# Supplementary Material for 'Matching Using Sufficient Dimension Reduction for Causal Inference'

## A Order determination

As mentioned in the main text, we use the ladle estimator proposed in Luo and Li (2016) to determine the dimensions of the central subspaces. As an illustration, we consider the central subspace $\mathcal{S}_{Y(0)|X}$ in the control group, and denote its dimension by $d$ for simplicity of notation.

Under the linearity condition (7) in the main text, the column space of matrix-valued parameter $M_0$ defined in Step 1 of the implementation procedure of the proposed method is identical to $\mathcal{S}_{Y(0)|X}$, so $d$ can be equivalently treated as the rank of $M_0$. The ladle estimator uses the information contained in both the eigenvectors and eigenvalues of $\hat{M}_0$ about the rank of $M_0$. It constructs an objective function, which we denote by $g(\cdot)$, of the candidate rank $k$ ranged from 0 to $p-1$. The function is a summation of two parts, $f_n(k)$, the bootstrap variation of the linear space spanned by the first $k$ eigenvectors of $\hat{M}_0$, and $\phi_n(k) = \hat{\lambda}_{k+1}/(1+\sum_{i=1}^{p}\hat{\lambda}_i)$, where $\hat{\lambda}_i$ denotes the $i$th largest eigenvalue of $\hat{M}_0$.

For a working dimension $k$, if $k > d$, then the space spanned by the first $k$ eigenvectors of $\hat{M}_0$ always contains some direction that falls in the null space of $M_0$ at the population level. Under some general regularity conditions, this direction will be arbitrarily selected from the null space of $M_0$ asymptotically, leading to a large variation of the space spanned by the first $k$ eigenvectors of $\hat{M}_0$. Using bootstrap resampling and applying $\hat{M}_0$ on each bootstrap sample, $f_n(\cdot)$, the bootstrap variation of the $k$-dimensional spaces, mimics the variation in the full-sample level and has a large value. When $k = d$, the space spanned by the first $k$ eigenvectors of $\hat{M}_0$ consistently estimates $\mathcal{S}_{Y(0)|X}$, so it has a small variation. Thus $f_n(d)$ is small.

On the other hand, $\hat{\lambda}_k$ conveys the scree plot, which means $\phi_n(k)$ is large if $k < d$ and small otherwise. Consequently, $g(k) = f_n(k) + \phi_n(k)$ is large when $k < d$ due to the large $\phi_n(k)$, small when $k = d$, and large again when $k > d$ due to the large $f_n(k)$. Thus, the ladle estimator, defined as the minimizer of $g(\cdot)$, consistently estimates $d$. More details can be found in Luo and Li (2016).

# B Proof of theorems

## B.1 Proof of Theorem 1

Without loss of generality, we denote $f(y(t)|X, T = t)$ as the density function of the conditional distribution $Y(t)|(X, T = t)$ with respect to a $\sigma$-finite measure $\nu$, for $t = 0, 1$. For any random element $R$, denote $\Omega(R|T = t)$ and $\Omega(R)$ as the support of $R|T = t$ and $R$, respectively. For any $a \in \Omega(X'\beta_t)$, if $a \notin \Omega(X'\beta_t|T = t)$, then $X'\beta_t = a$ implies that $T = 1 - t$ almost surely, which automatically implies that

$$Y(t) \perp\!\!\!\perp T \mid X'\beta_t = a. \tag{1}$$

Thus we only need to show (1) for any $a \in \Omega(X'\beta_t|T = t)$. That is, $f(y(t)|X'\beta_t = a, T = t) = f(y(t)|X'\beta_t = a)$. By the definition of $\beta_t$, we have

$$Y(t) \perp\!\!\!\perp X \mid X'\beta_t,$$

which means that $f(y(t)|X) = f(y(t)|X'\beta_t)$. For any $x \in \Omega(X|T = t)$ such that $X'\beta_t = a$, $f(y(t)|X = x, T = t) = f(y(t)|X = x) = f(y(t)|X'\beta_t = a)$, where the first equality is due to the ignorability assumption. Thus we have

$$\begin{aligned} f(y(t)|X'\beta_t = a, T = t) &= E\{f(y(t)|X = x, T = t)|X'\beta_t = a, T = t\} \\ &= E\{f(y(t)|X'\beta_t = a)|X'\beta_t = a, T = t\} \\ &= f(y(t)|X'\beta_t = a). \end{aligned}$$

This completes the proof.

## B.2 Proof of Theorem 2

For simplicity, we only prove the theorem for $t = 0$. The case for $t = 1$ can be shown in the same manner. First, we show that $\mathcal{S}^D_{Y(0)|X} \subseteq \mathcal{S}_{Y(0)|X}$, which is equivalent to that

$$Y(0) \perp\!\!\!\perp X \mid (X'\beta_0, T = 0) \tag{2}$$

Following the notations in the proof of Theorem 1, we denote $f(\cdot|X)$ as the conditional density function of $Y(0)|X$ and $f(\cdot|X, T = 0)$ as the conditional density function of $Y(0)|X$ when $T = 0$. (2) is equivalent to that $f(\cdot|X = x, T = 0) = f(\cdot|X'\beta_0 = x'\beta_0, T = 0)$ for any $x \in \Omega(X|T = 0)$. By the ignorability assumption, $f(\cdot|X = x, T = 0) = f(\cdot|X = x)$, and by the definition of $\mathcal{S}_{Y(0)|X}$, $f(\cdot|X = x) = f(\cdot|X'\beta_0 = x'\beta_0)$. Thus $f(\cdot|X = x, T = 0)$ is measurable with respect to $x'\beta_0$, which means that

$$\begin{aligned} f(\cdot|X = x, T = 0) &= E\{f(\cdot|X, T = 0)|X'\beta_0 = x'\beta_0, T = 0\} \\ &= f(\cdot|X'\beta_0 = x'\beta_0, T = 0). \end{aligned}$$

Hence (2) holds. Conversely, to show that $\mathcal{S}_{Y(0)|X} \subseteq \mathcal{S}^D_{Y(0)|X}$, note that it is equivalent to

$$Y(0) \perp\!\!\!\perp X \mid X'\beta^D_0, \tag{3}$$

which holds if $f(\cdot|X = x) = f(\cdot|X'\beta^D_0 = x'\beta^D_0)$ for any $x \in \Omega(X)$. Because $\Omega(X'\beta_0|T = 0) = \Omega(X'\beta_0)$, there exists $x^* \in \Omega(X|T = 0)$ such that $x'\beta_0 = x^{*\prime}\beta_0$. By the definition of $\mathcal{S}_{Y(0)|X}$, we have

$$f(\cdot|X = x) = f(\cdot|X'\beta_0 = x'\beta_0) = f(\cdot|X'\beta_0 = x^{*\prime}\beta_0) = f(\cdot|X = x^*).$$

Since $x^* \in \Omega(X|T = 0)$, by the ignorability assumption, $f(\cdot|X^* = x^*) = f(\cdot|X^* = x^*, T = 0)$, which, by the definition of $\mathcal{S}^D_{Y(0)|X}$, further implies that $f(\cdot|X^* = x^*) = f(\cdot|X^{*\prime}\beta^D_0 = x^{*\prime}\beta^D_0, T = 0)$. Thus $f(\cdot|X^* = x^*)$ is measurable with respect to $x^{*\prime}\beta^D_0$, which, similar to the above, implies that $f(\cdot|X^* = x^*) = f(\cdot|X^{*\prime}\beta^D_0 = x^{*\prime}\beta^D_0)$. Because $\mathcal{S}^D_{Y(0)|X} \subseteq \mathcal{S}_{Y(0)|X}$, we have $x'\beta^D_0 = x^{*\prime}\beta^D_0$. Hence $f(\cdot|X = x) = f(\cdot|X'\beta^D_0 = x'\beta^D_0)$, which implies (3). This completes the proof.

## B.3 Proof of Theorem 3

We first show a mathematical property that will be used in the proof. Let $r, s > 0$ be arbitrary positive real numbers. We show that, as $n \to \infty$,

$$n\left[\{1 - (r + s)n^{-1}\}^n - (1 - rn^{-1})^n(1 - sn^{-1})^n\right] \to -rs\exp\{-(r + s)\}, \tag{4}$$

which is a complement of the well known result $\lim_{n\to\infty}(1-rn^{-1})^n = e^{-r}$. To see why it holds, re-write the left-hand side as

$$n[\{1-(r+s)n^{-1}\}^n - \{1-(r+s)n^{-1}+rsn^{-2}\}^n],$$

and divide it by $\{1-(r+s)n^{-1}\}^n$, which converges to $\exp\{-(r+s)\}$. Then (4) is equivalent to

$$n\left[1-[1+rsn^{-2}\{1-(r+s)n^{-1}\}^{-1}]^n\right] \to -rs.$$

Denote the left-hand side above by $\epsilon_n$. By simple algebra,

$$\lim_{n\to\infty}(1-\epsilon_n/n)^n = e^{rs},$$

which means that $\epsilon_n \to -(rs)$, and (4) holds. For simplicity, we prove the theorem for $\hat{\mu}_X$ and $m=1$, and denote $J_1(i)$ by $J(i)$ for each subject $i$. The general case can be shown similarly. Denote $n_0$ and $n_1$ as the number of control and treated subjects, respectively. As we regard $T$ to be random, so are $n_0$ and $n_1$. By the law of large numbers, for $t=0,1$, $n_t n^{-1} \to (1-t)+(2t-1)P(T=1)$ in probability. Thus, for any $w \in \mathbb{R}$, $O_P(n_t^w) = O_P(n^w)$. For efficiency of presentation, without loss of generality, we treat each $T^{(i)}$ as given, as well as $n_0$ and $n_1$, in the rest of the proof. Following equation (7) in Abadie and Imbens (2006), let $K(i)$ be the number of times subject $i$ is used to match the others. $\hat{\mu}_X$ can be decomposed as

$$\begin{aligned}
\hat{\mu}_X - \mu &= n^{-1}\textstyle\sum_{i=1}^n[E\{Y(1)|X^{(i)}\} - E\{Y(0)|X^{(i)}\} - \mu] \\
&+ n^{-1}\textstyle\sum_{i=1}^n(2T^{(i)}-1)\{1+K(i)\}[Y^{(i)}(T^{(i)}) - E\{Y(T)|X^{(i)}\}] \\
&+ n^{-1}\textstyle\sum_{i=1}^n(2T^{(i)}-1)[E\{Y(1-T)|X^{(i)}\} - E\{Y(1-T)|X^{(J(i))}\}],
\end{aligned}$$

in which the first two terms on the right hand side are seen to be $O_P(n^{-1/2})$ with mean zero (Abadie and Imbens, 2006). The third term represents the bias in matching, which we denote by $B_n$. To show the result about $\hat{\mu}_X$, it suffices to show that $E(B_n) = O(n^{-2/p})$ and $\text{var}(B_n) = O(n^{-\min\{1+2/p,6/p\}})$. By Assumption 1 in the main text and Taylor's expansion,

$$\begin{aligned}
E\{Y(1-T)|X^{(i)}\} - E\{Y(1-T)|X^{(J(i))}\} = (X^{(i)} - X^{(J(i))})' g_i^{(i)} \\
+ (X^{(i)} - X^{(J(i))})' g_2^{(i)} (X^{(i)} - X^{(J(i))}) + g_3^{(i,J(i))} \|X^{(i)} - X^{(J(i))}\|^3,
\end{aligned}$$

where $g_1^{(i)}$ and $g_2^{(i)}$ are the gradient and the Hessian matrix of $E\{Y(1-T) \mid X\}$ at $X^{(i)}$, and $g_3^{(i,J(i))}$ is a bounded random element. Thus, for the desired result about $E(B_n)$ and $\mathrm{var}(B_n)$, it suffices to show that, for an arbitrary pair of subjects $(i,j)$,

$$\begin{aligned} & X^{(J(i))} - X^{(i)} = O_P(n^{-1/p}), E(X^{(J(i))} - X^{(i)}) = O(n^{-2/p}), \\ & \mathrm{cov}\{(X^{(J(i))} - X^{(i)})^{\otimes k}, (X^{(J(j))} - X^{(j)})^{\otimes l}\} = O(n^{-1-2/p}). \end{aligned} \tag{5}$$

in which $k,l \in \{1,2\}$, and $v^{\otimes 1} = v$ and $vv^{\otimes 2} = vv^{\intercal}$ for any real vector $v$. To show (5), we first suppose that both $i$ and $j$ belong to the same treatment group $T = t$. For any $a, b \in \Omega(X|T=t)$, let $(u,v) = n_{1-t}^{1/p}(X^{(J(a))} - a, X^{(J(b))} - b)$, in which $J(a)$ denotes the subject whose covariates are closest to a, and $J(b)$ likewise. From the proof of Theorem 1 in Abadie and Imbens (2006), let $f(\cdot)$ be the density function of random elements measurable with respect to $\{X^{(i)}, i = 1, \ldots, n\}$, we have

$$\begin{aligned} f(u) &= f(a + un_{1-t}^{-1/p})\{1 - P(\|X-a\| \le \|u\|n_{1-t}^{-1/p})\}^{n_{1-t}-1} \\ &= \{f(a) + f^{*\prime}(a,u)un_{1-t}^{-1/p}\}\{1 - P(\|X-a\| \le \|u\|n_{1-t}^{-1/p})\}^{n_{1-t}-1}, \end{aligned} \tag{6}$$

in which $f^*(a,u)$ is defined so as to make the equation holds. By Assumption 1, $f^*(a,u)$ is bounded. As shown in Abadie and Imbens (2006), $\{1 - P(\|X-a\| \le \|u\|n_{1-t}^{-1/p})\}^{n_{1-t}}$ converges to $\exp[-2\pi^{p/2}\|u\|^p f(a)/\{p\Gamma(p/2)\}]$. Thus $u = O_P(1)$, which implies that $X^{(J(a))} - a = O_P(n^{-1/p})$. Using the symmetry of $\{1 - P(\|X-a\| \le \|u\|n_{1-t}^{-1/p})\}^{n_{1-t}}$ about the origin in $\mathbb{R}^p$, Abadie and Imbens (2006) further showed that $E(X^{(J(a))} - a) = O(n^{-2/p})$. By the compactness of $\Omega(X)$, such convergence is uniform for $a$, thus $X^{(J(i))} - X^{(i)} = O_P(n^{-1/p})$ and $E(X^{(J(i))} - X^{(i)}) = O(n^{-2/p})$. Similar to (6), we further have

$$\begin{aligned} f(u,v) &= \{n_{1-t}(n_{1-t}-1)\}n_{1-t}^{-2} f(a + un_{1-t}^{-1/p}) f(b + vn_{1-t}^{-1/p}) \\ &\quad \{1 - P(\|X-a\| \le \|u\|n_{1-t}^{-1/p} \text{ or } \|X-b\| \le \|v\|n_{1-t}^{-1/p})\}^{n_{1-t}-2} \end{aligned}$$

in which $P(\|X-a\| \le \|u\|n_{1-t}^{-1/p} \text{ or } \|X-b\| \le \|v\|n_{1-t}^{-1/p})$ can be written as

$$\begin{aligned} & P(\|X-a\| \le \|u\|n_{1-t}^{-1/p}) + P(\|X-b\| \le \|v\|n_{1-t}^{-1/p}) \\ & \quad - P(\|X-a\| \le \|u\|n_{1-t}^{-1/p} \text{ and } \|X-b\| \le \|v\|n_{1-t}^{-1/p}) \\ & \equiv I + II + III \end{aligned}$$

We have $III = 0$ for all large $n$. Let $r_n = n_{1-t}P(\|X-a\| \leq \|u\|n_{1-t}^{-1/p})$, and $s_n = n_{1-t}P(\|X-b\| \leq \|v\|n_{1-t}^{-1/p})$. Then

$$f(u,v) = (1-n_{1-t}^{-1})f(a+un_{1-t}^{-1/p})f(b+vn_{1-t}^{-1/p})(1-r_n n_{1-t}^{-1} - s_n n_{1-t}^{-1})^{n_{1-t}-2}$$

for all large $n$, and $r_n \to r$ and $s_n \to s$, in which $r = 2\pi^{p/2}\|u\|^p f(a)/\{p\Gamma(p/2)\}$ and $s = 2\pi^{p/2}\|v\|^p f(b)/\{p\Gamma(p/2)\}$. Since

$$f(u)f(v) = f(a+un_{1-t}^{-1/p})f(b+vn_{1-t}^{-1/p})(1-r_n n_{1-t}^{-1})^{n_{1-t}-1}(1-s_n n_{1-t}^{-1})^{n_{1-t}-1},$$

we have, for $k,l \in \{1,2\}$,

$$\begin{aligned} & E\{u^{\otimes k}(v^{\otimes l})'\} - E(u^{\otimes k})E'(v^{\otimes l}) \\ & \quad = \int_{\mathbb{R}^p\times\mathbb{R}^p} u^{\otimes k}(v^{\otimes l})'\{f(u,v)-f(u)f(v)\}dudv \\ & \quad = \int_{\mathbb{R}^p\times\mathbb{R}^p} u^{\otimes k}(v^{\otimes l})' f(a+un_{1-t}^{-1/p})f(b+vn_{1-t}^{-1/p})\eta_{a,b}dudv + O(n^{-1}), \end{aligned}$$

in which $\eta_{a,b} = \{1-r_n n_{1-t}^{-1} - s_n n_{1-t}^{-1}\}^{n_{1-t}} - (1-r_n n_{1-t}^{-1})^{n_{1-t}}(1-s_n n_{1-t}^{-1})^{n_{1-t}}$. By (4), $\eta_{a,b} \to -{n_{1-t}}^{-1} rs\exp\{-(r+s)\}$. Since $\|u\|^2\|v\|^2 rs\exp\{-(r+s)\}$ is integrable on $(u,v) \in \mathbb{R}^p\times\mathbb{R}^p$, similar to Abadie and Imbens (2006), we have $E\{u^{\otimes k}(v^{\otimes l})'\} - E(u^{\otimes k})E'(v^{\otimes l}) = O(n^{-1})$. Next, suppose that $i$ and $j$ are control and treated subjects, respectively. Conditioning on $(X^{(i)}, X^{(j)}) = (a,b)$, let $u = n_1^{1/p}(X^{(J(a))} - a)$ and $v = n_0^{1/p}(X^{(J(b))} - b)$, we have

$$\begin{aligned} f(u,v) &= I(\max\{n_1^{-1/p}\|u\|, n_0^{-1/p}\|v\|\} < \|a-b\|)(1-n_1^{-1})(1-n_0^{-1}) \\ & \quad f(a+un_1^{-1/p})f(b+vn_0^{-1/p})(1-r_n)^{n_1-2}(1-s_n)^{n_0-2} \\ & \quad + I(n_1^{-1/p}\|u\| < \|a-b\|)\delta_{\|a-b\|}(\|v\|)(1-n_1^{-1})n_0^{-1}f(a+un_1^{-1/p}) \\ & \qquad (1-r_n)^{n_1-2}(1-s_n)^{n_0-1} \\ & \quad + I(n_0^{-1/p}\|v\| < \|a-b\|)\delta_{\|a-b\|}(\|u\|)(1-n_0^{-1})n_1^{-1}f(b+vn_0^{-1/p}) \\ & \qquad (1-s_n)^{n_0-2}(1-r_n)^{n_1-1} \\ & \quad + \delta_{(\|a-b\|,\|a-b\|)}((\|u\|,\|v\|))n_1^{-1}n_0^{-1}(1-r_n)^{n_1-1}(1-s_n)^{n_0-1}, \end{aligned}$$

in which $\delta_w(x)$ is the Dirac function of $x$ such that it is zero whenever $x \neq w$ and $\int_{\mathbb{R}^{dim(x)}} \delta_w(x)h(x)dx = h(w)$ for any function $h$ of $x$. Similar to the above, we can show that $\mathrm{cov}(u^{\otimes k}, v^{\otimes l}) = O(n^{-1})$. By the compactness of $\Omega(X)$, this convergence is uniform on $(a,b)$, which means that $\mathrm{cov}\{(X^{(J(i))} - X^{(i)})^{\otimes k}, (X^{(J(j))} - X^{(j)})^{\otimes l}\} = O(n^{-1-2/p})$. Hence (5) holds, which completes the proof.

## B.4 Proof of Theorem 4

The form of the variance follows directly from Theorem 5 of Abadie and Imbens (2006) and the sufficiency of $X'\beta_t$ for $Y(t)|X$. Let $\sigma_t^2(R) = \mathrm{var}\{Y(t)|R\}$ for any random element $R$. To see that $\mathrm{var}(\hat{\mu}_\pi) \geq \mathrm{var}(\hat{\mu}_r)$ for all large $n$, following Theorem 5 of Abadie and Imbens (2006), under Assumption 2 in the main text, we have

$$\begin{aligned} n\mathrm{var}(\hat{\mu}_\pi) \to V_{\pi(X)} + \sum_{t=0}^{1} \Bigg[ E\left\{ \frac{\sigma_t^2(\pi(X))}{h(\pi(X),t)} \right\} \\ + \frac{1}{2m} E\left[ \left\{ \frac{1}{h(\pi(X),t)} - h(\pi(X),t) \right\} \sigma_t^2(\pi(X)) \right] \Bigg]. \end{aligned}$$

We write $n\mathrm{var}(\hat{\mu}_\pi) = V_\pi + I_\pi + II_\pi$ and $n\mathrm{var}(\hat{\mu}_\pi) = V_X + I_r + II_r$, and additionally introduce $I_X$ and $II_X$, in which

$$\begin{aligned}
I_\pi &= E\left\{ \frac{\sigma_1^2(\pi(X))}{\pi(X)} + \frac{\sigma_0^2(\pi(X))}{1-\pi(X)} \right\} \\
II_\pi &= \frac{1}{2m} E\left[ \left\{ \frac{1}{\pi(X)} - \pi(X) \right\} \sigma_1^2(\pi(X)) + \left\{ \frac{1}{1-\pi(X)} - \{1-\pi(X)\} \right\} \sigma_0^2(\pi(X)) \right] \\
I_r &= E\left\{ \frac{\sigma_1^2(X)}{\pi(X'\beta_1)} + \frac{\sigma_0^2(X)}{1-\pi(X'\beta_0)} \right\} \\
II_r &= \frac{1}{2m} E\left[ \left\{ \frac{1}{\pi(X'\beta_1)} - \pi(X'\beta_1) \right\} \sigma_1^2(X) + \left\{ \frac{1}{1-\pi(X'\beta_0)} - \{1-\pi(X'\beta_0)\} \right\} \sigma_0^2(X) \right]. \\
I_X &= E\left\{ \frac{\sigma_1^2(X)}{\pi(X)} + \frac{\sigma_0^2(X)}{1-\pi(X)} \right\} \\
II_X &= \frac{1}{2m} E\left[ \left\{ \frac{1}{\pi(X)} - \pi(X) \right\} \sigma_1^2(X) + \left\{ \frac{1}{1-\pi(X)} - \{1-\pi(X)\} \right\} \sigma_0^2(X) \right].
\end{aligned}$$

Then the inequality $n\mathrm{var}(\hat{\mu}_\pi) \geq n\mathrm{var}(\hat{\mu}_r)$ can be implied if

$$V_\pi + I_\pi \geq V_X + I_X, \quad II_\pi \geq II_X, \quad I_X \geq I_r, \quad II_X \geq II_r. \tag{7}$$

Let $\mu_{c,t}(X) = E\{Y(t)|X\} - E\{Y(t)|\pi(X)\}$ for $t = 0, 1$. By definition,

$$\begin{aligned} V_X - V_\pi &= \mathrm{var}[E\{Y(1) - Y(0)|X\}] - \mathrm{var}[E\{Y(1) - Y(0)|\pi(X)\}] \\ &= E\{\mu_{c,1}(X)\}^2 + E\{\mu_{c,0}(X)\}^2 - 2E\{\mu_{c,1}(X)\mu_{c,0}(X)\}. \end{aligned} \tag{8}$$

Let $\pi^*(X) = \pi(X)/\{1-\pi(X)\}$, the logit function of $\pi(X)$. By the triangle inequality,

$$\begin{aligned}
-2E\{\mu_{c,1}(X)\mu_{c,0}(X)\} &= -2E\left[[\mu_{c,1}(X)\{\pi^*(X)\}^{-1/2}]\,[\mu_{c,0}(X)\{\pi^*(X)\}^{1/2}]\right] \\
&\leq E\{\mu_{c,1}^2(X)/\pi^*(X)\} + E\{\mu_{c,0}^2(X)\pi^*(X)\}.
\end{aligned}$$

By plugging it back to (8), we have

$$V_X - V_\pi \leq E[\mu_{c,1}(X)^2/\pi(X)] + E[\mu_{c,0}(X)^2/\{1-\pi(X)\}]$$

On the other hand, we have

$$\begin{aligned}
I_\pi - I_X &= E[\{\sigma_1^2(\pi(X)) - \sigma_1^2(X)\}/\pi(X)] + E[\{\sigma_0^2(\pi(X)) - \sigma_0^2(X)\}/\{1-\pi(X)\}] \\
&= E[\mu_{c,1}^2(X)/\pi(X)] + E[\mu_{c,0}^2(X)/\{1-\pi(X)\}].
\end{aligned}$$

Hence $V_X - V_\pi \leq I_\pi - I_X$, which implies the first inequality in (7). Because $E\{\sigma_1^2(X)|\pi(X)\} = E[\text{var}\{Y(1)|X\}|\pi(X)] \leq \text{var}\{Y(1)|\pi(X)\} = \sigma_1^2(\pi(X))$, we have

$$E[\{\pi(X)^{-1} - \pi(X)\}\sigma_1(X)] \leq E[\{\pi(X)^{-1} - \pi(X)\}\sigma_1(\pi(X))].$$

Similarly, we can show the corresponding inequality for the part when $t = 0$ with $1 - \pi(X)$ in place of $\pi(X)$. Thus $II_\pi \geq II_X$. Next, let $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ be that $\phi(x) = x^{-1}$. Then $\phi$ is a convex function. By Jensen's inequality, we have $\pi(X'\beta_1)^{-1} = [E\{\pi(X)|X'\beta_1\}]^{-1} \leq E\{\pi^{-1}(X)|X'\beta_1\}$, which, together with that $\sigma_1^2(X) = \sigma_1^2(X'\beta_1)$, implies that

$$E\{\sigma_1^2(X)/\pi(X'\beta_1)\} \leq E[\sigma_1^2(X'\beta_1)E\{\pi^{-1}(X)|X'\beta_1\}] = E\{\sigma_1^2(X)/\pi(X)\}.$$

Similarly, we can show the corresponding inequality for the part when $t = 0$ with $1 - \pi(X)$ in place of $\pi(X)$. Hence $I_X \geq I_r$. Finally, let $\psi : \mathbb{R}^+ \to \mathbb{R}$ be that $\psi(x) = x^{-1} - x$, then $\psi$ is also a convex function. Thus the fourth inequality in (7) can be shown similarly. From the arguments above, the equality between the limits of $n\text{var}(\hat{\mu}_\pi)$ and $n\text{var}(\hat{\mu}_r)$ holds if and only if all the four equalities in (7) hold, which means that

(a) $\mu_{c,1}(X)\pi^*(X)^{-1/2} = -\mu_{c,0}(X)\pi^*(X)^{1/2}$,

(b) $E\{Y(t)|X\} = E\{Y(t)|\pi(X)\}$ for $t = 0, 1$,

(c) $\pi(X) = \pi(X'\beta_1) = \pi(X'\beta_0)$.

Note that (b) is equivalent to that $\mu_{c,t}(X) = 0$ for $t = 0, 1$, which implies (a). If $\mathcal{S}_{Y(0)|X} = \mathcal{S}_{Y(1)|X}$, then (c) is equivalent to that $\pi(X) = \pi(X'\beta_1)$. Otherwise, it means that $\pi(X)$ is a constant, which, together with (b), indicates that $E\{Y(t)|X\} \equiv E\{Y(t)\}$ for $t = 0, 1$. This completes the proof.

# C More simulation results

The simulation results for estimating $ACET$ in Case 1 and Case 3 are displayed in Table 1 and Table 2 respectively, and the notations in these tables follow those in Table 1 of the main text.

# D R codes for data analysis

```
##### Prepare the data
library(MatchIt)
data(lalonde)
lalonde$u74<-ifelse(lalonde$re74>0,0,1)
lalonde$u75<-ifelse(lalonde$re75>0,0,1)
attach(lalonde)
W<-cbind(age,educ,black,hispan,married,nodegree,re74,re75,u74,u75)
Y <- lalonde$re78
Tr <- lalonde$treat

##### Perform dimension reduction on the covariates
library(dr)
s0<-dr(Y[Tr==0]~W[Tr==0,],method="sir",nslices=5)
s1<-dr(Y[Tr==1]~W[Tr==1,],method="sir",nslices=5)
```

Table 1: Simulation Results for $ACET$ in Case 1

| | I | II | III | IV | V |
|---|---|---|---|---|---|
| | | | BIAS | | |
| Abadie-Imbens | | | | | |
| Ambient | 0.4877 | 0.0815 | 0.3602 | 0.0513 | 0.0020 |
| Estimated PS | 0.0305 | 0.0206 | 0.0089 | -0.0030 | 0.0010 |
| True PS | 0.0680 | 0.0097 | 0.0125 | -0.0019 | 0.0041 |
| Prognostic Score | 0.0231 | 0.0018 | 0.0153 | 0.0034 | -0.0029 |
| Active Set (Oracle) | 0.0230 | 0.0152 | 0.0875 | 0.0513 | -0.0005 |
| SDR (Oracle) | 0.0230 | 0.0001 | 0.0086 | 0.0025 | -0.0005 |
| Proposed | 0.0294 | 0.0026 | 0.0139 | 0.0043 | 0.0014 |
| Genetic Matching | 0.3668 | 0.0263 | 0.2158 | 0.0257 | 0.0046 |
| Entropy Balancing | -0.0081 | 0.0115 | 0.0010 | -0.0011 | 0.0009 |
| Double Robust | 0.0161 | -0.0100 | 0.0190 | -0.0005 | 0.0000 |
| TMLE | -0.0163 | 0.0182 | -0.0012 | -0.0003 | 0.0004 |
| | | | SD | | |
| Abadie-Imbens | | | | | |
| Ambient | 0.5050 | 0.1025 | 0.1612 | 0.0631 | 0.0799 |
| Estimated PS | 0.6378 | 0.1203 | 0.2033 | 0.0640 | 0.1448 |
| True PS | 1.1372 | 0.2007 | 0.2520 | 0.0607 | 0.1477 |
| Prognostic Score | 0.2446 | 0.0645 | 0.1224 | 0.0606 | 0.0756 |
| Active Set (Oracle) | 0.2250 | 0.0642 | 0.1218 | 0.0631 | 0.0574 |
| SDR (Oracle) | 0.225 | 0.0593 | 0.1161 | 0.0610 | 0.0574 |
| Proposed | 0.2432 | 0.0654 | 0.1224 | 0.0603 | 0.0737 |
| Genetic Matching | 0.4000 | 0.0777 | 0.1509 | 0.0630 | 0.1047 |
| Entropy Balancing | 0.3783 | 0.0625 | 0.1110 | 0.0472 | 0.0887 |
| Double Robust | 0.2800 | 0.0666 | 0.1241 | 0.0507 | 0.0786 |
| TMLE | 0.2205 | 0.0670 | 0.1137 | 0.0512 | 0.0886 |
| | | | RMSE | | |
| Abadie-Imbens | | | | | |
| Ambient | 0.7019 | 0.1309 | 0.3946 | 0.0813 | 0.0798 |
| Estimated PS | 0.6382 | 0.1220 | 0.2034 | 0.0641 | 0.1447 |
| True PS | 1.1386 | 0.2009 | 0.2521 | 0.0607 | 0.1477 |
| Prognostic Score | 0.2456 | 0.0645 | 0.1233 | 0.0606 | 0.0756 |
| Active Set (Oracle) | 0.2261 | 0.0659 | 0.1499 | 0.0813 | 0.0574 |
| SDR (Oracle) | 0.2261 | 0.0592 | 0.1164 | 0.0610 | 0.0574 |
| Proposed | 0.2448 | 0.0655 | 0.1231 | 0.0604 | 0.0737 |
| Genetic Matching | 0.5426 | 0.0820 | 0.2633 | 0.0680 | 0.1048 |
| Entropy Balancing | 0.3783 | 0.0635 | 0.1110 | 0.0472 | 0.0887 |
| Double Robust | 0.2803 | 0.0673 | 0.1255 | 0.0507 | 0.0786 |
| TMLE | 0.2209 | 0.0694 | 0.1137 | 0.0511 | 0.0886 |

Table 2: Simulation Results for $ACET$ in Case 3

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| | | | | BIAS | | | |
| Abadie-Imbens | | | | | | | |
| Ambient | 0.0557 | 0.0487 | 0.0484 | 0.0578 | 0.0484 | 0.0600 | 0.0510 |
| Estimated PS | 0.0044 | 0.0039 | -0.0128 | -0.0072 | -0.0042 | -0.0071 | -0.0113 |
| True PS | 0.0053 | 0.0025 | 0.0033 | 0.0053 | 0.0043 | 0.0064 | -0.0038 |
| Prognostic Score | 0.0071 | 0.0062 | 0.0055 | 0.0095 | 0.0079 | 0.0092 | 0.0060 |
| SDR (Oracle) | 0.0024 | 0.0016 | 0.0020 | 0.0022 | 0.0016 | 0.0019 | 0.0017 |
| Active Set (Oracle) | 0.0257 | 0.0195 | 0.0155 | 0.0304 | 0.0238 | 0.0325 | 0.0200 |
| Proposed | 0.0035 | 0.0015 | 0.0028 | 0.0041 | 0.0036 | 0.0047 | 0.0018 |
| Genetic Matching | 0.0299 | 0.0216 | 0.0202 | 0.0291 | 0.0199 | 0.0340 | 0.0251 |
| Entropy Balancing | 0.0013 | 0.0098 | 0.0017 | 0.0018 | 0.0011 | 0.0017 | 0.0010 |
| Double Robust | -0.0129 | -0.0058 | -0.0408 | -0.0086 | -0.0015 | -0.0060 | -0.0092 |
| TMLE | 0.0012 | 0.0005 | 0.0012 | 0.0015 | 0.0007 | -0.0003 | 0.0025 |
| | | | | SD | | | |
| Abadie-Imbens | | | | | | | |
| Ambient | 0.0521 | 0.0525 | 0.0515 | 0.0554 | 0.0550 | 0.0538 | 0.0542 |
| Estimated PS | 0.0902 | 0.0877 | 0.0896 | 0.0947 | 0.0955 | 0.0948 | 0.0886 |
| True PS | 0.1159 | 0.1117 | 0.1432 | 0.1196 | 0.1189 | 0.1202 | 0.1468 |
| Prognostic Score | 0.0465 | 0.0462 | 0.0450 | 0.0492 | 0.0472 | 0.0483 | 0.0463 |
| SDR (Oracle) | 0.0367 | 0.0379 | 0.0372 | 0.0368 | 0.0376 | 0.0375 | 0.0379 |
| Active Set (Oracle) | 0.0680 | 0.0668 | 0.0667 | 0.0702 | 0.0705 | 0.0744 | 0.0754 |
| Proposed | 0.0449 | 0.0414 | 0.0414 | 0.0448 | 0.0438 | 0.0457 | 0.0431 |
| Genetic Matching | 0.0639 | 0.0631 | 0.0660 | 0.0663 | 0.0661 | 0.0643 | 0.0607 |
| Entropy Balancing | 0.0393 | 0.0371 | 0.0358 | 0.0416 | 0.0403 | 0.0409 | 0.0371 |
| Double Robust | 0.0418 | 0.0404 | 0.0488 | 0.0438 | 0.0440 | 0.0437 | 0.0484 |
| TMLE | 0.0392 | 0.0388 | 0.0468 | 0.0423 | 0.0423 | 0.0425 | 0.0471 |
| | | | | RMSE | | | |
| Abadie-Imbens | | | | | | | |
| Ambient | 0.0762 | 0.0716 | 0.0706 | 0.0801 | 0.0733 | 0.0806 | 0.0744 |
| Estimated PS | 0.0902 | 0.0877 | 0.0905 | 0.0950 | 0.0956 | 0.0950 | 0.0893 |
| True PS | 0.1159 | 0.1117 | 0.1432 | 0.1197 | 0.1189 | 0.1203 | 0.1468 |
| Prognostic Score | 0.0470 | 0.0466 | 0.0453 | 0.0500 | 0.0478 | 0.0491 | 0.0466 |
| SDR (Oracle) | 0.0368 | 0.0379 | 0.0372 | 0.0369 | 0.0376 | 0.0376 | 0.0379 |
| Active Set (Oracle) | 0.0726 | 0.0696 | 0.0685 | 0.0765 | 0.0744 | 0.0812 | 0.0780 |
| Proposed | 0.0450 | 0.0414 | 0.0415 | 0.0449 | 0.0439 | 0.0459 | 0.0431 |
| Genetic Matching | 0.0706 | 0.0666 | 0.0690 | 0.0724 | 0.0690 | 0.0727 | 0.0657 |
| Entropy Balancing | 0.0393 | 0.0371 | 0.0359 | 0.0417 | 0.0403 | 0.0409 | 0.0371 |
| Double Robust | 0.0437 | 0.0408 | 0.0635 | 0.0446 | 0.0440 | 0.0441 | 0.0493 |
| TMLE | 0.0392 | 0.0388 | 0.0468 | 0.0424 | 0.0423 | 0.0425 | 0.0471 |

```
##### To obtain the dimension of the central subspaces by ladle (Luo and Li, 2016),
please contact the authors of Luo and Li (2016) for the R codes. For this dataset,
we find that the dimension of the central subspaces in both treatment and control
groups is 2.
rx0<-W%*%s0$evectors[,1:2]
rx1<-W%*%s1$evectors[,1:2]

##### Plot Figure 1 in the main text
par(mfrow=c(1,2))

plot(rx0[Tr==0,1],Y[Tr==0],xlab="1st reduced covariate",ylab="Y")
l1<-loess(Y[Tr==0]~rx0[Tr==0,1])
p1<-predict(l1,se=TRUE)
f1<-p1$fit
f1u<-f1+p1$se.fit*2
f1l<-f1-p1$se.fit*2
rx00<-rx0[Tr==0,1]
lines(rx00[order(rx00)],f1[order(rx00)],col=2)
lines(rx00[order(rx00)],f1u[order(rx00)],col=4,lty="dashed")
lines(rx00[order(rx00)],f1l[order(rx00)],col=4,lty="dashed")

plot(rx0[Tr==0,2],Y[Tr==0],xlab="2nd reduced covariate",ylab="Y")
l1<-loess(Y[Tr==0]~rx0[Tr==0,2])
p1<-predict(l1,se=TRUE)
f1<-p1$fit
f1u<-f1+p1$se.fit*2
f1l<-f1-p1$se.fit*2
rx00<-rx0[Tr==0,2]
lines(rx00[order(rx00)],f1[order(rx00)],col=2)
```

```
lines(rx00[order(rx00)],f1u[order(rx00)],col=4,lty="dashed")
lines(rx00[order(rx00)],f1l[order(rx00)],col=4,lty="dashed")

##### Plot Figure 2 in the main text
par(mfrow=c(1,2))
boxplot(rx0[,1]~Tr,ylab="1st reduced covariate")
boxplot(rx0[,2]~Tr,ylab="2nd reduced covariate")

##### Plot Figure 3 in the main text
par(mfrow=c(2,1))
hist(rx0[Tr==1,1],xlab="1st reduced covariate",ylim=c(0,200),main="Histogram")
hist(rx0[Tr==0,1],add=TRUE,lty="dashed")
legend(6.5,150,c("Treat","Control"),lty=c("solid","dashed"))

par(mfrow=c(2,1))
hist(rx0[Tr==1,1],xlab="1st reduced covariate",ylim=c(0,200),main="Histogram")
hist(rx0[Tr==0,1],add=TRUE,lty="dashed")
legend(6.5,200,c("Treat","Control"),lty=c("solid","dashed"))

hist(rx0[Tr==1,2],xlab="2nd reduced covariate",ylim=c(0,250),main="Histogram")
hist(rx0[Tr==0,2],add=TRUE,lty="dashed")
legend(-0.9,250,c("Treat","Control"),lty=c("solid","dashed"))

##### Estimate average causal effect among the treated
library(Matching)
Match(Y=Y, Tr=Tr, X=rx0,estimand="ATT", M=1,replace=TRUE)$est
```

# References


Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* **74,** 235–267.

Luo, W. and Li, B. (2016). Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika* **103,** 875–887.