Supplement of "Outcome Regression-based Estimation of Conditional Average Treatment Effect"

Lu Li[1] · Niwen Zhou[2] · Lixing Zhu[2,3*]

1 Theoretical Results

Give some notations first.

(1) C and M stand for two generic bounded constants, Ξ is the σ-field generated by $X_{11}, \ldots, X_{1n}$.
(2) $\epsilon_{ti} = Y_i - E(Y_{(t)}|X_i)$, $\tau_t(x_1) = E[E\{Y|D=t,X\}|X_1 = x_1]$, $Z^t = \boldsymbol{\beta}_t^\top X$ for $t = 0, 1$ and $i = 1, \ldots, n$.
(3) Write $K_1\left(\frac{X_{1i}-X_1}{h_1}\right)$ as $K_{1h}(X_{1i})$; $K_2\left(\frac{X_i - X_j}{h_2}\right)$ as $K_{2h}(X_i - X_j)$, and $K_{4h}(Z_i - Z_j)$ as $K_4\left(\frac{Z_i - Z_j}{h_4}\right)$.

In the two-step estimation procedure for CATE, the second step involves, for $i = 1, \ldots, n$, the quantities:

$$\widehat{K}_{1h}(X_{1i}) = \sum_{j:j\neq i} w_{ij} K_{1h}(X_{1j}).$$

We call it the estimator of $K_{1h}(X_{1j})$. In different circumstances, w_{ij} can be different. Take OR-N as an example, and write w_{ij} as w_{ij}^N:

$$w_{ij}^N = \frac{\frac{1}{nh_2^p} K_{2h}(X_i - X_j)}{\frac{1}{nh_2^p} \sum_{i=1}^{n} K_{2h}(X_i - X_j)\mathbb{1}(D_i = 1)}$$

that depends on $X_1, \ldots, X_n$ only.

---

1 School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, 200240, China

2 Center for Statistics and Data Science, Beijing Normal University, Zhuhai,519087,China

3 Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

Lemma 1 *Given assumptions (C1) - (C4) in Subsection 2.1 and (A1) - (A4) in Subsections 2.2-2.3,*

$$|w_{ij}^N - w_{ji}^N| = \frac{O_p(h_2)}{nh_2^p} \left| K_{2h}\left(X_i - X_j\right) \right|, \tag{1}$$

Proof of Lemma 1. By assumption (A2), $w_{ij}^N = w_{ji}^N = 0$ for $||X_j - X_i||_\infty > h_2$ (Abrevaya, Hsu and Lieli 2015). Suppose that $||X_j - X_i||_\infty \le h_2$. For all j, we define

$$\widehat{f}(X_j) = \frac{1}{nh_2^p} \sum_{i: i \neq j}^{n} K_{2h}\left(X_i - X_j\right).$$

It is clear that

$$\begin{aligned} w_{ij}^N &= \frac{\frac{1}{nh_2^p} K_{2h}\left(X_i - X_j\right)}{\frac{1}{nh_2^p} \sum_{i=1}^{n} K_{2h}\left(X_i - X_j\right) \mathbb{1}(D_i = 1)} = \frac{\frac{1}{nh_2^p} K_{2h}\left(X_i - X_j\right)}{\frac{1}{nh_2^p} \sum_{i=1}^{n} K_{2h}\left(X_i - X_j\right)} \times \frac{\frac{1}{nh_2^p} \sum_{i=1}^{n} K_{2h}\left(X_i - X_j\right)}{\frac{1}{nh_2^p} \sum_{i=1}^{n} K_{2h}\left(X_i - X_j\right) \mathbb{1}(D_i = 1)} \\ &= \frac{\frac{1}{nh_2^p} K_{2h}\left(X_i - X_j\right)}{\widehat{f}(X_j) \widehat{p}(X_j)}. \end{aligned} \tag{2}$$

Then we have

$$\begin{aligned} &|w_{ij}^N - w_{ji}^N| \\ &= \frac{1}{nh_2^p} \left| \frac{K_{2h}\left(X_i - X_j\right)}{\widehat{p}(X_j)\widehat{f}(X_j)} - \frac{K_{2h}\left(X_j - X_i\right)}{\widehat{p}(X_i)\widehat{f}(X_i)} \right| = \frac{1}{nh_2^p} \left| K_{2h}\left(X_i - X_j\right) \right| \left| \frac{1}{\widehat{p}(X_j)\widehat{f}(X_j)} - \frac{1}{\widehat{p}(X_i)\widehat{f}(X_i)} \right| \\ &\le \frac{1}{nh_2^p} \left| K_{2h}\left(X_i - X_j\right) \right| \left\{ \left| \frac{1}{\widehat{p}(X_j)\widehat{f}(X_j)} - \frac{1}{p(X_j)f(X_j)} \right| + \left| \frac{1}{\widehat{p}(X_i)\widehat{f}(X_i)} - \frac{1}{p(X_i)f(X_i)} \right| \right. \\ &\quad \left. + \left| \frac{1}{p(X_j)f(X_j)} - \frac{1}{p(X_i)f(X_i)} \right| \right\} \\ &= \frac{1}{nh_2^p} \left| K_{2h}\left(X_i - X_j\right) \right| \left\{ \left| \frac{\widehat{p}(X_j)\widehat{f}(X_j) - p(X_j)f(X_j)}{\widehat{p}(X_j)p(X_j)\widehat{f}(X_j)f(X_j)} \right| + \left| \frac{\widehat{p}(X_i)\widehat{f}(X_i) - p(X_i)f(X_i)}{\widehat{p}(X_i)p(X_i)\widehat{f}(X_i)f(X_i)} \right| \right. \\ &\quad \left. + \left| \frac{p(X_i)f(X_i) - p(X_j)f(X_j)}{p(X_i)p(X_j)f(X_i)f(X_j)} \right| \right\}. \end{aligned} \tag{3}$$

Under conditions (C1)-(C4) and (A1)-(A4) for nonparametric estimation,

$$\sup_i |\widehat{f}(X_i) - f(X_i)| = O_p\left(h_2^{s_2} + \sqrt{\frac{\log n}{nh_2^p}} \right),$$

$$\sup_i |\widehat{p}(X_i) - p(X_i)| = O_p\left(h_2^{s_2} + \sqrt{\frac{\log n}{nh_2^p}} \right).$$

Since $s_2 \ge p \ge 2$, assumption (A3) implies that $\sup_i |\widehat{f}(X_i) - f(X_i)| = o_p(h_2)$ and $\sup_i |\widehat{p}(X_i) - p(X_i)| = o_p(h_2)$. By the mean value theorem,

$$\sup_j \left| \frac{\widehat{p}(X_j)\widehat{f}(X_j) - p(X_j)f(X_j)}{\widehat{p}(X_j)p(X_j)\widehat{f}(X_j)f(X_j)} \right| \le \sup_j \frac{1}{\widetilde{p}^2(X_j)\widetilde{f}^2(X_j)} \sup_j \left| \widehat{p}(X_j)\widehat{f}(X_j) - p(X_j)f(X_j) \right|,$$

where $\widetilde{p}(X_j)$ is a quantity between $\widehat{p}(X_j)$ and $p(X_j)$, similarly, $\widetilde{f}(X_j)$ is also a quantity between $\widehat{f}(X_j)$ and $f(X_j)$. Owing to that f and p are bounded away from zero, $\sup_j \widetilde{p}_j^{-2} \widetilde{f}_j^{-2} = O_p(1)$. After a simple calculation, we have

$$\sup_j \left|\widehat{p}(X_j)\widehat{f}(X_j) - p(X_j)f(X_j)\right| = O_p\left(h_2^{s_2} + \sqrt{\frac{\log n}{nh_2^p}}\right) = o_p(h_2).$$

Therefore,

$$\sup_j \left|\frac{\widehat{p}(X_j)\widehat{f}(X_j) - p(X_j)f(X_j)}{\widehat{p}(X_j)p(X_j)\widehat{f}(X_j)f(X_j)}\right| = o_p(h_2), \quad \sup_i \left|\frac{\widehat{p}(X_i)\widehat{f}(X_i) - p(X_i)f(X_i)}{\widehat{p}(X_i)p(X_i)\widehat{f}(X_i)f(X_i)}\right| = o_p(h_2).$$

As for the last term in (3), noticing that f and p are continuously differentiable on its compact support and bounded away from zero, we have $\left|\frac{1}{f(x_1)p(x_1)} - \frac{1}{f(x_2)p(x_2)}\right| \leq M||x_1 - x_2||_\infty$ for all $x_1, x_2 \in \mathcal{X}$ and a constant $M > 0$. $||X_j - X_i||_\infty \leq h_2$ leads to $\left|\frac{1}{f(X_i)p(X_i)} - \frac{1}{f(X_j)p(X_j)}\right| = O(h_2)$. Combining all results yields (1). □

Proof of Theorem 2. We can rewrite $\widehat{m}_1(X_i) - \widehat{m}_0(X_i) - \tau(x_1)$ as $\{\widehat{m}_1(X_i) - \tau_1(x_1)\} - \{\widehat{m}_0(X_i) - \tau_0(x_1)\}$. Then based on (2.2),

$$\begin{aligned}
&\sqrt{nh_1^k}(\widehat{\tau}(x_1) - \tau(x_1)) \\
=&\frac{\frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n K_{1h}(X_{1i}) \{[\widehat{m}_1(X_i) - \tau_1(x_1)] - [\widehat{m}_0(X_i) - \tau_0(x_1)]\}}{\frac{1}{nh_1^k} \sum_{i=1}^n K_{1h}(X_{1i})} \\
=&\frac{\frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n K_{1h}(X_{1i}) \{[\widehat{m}_1(X_i) - \tau_1(x_1)] - [\widehat{m}_0(X_i) - \tau_0(x_1)]\}}{f(x_1)}(1 + o_p(1)),
\end{aligned} \tag{4}$$

as

$$\sup_{x_1} \left|\frac{1}{nh_1^k} \sum_{i=1}^n K_{1h}(X_{1i}) - f(x_1)\right| = o_p(1).$$

First, deal with $\{\widehat{m}_1(X_i) - \tau_1(x_1)\}$ in (4). It is clear that

$$\begin{aligned}
&\frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n K_{1h}(X_{1i}) [\widehat{m}_1(X_i) - \tau_1(x_1)] \\
=& \frac{1}{\sqrt{nh_1^k}} \left\{ \sum_{i=1}^n K_{1h}(X_{1i}) [\widehat{m}_1(X_i) - m_1(X_i)] + \sum_{i=1}^n K_{1h}(X_{1i}) [m_1(X_i) - \tau_1(x_1)] \right\} \\
=:& \frac{1}{\sqrt{nh_1^k}} (I_{n,1} + I_{n,2}).
\end{aligned} \tag{5}$$

A simple calculation yields that

$$\left|\frac{1}{\sqrt{nh_1^k}} I_{n,1}\right| \leq \sup_x |\widehat{m}_1(X_i) - m_1(X_i)| \frac{1}{nh_1^k} \sum_{i=1}^n |K_{1h}(X_{1i})|.$$

As $h_1 \to 0$ and $\frac{1}{nh_1^k}\sum_{i=1}^{n}|K_{1h}(X_{1i})| = O_p(1)$, we then have $\frac{1}{\sqrt{nh_1^k}} I_{n,1} = O_p(\sqrt{h_1^k}) = o_p(1)$. Thus, equation (5) becomes

$$\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[\widehat{m}_1(X_i)-\tau_1(x_1)\right] = \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[m_1(X_i)-\tau_1(x_1)\right] + o_p(1).$$

Similarly,

$$\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[\widehat{m}_0(X_i)-\tau_0(x_1)\right] = \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[m_0(X_i)-\tau_0(x_1)\right] + o_p(1).$$

Altogether, the asymptotically linear representation of $\widehat{\tau}(x_1)$ is

$$\begin{aligned}\sqrt{nh_1^k}\{\widehat{\tau}(x_1)-\tau(x_1)\} &= \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\{m_1(X_i)-m_0(X_i)-\tau(x_1)\}}{f(x_1)}(1+o_p(1)) \\ &= \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\{m_1(X_i)-m_0(X_i)-\tau(x_1)\}}{f(x_1)} + o_p(1).\end{aligned}$$

The second equation is due to the asymptotic finiteness of the leading term that is asymptotically normal shown below. As it is the sum of independent variables, the asymptotic normality is easy to derive. Specifically, noticing that the random variables

$$\{K_{1h}\left(X_{1i}\right)\left[m_1(X_i)-m_0(X_i)-\tau(X_{1i})\right]\}_{i=1}^{n}$$

are i.i.d., then we can apply Lyapunov's central limit theorem to obtain the asymptotic distribution shown in Theorem 2. Under the assumptions (C1)-(C4) and (A1), we derive that

$$\sqrt{nh_1^k}\left\{\widehat{\tau}(x_1)-\tau(x_1)\right\} \xrightarrow{d} N\left(0, \frac{||K_1||_2^2\sigma_P^2(x_1)}{f(x_1)}\right),$$

we now give the formula of $\sigma_P^2(x_1)$. It is easy to see that when $n\to\infty$, the variance of

$$\frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\{m_1(X_i)-m_0(X_i)-\tau(x_1)\}}{f(x_1)}$$

converges to

$$\sigma_P^2(x_1) := E[\{m_1(X)-m_0(X)-\tau(x_1)\}^2|X_1=x_1].$$

The proof of Theorem 2 is finished. □

Proof of Theorem 3. First, we have

$$\begin{aligned}&\sqrt{nh_1^k}(\widehat{\tau}(x_1)-\tau(x_1)) \\ &= \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[\widehat{m}_1(X_i)-\tau_1(x_1)\right]}{\frac{1}{nh_1^k}\sum_{i=1}^{n} K_{1h}(X_{1i})} - \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[\widehat{m}_0(X_i)-\tau_0(x_1)\right]}{\frac{1}{nh_1^k}\sum_{i=1}^{n} K_{1h}(X_{1i})},\end{aligned} \tag{6}$$

where

$$\widehat{m}_1(X_i) = \frac{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_{2h}\left(X_j - X_i\right) Y_{1j} \mathbb{1}(D_j = 1)}{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_{2h}\left(X_j - X_i\right) \mathbb{1}(D_j = 1)},$$

$$\widehat{m}_0(X_i) = \frac{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_{2h}\left(X_j - X_i\right) Y_{0j} \mathbb{1}(D_j = 0)}{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_{2h}\left(X_j - X_i\right) \mathbb{1}(D_j = 0)}.$$

Similarly as the proof for Theorem 2, we have the following decomposition:

$$\begin{aligned}
&\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[m_1(X_i) - \tau_1(x_1)\right] \\
&= \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} \epsilon_{1i}\mathbb{1}(D_i = 1)\frac{K_{1h}\left(X_{1i}\right)}{p(X_i)} + \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[E(Y_{(1)}|X_i) - \tau_1(x_1)\right] \\
&\quad + \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} \epsilon_{1i}\mathbb{1}(D_i = 1)\sum_{j=1}^{n} K_{1h}\left(X_{1j}\right)(w_{ij}^N - w_{ji}^N) \\
&\quad + \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} \epsilon_{1i}\mathbb{1}(D_i = 1)\left[\sum_{j=1}^{n} K_{1h}\left(X_{1j}\right) w_{ji}^N - \frac{K_{1h}\left(X_{1i}\right)}{p(X_i)}\right] \\
&\quad + \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[\frac{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_{2h}\left(X_j - X_i\right)\mathbb{1}(D_j = 1) EY_{(1)}|X_j}{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_{2h}\left(X_j - X_i\right)\mathbb{1}(D_j = 1)} - EY_{(1)}|X_i\right] \\
&=: I_{n,3} + I_{n,4} + I_{n,5} + I_{n,6} + I_{n,7},
\end{aligned} \tag{7}$$

where

$$w_{ij}^N = \frac{\frac{1}{nh_2^p} K_{2h}\left(X_i - X_j\right)}{\frac{1}{nh_2^p}\sum_{i=1}^{n} K_{2h}\left(X_i - X_j\right)\mathbb{1}(D_i = 1)}, \quad \epsilon_{1i} = Y_i - E(Y_{(1)}|X_i).$$

Note that $I_{n,3}$ and $I_{n,4}$ in equation (7) yield the final expression in Theorem 3. Therefore, we need to show that $I_{n,5}$, $I_{n,6}$ and $I_{n,7}$ in equation (7) are all $o_p(1)$.

First show that $I_{n,5} = o_p(1)$. From Lemma 1,

$$\begin{aligned}
&\frac{1}{\sqrt{h_1^k}}\sup_i\left|\sum_{j=1}^{n} K_{1h}\left(X_{1j}\right)(w_{ij}^N - w_{ji}^N)\right| \le \frac{1}{\sqrt{h_1^k}}\sup_i\sum_{j:j\neq i}(w_{ij}^N - w_{ji}^N)\left|K_{1h}\left(X_{1j}\right)\right| \\
&\le \frac{MC}{h_2} \times \frac{h_2}{\sqrt{h_1^k}} \times \sup_i\sum_{j:j\neq i}\frac{1}{nh_2^p}\left|K_{2h}\left(X_i - X_j\right)\right| = O_p(1) \times o_p(1) \times O_p(1) = o_p(1),
\end{aligned}$$

Further, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{1i}\mathbb{1}(D_i=1)$ has finite limit and thus, is bounded by $O_p(1)$ and then $I_{n,5}=o_p(1)$.

Deal with $I_{n,6}$. As

$$\sum_{j=1}^{n}K_{1h}\left(X_{1j}\right)w_{ji}^{N}=\frac{\frac{1}{nh_2^p}\sum_{j=1}^{n}K_{2h}\left(X_j-X_i\right)}{\frac{1}{nh_2^p}\sum_{j=1}^{n}K_{2h}\left(X_j-X_i\right)\mathbb{1}(D_j=1)}\frac{\frac{1}{nh_2^p}\sum_{j=1}^{n}K_{2h}\left(X_j-X_i\right)K_{1h}\left(X_{1j}\right)}{\frac{1}{nh_2^p}\sum_{j=1}^{n}K_{2h}\left(X_j-X_i\right)},$$

we can then regard $\sum_{j=1}^{n}K_{1h}\left(X_{1j}\right)w_{ji}^{N}$ as an estimator of $\frac{K_{1h}(X_{1i})}{p(X_i)}$. Consider

$$\left[\sum_{j=1}^{n}K_{1h}\left(X_{1j}\right)w_{ji}^{N}-\frac{K_{1h}\left(X_{1i}\right)}{p(X_i)}\right],$$

which is the bias of $\frac{K_{1h}(X_{1i})}{p(X_i)}$ to $\frac{K_{1h}(X_{1i})}{p(X_i)}$. Write $X=(X_1,X_{(2)})$ and

$$K_{2h}\left(X-X_j\right)=K_{21}\left(\frac{X_1-X_{1j}}{h_2}\right)K_{22}\left(\frac{X_{(2)}-X_{2j}}{h_2}\right).$$

Since $\widehat{f}-f=o_p(1)$, and the kernel function is s^* $(\geq s_2)$ times continuously differentiable, we have

$$\begin{aligned}
&E\left\{\sum_{j=1}^{n}K_{1h}\left(X_{1j}\right)w_{ji}^{N}\middle|X_i\right\}\\
&=\frac{1+o_p(1)}{h_2^pf(X_i)p(X_i)}\int K_{21}\left(\frac{u_{1j}-X_{1i}}{h_2}\right)K_{22}\left(\frac{u_{2j}-X_{2i}}{h_2}\right)K_{1h}\left(u_{1j}\right)f(u_i)du\\
&=\frac{1+o_p(1)}{f(X_i)p(X_i)}\int K_{21}(v_1)K_{22}(v_2)K_1\left(\frac{X_{1i}-X_1}{h_1}+v_1\frac{h_2}{h_1}\right)f(X_i+h_2v)dv\\
&=\frac{K_{1h}\left(X_{1i}\right)}{p(X_i)}+O_p\left(\frac{h_2^{s_2}}{h_1^{s_2}}\right).
\end{aligned}\tag{8}$$

Note that

$$\begin{aligned}
&\frac{\widehat{K}_{1h}\left(X_{1i}\right)}{\widehat{p}(X_i)}-\frac{K_{1h}\left(X_{1i}\right)}{p(X_i)}\\
&=\left\{\frac{1}{\widehat{p}(X_i)}-\frac{1}{p(X_i)}+\frac{1}{p(X_i)}\right\}\left\{\widehat{K}_{1h}\left(X_{1i}\right)-K_{1h}\left(X_{1i}\right)+K_{1h}\left(X_{1i}\right)\right\}\\
&-\frac{K_{1h}\left(X_{1i}\right)}{p(X_i)}=\left\{\frac{1}{\widehat{p}(X_i)}-\frac{1}{p(X_i)}\right\}\left\{\widehat{K}_{1h}\left(X_{1i}\right)-K_{1h}\left(X_{1i}\right)\right\}\\
&+\frac{1}{p(X_i)}\left\{\widehat{K}_{1h}\left(X_{1i}\right)-K_{1h}\left(X_{1i}\right)\right\}+\left\{\frac{1}{\widehat{p}(X_i)}-\frac{1}{p(X_i)}\right\}K_{1h}\left(X_{1i}\right)\\
&=O_p\left(\frac{h_2^{s_2}}{h_1^{s_2}}+h_2^{s_2}+\sqrt{\frac{\log n}{nh_2^p}}\right)=O_p\left(\frac{h_2^{s_2}}{h_1^{s_2}}\right).
\end{aligned}$$

Thus, $\sup_i \left| \sum_{j=1}^{n} K_{1h}(X_{1j}) w_{ji}^N - \frac{K_{1h}(X_{1i})}{p(X_i)} \right| = O_p\left(\frac{h_2^{s_2}}{h_1^{s_2}}\right)$. Owing to assumption (A4) that $\frac{h_2^{2s_2}}{h_1^{2s_2+k}} \to 0$, we have

$$\sup_i \left| \frac{1}{\sqrt{h_1^k}} \left[\sum_{j=1}^{n} K_{1h}(X_{1j}) w_{ji}^N - \frac{K_{1h}(X_{1i})}{p(X_i)} \right] \right| = O_p\left(\frac{h_2^{2s_2}}{h_1^{2s_2+k}}\right) = o_p(1).$$

Since $\epsilon_{1i} = Y_i - E(Y_{(1)}|X_i)$ are mutually independent, we have $I_{n,6} = o_p(1)$ in equation (7). Finally, to show that $I_{n,7} = o_p(1)$ of equation (7). Note that

$$\begin{aligned}
&\frac{\frac{1}{nh_2^p} \sum_{j=1}^{n} K_{2h}(X_j - X_i) \mathbb{1}(D_j = 1) EY_{(1)}|X_j}{\frac{1}{nh_2^p} \sum_{j=1}^{n} K_{2h}(X_j - X_i) \mathbb{1}(D_j = 1)} \\
&= \frac{\frac{1}{nh_2^p} \sum_{j=1}^{n} K_{2h}(X_j - X_i) \mathbb{1}(D_j = 1) EY_{(1)}|X_j}{\frac{1}{nh_2^p} \sum_{j=1}^{n} K_{2h}(X_j - X_i)} \cdot \frac{\frac{1}{nh_2^p} \sum_{j=1}^{n} K_{2h}(X_j - X_i)}{\frac{1}{nh_2^p} \sum_{j=1}^{n} K_{2h}(X_j - X_i) \mathbb{1}(D_j = 1)},
\end{aligned}$$

which can be viewed as an estimator of $\frac{E\{\mathbb{1}(D=1)Y_{(1)}|X_i\}}{p(X_i)}$. Denote $A(X_i) = E\{\mathbb{1}(D = 1)Y_{(1)}|X_i\}$. We can derive easily that

$$\begin{aligned}
&\frac{\widehat{A}(X_i)}{\widehat{p}(X_i)} - \frac{A(X_i)}{p(X_i)} \\
&= \left\{\widehat{A}(X_i) - A(X_i) + A(X_i)\right\} \left\{\frac{1}{\widehat{p}(X_i)} - \frac{1}{p(X_i)} + \frac{1}{p(X_i)}\right\} - \frac{A(X_i)}{p(X_i)} \\
&= \left\{\widehat{A}(X_i) - A(X_i)\right\} \left\{\frac{1}{\widehat{p}(X_i)} - \frac{1}{p(X_i)}\right\} + A(X_i) \left\{\frac{1}{\widehat{p}(X_i)} - \frac{1}{p(X_i)}\right\} \\
&\quad + \left\{\widehat{A}(X_i) - A(X_i)\right\} \frac{1}{p(X_i)} = O_p\left(h_2^{s_2} + \sqrt{\frac{\log n}{nh_2^p}}\right).
\end{aligned}$$

Thus

$$\sup_i \left| \frac{\frac{1}{nh_2^p} \sum_{j=1}^{n} K_{2h}(X_j - X_i) \mathbb{1}(D_j = 1) EY_{(1)}|X_j}{\frac{1}{nh_2^p} \sum_{j=1}^{n} K_{2h}(X_j - X_i) \mathbb{1}(D_j = 1)} - EY_{(1)}|X_i \right| = O_p\left(h_2^{s_2} + \sqrt{\frac{\log n}{nh_2^p}}\right).$$

Then, we can bound $I_{n,7}$ as follows:

$$
\begin{aligned}
|I_{n,7}| &= \left| \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n K_{1h}(X_{1i}) \left[\frac{\frac{1}{nh_2^p}\sum_{j=1}^n K_{2h}(X_j - X_i)\mathbb{1}(D_j=1)EY_{(1)}|X_j}{\frac{1}{nh_2^p}\sum_{j=1}^n K_{2h}(X_j - X_i)\mathbb{1}(D_j=1)} - EY_{(1)}|X_i \right] \right| \\
&\le \sqrt{nh_1^k} \sup_i \left| \frac{\frac{1}{nh_2^p}\sum_{j=1}^n K_{2h}(X_j - X_i)\mathbb{1}(D_j=1)EY_{(1)}|X_j}{\frac{1}{nh_2^p}\sum_{j=1}^n K_{2h}(X_j - X_i)\mathbb{1}(D_j=1)} - EY_{(1)}|X_i \right| \frac{1}{nh_1^k}\sum_{i=1}^n |K_{1h}(X_{1i})| \\
&= \sqrt{nh_1^k} O_p\left(h_2^{s_2} + \sqrt{\frac{\log n}{nh_2^p}} \right) \cdot O_p(1) = o_p(1) \cdot O_p(1) = o_p(1),
\end{aligned}
$$

where assumption (A4) is used for the second equation. Thus, together with $I_{n,5} = o_p(1)$, $I_{n,6} = o_p(1)$ and $I_{n,7} = o_p(1)$, equation (7) becomes

$$
\begin{aligned}
&\frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n K_{1h}(X_{1i}) \left[\frac{\frac{1}{nh_2^p}\sum_{j=1}^n K_{2h}(X_j - X_i) Y_{1j}\mathbb{1}(D_j=1)}{\frac{1}{nh_2^p}\sum_{j=1}^n K_{2h}(X_j - X_i)\mathbb{1}(D_j=1)} - \tau_1(x_1) \right] \\
&= I_{n,3} + I_{n,4} + o_p(1).
\end{aligned}
$$

Similarly, we can also deal with $\widehat{m}_0(X_i) - \tau_0(x_1)$ of (6) to have

$$
\begin{aligned}
&\frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n K_{1h}(X_{1i}) \left[\frac{\frac{1}{nh_2^p}\sum_{j=1}^n K_{2h}(X_j - X_i) Y_{0j}\mathbb{1}(D_j=0)}{\frac{1}{nh_2^p}\sum_{j=1}^n K_{2h}(X_j - X_i)\mathbb{1}(D_j=0)} - \tau_1(x_1) \right] \\
&:= I_{n,8} + I_{n,9} + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
&I_{n,8} = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n \epsilon_{0i}\mathbb{1}(D_i=0)\frac{K_{1h}(X_{1i})}{1-p(X_i)}, \quad I_{n,9} = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n K_{1h}(X_{1i})\, EY_{(0)}|X_i, \\
&\epsilon_{0i} = Y_i - EY_{(0)}|X_i.
\end{aligned}
$$

Hence, we get the asymptotic linear representation of $\widehat{\tau}(x_1)$ as

$$
\sqrt{nh_1^k}\{\widehat{\tau}(x_1) - \tau(x_1)\} = \frac{1}{\sqrt{nh_1^k}} \frac{1}{f(x_1)} \sum_{i=1}^n \{\Psi_1(X_i, Y_i, D_i) - \tau(x_1)\} K_{1h}(X_{1i}) + o_p(1),
$$

which can be asymptotically normal. Again, we compute its asymptotic variance. Similarly as the proof for Theorem 2, we have

$$
\mathrm{Var}\{\widehat{\tau}(x_1)\} = \frac{1}{nh_1^k}\frac{||K_1||_2^2\sigma_N^2(x_1)}{f(x_1)} + o\left(\frac{1}{nh_1^k}\right).
$$

Then by assumptions (C1)– (C4) and (A1) – (A4) for some $s^* \ge s_2 \ge p$, we can derive that

$$
\sqrt{nh_1^k}\,\{\widehat{\tau}(x_1) - \tau(x_1)\} \xrightarrow{d} N\left(0, \frac{||K_1||_2^2\sigma_N^2(x_1)}{f(x_1)}\right),
$$

where

$$\sigma_N^2(x_1) \equiv E[\{\Psi_1(X,Y,D) - \tau(x_1)\}^2 | X_1 = x_1].$$

The proof is concluded. □

PROOF OF THEOREM 4. Inspired by the proof of Theorem 2 of Luo, Zhu and Ghosh (2017), we have

$$\begin{aligned}
&\sqrt{nh_1^k}(\widehat{\tau}(x_1) - \tau(x_1)) \\
&= \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}(X_{1i})\left[\widehat{m}_1(\widehat{\boldsymbol{\beta}}_1^\top X) - \tau_1(x_1)\right]}{\frac{1}{nh_1^k}\sum_{i=1}^n K_{1h}(X_{1i})} - \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}(X_{1i})\left[\widehat{m}_0(\widehat{\boldsymbol{\beta}}_0^\top X) - \tau_0(x_1)\right]}{\frac{1}{nh_1^k}\sum_{i=1}^n K_{1h}(X_{1i})} \\
&= \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}(X_{1i})\left[\widehat{m}_1(\boldsymbol{\beta}_1^\top X) - \tau_1(x_1)\right]}{\frac{1}{nh_1^k}\sum_{i=1}^n K_{1h}(X_{1i})} - \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}(X_{1i})\left[\widehat{m}_0(\boldsymbol{\beta}_0^\top X) - \tau_0(x_1)\right]}{\frac{1}{nh_1^k}\sum_{i=1}^n K_{1h}(X_{1i})} \\
&\quad + O_p(\sqrt{nh_1^k}||\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1|| + \sqrt{nh_1^k}||\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0||),
\end{aligned} \tag{9}$$

where

$$\widehat{m}_1(\widehat{\boldsymbol{\beta}}_1^\top X) = \frac{\frac{1}{nh_4^{r(1)}}\sum_{j=1}^n K_{4h}\left(\widehat{Z}_j^1 - \widehat{Z}_i^1\right) Y_{1j}\mathbb{1}(D_j = 1)}{\frac{1}{nh_4^{r(1)}}\sum_{j=1}^n K_{4h}\left(\widehat{Z}_j^1 - \widehat{Z}_i^1\right)\mathbb{1}(D_j = 1)}, \quad \widehat{Z}^1 = \widehat{\boldsymbol{\beta}}_1^\top X,$$

$$\widehat{m}_0(\widehat{\boldsymbol{\beta}}_0^\top X) = \frac{\frac{1}{nh_4^{r(1)}}\sum_{j=1}^n K_{4h}\left(\widehat{Z}_j^0 - \widehat{Z}_i^0\right) Y_{0j}\mathbb{1}(D_j = 0)}{\frac{1}{nh_4^{r(1)}}\sum_{j=1}^n K_{4h}\left(\widehat{Z}_j^0 - \widehat{Z}_i^0\right)\mathbb{1}(D_j = 0)}, \quad \widehat{Z}^0 = \widehat{\boldsymbol{\beta}}_0^\top X.$$

Under assumptions (A8), $O_p(\sqrt{nh_1^k}||\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1|| + \sqrt{nh_1^k}||\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0||) = O_p(\sqrt{h_1^k}) = o_p(1)$ as $h_1 \to 0$. Therefore, equation (9) becomes

$$\begin{aligned}
&\sqrt{nh_1^k}(\widehat{\tau}(x_1) - \tau(x_1)) \\
&= \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}(X_{1i})\left[\widehat{m}_1(\boldsymbol{\beta}_1^\top X) - \tau_1(x_1)\right]}{\frac{1}{nh_1^k}\sum_{i=1}^n K_{1h}(X_{1i})} - \frac{\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}(X_{1i})\left[\widehat{m}_0(\boldsymbol{\beta}_0^\top X) - \tau_0(x_1)\right]}{\frac{1}{nh_1^k}\sum_{i=1}^n K_{1h}(X_{1i})} \\
&\quad + o_p(1).
\end{aligned} \tag{10}$$

Similarly as the proof for Theorem 3, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}\left(X_{1i}\right)\left[\widehat{m}_1(\boldsymbol{\beta}_1^\top X)-\tau_1(x_1)\right]\\
=&\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}\left(X_{1i}\right)\left[EY_{(1)}|X_i-\tau_1(x_1)\right]+\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n \epsilon_{1i}\mathbb{1}(D_i=1)\sum_{j=1}^n K_{1h}\left(X_{1j}\right)(w_{ij}^{S_1}-w_{ji}^{S_1})\\
&+\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n \epsilon_{1i}\mathbb{1}(D_i=1)\sum_{j=1}^n K_{1h}\left(X_{1j}\right)w_{ji}^{S_1}\\
&+\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}\left(X_{1i}\right)\left[\frac{\frac{1}{nh_4^{r(1)}}\sum_{j=1}^n K_{4h}\left(Z_j^1-Z_i^1\right)\mathbb{1}(D_j=1)EY_{(1)}|X_j}{\frac{1}{nh_4^{r(1)}}\sum_{j=1}^n K_{4h}\left(Z_j^1-Z_i^1\right)\mathbb{1}(D_j=1)}-EY_{(1)}|X_i\right]\\
=:&\, I_{n,10}+I_{n,11}+I_{n,12}+I_{n,13},
\end{aligned}
$$

where

$$
w_{ij}^{S_1}=\frac{\frac{1}{nh_4^{r(1)}}K_{4h}\left(Z_i^1-Z_j^1\right)}{\frac{1}{nh_4^{r(1)}}\sum_{i=1}^n K_{4h}\left(Z_i^1-Z_j^1\right)\mathbb{1}(D_i=1)},\quad \epsilon_{1i}=Y_i-EY_{(1)}|X_i.
$$

Similarly, we can decompose $\widehat{m}_0(\boldsymbol{\beta}_0^\top X)-\tau_0(x_1)$ as

$$
\begin{aligned}
&\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}\left(X_{1i}\right)\left[\widehat{m}_0(\boldsymbol{\beta}_0^\top X)-\tau_0(x_1)\right]\\
=&\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}\left(X_{1i}\right)\left[EY_{(0)}|X_i-\tau_0(x_1)\right]+\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n \epsilon_{0i}\mathbb{1}(D_i=0)\sum_{j=1}^n K_{1h}\left(X_{1j}\right)(w_{ij}^{S_2}-w_{ji}^{S_2})\\
&+\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n \epsilon_{0i}\mathbb{1}(D_i=0)\sum_{j=1}^n K_{1h}\left(X_{1j}\right)w_{ji}^{S_2}\\
&+\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n K_{1h}\left(X_{1i}\right)\left[\frac{\frac{1}{nh_4^{r(0)}}\sum_{j=1}^n K_{4h}\left(Z_j^0-Z_i^0\right)\mathbb{1}(D_j=0)EY_{(0)}|X_j}{\frac{1}{nh_4^{r(0)}}\sum_{j=1}^n K_{4h}\left(Z_j^0-Z_i^0\right)\mathbb{1}(D_j=0)}-EY_{(0)}|X_i\right]\\
=:&\, I'_{n,10}+I'_{n,11}+I'_{n,12}+I'_{n,13},
\end{aligned}
$$

where

$$
w_{ij}^{S_2}=\frac{\frac{1}{nh_4^{r(0)}}K_{4h}\left(Z_i^0-Z_j^0\right)}{\frac{1}{nh_4^{r(0)}}\sum_{i=1}^n K_{4h}\left(Z_i^0-Z_j^0\right)\mathbb{1}(D_i=0)},\quad \epsilon_{0i}=Y_i-EY_{(0)}|X_i.
$$

It is easy to show that $I_{n,11}$, $I'_{n,11}$, $I_{n,13}$ and $I'_{n,13}$ are $o_p(1)$ following the same arguments for proving that $I_{n,5}=o_p(1)$ and $I_{n,7}=o_p(1)$ for Theorem 3. The details are omitted here. We now deal with $I_{n,12}$ and $I'_{n,12}$.

Lemma 2 *Suppose assumptions (C1) – (C4), (A1) and (A5) – (A7) are satisfied. Then, for each point x_1 in the support of X_1,*
(1) If $X_1 \subset^{k-q} \boldsymbol{\beta}_1^\top X$ and $X_1 \subset^{k-q} \boldsymbol{\beta}_0^\top X$ with $s_4(2-k/q)+k>0$ and $0<q\le k$, we have

$$I_{n,12} = o_p(1), \quad I'_{n,12} = o_p(1). \tag{11}$$

The corresponding asymptotically linear representation is then

$$\sqrt{nh_1^k}\{\widehat{\tau}(x_1)-\tau(x_1)\} = \frac{1}{\sqrt{nh_1^k}}\frac{1}{f(x_1)}\sum_{i=1}^n\{m_1(X_i)-m_0(X_i)-\tau(x_1)\}K_{1h}(X_{1i})+o_p(1).$$

(2) If $X_1 \subset \boldsymbol{\beta}_1^\top X$ and $X_1 \subset^{k-q} \boldsymbol{\beta}_0^\top X$ with $s_4(2-k/q)+k>0$ and $0<q\le k$, we have

$$I_{n,12} = \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n \epsilon_{1i}\mathbb{1}(D_i=1)\frac{K_{1h}(X_{1i})}{p(X_i)}+o_p(1), \quad I'_{n,12}=o_p(1). \tag{12}$$

Then we have

$$\sqrt{nh_1^k}\{\widehat{\tau}(x_1)-\tau(x_1)\} = \frac{1}{\sqrt{nh_1^k}}\frac{1}{f(x_1)}\sum_{i=1}^n\{\Psi_2(X_i,Y_i,D_i)-\tau(x_1)\}K_{1h}(X_{1i})+o_p(1).$$

(3) If $X_1 \subset^{k-q} \boldsymbol{\beta}_1^\top X$ and $X_1 \subset \boldsymbol{\beta}_0^\top X$ with $s_4(2-k/q)+k>0$ and $0<q\le k$, we have

$$I_{n,12} = o_p(1), \quad I'_{n,12} = \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n \epsilon_{0i}\mathbb{1}(D_i=0)\frac{K_{1h}(X_{1i})}{p(X_i)}+o_p(1). \tag{13}$$

The corresponding asymptotically linear representation is

$$\sqrt{nh_1^k}\{\widehat{\tau}(x_1)-\tau(x_1)\} = \frac{1}{\sqrt{nh_1^k}}\frac{1}{f(x_1)}\sum_{i=1}^n\{\Psi_3(X_i,Y_i,D_i)-\tau(x_1)\}K_{1h}(X_{1i})+o_p(1).$$

(4) If $X_1 \subset \boldsymbol{\beta}_1^\top X$ and $X_1 \subset \boldsymbol{\beta}_0^\top X$, we have

$$I_{n,12} = \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n \epsilon_{1i}\mathbb{1}(D_i=1)\frac{K_{1h}(X_{1i})}{p(X_i)}+o_p(1), \quad I'_{n,12} = \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^n \epsilon_{0i}\mathbb{1}(D_i=0)\frac{K_{1h}(X_{1i})}{p(X_i)}+o_p(1). \tag{14}$$

We have

$$\sqrt{nh_1^k}\{\widehat{\tau}(x_1)-\tau(x_1)\} = \frac{1}{\sqrt{nh_1^k}}\frac{1}{f(x_1)}\sum_{i=1}^n\{\Psi_4(X_i,Y_i,D_i)-\tau(x_1)\}K_{1h}(X_{1i})+o_p(1).$$

PROOF OF LEMMA 2. We need to show that $I_{n,12} = o_p(1)$ if $X_1 \subset^{k-q} \boldsymbol{\beta}_1^\top X$ with $s_4(2-k/q)+k>0$ and $0<q\le k$. Let $X_1 = v_1$, $\boldsymbol{\beta}_1^\top X = v_2$, and denote $\left(\frac{v_1-v_{1i}}{h_4}, \frac{v_2-v_{2i}}{h_4}\right)$ as

(t_1, t_2). We have

$$
\begin{aligned}
& E\left\{\sum_{j=1}^{n} K_{1h}\left(X_{1j}\right) w_{ji}^{S_1} \middle| X_i\right\} \\
= & \frac{1+o_p(1)}{h_4^{r(1)} f(v_{2i})p(v_{2i})} \int K_4\left(\frac{v_{2j}-\boldsymbol{\beta}_1^{\top} X_i}{h_4}\right) K_1\left(\frac{v_{1j}-X_1}{h_1}\right) f(v_i)dv \\
= & h_4 \frac{1+o_p(1)}{f(v_{2i})p(v_{2i})} \int K_4(t_2) K_1\left(\frac{v_{1i}-X_1}{h_1}+t_1\frac{h_4}{h_1}\right) f_{12}(v_{1i}+h_4t_1, v_{2i}+h_4t_2)dt_1dt_2 \\
= & h_4^q K_1\left(\frac{v_{1i}-X_1}{h_1}\right) \frac{f_{12}(v_{1i},v_{2i})}{f(v_{2i})p(v_{2i})} \int K_4(t_2)dt_1dt_2 \\
& + \frac{h_4^{q+1}}{h_1} K_1'\left(\frac{v_{1i}-X_1}{h_1}\right) \frac{f_{12}(v_{1i},v_{2i})}{f(v_{2i})p(v_{2i})} \int t_1 K_4(t_2)dt_1dt_2 + o_p\left(\frac{h_4^2}{h_1}\right),
\end{aligned}
$$

where $f_{12}(v_{1i}, v_{2i})$ is the joint density function of $(X_1, \boldsymbol{\beta}_1^{\top} X)$. Under assumptions (A5) – (A7), we have

$$
E\left\{\sum_{j=1}^{n} K_{1h}\left(X_{1j}\right) w_{ji}^{S_1} \middle| X_i\right\} = C_2 h_4^q K_{1h}\left(X_{1i}\right) \frac{f_{12}(X_{1i}, \boldsymbol{\beta}_1^{\top} X_i)}{f(X_i)p(\boldsymbol{\beta}_1^{\top} X_i)} + O_p\left(\frac{h_4^{q+1}}{h_1}\right) = O_p\left(h_4^q + \frac{h_4^{q+1}}{h_1}\right).
$$

Hence, under assumptions (A6), (A7), $s_4(2-k/q)+k>0$ and $0<q\le k$,

$$
\frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \epsilon_{1i} \mathbb{1}(D_i=1) \sum_{j=1}^{n} K_{1h}\left(X_{1j}\right) w_{ji}^{S_1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_{1i} \mathbb{1}(D_i=1) O_p\left(\frac{h_4^q}{h_1^{k/2}} + \frac{h_4^{q+1}}{h_1^{k/2+1}}\right) = o_p(1).
$$

Analogously, we get $I_{n,12}' = o_p(1)$ if $X_1 \subset^{k-q} \boldsymbol{\beta}_0^{\top} X$. Next, we prove that

$$
I_{n,12} = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \epsilon_{1i} \mathbb{1}(D_i=1) \frac{K_{1h}\left(X_{1i}\right)}{p(X_i)} + o_p(1),
$$

if $X_1 \subset \boldsymbol{\beta}_1^{\top} X$. As that case that $X_1 \subset \boldsymbol{\beta}_1^{\top} X$ is similar to that $X_1 \subset X$ in nonparametric case, then parallelling to derive equation (8), we get the desired result. Similarly, we have $I_{n,12}' = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \epsilon_{0i} \mathbb{1}(D_i=0) \frac{K_{1h}(X_{1i})}{p(X_i)} + o_p(1)$ if $X_1 \subset \boldsymbol{\beta}_0^{\top} X$. The proof for Lemma 2 is concluded.

Proof of Corollary 2. Consider the case where $X_1 \not\subset \widetilde{X} \in \mathbb{R}^q$. Similarly as before, we derive that

$$
\begin{aligned}
& \sqrt{nh_1^k}(\widehat{\tau}(x_1)-\tau(x_1)) \\
= & \frac{\frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[\widehat{m}_1(\widetilde{X}_i)-\tau_1(x_1)\right]}{\frac{1}{nh_1^k} \sum_{i=1}^{n} K_{1h}(X_{1i})} - \frac{\frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[\widehat{m}_0(\widetilde{X}_i)-\tau_0(x_1)\right]}{\frac{1}{nh_1^k} \sum_{i=1}^{n} K_{1h}(X_{1i})},
\end{aligned}
\tag{15}
$$

where

$$
\widehat{m}_1(\widetilde{X}_i) = \frac{\frac{1}{nh_2^q} \sum_{j=1}^{n} K_{2h}\left(\widetilde{X}_j-\widetilde{X}_i\right) Y_{1j} \mathbb{1}(D_j=1)}{\frac{1}{nh_2^q} \sum_{j=1}^{n} K_{2h}\left(\widetilde{X}_j-\widetilde{X}_i\right) \mathbb{1}(D_j=1)}, \quad \widehat{m}_0(\widetilde{X}_i) = \frac{\frac{1}{nh_2^q} \sum_{j=1}^{n} K_{2h}\left(\widetilde{X}_j-\widetilde{X}_i\right) Y_{0j} \mathbb{1}(D_j=0)}{\frac{1}{nh_2^q} \sum_{j=1}^{n} K_{2h}\left(\widetilde{X}_j-\widetilde{X}_i\right) \mathbb{1}(D_j=0)}.
$$

Some similar calculations lead to $\widehat{m}_1(\widetilde{X}_i) - \tau_1(x_1)$.

$$
\begin{aligned}
&\frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[\widehat{m}_1(\widetilde{X}_i) - \tau_1(x_1)\right] \\
=& \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[EY_{(1)}|X_i - \tau_1(x_1)\right] + \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n}\epsilon_{1i}\mathbb{1}(D_i=1)\sum_{j=1}^{n}K_{1h}\left(X_{1j}\right)\left(w_{ij}^{N_1} - w_{ji}^{N_1}\right) \\
&+ \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n}\epsilon_{1i}\mathbb{1}(D_i=1)\sum_{j=1}^{n}K_{1h}\left(X_{1j}\right)w_{ji}^{N_1} \\
&+ \frac{1}{\sqrt{nh_1^k}}\sum_{i=1}^{n} K_{1h}\left(X_{1i}\right)\left[\frac{\frac{1}{nh_2^q}\sum_{j=1}^{n}K_{2h}\left(\widetilde{X}_j - \widetilde{X}_i\right)\mathbb{1}(D_j=1)EY_{(1)}|X_j}{\frac{1}{nh_2^q}\sum_{j=1}^{n}K_{2h}\left(\widetilde{X}_j - \widetilde{X}_i\right)\mathbb{1}(D_j=1)} - EY_{(1)}|X_i\right] \\
=:& I_{n,14} + I_{n,15} + I_{n,16} + I_{n,17},
\end{aligned}
$$

where

$$
w_{ij}^{N_1} = \frac{\frac{1}{nh_2^q}K_{2h}\left(\widetilde{X}_i - \widetilde{X}_j\right)}{\frac{1}{nh_2^q}\sum_{i=1}^{n}K_{2h}\left(\widetilde{X}_i - \widetilde{X}_j\right)\mathbb{1}(D_i=1)}.
$$

Then we can prove that $I_{n,15}$ and $I_{n,17}$ are $o_p(1)$ by the same arguments as those used to handle $I_{n,5}$ and $I_{n,7}$ for proving Theorem 3. Owing to $X_1 \not\subset \widetilde{X}$, similar arguments for proving Lemma 2 implies that $I_{n,16} = o_p(1)$. The proof for Corollary 1 is concluded. □

Proof of Corollary 3. From the proof for Theorem 3, we can see that

$$
E\left\{\sum_{j=1}^{n}K_{1h}\left(X_{1j}\right)w_{ji}^{N}\middle| X_i\right\} = O_p\left(h_2 + \frac{h_2^{s_2}}{h_1^{s_2}}\right),
$$

by the condition $\sqrt{nh_1^k}\left(h_2^s + \sqrt{\log(n)/nh_2^p}\right) = o(1)$. Then OR-N shares the same asymptotic distribution as OR-P. For OR-S, we can use similar arguments to show the same result. The proof is finished. □

References

Abrevaya, J., Hsu, Y. C., & Lieli, R. P. (2015). Estimating conditional average treatment effects. *Journal of Business & Economic Statistics* **33**, 485-505.

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association* **89**, 81-87.

Cook, R. D., & Li, B. (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics* **30**, 455-474.

Fan, Q. , Hsu, Y.C. , Lieli, R.P., & Zhang, Y.(2020). Estimation of conditional average treatment effects with high-dimensional data. *Journal of Business & Economic Statistics*, Online.

Feng, Z., Wen, X. M., Yu, Z., & Zhu, L. (2013). On partial sufficient dimension reduction with applications to partially linear multi-index models. *Journal of the American Statistical Association* **108**, 237-246.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66**, 315-331.

HAMMER, S. M., KATZENSTEIN, D. A., HUGHES, M. D., GUNDACKER, H., SCHOOLEY, R. T., HAUBRICH, R. H., ... & HIRSCH, M. S. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. New England Journal of Medicine. *Econometrica* **335**, 1081-1090.

HEALY, M., & WESTMACOTT, M. (1956). Missing values in experiments analysed on automatic computers. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **5**, 203-206.

HIRANO, K., IMBENS, G. W., & RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71**, 1161-1189.

LI, Q., & RACINE, J. S. (2007). Nonparametric econometrics: theory and practice. *Princeton University Press.*

LUO, W., WU, W., & ZHU, Y. (2019). Learning Heterogeneity in Causal Inference Using Sufficient Dimension Reduction. Published online.

LUO, W., ZHU, Y., & GHOSH, D (2017). On estimating regression-based causal effects using sufficient dimension reduction. *Biometrika* **104**, 51-65.

MA, S., ZHU, L., ZHANG, Z., TSAI, C. L., & CARROLL, R. J. (2019). A robust and efficient approach to causal inference based on sparse sufficient dimension reduction. *The Annals of Statistics* **47**, 1505-1535.

MATLOFF, N. S. (1981). Use of regression functions for improved estimation of means. *Biometrika* **68**, 685-689.

NADARAYA, E. A. (1964). On estimating regression. *Theory of Probability* & *Its Applications* **9)**, 141-142.

PAGAN, A., & ULLAH, A. (1999). Nonparametric econometrics. *Cambridge university press*

RAO, J. N. K. (1996). On variance estimation with imputed survey data. *Journal of the American Statistical Association* **91**, 499-506.

ROSENBAUM, P. R., & RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41-55.

ROSENBAUM, P. R., & RUBIN, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* **39**, 33-38.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* **66**, 688.

SONG, X., & MA, S. (2008). Multiple augmentation for interval censored data with measurement error. *Statistics in medicine* **27**, 3178-3190.

WANG, Q., LINTON, O. & HÄRDLE, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association* **99**, 334-345.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A* **26**, 359-372.

YIN, J., GENG, Z., LI, R. & WANG, H. (2010). Nonparametric covariance model. *Statistica Sinica* **20**, 469-479.

XIA, Y., LI, W.K., TONG, H. & ZHU, L. X. (2002). An adaptive estimation of dimension reduction space. *Journal of Royal Statistical Society, Series B*, **64**, 363-410.

ZHOU, N. W. & ZHU L. X. (2019). Estimation of conditional average treatment effect. Submitted.