leq\ & \Pr\left(\frac{U}{\sqrt{1-U^2/n}} > \sqrt{\frac{n-1}{n}} z_{\alpha/2} \big| \mathcal{I}, \{O_i\}_{i\in\mathcal{I}^c}\right) = \Pr\left(U > \bar{c}(\alpha, n) \big| \mathcal{I}, \{O_i\}_{i\in\mathcal{I}^c}\right).
\end{aligned} \tag{S.5}$$

where U is the self-normalized sum $\sqrt{n}\bar{\xi}_{\mathcal{I}}/\sqrt{\sum_{i\in\mathcal{I}}\xi_i^2/n}$, and

$$\bar{c}(\alpha, n) = \frac{\sqrt{(n-1)/n}\, z_{\alpha/2}}{\sqrt{1+(n-1)z^2_{\alpha/2}/n^2}}.$$

Under (A3) and the condition $\mathrm{E}|Y|^3 = O(1)$, it follows from Theorem 7.4 in Peña et al.

(2008) that

$$\Pr\left(U > \bar{c}(\alpha, n) \big| \mathcal{I}, \{O_i\}_{i \in \mathcal{I}^c}\right) = \bar{\Phi}(c(\alpha, n)) + O\left(\frac{1}{\sqrt{n}\sigma_0^3(\hat{d}_{\mathcal{I}^c})}\right). \tag{S.6}$$

On the set $\sigma_0(\hat{d}_{\mathcal{I}^c}) \geq \varepsilon\delta_n$, it follows from (S.6) that

$$\Pr\left(U > \bar{c}(\alpha, n) \big| \mathcal{I}, \{O_i\}_{i \in \mathcal{I}^c}, \mathcal{A}\right) = \bar{\Phi}(c(\alpha, n)) + O(1)\left(\frac{1}{\varepsilon^3\sqrt{n}\delta_n^3}\right).$$

By (S.5), this further implies

$$\Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2} \big| \mathcal{I}, \{O_i\}_{i \in \mathcal{I}^c}, \mathcal{A}\right) \leq \bar{\Phi}(c(\alpha, n)) + \frac{\bar{c}}{\varepsilon^3\sqrt{n}\delta_n^3}, \tag{S.7}$$

for some constant $\bar{c} > 0$. Take the conditional expectation of $\mathcal{I}, \{O_i\}_{i \in \mathcal{I}^c}$ given the event $\mathcal{A}$ on both sides of (S.7), we obtain

$$\Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2} \big| \mathcal{A}\right) \leq \bar{\Phi}(c(\alpha, n)) + \frac{\bar{c}}{\varepsilon^3\sqrt{n}\delta_n^3}.$$

Since $c(\alpha, n) \to z_{\alpha/2}$, under the condition that $\delta_n \gg n^{-1/6}$, we obtain (S.3).

Proof of (S.4). It follows from Chebyshev's inequality that

$$\begin{aligned}
& \Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2} \big| \mathcal{I}, \{O_i\}_{i \in \mathcal{I}^c}\right) \\
\leq \; & \Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\delta_n} > z_{\alpha/2} \big| \mathcal{I}, \{O_i\}_{i \in \mathcal{I}^c}\right) \\
\leq \; & \mathrm{E}\left(\frac{n\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}^2}{z_{\alpha/2}^2\delta_n^2} \big| \mathcal{I}, \{O_i\}_{i \in \mathcal{I}^c}\right) = \frac{\sigma_0^2(\hat{d}_{\mathcal{I}^c})}{z_{\alpha/2}^2\delta_n^2}.
\end{aligned}$$

Under the event $\mathcal{A}^c$, we have

$$\frac{\sigma_0^2(\hat{d}_{\mathcal{I}^c})}{z_{\alpha/2}^2\delta_n^2} \leq \frac{\varepsilon^2}{z_{\alpha/2}^2},$$

and hence

$$\Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2} \big| \mathcal{I}, \{O_i\}_{i \in \mathcal{I}^c}, \mathcal{A}^c\right) \leq \frac{\varepsilon^2}{z_{\alpha/2}^2}. \tag{S.8}$$

Take the conditional expectation of $\mathcal{I}, \{O_i\}_{i\in\mathcal{I}^c}$ given the event $\mathcal{A}^c$ on both sides of (S.8), we obtain (S.4).

Assume

$$\mathrm{Var}\left\{\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right) Y\{1-\hat{d}_{\mathcal{I}_j}(X)\} \mid \{O_i\}_{i\in\mathcal{I}_j}\right\} = o_p(\delta_n),$$

for $j = 1, 2$. Similar to the proof of (S.4), we can show that

$$\Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2}\right) \leq \frac{\varepsilon^2}{z_{\alpha/2}^2},$$

for any $\varepsilon > 0$. Since ε can be arbitrarily small, we have

$$\Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2}\right) \to 0.$$

By Bonferroni's inequality, we obtain

$$\Pr(\widehat{T}_{CV} > z_{\alpha/2}) \to 0.$$

The proof is hence completed.

D.3 Proof of Theorem 3.2

We first show under Assumptions (C1) and (C2),

$$V(\hat{d}_{\mathcal{I}}) = V(d^{opt}) + o_p(n^{-1/2}), \tag{S.9}$$

for an arbitrary subset $\mathcal{I}$ of $\{1, \dots, N\}$ of size n. For a given $\varepsilon > 0$, it follows from Condition (C2) and Markov's inequality that

$$\Pr\left(\mathrm{E}^X|\hat{\tau}_{\mathcal{I}}(X) - \tau(X)|^2 > \varepsilon n^{-(2+\gamma)/(2+2\gamma)}\right) \leq \frac{\mathrm{E}\mathrm{E}^X|\hat{\tau}_{\mathcal{I}}(X) - \tau(X)|^2}{\varepsilon n^{-(2+\gamma)/(2+2\gamma)}} = \frac{\mathrm{E}|\hat{\tau}_{\mathcal{I}}(X) - \tau(X)|^2}{\varepsilon n^{-(2+\gamma)/(2+2\gamma)}} \to 0,$$

where the expectation E^X is taken with respect to X independent of the training samples $\{O_i\}_{i=1,\dots,N}$ and $\mathcal{I}_1, \mathcal{I}_2$. Since ε can be arbitrarily small, we have

$$\mathrm{E}^X|\hat{\tau}_{\mathcal{I}_j}(X) - \tau(X)|^2 = o_p\left(n^{-(2+\gamma)/(2+2\gamma)}\right), \tag{S.10}$$

for $j = 1, 2$. In view of (S.10), Under (C1) and (C2), (S.9) follows from an application of Theorem 8 in Luedtke and van der Laan (2016).

We now show $\Pr(\widehat{T}_{CV} > z_{\alpha/2}) \to 1$ when $h_n \gg n^{-1/2}$. By definition, we have

$$\begin{aligned}\Pr(\widehat{T}_{CV} > z_{\alpha/2}) &= \Pr\left\{\max_{j=1,2}\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2})}{\max(\hat{\sigma}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2}), \delta_n)}, \frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_2}(\hat{d}_{\mathcal{I}_1})}{\max(\hat{\sigma}_{\mathcal{I}_2}(\hat{d}_{\mathcal{I}_1}), \delta_n)}\right) > z_{\alpha/2}\right\} \\ &\geq \Pr\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2}\right),\end{aligned}$$

for a random subset $\mathcal{I}$ of $\{1, \dots, N\}$ of size n. Hence, it suffices to show

$$\Pr\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} > z_{\alpha/2}\right) \to 1,$$

or equivalently

$$\Pr\left(\frac{\sqrt{n}\{-\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) + \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} \geq \frac{\sqrt{n}\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} - z_{\alpha/2}\right) \to 0. \tag{S.11}$$

It follows from (S.9) that

$$\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) = V(\hat{d}_{\mathcal{I}^c}) - V(1) = V(\hat{d}_{\mathcal{I}^c}) - V(d^{opt}) + V(d^{opt}) - V(1) = h_n + o_p(n^{-1/2}).$$

Under the assumption $h_n \gg n^{-1/2}$, for sufficiently large n, we have $h_n + o(n^{-1/2}) \geq h_n/2$. Therefore, we have

$$\begin{aligned}&\Pr\left(\frac{\sqrt{n}\{-\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) + \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} \geq \frac{\sqrt{n}\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} - z_{\alpha/2}\right) \\ \leq\ &\Pr\left(\frac{\sqrt{n}\{-\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) + \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} \geq \frac{\sqrt{n}h_n/2}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} - z_{\alpha/2}\right) + o(1).\end{aligned}$$

Therefore, it suffices to show

$$\Pr\left(\frac{\sqrt{n}\{-\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) + \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} \geq \frac{\sqrt{n}h_n/2}{\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}), \delta_n)} - z_{\alpha/2}\right) \to 0.$$

By the dominated convergence theorem, it suffices to show

$$\Pr\left(\sqrt{n}\{\widehat{\mathrm{EVD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})-\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}\geq\frac{\sqrt{n}h_n}{2}-z_{\frac{\alpha}{2}}\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}),\delta_n)\mid\{O_i\}_{i\in\mathcal{I}^c},\mathcal{I}\right)=o_p(1)\tag{S.12}$$

Given $\{O_i\}_{i\in\mathcal{I}^c}$ and $\mathcal{I}$, the estimated treatment regime $\hat{d}_{\mathcal{I}^c}$ is fixed. Thus, conditional on $\{O_i\}_{i\in\mathcal{I}^c}$ and $\mathcal{I}$, it follows from the law of large numbers that,

$$\frac{1}{n}\sum_{i\in\mathcal{I}}\xi_i^2\xrightarrow{P}\mathrm{E}\xi_1^2\ \text{ and }\ \frac{1}{n}\sum_{i\in\mathcal{I}}\xi_i\xrightarrow{P}0,\tag{S.13}$$

where ξ_i is defined in the proof of Theorem 1. By (S.13), we have that conditional on $\{O_i\}_{i\in\mathcal{I}^c}$ and $\mathcal{I}$,

$$\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\xrightarrow{P}\sigma_0(\hat{d}_{\mathcal{I}^c}).\tag{S.14}$$

Under (A3) and the condition $\mathrm{E}|Y|^3=O(1)$, we have

$$\begin{aligned}\sigma_0(\hat{d}_{\mathcal{I}^c})^2 &\leq \mathrm{E}\left\{\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\{1-\hat{d}_{\mathcal{I}^c}\}\right\}^2\leq\mathrm{E}\left\{\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\right\}^2\\ &\leq \left\{\mathrm{E}\left|\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\right|^3\right\}^{2/3}=O(1).\end{aligned}\tag{S.15}$$

Combining this together with (S.14) and the conditions $h_n\gg n^{-1/2}$, $\delta_n\to 0$, we obtain that

$$\Pr\left(\frac{\sqrt{n}h_n}{2}-z_{\alpha/2}\max(\hat{\sigma}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}),\delta_n)\geq\frac{\sqrt{n}h_n}{4}\mid\{O_i\}_{i\in\mathcal{I}^c},\mathcal{I}\right)\xrightarrow{P}1.$$

In view of (S.12), it suffices to show

$$\Pr\left(\sqrt{n}\{-\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})+\widehat{\mathrm{EVD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}\geq\frac{\sqrt{n}h_n}{4}\mid\{O_i\}_{i\in\mathcal{I}^c},\mathcal{I}\right)=o_p(1).$$

However, this is immediate to see since

$$\begin{aligned}&\Pr\left(\sqrt{n}\{-\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})+\widehat{\mathrm{EVD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}\geq\frac{\sqrt{n}h_n}{4}\mid\{O_i\}_{i\in\mathcal{I}^c},\mathcal{I}\right)\\ \leq\ &\mathrm{E}\left(\frac{16n\{-\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})+\widehat{\mathrm{EVD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})\}^2}{nh_n^2}\mid\{O_i\}_{i\in\mathcal{I}^c},\mathcal{I}\right)=\frac{16\sigma_0^2(\hat{d}_{\mathcal{I}^c})}{nh_n^2}=o(1),\end{aligned}$$

where the first inequality is due to the Chebyshev's inequality and the last equality is due to (S.15) and the condition $h_n \gg n^{-1/2}$.

We now show the power of our test statistic in the regular cases where $\Pr(\tau(X)=0)=0$. We begin by providing an upper bound for $\mathrm{E}^X|\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)|$. Since $\Pr(\tau(X)=0)=0$, it follows from Assumption (C1) that

$$\Pr(|\tau(X)|\le t)=O(t^{\gamma}) \qquad \forall 0<t\le\delta_0. \tag{S.16}$$

Observe that

$$\begin{aligned}\mathrm{E}^X|\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)| &= \mathrm{E}^X|\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)|I(|\tau(X)|\le t) \\ &+ \mathrm{E}^X|\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)|I(|\tau(X)|>t) \triangleq I_1+I_2.\end{aligned} \tag{S.17}$$

For any $t\le\delta_0$, it follows from (S.16) that $I_1=O(t^{\gamma})$. Note that for any $a,b\in\mathbb{R}$, we have $I(a>0)=I(b>0)$ when $|a-b|>|a|$. Therefore, we have

$$\begin{aligned}I_2 &\le \mathrm{E}^X I(|\tau(X)-\hat{\tau}_{\mathcal{I}^c}(X)|>|\tau(X)|)I(|\tau(X)|>t)\le \mathrm{E}^X\frac{|\tau(X)-\hat{\tau}_{\mathcal{I}^c}(X)|^2}{|\tau(X)|^2}I(|\tau(X)|>t) \\ &\le \frac{1}{t^2}\mathrm{E}^X|\tau(X)-\hat{\tau}_{\mathcal{I}^c}(X)|^2 I(|\tau(X)|>t)\le\frac{1}{t^2}\mathrm{E}^X|\tau(X)-\hat{\tau}_{\mathcal{I}^c}(X)|^2,\end{aligned}$$

where the second inequality is due to Markov's inequality.

Set $t=\left(\mathrm{E}^X|\tau(X)-\hat{\tau}_{\mathcal{I}^c}(X)|^2\right)^{1/(2+\gamma)}$, we obtain

$$I_1+I_2=\left(\mathrm{E}^X|\tau(X)-\hat{\tau}_{\mathcal{I}^c}(X)|^2\right)^{\gamma/(2+\gamma)}=o_p\left(n^{-\gamma/(2+2\gamma)}\right),$$

where the last equality is due to (S.10). By (S.17), this implies

$$\mathrm{E}^X|\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)|=o_p\left(n^{-\gamma/(2+2\gamma)}\right)=o_p(1). \tag{S.18}$$

Therefore, the estimated treatment regime is consistent to d^{opt} in the regular cases.

As a result, it follows from Hölder's inequality that

$$
\begin{aligned}
& \mathrm{E}^{X,A,Y}\left|\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\{\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)\}\right|^2 \qquad \text{(S.19)}\\
\leq & \left\{\mathrm{E}\left|\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\right|^{3/2}\right\}^{2/3}\left(\mathrm{E}^X\left|\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)\right|^6\right)^{1/3}\\
= & \left\{\mathrm{E}\left|\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\right|^{3/2}\right\}^{2/3}\left(\mathrm{E}^X\left|\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)\right|\right)^{1/3}=o_p(1),
\end{aligned}
$$

where the last equality is due to (A3), the condition $\mathrm{E}|Y|^3=O(1)$ and (S.18).

Similarly, we can show

$$
\begin{aligned}
& \mathrm{E}\left\{\mathrm{E}^{X,A,Y}\left|\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\{\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)\}\right|^2\right\}^{3/2}\\
\leq & \;\mathrm{E}\left\{\mathrm{E}^{X,A,Y}\left|\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\right|^2\right\}^{3/2}\leq \mathrm{E}\mathrm{E}^{X,A,Y}\left|\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\right|^3\\
= & \;\mathrm{E}\left|\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\right|^3=O(1).
\end{aligned}
$$

This implies LHS of (S.19) is uniformly integrable. Hence, it follows from (S.19) that

$$
\mathrm{E}\left|\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)Y\{\hat{d}_{\mathcal{I}^c}(X)-d^{opt}(X)\}\right|^2=o(1). \qquad \text{(S.20)}
$$

Similarly, we can show $V(\hat{d}_{\mathcal{I}^c})-V(d^{opt})$ is uniformly integrable and hence by (S.9),

$$
\left|\mathrm{E}V(\hat{d}_{\mathcal{I}^c})-\mathrm{E}V(d^{opt})\right|=\left|\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})-\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(d^{opt})\right|=o(n^{-1/2}). \qquad \text{(S.21)}
$$

Combing (S.21) with Cauchy-Schwarz inequality, we obtain that

$$
\begin{aligned}
& \mathrm{E}\left|\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})-\widehat{\mathrm{VD}}_{\mathcal{I}}(d^{opt})\}\right|^2\\
\leq & \; 2\mathrm{E}\left|\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})-\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})-\widehat{\mathrm{VD}}_{\mathcal{I}}(d^{opt})+\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(d^{opt})\}\right|^2\\
+ & \; 2n\left|\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c})-\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(d^{opt})\right|^2\leq\frac{2}{n}\sum_{i\in\mathcal{I}}\mathrm{Var}(\eta_i)+\frac{4}{n}\sum_{i\neq j}\mathrm{cov}(\eta_i,\eta_j)+o(1), \qquad \text{(S.22)}
\end{aligned}
$$

where

$$\eta_i = \left(\frac{A_i}{\pi(X_i)} - \frac{1 - A_i}{1 - \pi(X_i)} \right) Y_i \{ \hat{d}_{\mathcal{I}^c}(X_i) - d^{opt}(X_i) \}.$$

Notice that conditional on $\hat{d}_{\mathcal{I}^c}$, η_i and η_j are independent for $i, j \in \mathcal{I}, i \neq j$. This implies

$$\frac{4}{n} \sum_{i \neq j} \mathrm{cov}(\eta_i, \eta_j) = \frac{4}{n} \sum_{i \neq j} \mathrm{Ecov}(\eta_i, \eta_j | \hat{d}_{\mathcal{I}^c}) = 0.$$

In view of (S.20) and (S.22), we obtain

$$\mathrm{E} \left| \sqrt{n} \{ \widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \widehat{\mathrm{VD}}_{\mathcal{I}}(d^{opt}) \} \right|^2 \leq \frac{2}{n} \sum_{i \in \mathcal{I}} \mathrm{Var}(\eta_i) + o(1) = 2\mathrm{Var}(\eta_1) + o(1) \leq 2\mathrm{E}\eta_1^2 + o(1) = o(1).$$

By Chebyshev's inequality, this implies

$$|\sqrt{n} \{ \widehat{\mathrm{VD}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}^c}) - \widehat{\mathrm{VD}}_{\mathcal{I}}(d^{opt}) \}| \xrightarrow{P} 0,$$

and hence

$$|\sqrt{n} \{ \widehat{\mathrm{VD}}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c}) - \widehat{\mathrm{VD}}_{\mathcal{I}_j}(d^{opt}) \}| \xrightarrow{P} 0, \tag{S.23}$$

for $j = 1, 2$.

Similarly, we can show

$$\left| \frac{1}{n} \sum_{i \in \mathcal{I}_j} \left(\frac{A_i}{\pi_i} - \frac{1 - A_i}{1 - \pi_i} \right)^2 Y_i^2 \{1 - \hat{d}_{\mathcal{I}_j^c}(X_i)\}^2 - \frac{1}{n} \sum_{i \in \mathcal{I}_j} \left(\frac{A_i}{\pi_i} - \frac{1 - A_i}{1 - \pi_i} \right)^2 Y_i^2 \{1 - d^{opt}(X_i)\}^2 \right| \xrightarrow{P} 0,$$

which together with (S.23) implies that

$$\left| \hat{\sigma}_{\mathcal{I}_j}^2(\hat{d}_{\mathcal{I}_j^c}) - \hat{\sigma}_{\mathcal{I}_j}^2(d^{opt}) \right| \xrightarrow{P} 0,$$

for $j = 1, 2$. This immediately implies that

$$\begin{aligned} \left| \hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c}) - \hat{\sigma}_{\mathcal{I}_j}(d^{opt}) \right| &\leq \left(\left| \hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c}) - \hat{\sigma}_{\mathcal{I}_j}(d^{opt}) \right| \left| \hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c}) + \hat{\sigma}_{\mathcal{I}_j}(d^{opt}) \right| \right)^{1/2} \\ &= \left(\left| \hat{\sigma}_{\mathcal{I}_j}^2(\hat{d}_{\mathcal{I}_j^c}) - \hat{\sigma}_{\mathcal{I}_j}^2(d^{opt}) \right| \right)^{1/2} \xrightarrow{P} 0. \end{aligned} \tag{S.24}$$

Similar to (S.14), we can show

$$\left|\hat{\sigma}_{\mathcal{I}_j}(d^{opt}) - \sigma_0\right| \xrightarrow{P} 0. \tag{S.25}$$

This together with (S.24) and the condition $\liminf_n \sigma_0 > 0$ implies that there exists some constant $\bar{c} > 0$ such that

$$\Pr(\cup_{j=1,2}\hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c}) > \bar{c}) \to 1, \qquad \Pr(\cup_{j=1,2}\hat{\sigma}_{\mathcal{I}_j}(d^{opt}) > \bar{c}) \to 1. \tag{S.26}$$

Since $\delta_n \to 0$, we have

$$\Pr\left(\max(\hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c}), \delta_n) = \hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c})\right) \to 1,$$

for $j = 1, 2$. This further implies

$$\Pr(\widehat{T}_{CV} > z_{\alpha/2}) = \Pr\left\{\max\left(\frac{\sqrt{n}\widehat{\text{VD}}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2})}{\hat{\sigma}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2})}, \frac{\sqrt{n}\widehat{\text{VD}}_{\mathcal{I}_2}(\hat{d}_{\mathcal{I}_1})}{\hat{\sigma}_{\mathcal{I}_2}(\hat{d}_{\mathcal{I}_1})}\right) > z_{\alpha/2}\right\} + o(1). \tag{S.27}$$

Besides, under (A3) and the condition $\text{E}|Y^3| = O(1)$, we have

$$\begin{aligned}
\text{Var}\left(\sqrt{n}\widehat{\text{VD}}_{\mathcal{I}_j}(d^{opt})\right) &= \sigma_0^2 \le \text{E}\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right)^2 Y^2 \\
&\le \left(\text{E}\left|\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right|^3 |Y|^3\right)^{2/3} = O(1).
\end{aligned}$$

By Chebyshev's inequality, we obtain $\sqrt{n}\{\widehat{\text{VD}}_{\mathcal{I}_j}(d^{opt}) - \text{E}\widehat{\text{VD}}_{\mathcal{I}_j}(d^{opt})\} = O_p(1)$. By assumption, we have $\sqrt{n}\text{E}\widehat{\text{VD}}_{\mathcal{I}_j}(d^{opt}) = \sqrt{n}h_n = O(1)$. Therefore, we obtain

$$\sqrt{n}\widehat{\text{VD}}_{\mathcal{I}_j}(d^{opt}) = O_p(1), \tag{S.28}$$

for $j = 1, 2$.

It follows from (S.23), (S.24), (S.26) and (S.28) that

$$\begin{aligned}
\left|\frac{\sqrt{n}\widehat{\text{VD}}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c})}{\hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c})} - \frac{\sqrt{n}\widehat{\text{VD}}_{\mathcal{I}_j}(d^{opt})}{\hat{\sigma}_{\mathcal{I}_j}(d^{opt})}\right| &\le \left|\frac{\sqrt{n}\{\widehat{\text{VD}}_{\mathcal{I}_j}(d^{opt}) - \widehat{\text{VD}}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c})\}}{\hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c})}\right| \\
+ \quad \sqrt{n}\left|\widehat{\text{VD}}_{\mathcal{I}_j}(d^{opt})\right| \left|\frac{\hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c}) - \hat{\sigma}_{\mathcal{I}_j}(d^{opt})}{\hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c})\hat{\sigma}_{\mathcal{I}_j}(d^{opt})}\right| &= o_p(1),
\end{aligned} \tag{S.29}$$

for $j = 1, 2$.

Besides, it follows from (S.25), and the conditions $\liminf_n \sigma_0 > 0$, $\sqrt{n}h_n = O(1)$ that

$$\frac{\sqrt{n}h_n}{\hat{\sigma}_{\mathcal{I}_j}(d^{opt})} - \frac{\sqrt{n}h_n}{\sigma_0} \xrightarrow{P} 0.$$

This together with (S.29) implies that

$$\left| \frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c})}{\hat{\sigma}_{\mathcal{I}_j}(\hat{d}_{\mathcal{I}_j^c})} - \frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_j}(d^{opt}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_j}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_j}(d^{opt})} - \frac{\sqrt{n}h_n}{\sigma_0} \right| \xrightarrow{P} 0. \tag{S.30}$$

For any random variables V_1, V_2, V_3, V_4, we have

$$\begin{aligned} & |\max(V_1, V_2) - \max(V_3, V_4)| = |\max(0, V_2 - V_1) - \max(0, V_4 - V_3) + V_1 - V_3| \\ \leq\ & |V_1 - V_3| + |\max(0, V_2 - V_1) - \max(0, V_4 - V_3)| \leq |V_1 - V_3| + |V_2 - V_1 - V_4 + V_3| \\ \leq\ & 2|V_1 - V_3| + |V_2 - V_4|. \end{aligned}$$

Therefore, it follows from (S.29) that

$$\begin{aligned} \Bigg| \max\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_1}(d^{opt})} + \frac{\sqrt{n}h_n}{\sigma_0}, \frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_2}(d^{opt}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_2}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_2}(d^{opt})} + \frac{\sqrt{n}h_n}{\sigma_0} \right) \\ - \max\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2})}{\hat{\sigma}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2})}, \frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_2}(\hat{d}_{\mathcal{I}_1})}{\hat{\sigma}_{\mathcal{I}_2}(\hat{d}_{\mathcal{I}_1})} \right) \Bigg| \xrightarrow{P} 0. \end{aligned}$$

For any $\varepsilon > 0$, this together with (S.27) gives

$$\begin{aligned} \Pr\Bigg(\max\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_1}(d^{opt})}, \frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_2}(d^{opt}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_2}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_2}(d^{opt})} \right) \\ > z_{\alpha/2} - \varepsilon - \frac{\sqrt{n}h_n}{\sigma_0} \Bigg) \leq \Pr(\widehat{T}_{CV} > z_{\alpha/2}) + o(1), \end{aligned} \tag{S.31}$$

and

$$\begin{aligned} \Pr\Bigg(\max\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_1}(d^{opt})}, \frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_2}(d^{opt}) - \mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_2}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_2}(d^{opt})} \right) \\ > z_{\alpha/2} + \varepsilon - \frac{\sqrt{n}h_n}{\sigma_0} \Bigg) \geq \Pr(\widehat{T}_{CV} > z_{\alpha/2}) + o(1). \end{aligned} \tag{S.32}$$

Since $\{\mathcal{O}_i\}_{i\in\mathcal{I}_1}$ and $\{\mathcal{O}_i\}_{i\in\mathcal{I}_2}$ are independent, we have,

$$\begin{aligned}
&\Pr\left(\max\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt})-\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_1}(d^{opt})},\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_2}(d^{opt})-\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_2}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_2}(d^{opt})}\right)\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\left.>z_{\alpha/2}+\varepsilon-\frac{\sqrt{n}h_n}{\sigma_0}\right)\\
&\qquad\qquad=2\Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt})-\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_1}(d^{opt})}>z_{\alpha/2}+\varepsilon-\frac{\sqrt{n}h_n}{\sigma_0}\right)\\
&\qquad\qquad\quad-\Pr^2\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}}(d^{opt})-\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}}(d^{opt})}>z_{\alpha/2}+\varepsilon-\frac{\sqrt{n}h_n}{\sigma_0}\right),
\end{aligned}$$

for a simple random sample $\mathcal{I}$ of size n.

Assume for now, we have for any $|\varepsilon|\le 1$,

$$\begin{aligned}
&\Pr\left(\frac{\sqrt{n}\{\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt})-\mathrm{E}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d^{opt})\}}{\hat{\sigma}_{\mathcal{I}_1}(d^{opt})}>z_{\alpha/2}+\varepsilon-\frac{\sqrt{n}h_n}{\sigma_0}\right)\\
=\;&\bar{\Phi}\left(z_{\alpha/2}+\varepsilon-\frac{\sqrt{n}h_n}{\sigma_0}\right)+o(1), \qquad (S.33)
\end{aligned}$$

where the little-o term is uniform in ε. In view of (S.31) and (S.32), we have

$$\Pr(\widehat{T}_{CV}>z_{\alpha/2})\le 2\bar{\Phi}\left(z_{\alpha/2}+\varepsilon-\frac{\sqrt{n}h_n}{\sigma_0}\right)-\bar{\Phi}^2\left(z_{\alpha/2}+\varepsilon-\frac{\sqrt{n}h_n}{\sigma_0}\right)+o(1),$$

and

$$\Pr(\widehat{T}_{CV}>z_{\alpha/2})\ge 2\bar{\Phi}\left(z_{\alpha/2}-\varepsilon-\frac{\sqrt{n}h_n}{\sigma_0}\right)-\bar{\Phi}^2\left(z_{\alpha/2}-\varepsilon-\frac{\sqrt{n}h_n}{\sigma_0}\right)+o(1),$$

for any $0<\varepsilon\le 1$. Let $\varepsilon\to 0$, we obtain

$$\Pr(\widehat{T}_{CV}>z_{\alpha/2})=2\bar{\Phi}\left(z_{\alpha/2}-\frac{\sqrt{n}h_n}{\sigma_0}\right)-\bar{\Phi}^2\left(z_{\alpha/2}-\frac{\sqrt{n}h_n}{\sigma_0}\right)+o(1).$$

This gives the asymptotic power function of $\widehat{T}_{CV}$. Therefore, it remains to show (S.33). Similar to (S.5), we can show LHS of (S.33) is equal to

$$\Pr(V>\bar{c}(\alpha,n,\varepsilon,h_n)),$$

where V is the self-normalized sum $\sqrt{n}\bar{\zeta}_{\mathcal{I}_1}/\{\sum_{i\in\mathcal{I}_1}\zeta_i^2/n\}$ where

$$\zeta_i = \left(\frac{A_i}{\pi(X_i)} - \frac{1-A_i}{1-\pi(X_i)}\right)Y_i\{1-d^{opt}(X_i)\} - \mathrm{E}\left(\frac{A_1}{\pi(X_1)} - \frac{1-A_1}{1-\pi(X_1)}\right)Y_1\{1-d^{opt}(X_1)\},$$

$\bar{\zeta} = \sum_{i\in\mathcal{I}_1}\zeta_i^2/n$, and

$$\bar{c}(\alpha, n, \varepsilon, h_n) = \frac{\sqrt{(n-1)/n}(z_{\alpha/2}+\varepsilon+\sqrt{n}h_n/\sigma_0)}{\sqrt{1+(n-1)(z_{\alpha/2}+\varepsilon+\sqrt{n}h_n/\sigma_0)/n^2}}.$$

Under the condition $\liminf_n \sigma_0 > 0$, $\sqrt{n}h_n = O(1)$, $\bar{c}(\alpha, n, \varepsilon, h_n)$ is equivalent to $z_{\alpha/2}+\varepsilon+\sqrt{n}h_n/\sigma_0$. Moreover, it follows from Theorem 7.4 in Peña et al. (2008) that

$$\Pr(V > \bar{c}(\alpha, n, \varepsilon, h_n) \mid \mathcal{I}_1) = \bar{\Phi}(\bar{c}(\alpha, n, \varepsilon, h_n)) + O(1)\left(\frac{1}{\sqrt{n}\sigma_0^3}\right),$$

where the $O(1)$ term is bounded by an absolute constant. Therefore, we have

$$\Pr(V > \bar{c}(\alpha, n, \varepsilon, h_n)) = \bar{\Phi}(\bar{c}(\alpha, n, \varepsilon, h_n)) + o(1) = \bar{\Phi}(z_{\alpha/2}+\varepsilon+\sqrt{n}h_n/\sigma_0) + o(1).$$

This proves (S.33). The proof is hence completed.

D.4 Proof of Theorem 3.4

D.4.1 Consistency of the test

Recall that $h_n^* = V(d_{S_*}^{opt}) - V(1)$. We first show

$$\Pr\left(\widehat{T}_{SRP} > z_{\alpha/2}\right) \to 1,$$

when $h_n^* \gg \max(\sqrt{\log B}/\sqrt{n}, n^{-r_0/2}\sqrt{\log n})$. By definition, it suffices to show that

$$\Pr\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2}^{S_{\mathcal{I}_2}})}{\max(\hat{\sigma}_{\mathcal{I}_1}(\hat{d}_{\mathcal{I}_2}^{S_{\mathcal{I}_2}}), \delta_n)} > z_{\alpha/2}\right) \to 1.$$

Similar to the proof of Theorem 3.2, it suffices to show that with probability tending to 1,

$$\mathrm{VD}(\hat{d}_{\mathcal{I}_2}^{S_{\mathcal{I}_2}}) \gg \max(n^{-1/2}\sqrt{\log B}, n^{-r_0/2}\sqrt{\log n}). \tag{S.34}$$

According to Algorithm 2, we generate $S_1, S_2, \ldots, S_B$ according as S_0. Let $\{\mathcal{I}_j^{(k)}\}_{k=1}^{\mathbb{K}}$

be a partition of $\mathcal{I}_j$ for $j=1,2$, and let $\mathcal{I}_j^{(k)-}$ be the subset of $\mathcal{I}_j$ excluding $\mathcal{I}_j^{(k)}$. For any $k=1,\ldots,\mathbb{K}$, it follows from Lemma A.2 in Chernozhukov et al. (2014) that for every $t>0$, we have

$$
\begin{aligned}
& \Pr\left\{\max_{b=1}^{B}\left|n\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})-\mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right)\right|\right. \\
\geq\ & \left. 2\mathrm{E}\left|n\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})-\mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right)\right|+t \mid \{S_b\}_{b=1}^B,\{O_i\}_{i\in\mathcal{I}_2^{(k)-}}\right\}\leq\exp\left(-\frac{t^2}{3n\sigma^2}\right) \\
+\ & \frac{C_0 n}{t^3}\mathrm{E}\left\{\max_{b=1}^{B}\left|\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right|^3|Y|^3\left|1-\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X)\right|^3|\{S_b\}_{b=1}^B,\{O_i\}_{i\in\mathcal{I}_2^{(k)-}}\right\},
\end{aligned}
$$

for some universal constant $C_0>0$, where

$$
\sigma^2=\mathrm{E}\left\{\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)^2 Y^2\left|1-\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X)\right|^2|\{S_b\}_{b=1}^B,\{O_i\}_{i\in\mathcal{I}_2^{(k)-}}\right\}.
$$

Under (A3) and the condition $\mathrm{E}|Y|^3=O(1)$, we have

$$
\mathrm{E}\left(\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right)^2 Y^2\leq\left\{\mathrm{E}\left|\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right|^3|Y|^3\right\}^{2/3}=O(1).
$$

This implies

$$
\sigma^2=O(1), \tag{S.35}
$$

where the big-O term is independent of $\{S_b\}_{b=1}^B$ and $\{O_i\}_{i\in\mathcal{I}_2^{(k)-}}$. In addition,

$$
\begin{aligned}
& \mathrm{E}\left\{\max_{b=1}^{B}\left|\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right|^3|Y|^3\left|1-\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X)\right|^3\mid\{S_b\}_{b=1}^B,\{O_i\}_{i\in\mathcal{I}_2^{(k)-}}\right\} \\
\leq\ & \mathrm{E}\left|\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right|^3|Y|^3=O(1),
\end{aligned}
$$

where the big-O term is independent of $\{S_b\}_{b=1}^B$ and $\{O_i\}_{i\in\mathcal{I}_2^{(k)-}}$. Therefore, we obtain

$$
\begin{aligned}
\Pr\left\{\max_{b=1}^{B}\left|n\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})-\mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right)\right|\geq 2\mathrm{E}\left|n\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})-\mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right)\right|+t\right\} \\
\leq\exp\left(-C_1t^2/n\right)+\frac{C_2n}{t^3} \quad \text{(S.36)}
\end{aligned}
$$

for some constant $C_1, C_2 > 0$.

Moreover, it follows from (S.35) and Lemma A3 in Chernozhukov et al. (2014) that

$$\begin{aligned}
& \mathrm{E}\left\{\max_{b=1}^{B}\left|n\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right)\right| \mid \{S_b\}_{b=1}^{B}, \{O_i\}_{i\in\mathcal{I}_2^{(k)-}}\right\} \\
\leq\ & C_3(\sigma\sqrt{n\log B} + \sqrt{\mathrm{E}M^2}\log B) \leq C_4(\sqrt{n\log B} + \sqrt{\mathrm{E}M^2}\log B), \qquad \text{(S.37)}
\end{aligned}$$

where C_4 is some universal constant and

$$M = \max_{i\in\mathcal{I}_2^{(k)}}\max_{b=1}^{B}\left|\left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right)Y_i\{1-\hat{d}^{S_b}_{\mathcal{I}_2^{(k)}}(X_i)\}\right|.$$

Under (A3) and the condition $\mathrm{E}|Y|^3 = O(1)$, we have

$$\begin{aligned}
\mathrm{E}M^2 \ &\leq\ \mathrm{E}\max_{i=1}^{n}\left|\left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right)Y_i\right|^2 \leq \mathrm{E}\left\{\max_{i=1}^{n}\left|\left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right)Y_i\right|^3\right\}^{2/3} \\
&\leq\ \mathrm{E}\left\{\sum_{i=1}^{n}\left|\left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right)Y_i\right|^3\right\}^{2/3} \leq C_5 n^{2/3},
\end{aligned}$$

for some constant $C_5 > 0$. Combining this together with (S.37), we obtain

$$\begin{aligned}
& \mathrm{E}\left\{\max_{b=1}^{B}\left|n\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right)\right| \mid \{S_b\}_{b=1}^{B}, \{O_i\}_{i\in\mathcal{I}_2^{(k)-}}\right\} \\
& \qquad \leq C_6(n^{1/3}\log B + n^{1/2}\sqrt{\log B}) \leq 2C_6 n^{1/2}\sqrt{\log B},
\end{aligned}$$

for some constant $C_6 > 0$, since $\log B = o(n^{1/3})$. Therefore, we obtain

$$\mathrm{E}\max_{b=1}^{B}\left|n\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right)\right| \leq 2C_6 n^{1/2}\sqrt{\log B}.$$

Combining this together with (S.36), we obtain

$$\begin{aligned}
& \Pr\left\{\max_{b=1}^{B}\left|n\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right)\right| \geq 4C_6 n^{1/2}\sqrt{\log B} + t\right\} \\
& \qquad \leq \exp\left(-\frac{C_1 t^2}{n}\right) + \frac{C_2 n}{t^3}, \qquad \text{(S.38)}
\end{aligned}$$

Condition (A5) requires $B \to \infty$. Take $t = n^{1/2}\sqrt{\log B}$ in (S.38), we obtain that with probability tending to 1,

$$\max_{b=1}^{B} \left| n \left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) \right) \right| \le C_7 n^{1/2} \sqrt{\log B},$$

for some constant $C_7 > 0$. Define

$$\mathcal{A}_0^{(2)} = \left\{ \max_{k=1}^{\mathbb{K}} \max_{b=1}^{B} \left| n \left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) \right) \right| \le C_7 n^{1/2} \sqrt{\log B} \right\}.$$

Since $\mathbb{K}$ is fixed, by Bonferroni's inequality, we obtain

$$\Pr\left(\mathcal{A}_0^{(2)}\right) \ge \sum_{k=1}^{\mathbb{K}} \Pr\left\{ \max_{b=1}^{B} \left| n \left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) \right) \right| \le C_7 n^{1/2} \sqrt{\log B} \right\} \to 1. \tag{S.39}$$

Below, we show

$$B \Pr\left\{ d^{\tau}(S_0, S^*) \le c_n n^{-1/2} \right\} \to \infty, \tag{S.40}$$

for some sequence $c_n \to 0$, where S^* in defined in Condition (A5). Notice that q, s are fixed. By (9), it suffices to show

$$B \Pr\left(\sum_{j=1}^{q} \|S_0^{(j)} - S^{*(j)}\|_2^2 \le s^2 q c_n^* n^{-1} \right) \to \infty, \tag{S.41}$$

for some sequence $c_n^* \to 0$. By definition, $\|S^{*(j)}\|_2 = 1$, $\|S^{*(j)}\|_0 \le s$, $\forall j = 1, \dots, q$. Let $\mathcal{M}_j^*$ be the support of $S^{*(j)}$. According to Algorithm 2, we have

$$\sum_{j=1}^{q} \|S^{(j)} - S^{*(j)}\|_2^2 = \sum_{j=1}^{q} \left\| \frac{g_j}{\|g_j\|_2} - S^{*(j)}_{\mathcal{M}_j} \right\|_2^2 + \sum_{j=1}^{q} \|S^{*(j)}_{\mathcal{M}_j^* \cap \mathcal{M}_j^c}\|_2^2.$$

Notice that

$$\Pr\left(\bigcap_{j=1}^{q} \{\mathcal{M}_j^* \subseteq \mathcal{M}_j\} \right) = \frac{1}{\binom{p}{s}^q} \ge \frac{1}{p^{sq}}. \tag{S.42}$$

Under the event defined in (S.42), we have

$$\sum_{j=1}^{q} \|S^{(j)} - S^{*(j)}\|_2^2 = \sum_{j=1}^{q} \left\| \frac{g_j}{\|g_j\|_2} - S^{*(j)}_{\mathcal{M}_j} \right\|_2^2. \tag{S.43}$$

The random vector $g_j/\|g_j\|_2$ follows a uniform distribution on the unit sphere. By Equation (4.11) in Fang et al. (1990), we have

$$\sum_{j=1}^{q} \left\| \frac{g_j}{\|g_j\|_2} - S_{\mathcal{M}_j}^{*(j)} \right\|_2^2 \overset{d}{=} \sum_{j=1}^{q} \left\| \frac{g_j}{\|g_j\|_2} - (\underbrace{0, 0, \dots, 0}_{s-1}, 1)^T \right\|_2^2. \tag{S.44}$$

Consider the variable transformation:

$$\frac{g_j}{\|g_j\|_2} = \{\cos(\Theta_1), \sin(\Theta_1)\cos(\Theta_2), \cdots, \prod_{i=1}^{s-2} \sin(\Theta_i)\cos(\Theta_{s-1}), \prod_{i=1}^{s-1} \sin(\Theta_i)\}^T,$$

for $0 \le \Theta_i \le \pi, \forall 1 \le i \le s-2$ and $0 \le \Theta_{s-1} \le 2\pi$. Notice that

$$\left| \prod_{i=1}^{k-1} \sin(\Theta_i)\cos(\Theta_j) \right| \le |\cos(\Theta_j)| = \left| \sin\left(\frac{\pi}{2} - \Theta_j\right) \right| \le \left| \frac{\pi}{2} - \Theta_j \right|, \quad \forall 1 \le k \le s-1,$$

and

$$\left| \prod_{i=1}^{s-1} \sin(\Theta_i) - 1 \right| \le \sum_{k=1}^{s-1} \left| \prod_{i=1}^{k} \sin(\Theta_i) - \prod_{i=1}^{k-1} \sin(\Theta_{i-1}) \right| \le \sum_{k=1}^{s-1} |\sin(\Theta_k) - 1|$$

$$= \sum_{k=1}^{s-1} \left| \cos\left(\frac{\pi}{2} - \Theta_k\right) - 1 \right| = 2\sum_{k=1}^{s-1} \sin^2\left(\frac{\pi}{4} - \frac{\Theta_k}{2}\right) \le 2\sum_{k=1}^{s-1} \sin\left|\frac{\pi}{4} - \frac{\Theta_k}{2}\right| \le \sum_{k=1}^{s-1} \left|\frac{\pi}{2} - \Theta_k\right|.$$

Since s is bounded, it follows from Equation (4.15) in Fang et al. (1990) that the density function of $\Theta_1, \Theta_2, \dots, \Theta_{s-1}$ around $\pi/2$ are bounded from below. In addition, these $s-1$ random variables are independent. By (S.44), we have

$$\begin{aligned} \Pr\left(\bigcap_{j=1}^{q} \left\{ \left\| \frac{g_j}{\|g_j\|_2} - S_{\mathcal{M}_j}^{*(j)} \right\|_2 \le \frac{s\sqrt{c_n^*}}{\sqrt{n}} \right\} \right) &= \Pr^q\left(\left\{ \left\| \frac{g_j}{\|g_j\|_2} - S_{\mathcal{M}_j}^{*(j)} \right\|_2 \le \frac{s\sqrt{c_n^*}}{\sqrt{n}} \right\} \right) \\ &\ge \Pr^q\left(\bigcap_{i=1}^{s-1} \left| \frac{\pi}{2} - \Theta_i \right| \le \frac{\sqrt{c_n^*}}{\sqrt{n}} \right) \ge \bar{c}_{**} \left(\frac{c_n^*}{n} \right)^{q(s-1)/2}, \end{aligned}$$

for some constant $\bar{c}_{**} > 0$. This together with (S.43) and (S.42) yields that

$$\Pr\left(\sum_{j=1}^{q} \|S^{(j)} - S^{*(j)}\|_2^2 \le \frac{qs^2 c_n^*}{n} \right) \ge \frac{\bar{c}_{**} c_n^{*q(s-1)/2}}{(p\sqrt{n})^{q(s-1)}}.$$

Therefore, it is immediate to see (S.41) holds with proper choice of c_n^* when the number of sketching matrix B satisfies $B \gg (p\sqrt{n})^{q(s-1)}$. Thus, (S.40) is also satisfied.

For any $x > 0$, we have $1 - x \leq \exp(-x)$. Given (S.40), we have

$$\begin{aligned} & \Pr\left(\min_{b=1}^{B} d^{\tau}(S_b, S^*) > c_n n^{-1/2}\right) = \left\{\Pr\left(d^{\tau}(S_0, S^*) > c_n n^{-1/2}\right)\right\}^B \\ = & \left[1 - \left\{\Pr\left(d^{\tau}(S_0, S^*) \leq c_n n^{-1/2}\right)\right\}\right]^B \leq \exp\left\{-B\Pr\left(d^{\tau}(S_0, S^*) \leq c_n n^{-1/2}\right)\right\} \to 0. \end{aligned}$$

This implies that with probability tending to 1, we have

$$\Pr\left\{\min_{b=1}^{B} \mathrm{E}^X \left|\tau^{S_b}(S_b X) - \tau^{S^*}(S^* X)\right|^2 \leq c_n^2/n\right\} \to 1. \tag{S.45}$$

Notice that $c_n \to 0$, we obtain that

$$\Pr\left\{\min_{b=1}^{B} \mathrm{E}^X \left|\tau^{S_b}(S_b X) - \tau^{S^*}(S^* X)\right|^2 = o\left(n^{-1}\right)\right\} \to 1. \tag{S.46}$$

Besides, for any sketching matrix S, we have

$$\begin{aligned} & |\mathrm{VD}(d^{opt}_{S^*}) - \mathrm{VD}(d^{opt}_{S})| = |\mathrm{E}\tau(X)\{I(\tau^{S^*}(S^*X) > 0) - I(\tau^{S}(SX) > 0)\}| \\ \leq & |\mathrm{E}\tau^{S^*}(S^*X)I(\tau^{S^*}(S^*X) > 0) - \mathrm{E}\tau^{S}(SX)I(\tau^{S}(SX) > 0)| \\ \leq & |\mathrm{E}\max(\tau^{S^*}(S^*X), 0) - \mathrm{E}\max(\tau^{S}(SX), 0)| \leq \mathrm{E}|\max(\tau^{S^*}(S^*X), 0) - \max(\tau^{S}(SX), 0)| \\ \leq & \mathrm{E}|\tau^{S^*}(S^*X) - \tau^{S}(SX)| \leq \sqrt{\mathrm{E}|\tau^{S^*}(S^*X) - \tau^{S}(SX)|^2}. \end{aligned}$$

In view of (S.46), this implies $\Pr(\mathcal{A}_1^{(2)}) \to 1$ where

$$\mathcal{A}_1^{(2)} = \left\{\max_{b=1}^{B} \mathrm{VD}(d^{opt}_{S_b}) - \mathrm{VD}(d^{opt}_{S^*}) = o(n^{-1/2})\right\}. \tag{S.47}$$

For any $k = 1, \dots, \mathbb{K}$, the number of observations in $\mathbb{I}_2^{(k)-}$, i.e, $|\mathbb{I}_2^{(k)-}|$ satisfies $|\mathbb{I}_2^{(k)-}| \geq (\mathbb{K}-1)n/\mathbb{K} \geq n/2$. Thus, it follows from Condition (A4) that

$$\Pr\left\{\max_{b=1}^{B} \mathrm{E}^X \left|\hat{\tau}^{S_b}_{\mathcal{I}_2^{(k)-}}(S_b X) - \tau^{S_b}(S_b X)\right|^2 = O(n^{-r_0}\log n)\right\} \to 1,$$

for any $1 \leq k \leq \mathbb{K}$. Since $\mathbb{K}$ is fixed, it follows Bonferroni's inequality that

$$\Pr\left\{\max_{k=1}^{\mathbb{K}} \max_{b=1}^{B} \mathrm{E}^X \left|\hat{\tau}^{S_b}_{\mathcal{I}_2^{(k)-}}(S_b X) - \tau^{S_b}(S_b X)\right|^2 = O(n^{-r_0}\log n)\right\} \to 1. \tag{S.48}$$

Notice that for any sketching matrix S, we have

$$\begin{aligned}
\mathrm{VD}(d_S^{opt}) - \mathrm{VD}(\hat{d}^S_{\mathcal{I}_2^{(k)-}}) &= \mathrm{E}^X \tau^S(SX)\left\{I\left(\tau^S(SX)>0\right) - I\left(\hat{\tau}^S_{\mathcal{I}_2^{(k)-}}(SX)>0\right)\right\}\\
&\le \mathrm{E}^X \tau^S(SX) I\left(\left|\hat{\tau}^S_{\mathcal{I}_2^{(k)-}}(SX) - \tau^S(SX)\right| \ge \left|\tau^S(SX)\right|\right)\\
&\le \mathrm{E}^X \left|\hat{\tau}^S_{\mathcal{I}_2^{(k)-}}(SX) - \tau^S(SX)\right| \le \sqrt{\mathrm{E}^X \left|\hat{\tau}^S_{\mathcal{I}_2^{(k)-}}(SX) - \tau^S(SX)\right|^2},
\end{aligned}$$

where the first inequality is due to the fact that for any $a, b \in \mathbb{R}$, if $I(a>0) \neq I(b>0)$, then we have $|a-b| \ge |a|$, and the second inequality is due to Markov's inequality.

Under the event defined in (S.48), we have $\Pr(\mathcal{A}_2^{(2)}) \to 1$ where

$$\mathcal{A}_2^{(2)} = \left\{\max_{k=1}^{\mathbb{K}} \max_{b=1}^{B} \left|\mathrm{VD}(d_{S_b}^{opt}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right| = O(n^{-r_0/2}\sqrt{\log n})\right\}. \tag{S.49}$$

Let $\hat{b} = \arg\max_{b=1}^{B} \sum_{k=1}^{\mathbb{K}} \widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})/\mathbb{K}$. We have $S_{\hat{b}} = S_{\mathcal{I}_2}$. With some calculation, we can show that

$$\begin{aligned}
&\mathbb{K}\left(\mathrm{VD}(d^{opt}_{S_{\mathcal{I}_2}}) - \mathrm{VD}(d^{opt}_{S^*})\right) \qquad\qquad \text{(S.50)}\\
\ge\ & \sum_{k=1}^{\mathbb{K}}\left(\mathrm{VD}(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(d^{opt}_{S^*})\right) - \mathbb{K}\underbrace{\max_{k=1}^{\mathbb{K}}\max_{b=1}^{B}\left|\mathrm{VD}(d^{opt}_{S_b}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right|}_{I_1}\\
\ge\ & \sum_{k=1}^{\mathbb{K}}\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(d^{opt}_{S^*})\right) - \sum_{k=1}^{\mathbb{K}}\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}})\right) - \mathbb{K}I_1\\
\ge\ & \max_{b=1}^{B}\sum_{k=1}^{\mathbb{K}}\left(\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(d^{opt}_{S^*})\right) - \mathbb{K}\max_{k=1}^{\mathbb{K}}\max_{b=1}^{B}\left|\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right| - \mathbb{K}I_1\\
\ge\ & \max_{b=1}^{B}\sum_{k=1}^{\mathbb{K}}\left(\mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(d^{opt}_{S^*})\right) - 2\mathbb{K}\underbrace{\max_{k=1}^{\mathbb{K}}\max_{b=1}^{B}\left|\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}_{\mathcal{I}_2^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right|}_{I_2} - \mathbb{K}I_1\\
\ge\ & \max_{b=1}^{B}\sum_{k=1}^{\mathbb{K}}\left(\mathrm{VD}(d^{opt}_{S_b}) - \mathrm{VD}(d^{opt}_{S^*})\right) - \mathbb{K}I_1 - 2\mathbb{K}I_2 - \mathbb{K}\max_{k=1}^{\mathbb{K}}\max_{b=1}^{B}\left|\mathrm{VD}(d^{opt}_{S_b}) - \mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})\right|,
\end{aligned}$$

Since $\Pr(\mathcal{A}_0^{(2)} \cap \mathcal{A}_1^{(2)} \cap \mathcal{A}_2^{(2)}) \to 1$, with probability tending to 1, we have

$$V(d^{opt}_{S_{\mathcal{I}_2}}) - V(d^{opt}_{S^*}) = O(\sqrt{\log B}/\sqrt{n} + n^{-r_0/2}\sqrt{\log n}). \tag{S.51}$$

Under the given conditions, we have that $\mathrm{VD}(d_{S^*}^{opt}) \gg n^{-r_0/2}\sqrt{\log n}+\sqrt{\log B}/\sqrt{n}$. This together with (S.51) implies that

$$\Pr\left(\mathrm{VD}(d_{S_{\mathcal{I}_2}}^{opt}) \gg n^{-r_0/2}\sqrt{\log n}+\frac{\sqrt{\log B}}{\sqrt{n}}\right) \geq \Pr\left(\mathrm{VD}(d_{S_{\mathcal{I}_2}}^{opt}) \geq \frac{1}{2}\mathrm{VD}(d_{S^*}^{opt})\right) \to 1. \quad \text{(S.52)}$$

In addition, similar to (S.49), we can show that

$$\Pr\left\{\max_{b=1}^{B}\left|\mathrm{VD}(d_{S_b}^{opt}) - \mathrm{VD}(\hat{d}_{\mathcal{I}_2}^{S_b})\right| = O(n^{-r_0/2}\sqrt{\log n})\right\},$$

and hence

$$\Pr\left(\left|\mathrm{VD}(d_{S_{\mathcal{I}_2}}^{opt}) - \mathrm{VD}(\hat{d}_{\mathcal{I}_2}^{S_{\mathcal{I}_2}})\right| = O(n^{-r_0/2}\sqrt{\log n}+n^{-1/2}\sqrt{\log B})\right) \to 1.$$

This together with (S.52) implies (S.34).

D.4.2 Asymptotic power function

We now derive the asymptotic distribution of our proposed test in the regular cases where $\Pr(\tau(X)=0)=0$. By (7), we have

$$\mathrm{E}\tau(X)\{I(\tau(X)>0) - I(\tau^{S^*}(S^*X)>0)\},$$

where S^* is defined in (S.45). Notice that $\tau(X)\{I(\tau(X)>0) - I(\tau^{S^*}(S^*X)>0)\} \geq 0$, for any realization of X. Thus, we have $\tau(X)\{I(\tau(X)>0) - I(\tau^{S^*}(S^*X)>0)\} = 0$, almost surely. In the regular cases, this further implies

$$I(\tau(X)>0) = I(\tau^{S^*}(S^*X)>0), \quad \text{(S.53)}$$

almost surely.

Notice that $n^{-1/2}\sqrt{\log B} \to 0$, $n^{-r_0}\log n \to 0$. By Condition (A6), the following event holds with probability tending to 1,

$$\Pr(0<|\tau^{S_{\mathcal{I}_2}}(S_{\mathcal{I}_2}X)|<t) = O(t^{\gamma}), \quad \forall 0<t\leq\delta_0. \quad \text{(S.54)}$$

Since $r_0 > (2+\gamma)/(2+2\gamma)$, similar to Theorem 8 in Luedtke and van der Laan (2016), we can show that

$$\Pr(\mathcal{A}_3^{(2)}) \to 1, \tag{S.55}$$

where

$$\mathcal{A}_3^{(2)} = \left\{ \left| \mathrm{VD}(d^{opt}_{S_{\mathcal{I}_2}}) - \mathrm{VD}(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2}) \right| = o(n^{-1/2}) \right\}.$$

Similarly, we have

$$\Pr\left\{ \max_{k=1}^{\mathbb{K}} \left| \mathrm{VD}(d^{opt}_{S_{\mathcal{I}_2}}) - \mathrm{VD}(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}) \right| = o(n^{-1/2}) \right\} \to 1. \tag{S.56}$$

Combining (S.55) with (S.51), we obtain that with probability tending to 1,

$$\mathrm{E}^X \tau(X)\{I(\tau^{S^*}(S^*X) > 0) - \hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2}(X)\} \le \varepsilon_n,$$

for some sequence $\varepsilon_n \to 0$. In view of (S.53), we have

$$\mathrm{E}^X \tau(X)\{I(\tau(X) > 0) - \hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2}(X)\} \le \varepsilon_n,$$

with probability tending to 1. Since $\tau(X)\{I(\tau(X) > 0) - I(\tau^{S_{\mathcal{I}_2}}(S_{\mathcal{I}_2}X) > 0)\}$ is nonnegative for any realization of X, we obtain

$$\mathrm{E}^X |\tau(X)||d^{opt}(X) - \hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2}(X)| \le \varepsilon_n,$$

or equivalently,

$$\mathrm{E}^X |\tau(X)||d^{opt}_{S^*}(X) - \hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2}(X)| \le \varepsilon_n, \tag{S.57}$$

with probability tending to 1. Under the event defined in (S.57), we have

$$\mathrm{E}^X |\tau(X)| I(|\tau(X)| \ge \varepsilon_n^{1/(1+\gamma)}) |d^{opt}_{S^*}(X) - \hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2}(X)| \le \varepsilon_n,$$

and hence

$$\mathrm{E}^X |d^{opt}_{S^*}(X) - \hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2}(X)| I(|\tau(X)| \ge \varepsilon_n^{1/(1+\gamma)}) \le \varepsilon_n^{\gamma/(1+\gamma)}. \tag{S.58}$$

By (A5), $\tau(X)$ also satisfies the margin condition. In the regular cases, we have $\Pr(|\tau(X)| \leq \varepsilon_n^{1/(1+\gamma)}) = O(\varepsilon_n^{\gamma/(1+\gamma)})$. Therefore,

$$\mathrm{E}^X |d_{S^*}^{opt}(X) - \hat{d}_{\mathcal{I}_2}^{S_{\mathcal{I}_2}}(X)| I(|\tau(X)| \leq \varepsilon_n^{1/(1+\gamma)}) = O(\varepsilon_n^{\gamma/(1+\gamma)}).$$

This together with (S.58) yields

$$\mathrm{E}^X |d_{S^*}^{opt}(X) - \hat{d}_{\mathcal{I}_2}^{S_{\mathcal{I}_2}}(X)| = O(\varepsilon_n^{\gamma/(1+\gamma)}), \tag{S.59}$$

with probability tending to 1. To summarize, we've shown

$$\Pr\left(\mathrm{E}^X \left|\hat{d}_{\mathcal{I}_2}^{S_{\mathcal{I}_2}}(X) - d_{S^*}^{opt}(X)\right| = o(1)\right) \to 1. \tag{S.60}$$

Similarly we can show

$$\Pr\left(\mathrm{E}^X \left|\hat{d}_{\mathcal{I}_1}^{S_{\mathcal{I}_1}}(X) - d_{S^*}^{opt}(X)\right| = o(1)\right) \to 1. \tag{S.61}$$

Assume for now, we've shown

$$\Pr\left(\mathrm{VD}(\hat{d}_{\mathcal{I}_j}^{S_{\mathcal{I}_j}}) - \mathrm{VD}(d_{S^*}^{opt}) = o(n^{-1/2})\right) \to 1, \tag{S.62}$$

for $j = 1, 2$. Using similar arguments in the proof of Theorem 3.2, it follows from (S.60)-(S.62) and the conditions $\sqrt{n} h_n = \sqrt{n}\mathrm{VD}(d^{opt}) = O(1)$, $\liminf_n \sigma_0^2 > 0$ that $\widehat{T}_{SRP}$ is equivalent to

$$\max\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_1}(d_{S^*}^{opt})}{\hat{\sigma}_{\mathcal{I}_1}(d_{S^*}^{opt})}, \frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathcal{I}_2}(d_{S^*}^{opt})}{\hat{\sigma}_{\mathcal{I}_2}(d_{S^*}^{opt})}\right).$$

The limiting distribution therefore follows from the central limit theorem for the self-normalized sums.

It remains to show (S.62). By (S.55), it suffices to show that with probability tending to 1,

$$\mathrm{VD}(d_{S_{\mathcal{I}_2}}^{opt}) = \mathrm{VD}(d_{S^*}^{opt}) + o(n^{-1/2}), \tag{S.63}$$

for some $S^* \in \mathcal{S}^*$. Define b_0 to be the smallest element in $[1, \dots, B]$ such that

$$\mathrm{E}^X \left| \tau^{S_{b_0}}(S_{b_0} X) - \tau^{S^*}(S^* X) \right|^2 \le c_n^2 / n.$$

Under the event defined in the LHS of (S.45), b_0 is well defined. Similar to (S.46)-(S.47), we can show that with probability tending to 1,

$$\mathrm{VD}(d^{opt}_{S_{b_0}}) = \mathrm{VD}(d^{opt}_{S^*}) + o(n^{-1/2}). \tag{S.64}$$

In addition, similar to (S.56), we have

$$\Pr\left\{ \max_{k=1}^{\mathbb{K}} \left| \mathrm{VD}(d^{opt}_{S_{b_0}}) - \mathrm{VD}(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}) \right| = o(n^{-1/2}) \right\} \to 1.$$

This together with (S.64) yields

$$\sum_{k=1}^{\mathbb{K}} \mathrm{VD}(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}) = \mathbb{K}\mathrm{VD}(d^{opt}_{S^*}) + o(n^{-1/2}), \tag{S.65}$$

with probability tending to 1.

Under the event in (S.65), we have

$$\begin{aligned}
& \sum_{k=1}^{\mathbb{K}} \left(\mathrm{VD}(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(d^{opt}_{S^*}) \right) \qquad \text{(S.66)} \\
\ge\ & \sum_{k=1}^{\mathbb{K}} \left(\mathrm{VD}(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}) - \mathrm{VD}(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}) \right) - \mathbb{K} \underbrace{\max_{k=1}^{\mathbb{K}} \left| \mathrm{VD}(d^{opt}_{S^*}) - \mathrm{VD}(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}) \right|}_{I_3} \\
\ge\ & \max_{b=1}^{B} \sum_{k=1}^{\mathbb{K}} \left(\widehat{\mathrm{VD}}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}) - \widehat{\mathrm{VD}}(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}) \right) - \mathbb{K} I_3 - 2\mathbb{K} I_2 \\
\ge\ & -\mathbb{K} I_3 - 2\mathbb{K} I_2 = O\left(\sqrt{\log B} / \sqrt{n} \right) = O(\sqrt{\log n}/\sqrt{n}),
\end{aligned}$$

where the last equality is due to the condition that $B = O(n^{\kappa_B})$ for some $\kappa_B > 0$. Similar to (S.59), we can show that with probability tending to 1,

$$\max_{k=1}^{\mathbb{K}} \mathrm{E}^X \left| \hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}(X) - d^{opt}_{S^*}(X) \right| = O\left(\frac{\log^{\kappa_0} n}{n^{\kappa_0}} \right), \tag{S.67}$$

where $\kappa_0 = \gamma/(2 + 2\gamma)$.

Define $\phi_n = (n \log n)^{1/3}$, we have that

$$\begin{aligned}
\Pr\left\{\max_{i=1}^{n}\left|\left(\frac{A_i}{\pi(X_i)} - \frac{1-A_i}{1-\pi(X_i)}\right)Y_i\right| > \phi_n\right\} &\leq \sum_{i=1}^{n}\Pr\left\{\left|\left(\frac{A_1}{\pi(X_1)} - \frac{1-A_1}{1-\pi(X_1)}\right)Y_1\right| > \phi_n\right\} \\
&\leq n\frac{1}{\phi_n^3}\mathrm{E}\left|\left(\frac{A_1}{\pi(X_1)} - \frac{1-A_1}{1-\pi(X_1)}\right)Y_1\right|^3 = o(1),
\end{aligned}$$

where the first inequality is due to Bonferroni's inequality, the second inequality is due to Markov's inequality, the last equality is due to (A3) and the condition $\mathrm{E}|Y|^3 = O(1)$. Hence, we obtain $\Pr(\mathcal{A}_5) \to 1$, where

$$\mathcal{A}_5 = \left\{\max_{i=1}^{n}\left|\left(\frac{A_i}{\pi(X_i)} - \frac{1-A_i}{1-\pi(X_i)}\right)Y_i\right| \leq \phi_n\right\}.$$

Let $Z_i = [(1-A_i)/\{1-\pi(X_i)\} - A_i/\pi(X_i)]Y_i I(|[(1-A_i)/\{1-\pi(X_i)\} - A_i/\pi(X_i)]Y_i| \leq \phi_n)$. It follows from Bernstein's inequality that conditional on $\{S_b\}_{b=1}^B$ and $\{O_i\}_{i\in\mathcal{I}_2^{(k)-}}$, for any $t_{b,k} > 0$,

$$\begin{aligned}
&\max_{b=1}^{B}\Pr\left\{\left|\sum_{i\in\mathcal{I}_2^{(k)}}\left(Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i) - d^{opt}_{S^*}(X_i)\} - \mathrm{E}Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i) - d^{opt}_{S^*}(X_i)\}\right)\right| > t_{b,k}\right\} \\
&\qquad\qquad \leq \exp\left(-\frac{t_{b,k}^2/2}{|\mathcal{I}_2^{(k)}|\sigma_{b,k}^2 + M_{b,k}t_{b,k}/3}\right),
\end{aligned}$$

where

$$\begin{aligned}
\sigma_{b,k}^2 &= \mathrm{Var}\left(Z\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X) - d^{opt}_{S^*}(X)\} \mid S_b, \{O_i\}_{i\in\mathcal{I}_2^{(k)-}}\right), \\
M_{b,k} &= \max_i |Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i) - d^{opt}_{S^*}(X_i)\} - \mathrm{E}Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i) - d^{opt}_{S^*}(X_i)\}|.
\end{aligned}$$

Similar to (S.19), we can show

$$\begin{aligned}
&\sigma_{b,k}^2 \leq \mathrm{E}\left(Z^2\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X) - d^{opt}_{S^*}(X)\}^2 \mid S_b, \{O_i\}_{i\in\mathcal{I}_2^{(k)-}}\right) \\
\leq\;& \mathrm{E}\left\{\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right)^2 Y^2\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X) - d^{opt}_{S^*}(X)\}^2 \mid S_b, \{O_i\}_{i\in\mathcal{I}_2^{(k)-}}\right\} \\
\leq\;& \bar{c}\left(\mathrm{E}^X\left|\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X) - d^{opt}_{S^*}(X)\right|\right)^{1/3},
\end{aligned}$$

for some constant $\bar{c} > 0$.

Besides, by the definition of Z_i, we have $|M_{b,k}| \leq \max_i |Z_i| + \mathrm{E} \max_i |Z_i| \leq 2\phi_n$. Therefore, we obtain that conditional on $\{S_b\}_{b=1}^B$ and $\{O_i\}_{i \in \mathcal{I}_2^{(k)-}}$,

$$\max_{b=1}^{B} \Pr \left\{ \left| \sum_{i \in \mathcal{I}_2^{(k)}} \left(Z_i \{ \hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(S_b X_i) - d^{opt}_{S^*}(S^* X_i) \} - \mathrm{E} Z_i \{ \hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i) - d^{opt}_{S^*}(X_i) \} \right) \right| > t_{b,k} \right\}$$
$$\leq \exp \left(- \frac{t_{b,k}^2/2}{\bar{c} n (\mathbb{K} - 1) \left(\mathrm{E}^X \left| \hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X) - d^{opt}_{S^*}(X) \right| \right)^{1/3} / \mathbb{K} + 2\phi_n t_{b,k}/3} \right).$$

Let

$$t_{b,k} = 4 \max \left[\left\{ \mathrm{E}^X \left| \hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X) - d^{opt}_{S^*}(X) \right| \right\}^{1/6} \sqrt{\bar{c} n (\mathbb{K} - 1)(\log B)/\mathbb{K}}, 4\phi_n (\log B)/3 \right].$$

We have that

$$t_{b,k}^2/2 \geq 8 \log B \times \bar{c} n (\mathbb{K} - 1) \left(\mathrm{E}^X \left| \hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X) - d^{opt}_{S^*}(X) \right| \right)^{1/3} / \mathbb{K},$$
$$t_{b,k}^2/2 \geq 4 \log B \times 2\phi_n t_{b,k}/3.$$

Therefore,

$$\frac{t_{b,k}^2/2}{\bar{c} n (\mathbb{K} - 1) \mathrm{E}^X \left| \hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X) - d^{opt}_{S^*}(X) \right| / \mathbb{K} + 2\phi_n t_{b,k}/3} \geq 2 \log B.$$

Therefore, we have that conditional on $\{S_b\}_{b=1}^B$ and $\{O_i\}_{i \in \mathcal{I}_2^{(k)-}}$,

$$\max_{b=1}^{B} \Pr \left\{ \left| \sum_{i \in \mathcal{I}_2^{(k)}} \left(Z_i \{ \hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i) - d^{opt}_{S^*}(X_i) \} - \mathrm{E} Z_i \{ \hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i) - d^{opt}_{S^*}(X_i) \} \right) \right| > t_{b,k} \right\}$$
$$\leq \exp(-2 \log B) = \frac{1}{B^2}.$$

It follows from Bonferroni's inequality that conditional on $\{S_b\}_{b=1}^B$ and $\{O_i\}_{i\in\mathcal{I}_2^{(k)-}}$,

$$\Pr\left[\bigcup_{b=1}^{B}\left\{\left|\sum_{i\in\mathcal{I}_2^{(k)}}\left(Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i)-d^{opt}_{S^*}(X_i)\}-\mathrm{E}Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i)-d^{opt}_{S^*}(X_i)\}\right)\right|>t_{b,k}\right\}\right]$$
$$\leq B\max_{b=1}^{B}\Pr\left\{\left|\sum_{i\in\mathcal{I}_2^{(k)}}\left(Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i)-d^{opt}_{S^*}(X_i)\}-\mathrm{E}Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i)-d^{opt}_{S^*}(X_i)\}\right)\right|>t_{b,k}\right\}$$
$$\leq\frac{1}{B}\to 0.$$

Since the probability bound $1/B$ is independent of $\{S_b\}_{b=1}^B$ and $\{O_i\}_{i\in\mathcal{I}_2^{(k)-}}$. The above inequality also holds marginally. Note that $\mathbb{K}$ is fixed. By Bonferroni's inequality, we obtain that $\Pr(\mathcal{A}_6^{(2)})\to 1$, where

$$\mathcal{A}_6^{(2)}=\bigcap_{b=1}^{B}\bigcap_{k=1}^{\mathbb{K}}\left\{\left|\sum_{i\in\mathcal{I}_2^{(k)}}\left(Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i)-d^{opt}_{S^*}(X_i)\}-\mathrm{E}Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i)-d^{opt}_{S^*}(X_i)\}\right)\right|\leq t_{b,k}\right\}.$$

It follows from Markov's inequality and the definition of ϕ_n that

$$\begin{aligned}
&\left|\mathrm{E}Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i)-d^{opt}_{S^*}(X_i)\}-\mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})+\mathrm{VD}(d^{opt}_{S^*})\right|\\
\leq\;&\mathrm{E}\left|\left(\frac{A_i}{\pi(X_i)}-\frac{1-A_i}{1-\pi(X_i)}\right)Y_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i)-d^{opt}_{S^*}(X_i)\}I\left\{\left|\left(\frac{A_i}{\pi(X_i)}-\frac{1-A_i}{1-\pi(X_i)}\right)Y_i\right|>\phi_n\right\}\right|\\
\leq\;&\mathrm{E}\left|\left(\frac{A_i}{\pi(X_i)}-\frac{1-A_i}{1-\pi(X_i)}\right)Y_i\right|I\left\{\left|\left(\frac{A_i}{\pi(X_i)}-\frac{1-A_i}{1-\pi(X_i)}\right)Y_i\right|>\phi_n\right\}\\
\leq\;&\frac{1}{\phi_n^2}\mathrm{E}\left|\left(\frac{A_i}{\pi(X_i)}-\frac{1-A_i}{1-\pi(X_i)}\right)Y_i\right|^3=o(n^{-1/2}).
\end{aligned}$$

Since $\Pr(\mathcal{A}_6^{(2)})\to 1$, we obtain $\Pr(\mathcal{A}_7^{(2)})\to 1$, where

$$\mathcal{A}_7^{(2)}=\bigcap_{b=1}^{B}\bigcap_{k=1}^{\mathbb{K}}\left\{\left|\sum_{i\in\mathcal{I}_2^{(k)}}\left(Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i)-d^{opt}_{S^*}(X_i)\}-\mathrm{VD}(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}})+\mathrm{VD}(d^{opt}_{S^*})\right)\right|\leq t_{b,k}+o(\sqrt{n})\right\}.$$

Under the event defined in $\mathcal{A}_5$, we have that $Z_i=[(1-A_i)/\{1-\pi(X_i)\}-A_i/\pi(X_i)]Y_iI(|[(1-$

$A_i)/\{1-\pi(X_i)\} - A_i/\pi(X_i)]Y_i$ and hence

$$\sum_{i\in\mathcal{I}_2^{(k)}} Z_i\{\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}(X_i) - d^{opt}_{S^*}(X_i)\} = \frac{n}{\mathbb{K}}\left\{\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}\right) - \widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(d^{opt}_{S^*}\right)\right\}.$$

Therefore, with probability tending to 1, we have that for any $b = 1,\dots,B$ and $k = 1,\dots,\mathbb{K}$,

$$\left|\frac{n}{\mathbb{K}}\left\{\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}\right) - \widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(d^{opt}_{S^*}\right) - \mathrm{VD}\left(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}\right) + \mathrm{VD}\left(d^{opt}_{S^*}\right)\right\}\right| \le t_{b,k} + o(1)\sqrt{n},$$

and hence,

$$\left|n\sum_{k=1}^{\mathbb{K}}\left\{\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}\right) - \widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(d^{opt}_{S^*}\right) - \mathrm{VD}\left(\hat{d}^{S_b}_{\mathcal{I}_2^{(k)-}}\right) + \mathrm{VD}\left(d^{opt}_{S^*}\right)\right\}\right| \le \mathbb{K}\sum_k t_{b,k} + o(1)\sqrt{n},$$

for $b = 1,\dots,B$. Define the above event to be $\mathcal{A}_8^{(2)}$. When it holds, we have

$$\left|n\sum_{k=1}^{\mathbb{K}}\left\{\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}\right) - \widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(d^{opt}_{S^*}\right) - \mathrm{VD}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}\right) + \mathrm{VD}\left(d^{opt}_{S^*}\right)\right\}\right| \le \mathbb{K}\sum_k t_{\hat{b},k} + o(\sqrt{n}).$$

It follows from (S.67) and the condition $B = O(n^{\kappa_B})$ that we have with probability tending to 1,

$$\max_k |t_{\hat{b},k}|/\sqrt{n} = O\left(n^{-\kappa_0/6}\log^{\kappa_0/6+1/2} n\right) + O\left(\frac{\phi_n \log n}{\sqrt{n}}\right).$$

Notice that the above expression is $o(1)$. Therefore, we obtain that with probability tending to 1,

$$\left|n\sum_{k=1}^{\mathbb{K}}\left\{\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}\right) - \widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(d^{opt}_{S^*}\right) - \mathrm{VD}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}\right) + \mathrm{VD}\left(d^{opt}_{S^*}\right)\right\}\right| = o(\sqrt{n}). \quad \text{(S.68)}$$

Recall that b_0 is the smallest element in $[1,\dots,B]$ such that

$$\mathrm{E}^X\left|\tau^{S_{b_0}}(S_{b_0}X) - \tau^{S^*}(S^*X)\right|^2 \le c_n^2/n.$$

Similar to (S.68), we can show that with probability tending to 1,

$$n\sum_{k=1}^{\mathbb{K}}\left\{\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}\right)-\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(d^{opt}_{S^*}\right)-\mathrm{VD}\left(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}\right)+\mathrm{VD}\left(d^{opt}_{S^*}\right)\right\}=o(\sqrt{n}).$$

In view of (S.68), we have with probability tending to 1 that

$$n\sum_{k=1}^{\mathbb{K}}\left\{\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}\right)-\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}\right)-\mathrm{VD}\left(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}\right)+\mathrm{VD}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}\right)\right\}=o(\sqrt{n}).$$

By definition, we have

$$n\sum_{k=1}^{\mathbb{K}}\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}\right)\le n\sum_{k=1}^{\mathbb{K}}\widehat{\mathrm{VD}}_{\mathcal{I}_2^{(k)}}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}\right).$$

Therefore, we obtain that with probability tending to 1,

$$n\sum_{k=1}^{\mathbb{K}}\mathrm{VD}\left(\hat{d}^{S_{b_0}}_{\mathcal{I}_2^{(k)-}}\right)\le n\sum_{k=1}^{\mathbb{K}}\mathrm{VD}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}\right)+o(\sqrt{n}),$$

which together with (S.65) implies that with probability tending to 1, we have

$$n\mathbb{K}\mathrm{VD}\left(d^{opt}_{S^*}\right)\le n\sum_{k=1}^{\mathbb{K}}\mathrm{VD}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2^{(k)-}}\right)+o(\sqrt{n}). \tag{S.69}$$

Under the events defined in (S.56) and (S.69), we have that,

$$\mathrm{VD}\left(d^{opt}_{S^*}\right)\le\mathrm{VD}\left(d^{opt}_{S_{\mathcal{I}_2}}\right)+o(\sqrt{n}).$$

Since $\mathrm{VD}\left(d^{opt}_{S^*}\right)\ge\mathrm{VD}\left(d^{opt}_{S_{\mathcal{I}_2}}\right)$, this implies that with probability tending to 1,

$$\mathrm{VD}\left(d^{opt}_{S^*}\right)=\mathrm{VD}\left(d^{opt}_{S_{\mathcal{I}_2}}\right)+o(\sqrt{n}).$$

Now, it follows from (S.55) that

$$\Pr\left\{\mathrm{VD}\left(d^{opt}_{S^*}\right)=\mathrm{VD}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2}\right)+o(\sqrt{n})\right\}\to 1,$$

and hence

$$\Pr\left\{\mathrm{VD}\left(d^{opt}_{S^*}\right)=\mathrm{VD}\left(\hat{d}^{S_{\mathcal{I}_2}}_{\mathcal{I}_2}\right)+o(\sqrt{n})\right\}\to 1,$$

Similarly, we can show that

$$\Pr\left\{\mathrm{VD}\left(d^{opt}_{S^*}\right) = \mathrm{VD}\left(\hat{d}^{S_{\mathcal{I}_1}}_{\mathcal{I}_1}\right) + o(\sqrt{n})\right\} \to 1.$$

This proves (S.62). The proof is thus completed.

E Additional simulation results

E.1 Dependent covariates

Table S.1: Rejection probabilities (%) of the sparse random projection-based test, dense random projection-based test, penalized least square-based test, step-wise selection-based test and the supremum-type test based on the desparsified Lasso estimator, with standard errors in parenthesis (%), under Scenarios 1 and 2 where $X \sim N(0, \{0.5^{|i-j|}\}_{i,j=1,\dots,p})$.

Scenario 1		VD = 0		VD = 30%		VD = 50%		VD = 70%	
		α level		α level		α level		α level	
	p	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
$\widehat{T}^{dr}_{SRP}$	50	0.2(0.2)	0.2(0.2)	9.4(1.3)	18(1.7)	39.4(2.2)	50.8(2.2)	63.4(2.2)	73.2(2)
	100	0(0)	0(0)	7(1.1)	16.2(1.6)	37.2(2.2)	51.2(2.2)	64(2.1)	74.2(2)
$\widehat{T}^{dr}_{RP}$	50	0(0)	0(0)	0(0)	0.2(0.2)	0(0)	0.4(0.3)	0.2(0.2)	1.4(0.5)
	100	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0.8(0.4)
$\widehat{T}_{PLS}$	50	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
	100	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
$\widehat{T}_{VS}$	50	0(0)	0(0)	0.4(0.3)	1(0.4)	2.4(0.7)	5.4(1)	5.6(1)	12.2(1.5)
	100	0(0)	0(0)	0.2(0.2)	0.6(0.3)	1.2(0.5)	3.2(0.8)	4.2(0.9)	9.4(1.3)
$\widehat{T}_{DL}$	50	4.8(1)	14.6(1.6)	4.8(1)	14.2(1.6)	5.4(1)	14.6(1.6)	5.6(1)	15.2(1.6)
	100	5.2(1)	20.6(1.8)	6.6(1.1)	20.4(1.8)	7(1.1)	21(1.8)	7.6(1.2)	20.2(1.8)
Scenario 2		VD = 0		VD = 30%		VD = 50%		VD = 70%	
		α level		α level		α level		α level	
	p	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
$\widehat{T}^{dr}_{SRP}$	50	1(0.4)	5.4(1)	32.8(2.1)	46(2.2)	61(2.2)	71.4(2)	67(2.1)	76.8(1.9)
	100	0.8(0.4)	5.8(1)	26.2(2)	37(2.2)	54.8(2.2)	64.2(2.1)	61.6(2.2)	71.2(2)
$\widehat{T}^{dr}_{RP}$	50	1(0.4)	5.6(1)	1(0.4)	5.6(1)	1.4(0.5)	7(1.1)	1.8(0.6)	6(1.1)
	100	0.8(0.4)	5.6(1)	0.8(0.4)	6.6(1.1)	0.8(0.4)	6.4(1.1)	0.8(0.4)	5(1)
$\widehat{T}_{PLS}$	50	1(0.4)	4.8(1)	2(0.6)	6(1.1)	1.6(0.6)	7(1.1)	1.6(0.6)	7.4(1.2)
	100	1.6(0.6)	6.8(1.1)	1.8(0.6)	7.2(1.2)	2.6(0.7)	6.6(1.1)	2.4(0.7)	7(1.1)
$\widehat{T}_{VS}$	50	1.8(0.6)	6.8(1.1)	1(0.4)	7.2(1.2)	0.8(0.4)	6.6(1.1)	1.2(0.5)	6.2(1.1)
	100	0.8(0.4)	5(1)	1.4(0.5)	5.6(1)	1.0(0.4)	4.2(0.9)	0.2(0.2)	4.6(0.9)
$\widehat{T}_{DL}$	50	0.8(0.4)	4.8(1)	4.2(0.9)	12.8(1.5)	5.2(1)	14.8(1.6)	5.4(1)	14.4(1.6)
	100	0.8(0.4)	4.4(0.9)	4.2(0.9)	15.2(1.6)	5.8(1)	17.8(1.7)	7.2(1.2)	18.2(1.7)

In this section, we examine the performance of the proposed test under the setting where covariates are dependent. Reported in Table S.1 and S.2 are the rejection probabilities of $\widehat{T}^{dr}_{SRP}$, $\widehat{T}^{dr}_{RP}$, $\widehat{T}_{PLS}$, $\widehat{T}_{VS}$ and $\widehat{T}_{DL}$ where $X \sim N(0, \{0.5^{|i-j|}\}_{i,j=1,\dots,p})$. Similar to the results reported in Table 1 and Table S.2, our proposed test in much more powerful than other competing tests under Scenarios 1 and 2 where the OITR is sparse and nonlinear. In Scenario 3 where the OITR is sparse and linear, the rejection probabilities of our test are slightly smaller than when compared to $\widehat{T}_{PLS}$, $\widehat{T}_{DL}$ and $\widehat{T}_{VS}$, but are much larger than those of $\widehat{T}^{dr}_{RP}$. In the last scenario, the rejection probabilities of $\widehat{T}_{DL}$ are much smaller than other test statistics. Our proposed test achieves the greatest power in nearly all settings.

Table S.2: Rejection probabilities (%) of the sparse random projection-based test, dense random projection-based test, penalized least square-based test, step-wise selection-based test and the supremum-type test based on the desparsified Lasso estimator, with standard errors in parenthesis (%), under Scenarios 3 and 4 where $X \sim N(0, \{0.5^{|i-j|}\}_{i,j=1,\dots,p})$.

Scenario 3		VD = 30%		VD = 50%		VD = 70%	
		α level		α level		α level	
	p	0.01	0.05	0.01	0.05	0.01	0.05
$\widehat{T}_{SRP}^{dr}$	50	79.8(1.8)	94(1.1)	95.4(0.9)	99.4(0.3)	98.2(0.6)	99.8(0.2)
	100	74(2)	92(1.2)	94.4(1)	98.8(0.5)	97.6(0.7)	99.6(0.3)
$\widehat{T}_{RP}^{dr}$	50	31.8(2.1)	60(2.2)	48.8(2.2)	77.2(1.9)	55.2(2.2)	84.2(1.6)
	100	12.8(1.5)	34.6(2.1)	27.8(2)	54.6(2.2)	29.8(2)	63.8(2.1)
$\widehat{T}_{PLS}$	50	92(1.2)	98.2(0.6)	97.8(0.7)	100(0)	99(0.4)	100(0)
	100	92.4(1.2)	98.8(0.5)	98(0.6)	100(0)	99(0.4)	100(0)
$\widehat{T}_{VS}$	50	86.2(1.5)	97.6(0.7)	98.2(0.6)	99.8(0.2)	99.2(0.4)	100(0)
	100	81(1.8)	96.2(0.9)	96(0.9)	99.8(0.2)	98.4(0.6)	99.8(0.2)
$\widehat{T}_{DL}$	50	97.4(0.7)	100(0)	99.8(0.2)	100(0)	100(0)	100(0)
	100	98.4(0.6)	100(0)	99.6(0.3)	100(0)	99.6(0.3)	100(0)
Scenario 4		**VD = 30%**		**VD = 50%**		**VD = 70%**	
		α level		α level		α level	
	p	0.01	0.05	0.01	0.05	0.01	0.05
$\widehat{T}_{SRP}^{dr}$	50	66.2(2.1)	89.4(1.4)	87.8(1.5)	98.4(0.6)	94(1.1)	99.2(0.4)
	100	59.6(2.2)	85.8(1.6)	87.4(1.5)	98.4(0.6)	94.6(1)	99.6(0.3)
$\widehat{T}_{RP}^{dr}$	50	44.8(2.2)	75.8(1.9)	60.4(2.2)	84.4(1.6)	68.4(2.1)	90(1.3)
	100	42.4(2.2)	74.4(2)	57.6(2.2)	82.6(1.7)	62.4(2.2)	87.6(1.5)
$\widehat{T}_{PLS}^{dr}$	50	54.6(2.2)	80.2(1.8)	67.4(2.1)	86.4(1.5)	72.4(2)	92(1.2)
	100	71.6(2)	89.8(1.4)	64(2.1)	88.4(1.4)	71.6(2)	89.8(1.4)
$\widehat{T}_{VS}^{dr}$	50	61(2.2)	84(1.6)	73.4(2)	93.4(1.1)	76.2(1.9)	92.2(1.2)
	100	50.6(2.2)	80.4(1.8)	65.2(2.1)	89.6(1.4)	66.8(2.1)	89.2(1.4)
$\widehat{T}_{DL}^{dr}$	50	3(0.8)	11.6(1.4)	4.2(0.9)	14(1.6)	4.6(0.9)	12(1.5)
	100	3.6(0.8)	11.8(1.4)	4(0.9)	12.8(1.5)	3.6(0.8)	13.8(1.5)

E.2 Choice of q

In this section, we examine the finite sample performance of the proposed test with different choices of the projected dimension q. We use the same model described in Section 4.1, i.e,

$$Y = 1 + (X^{(1)} - X^{(2)})/2 + A\tau(X) + e,$$

where $X \sim N(0, I_p)$, $A \sim \text{Binom}(1, 0.5)$ and $e \sim N(0, 0.5^2)$. We set $q = 50$ and $N = 600$.

Consider the following three scenarios. In the first scenario, we set

$$\tau(X) = \delta \left\{ \left(\frac{X^{(1)} + X^{(2)}}{\sqrt{2}} \right)^2 - \left(\frac{X^{(3)} + X^{(4)}}{\sqrt{2}} \right)^2 \right\} \left(\frac{\sum_{j=5}^{9} X^{(j)}}{\sqrt{5}} \right)^2,$$

for some $\delta \geq 0$. In the last two scenarios, we set

$$\tau(X) = \delta \left(\frac{\sum_{j=1}^{m} X^{(j)}}{\sqrt{m}} \right) \left(\frac{\sum_{j=m+1}^{m+5} X^{(j)}}{\sqrt{5}} \right)^2,$$

for some $\delta \geq 0$ and some positive integer m. More specifically, we set $m = 15$ in Scenario 2 and $m = 30$ in Scenario 3. Notice that the total number of variables involved in the OITR under Scenarios 1, 2 and 3 equals 4, 15 and 30, respectively.

As in Section 4.1, the parameter δ controls the degree of overall qualitative treatment effects. When $\delta = 0$, the null hypothesis holds. Otherwise, the alternative holds. In each scenario, we consider four cases by setting $\mathrm{VD}(d^{opt}) = 0, 0.3, 0.5$ and 0.7. In all three scenarios, the settings for $\mathrm{VD}(d^{opt}) = 0$ are the same. Hence, in Scenarios 2 and 3, we only report the simulation results for $\mathrm{VD}(d^{opt}) = 0.3, 0.5$ and 0.7.

We further consider four settings, corresponding to four choices of q. In the first three settings, we set $q = 1, 2$ and 3, respectively. In the last setting, we adaptively choose q when sampling $S_{\mathcal{I}_j}$. Specifically, for $b = 1, \ldots, B$, we first sample q uniformly from the set $\{1, \ldots, Q\}$, then sample s according as the random variable $s_0 = 2 + \mathrm{Binom}(p-2, 2/(p-2))$, and finally sample S_b according to Step 3 of Algorithm 2. We then output the sparse sketching matrix that maximizes the estimated value difference function. For completeness, we summarize the whole procedure for generating sparse sketching matrix in Algorithm 3. Here, in Step 2(i) of Algorithm 3, we set $Q = 3$. In all four settings, we set $B = 5 \times 10^5$ and estimate the projected contrast function as described in Section 3.3.2 and Section 4.1.

Results are reported in Table S.3 and S.4. It can be seen that the type-I error rates of all the four test statistics are well controlled. The powers of our tests vary across different choices of q. In Scenario 1 where the OITR involves 4 variables, the test with $q = 1$ is much powerful than those with $q = 2$ and 3. In Scenarios 2 and 3 where more than 15 variables are involved in the OITR, the rejection probabilities of the tests with $q = 2$ and 3 are much larger than those with $q = 1$.

In addition, the proposed adaptive method performs no worse than any fixed choice of q. In Scenarios 2 and 3, the test with adaptively chosen q achieves the greatest power in nearly all cases. In Scenario 1, when $\mathrm{VD} = 0.3$ and 0.5, the adaptive test has comparable performance with the test with $q = 2$. When VD=0.7, it achieves greater power when compared to the tests with $q = 2$ and 3.

Algorithm 3. Generate data-dependent sparse random sketching matrix with adaptively chosen q.

1. Input observations $\{O_i\}_{i \in \mathcal{I}}$, integers B, and $\mathbb{K} \geq 2$.
2. Generate i.i.d matrices $S_1, S_2, \dots, S_B$ according as S_0 whose distribution is described as follows.
   (i) Uniformly sample q from the set $\{1, \dots, Q\}$.
   (ii) Sample s according as the random variable $2 + \text{Binom}(p-2, 2/(p-2))$.
   (iii) For $j = 1, \dots, q$,
      (iii.1) Independently select a simple random sample $\mathcal{M}_j$ of size s from $\{1, \dots, p\}$;
      (iii.2) Independently generate a Gaussian random vector $g_j \sim N(0, I_s)$;
      (iii.3) Set $S^{(j)}_{0,\mathcal{M}_j^c} = 0$ and $S^{(j)}_{0,\mathcal{M}_j} = g_j / \|g_j\|_2$.
3. Randomly divide $\mathcal{I}$ into $\mathbb{K}$ subsets $\{\mathcal{I}^{(k)}\}_{k=1}^{\mathbb{K}}$ of equal sizes. Let $\mathcal{I}^{(k)-} = \mathcal{I} \cap (\mathcal{I}^{(k)})^c$.
4. For $b = 1, \dots, B$,
   (i) For $k = 1, \dots, \mathbb{K}$,
      (i.1) Obtain the estimator $\hat{\tau}^{S_b}_{\mathcal{I}^{(k)-}}$ and $\hat{d}^{S_b}_{\mathcal{I}^{(k)-}}(x) = I\{\hat{\tau}^{S_b}_{\mathcal{I}^{(k)-}}(S_b x) > 0\}$;
      (i.2) Evaluate the value difference $\widehat{\text{VD}}_{\mathcal{I}^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}^{(k)-}})$.
   (ii) Obtain the cross-validated estimator $\widehat{\text{VD}}^{S_b}_{CV} = \sum_k \widehat{\text{VD}}_{\mathcal{I}^{(k)}}(\hat{d}^{S_b}_{\mathcal{I}^{(k)-}})/\mathbb{K}$.
5. Output $S_{\hat{b}}$, where $\hat{b} = \arg\max_{b=1}^{B} \widehat{\text{VD}}^{S_b}_{CV}$.

Table S.3: Rejection probabilities (%) of the sparse random projection-based test with different choices of q, with standard errors in parenthesis, under Scenario 1.

	VD = 0		VD = 30%		VD = 50%		VD = 70%	
α level	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
$q = 1$	0.4(0.3)	4.8(1)	16.4(1.7)	33.8(2.1)	31.6(2.1)	57(2.2)	45.6(2.2)	72.2(2)
$q = 2$	0.8(0.4)	2.8(0.7)	9(1.3)	23(1.9)	22(1.9)	44.8(2.2)	29.6(2)	57.6(2.2)
$q = 3$	0.6(0.3)	5(1)	7.8(1.2)	19.2(1.8)	15.2(1.6)	36.2(2.1)	24.4(1.9)	47.4(2.2)
adaptive q	0.2(0.2)	3.6(0.8)	8.8(1.3)	21.6(1.8)	23.8(1.9)	43(2.2)	35.4(2.1)	62.2(2.2)

Table S.4: Rejection probabilities (%) of the sparse random projection-based test with different choices of q, with standard errors in parenthesis, under Scenarios 2 and 3.

	VD = 30%		VD = 50%		VD = 70%	
α level	0.01	0.05	0.01	0.05	0.01	0.05
			Scenario 2			
$q = 1$	11.2(1.4)	25.8(2)	19.4(1.8)	45.6(2.2)	26(2)	53.4(2.2)
$q = 2$	15.6(1.6)	29.4(2)	23.8(1.9)	51.8(2.2)	30.6(2.1)	59(2.2)
$q = 3$	14.6(1.6)	34.2(2.1)	21.8(1.8)	47.2(2.2)	30.4(2.1)	56.4(2.2)
adaptive q	14.6(1.6)	35.6(2.1)	20.6(1.8)	50.6(2.2)	31(2.1)	60.2(2.2)
			Scenario 3			
α level	0.01	0.05	0.01	0.05	0.01	0.05
$q = 1$	5.4(1)	16.8(1.7)	9.4(1.3)	26.2(2)	9.6(1.3)	27.6(2)
$q = 2$	6(1.1)	25.8(2)	12.6(1.5)	33.6(2.1)	15.2(1.6)	36.2(2.1)
$q = 3$	8.2(1.2)	21.4(1.8)	12.2(1.5)	37.8(2.2)	15.4(1.6)	40.2(2.2)
adaptive q	8.2(1.2)	26.6(2)	13.2(1.5)	33.8(2.1)	17(1.7)	36.6(2.2)

References

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Testing many moment inequalities. *arXiv preprint arXiv:1312.7614*.

de Boor, C. (1973). The quasi-interpolant as a tool in elementary polynomial spline theory.

Fang, K. T., S. Kotz, and K. W. Ng (1990). *Symmetric multivariate and related distributions*, Volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London.

Luedtke, A. R. and M. J. van der Laan (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist. 44*(2), 713–742.

Peña, V. H., T. L. Lai, and Q.-M. Shao (2008). *Self-normalized processes: Limit theory and Statistical Applications*. Springer Science and Business Media.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Zhang, Y., J. Duchi, and M. Wainwright (2013). Divide and conquer kernel ridge regression. In *Conference on Learning Theory*, pp. 592–617.

Zhou, S. (2009). Restricted eigenvalue conditions on subgaussian random matrices. *arXiv: 0912.4045*.

Zhou, S., X. Shen, and D. A. Wolfe (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist. 26*(5), 1760–1782.