DOUBLY ROBUST ESTIMATION FOR CONDITIONAL TREATMENT EFFECT: A STUDY ON ASYMPTOTICS

Chuyun Ye¹, Keli Guo² and Lixing Zhu¹,²

¹ Beijing Normal University, Beijing, China
² Hong Kong Baptist University, Hong Kong

Abstract: In this paper, we apply doubly robust approach to estimate, when some covariates are given, the conditional average treatment effect under parametric, semiparametric and nonparametric structure of the nuisance propensity score and outcome regression models. We then conduct a systematic study on the asymptotic distributions of nine estimators with different combinations of estimated propensity score and outcome regressions. The study covers the asymptotic properties with all models correctly specified; with either propensity score or outcome regressions locally / globally misspecified; and with all models locally / globally misspecified. The asymptotic variances are compared and the asymptotic bias correction under model-misspecification is discussed. The phenomenon that the asymptotic variance, with model-misspecification, could sometimes be even smaller than that with all models correctly specified is explored. We also conduct a numerical study to examine the theoretical results.
1. Introduction

To explore the heterogeneity of treatment effect under Rubin’s potential outcome framework (Rosenbaum and Rubin (1983)) to reveal the causality of a treatment, conditional average treatment effect (CATE) is useful, which is conditional on some covariates of interest. See Abrevaya et al. (2015) as an example. Shi et al. (2019) showed that the existence of optimal individualized treatment regime (OITR) has a close connection with CATE.

To estimate CATE, there are some standard approaches available in the literature. When either propensity score function or outcome regression functions or both are unknown, we need to estimate them first such that we can then estimate the CATE function. Regard these functions as nuisance models. Abrevaya et al. (2015) used the propensity score-based (PS-based) estimation under parametric (P-IPW) and nonparametric structure (N-IPW), and showed that N-IPW is asymptotically more efficient than P-IPW. Zhou and Zhu (2020) suggested the PS-based estimation under a semiparametric dimension reduction structure (S-IPW) to show the advantage of semiparametric estimation and Li et al. (2020) considered outcome
regression-based (OR-based) estimation under parametric (P-OR), semi-
parametric (S-OR) and nonparametric structure (N-OR) to derive their
asymptotic properties and suggested also the use of semiparametric method.
Both of the works together give an estimation efficiency comparison between
PS-based and OR-based estimators. A clear asymptotic efficiency ranking
was shown by Li et al. (2020) when the propensity score and outcome re-
gression models are all correctly specified and the underlying nonparametric
models is sufficiently smooth such that, with delicately selecting bandwidths
and kernel functions, the nonparametric estimation can achieve sufficiently
fast rates of convergence:

\[
\begin{align*}
\text{OR-based estimators} & : \quad O-OR \preceq P-OR \preceq S-OR \preceq N-OR \\
\text{PS-based CATE estimators} & : \quad N-IPW \preceq S-IPW \preceq P-IPW \cong O-IPW
\end{align*}
\]

where \( A \preceq B \) denotes the asymptotic efficiency advantage, with smaller
variance, of \( A \) over \( B \), \( A \cong B \) the efficiency equivalence and O-OR and
O-IPW stand for OR-based and PS-based estimator respectively assuming
the nuisance models are known with no need to estimate.

As well known, the doubly robust (DR) method that was first sug-
gested as the augmented inverse probability weighting (AIPW) estimation
proposed by Robins et al. (1994). Later developments provide the estima-
tion consistency (Scharfstein et al. (1999)) for more general doubly robust
estimation, not restricted to AIPW, that even has one misspecified in the
two involved models. For further discussion and introduction on DR estimation, readers can refer to, as an example, Seaman and Vansteelandt (2018). Like Abrevaya et al. (2015), Lee et al. (2017) brought up a two-step AIPW estimator of CATE also under parametric structure. For the cases with high-dimensional covariate, Fan et al. (2019) and Zimmert and Lechner (2019) combined such an estimator with statistical learning.

In the current paper, we focus on investigating the asymptotic efficiency comparisons among nine doubly robust estimators under parametric, semiparametric dimension reduction and nonparametric structure. To this end, we will give a systematic study to provide insight into which combinations may have merit in an asymptotic sense and in practice, which ones would be worth of recommendation for use. We also further consider the asymptotic efficiency when nuisance models are globally or locally misspecified, which will be defined later. Roughly speaking, local misspecification means that misspecified model can converge, at a certain rate, to the corresponding correctly specified model as the sample size $n$ goes to infinity, while globally misspecified model cannot. Denote $c_n$, $d_{1n}$ and $d_{0n}$ respectively the departure degrees of used models to the corresponding correctly specified models, and $V_i(x_1)$ for $i = 1, 2, 3, 4$, which will be clarified in Theorems 1, 2, 3 and 5 respectively, of the asymptotic variance functions of $x_1$ for all nine
estimators in difference scenarios. Here $V_1(x_1)$ is the asymptotic variance when all models are correctly specified, which is regarded as a benchmark for comparisons. We have that $V_1(x_1) \leq V_3(x_1)$, but $V_2(x_1)$ and $V_4(x_1)$ are not necessarily larger than $V_1(x_1)$. Here we display main findings in this paper.

- When all nuisance models are correctly specified, and the tuning parameters including the bandwidths in nonparametric estimations are delicately selected, the asymptotic variances are all equal to $V_1(x_1)$. Write all DR estimators as $DRCATE$. Together with (1.1), the asymptotic efficiency ranking is as:

$$
\begin{align*}
\text{OR-based estimators} & : O-OR \preceq P-OR \preceq S-OR \preceq N-OR \equiv DRCATE \equiv N-IPW \preceq S-IPW \preceq P-IPW \equiv O-IPW \\
\text{PS-based CATE estimators} & : O-IPW \preceq P-IPW \preceq S-IPW \preceq N-IPW \equiv DRCATE \equiv S-OR \preceq P-OR \preceq O-OR
\end{align*}
$$

- If only one of the nuisance models, either propensity score or outcome regressions, is (are) misspecified, the estimators remain unbiased as expectably. But globally misspecified outcome regressions or propensity score lead to asymptotic variance changes. We can give examples of propensity score to show that the variance can be even smaller than that with correctly specified models. Further, when the nuisance models are locally misspecified, the asymptotic efficiency remains the same as that with no misspecification.
Further, when all nuisance models are globally misspecified, we need to take care of estimation bias. When the misspecifications are all local, but the convergence rates $c_n d_{1n}$ and $c_n d_{0n}$ are all faster than the convergence rate of nonparametric estimation that will be specified later, the asymptotic distributions remain unchanged.

To give a quick access to the results about the asymptotic variances, we present a summary in Table 1. Denote $PS(P)$, $PS(N)$ and $PS(S)$ as estimators with parametrically, nonparametrically and semiparametrically estimated PS function respectively, $OR(P)$, $OR(N)$ and $OR(S)$ as estimators with parametrically, nonparametrically and semiparametrically estimated OR functions respectively. Dark cells mean no such combinations.

The remaining parts of this article are organized as follows. We first describe the Rubin’s potential outcome framework and the relevant notations in Section 2. Section 3 contains a general two-step estimation of CATE, while Section 4 describes the corresponding asymptotic properties under different situations. Section 5 presents the results of Monte Carlo simulations and Section 6 includes some concluding remarks. We would like to point out that such comparisons do not mean the estimations that are of asymptotic efficiency advantage are always worthwhile to recommend be-
Table 1: Asymptotic variance result summary

| Combination | All Correctly specified | Globally Misspecified PS | Locally Misspecified PS | Globally Misspecified OR | Locally Misspecified OR |
|-------------|-------------------------|--------------------------|-------------------------|--------------------------|-------------------------|
| \(PS(P) + OR(P)\) | \(V_1(x_1)\) | \(V_2(x_1)\) (Not necessarily enlarged) | \(V_1(x_1)\) | \(V_2(x_1)\) (Enlarged) | \(V_1(x_1)\) |
| \(PS(P) + OR(N)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) |
| \(PS(N) + OR(P)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) |
| \(PS(N) + OR(N)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) |
| \(PS(P) + OR(S)\) | \(V_1(x_1)\) | \(V_2(x_1)\) (Not necessarily enlarged) | \(V_1(x_1)\) | \(V_1(x_1)\) |
| \(PS(S) + OR(P)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_2(x_1)\) (Enlarged) | \(V_1(x_1)\) |
| \(PS(S) + OR(N)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) |
| \(PS(S) + OR(S)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) | \(V_1(x_1)\) |

cause, particularly, the nonparametric-based estimations may have severe difficulties to handle high- even moderate-dimensional models in practice. But the comparisons can provide a good insight into the nature of various estimations such that the practitioners can have a relatively complete pic-
ture about them and have idea for when and how to use these estimations.

2. Framework and Notation

For any individual, datum $W = (X^\top, Y, D)^\top$ is observable, including the observed effect $Y$, the treatment status $D$, and the $p$-dimensional covariates $X$. $D = 1$ implies that the individual is treated, and $D = 0$ means untreated. Denote $Y(1)$ and $Y(0)$ as the potential outcomes with and without treatment, respectively. The observed effect $Y$ can be expressed as $Y = DY(1) + (1 - D)Y(0)$. Denote that $p(X) = P(D = 1|X), m_1(X) = E(Y(1)|X), m_0(X) = E(Y(0)|X)$ as propensity score function and outcome regression functions. The following conditions are commonly used when we discuss the potential outcome framework.

(C1) (Sampling distribution) \( \{W_i\}_{i=1}^n \) is a set of identically distributed samples.

(C2) (Ignorability condition)

(i) (Unconfoundedness) \((Y(1), Y(0)) \perp D|X\)

(ii) Denote $\mathcal{X}$ as the support of $X$, where $\mathcal{X}$ is a Cartersian product of compact intervals. For any $x \in \mathcal{X}$, $p(x)$ is bounded away from 0 and 1.
Denote $\tau(x_1)$ as CATE:

$$
\tau(x_1) = E[Y(1) - Y(0)|X_1 = x_1]
$$

where $X_1$ is a strict subset of $X$. That is, $X_1$ is a $k$-dimension covariate, and $k < p$. Also denote $f(x_1)$ as the density function of $X_1$.

3. Doubly robust estimation

Rewrite $\tau(x_1)$ as

$$
\tau(x_1) = E \left\{ [m_1(X) - m_0(X)] | X_1 = x_1 \right\}
= E \left\{ \left[ \frac{DY}{p(X)} - \frac{(1 - D)Y}{1 - p(X)} \right] | X_1 = x_1 \right\}
= E \left\{ \left[ \frac{D}{p(X)}[Y - m_1(X)] - \frac{1 - D}{1 - p(X)}[Y - m_0(X)] + m_1(X) - m_0(X) \right] | X_1 = x_1 \right\}
$$

(3.2)

The first two equations in (3.2) show how OR and PS method work for estimating CATE. The third equation in (3.2) is an essential expression to construct a doubly robust estimator of $\tau(x_1)$. Under which, we propose a two-step estimation. In the first step, we estimate the function in (3.2):

$$
\frac{D}{p(X)}[Y - m_1(X)] - \frac{1 - D}{1 - p(X)}[Y - m_0(X)] + m_1(X) - m_0(X).
$$

To study the influence from estimating the nuisance functions, $p(X)$ and $m_1(X), m_0(X)$ under parametric, nonparametric, and semiparametric di-
mension reduction framework, we will construct the corresponding estimations below.

After this, we can then estimate the conditional expectation given $x_1$. This is a standard nonparametric estimation. We utilize the Nadaraya-Watson type estimator to define the resulting estimator:

$$
\hat{\tau}(x_1) = \frac{1}{nh_1^s} \sum_{i=1}^{n} \left[ \frac{D_i}{\hat{p}_i} (Y_i - \hat{m}_{1i}) - \frac{1-D_i}{1-\hat{p}_i} (Y_i - \hat{m}_{0i}) + \hat{m}_{1i} - \hat{m}_{0i} \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right),
$$

where $K_1(u)$ is a kernel function of order $s_1$, which is $s*$ times continuously differentiable, and $h_1$ is the corresponding bandwidth and $\hat{p}_i, \hat{m}_{1i}, \hat{m}_{0i}$ denote the estimators of $p(X_i), m_1(X_i), m_0(X_i)$ respectively, which are general notations and have different formulas under different model structures.

We now consider the estimations of the nuisance functions. Under the parametric structures with $\hat{p}(x; \beta), \hat{m}_1(x; \gamma_1)$ and $\hat{m}_0(x; \gamma_0)$ as the specified parametric models of $p(x), m_1(x)$ and $m_0(x)$ respectively, where $\beta, \gamma_1$ and $\gamma_0$ are unknown parameters. By maximum likelihood estimation, we can obtain $\hat{\beta}, \hat{\gamma}_1$ and $\hat{\gamma}_0$ so as to have $\hat{p}(X_i; \hat{\beta}), \hat{m}_1(X_i; \hat{\gamma}_1)$ and $\hat{m}_0(X_i; \hat{\gamma}_0)$ as the parametric estimators. Note that the specified models are not necessarily equal to true data generate mechanism. Now we further distinguish correctly specified, globally misspecified and locally misspecified case. For all $x \in \mathcal{X}$, there exist $\beta_0, \gamma_{10}, \gamma_{00}$, such that the true models have the rela-
tionship with the specified models:

\[ p(x) = \tilde{p}(x; \beta_0)[1 + c_n a(x)], \]
\[ m_1(x) = \tilde{m}_1(x; \gamma_{10}) + d_{1n} b_1(x), \]
\[ m_0(x) = \tilde{m}_0(x; \gamma_{00}) + d_{0n} b_0(x). \]

(3.3)

Take propensity score function as an example. If \( c_n = 0 \), then the parametric propensity score model \( \tilde{p}(x; \beta_0) \) is correctly specified, otherwise, it is not. If \( c_n \) converges to 0 as \( n \) goes to infinity, the parametric model is locally misspecified. If \( c_n \) remains a nonzero constant, it is a globally misspecified case. Similarly for the models with \( d_{1n} \) and \( d_{0n} \). Recall that \( \hat{\beta}, \hat{\gamma}_1 \) and \( \hat{\gamma}_0 \) are the maximum likelihood estimators of the corresponding unknown parameters. Denote \( \beta^*, \gamma_{1}^* \) and \( \gamma_{0}^* \) as the limits of \( \hat{\beta}, \hat{\gamma}_1 \) and \( \hat{\gamma}_0 \) as \( n \) goes to infinity.

Under the nonparametric structure, we utilize the kernel-based non-parametric estimators as

\[ \hat{p}(X_i) = \frac{\sum_{j=1}^{n} D_j K_2 \left( \frac{X_j - X_i}{h_2} \right)}{\sum_{i=1}^{n} K_2 \left( \frac{X_i - X_i}{h_2} \right)}, \]
\[ \hat{m}_1(X_i) = \frac{\sum_{j=1}^{n} D_j Y_j K_3 \left( \frac{X_j - X_i}{h_3} \right)}{\sum_{i=1}^{n} K_3 \left( \frac{X_i - X_i}{h_3} \right)}, \]
\[ \hat{m}_0(X_i) = \frac{\sum_{j=1}^{n} (1 - D_j) Y_j K_4 \left( \frac{X_j - X_i}{h_4} \right)}{\sum_{i=1}^{n} (1 - D_i) K_4 \left( \frac{X_i - X_i}{h_4} \right)}. \]
where \( K_2(u), K_3(u) \) and \( K_4(u) \) are kernels of order \( s_2 \geq d, s_3 \geq d \) and \( s_4 \geq d \), with the corresponding bandwidths \( h_2, h_3 \) and \( h_4 \). The conditions on the kernel functions and bandwidths will be listed in the supplement.

Under the semiparametric structure on the baseline covariate \( X \) for propensity score and outcome regressions, we have the following dimension reduction framework. Denote the matrix \( A \in \mathbb{R}^{d \times d_2} \) such that

\[
A^\top X, \quad \text{(3.4)}
\]

where \( d_2 \leq d \). The \( A \) spanned space \( \mathcal{S}_{E(D|X)} \) is called the central mean subspace if it is the intersection of all subspaces spanned by all \( A \) satisfy the above conditional independence. The dimension of \( \mathcal{S}_{E(D|X)} \) is called the structural dimension that is often smaller than or equal to \( d_2 \). Without confusion, still write it as \( d_2 \). Formula (3.4) implies that \( p(X) = E(D|X) = E(D|A^\top X) := g(A^\top X) \). Note that a nonparametric estimation of \( p(X) \) may have very slow rate of convergence when \( p \) is large. However, under (3.4) we can estimate the matrix \( A \) first to reduce the dimension \( d \) to \( d_2 \), the nonparametric estimation of \( E(D|A^\top X) \) can achieve a faster rate of convergence. The semiparametric estimator \( p(X_i) \) is then defined as, when \( A \) is root-\( n \) consistently by an estimator \( \hat{A} \),

\[
\hat{g}(\hat{A}^\top X_i) = \frac{\sum_{j=1}^{n} D_j K_5\left(\frac{\hat{A}^\top X_j - \hat{A}^\top X_i}{h_5}\right)}{\sum_{i=1}^{n} K_5\left(\frac{\hat{A}^\top X_i - \hat{A}^\top X_i}{h_5}\right)}.
\]
Similarly, for regression models, denote matrixes $B_1 \in \mathbb{R}^{d \times d_1}$ and $B_0 \in \mathbb{R}^{d \times d_0}$, such that

$$
E(Y(1)|X) \perp X|B_1^\top X,
$$
$$
E(Y(0)|X) \perp X|B_0^\top X.
$$

(3.5)

The corresponding dimension reduction subspaces are called the central mean subspaces (see Cook and Li 2002). Thus, $m_1(X) = E(Y(1)|X) = E(Y(1)|B_1^\top X)$ and $m_0(X) = E(Y(0)|X) = E(Y(0)|B_0^\top X)$ := $r_0(B_0^\top X)$. The semiparametric estimators $m_1(X_i)$ and $m_0(X_i)$ are defined as, with $\hat{B}_i$ being the estimators of $B_i$, $i = 0, 1$,

$$
\hat{r}_1(\hat{B}_1^\top X_i) = \frac{\sum_{j=1}^n D_j Y_j K_6 \left( \frac{\hat{B}_1^\top X_j - \hat{B}_1^\top X_i}{h_6} \right)}{\sum_{t=1}^n D_t K_6 \left( \frac{B_1^\top X_t - B_1^\top X_i}{h_6} \right)},
$$
$$
\hat{r}_0(\hat{B}_0^\top X_i) = \frac{\sum_{j=1}^n D_j Y_j K_7 \left( \frac{\hat{B}_0^\top X_j - \hat{B}_0^\top X_i}{h_7} \right)}{\sum_{t=1}^n D_t K_7 \left( \frac{B_1^\top X_t - B_1^\top X_i}{h_7} \right)}
$$

where $K_5(u)$, $K_6(u)$ and $K_7(u)$ are kernels of order $s_5 \geq d$, $s_6 \geq d$ and $s_7 \geq d$, with the corresponding bandwidths $h_5$, $h_6$ and $h_7$. 
4. Asymptotic Properties

Define the following functions

\[ \Psi_1(X, Y, D) := \frac{D[Y - m_1(X)]}{p(X)} - \frac{(1 - D)[Y - m_0(X)]}{1 - p(X)} + m_1(X) - m_0(X), \]
\[ \Psi_2(X, Y, D) := \frac{D\{Y - m_1(X)\}}{\tilde{p}(X; \beta^*)} - \frac{(1 - D)\{Y - m_0(X)\}}{1 - \tilde{p}(X; \beta^*)} + m_1(X) - m_0(X), \]
\[ \Psi_3(X, Y, D) := \frac{D\{Y - \tilde{m}_1(X; \gamma^*_1)\}}{p(X)} - \frac{(1 - D)\{Y - \tilde{m}_0(X; \gamma^*_0)\}}{1 - p(X)} + \tilde{m}_1(X; \gamma^*_1) - \tilde{m}_0(X; \gamma^*_0), \]
\[ \Psi_4(X, Y, D) := \frac{D\{Y - \tilde{m}_1(X; \gamma^*_1)\}}{\tilde{p}(X; \beta^*)} - \frac{(1 - D)\{Y - \tilde{m}_0(X; \gamma^*_0)\}}{1 - \tilde{p}(X; \beta^*)} + \tilde{m}_1(X; \gamma^*_1) - \tilde{m}_0(X; \gamma^*_0). \]

4.1 The Cases With No Model Misspecification

The following theorem shows all asymptotic distributions of the estimators are identical.

**Theorem 1.** Suppose Conditions (C1) – (C6), (A1), (A2) and (B1) are satisfied for \( s^* \geq s_2 \geq d, s^* \geq s_3 \geq d, s^* \geq s_4 \geq d, s^* \geq s_5 \geq d_2, s^* \geq s_6 \geq d_1, s^* \geq s_7 \geq d_0 \), and formulas 3.4 and 3.5 hold. Then, for each point \( x_1 \), we have

\[
\sqrt{nh_1^k} [\hat{\tau}(x_1) - \tau(x_1)] = \frac{1}{\sqrt{nh_1^k} f(x_1)} \sum_{i=1}^{n} \left[ \Psi_1(X_i, Y_i, D_i) - \tau(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) + o_p(1),
\]
and
\[
\sqrt{nh_k^k [\hat{\tau}(x_1) - \tau(x_1)]} \overset{d}{\to} N(0, V_1(x_1)),
\]
where
\[
V_1(x_1) = \frac{\sigma_1^2(x_1) \int K^2(u)du}{f(x_1)}, \quad \sigma_1^2(x_1) = E \{ [\Psi_1(X,Y,D) - \tau(x_1)]^2 | X_1 = x_1 \}.
\]

4.2 The Cases With Misspecified Models

Now we discuss the asymptotic behaviours of the proposed estimators if either outcome regression models or propensity score model is (are) misspecified. The following results show how global misspecification affects the asymptotic properties.

**Theorem 2.** Assume that the propensity score is globally misspecified in which \(c_n = C\) is a nonzero constant. Suppose conditions (C1) – (C6), (A1), (A2) and (B1) are satisfied for \(s^* \geq s_3 \geq d, s^* \geq s_4 \geq d, s^* \geq s_6 \geq d_1, s^* \geq s_7 \geq d_0, s_6 < (2s_6 + k)(d - d_1), s_7 < (2s_7 + k)(d - d_0)\).

1). When the outcome regression functions are estimated nonparametrically, then, for each value \(x_1\), we have
\[
\sqrt{nh_k^k [\hat{\tau}(x_1) - \tau(x_1)]} \overset{d}{\to} N(0, V_1(x_1)).
\]

2). When the outcome regression functions have a dimension reduction structure specified in \((3.5)\) or are correctly specified with \(d_{1n} = d_{0n} = 0\) with
4.2 The Cases With Misspecified Models

parametric estimation, for each value $x_1$, the asymptotic distributions are identical:

$$\sqrt{nh_1^k [\hat{\tau}(x_1) - \tau(x_1)]} \xrightarrow{d} N(0, V_2(x_1)),$$

where

$$V_2(x_1) = \frac{\sigma_2^2(x_1) \int K_2^2(u)du}{f(x_1)}, \quad \sigma_2^2(x_1) = E \{[\Psi_2(X,Y,D) - \tau(x_1)]^2 | X_1 = x_1\}.$$

Now we consider the cases with global misspecification of the outcome regression models.

**Theorem 3.** Assume that the outcome regression models are globally misspecified with fixed nonzero constants $d_{1n} = d_1$ and $d_{0n} = d_2$. Suppose conditions (C1) – (C6), (A1), (A2) and (B1) are satisfied for $s^* \geq s_2 \geq d$, $s^* \geq s_5 \geq d_2$, $s_5 < (2s_5 + k)(d - d_2)$.

1). When the propensity score is estimated nonparametrically, then, for each $x_1$,

$$\sqrt{nh_1^k [\hat{\tau}(x_1) - \tau(x_1)]} \xrightarrow{d} N(0, V_1(x_1)).$$

2). When the propensity score has a dimension reduction structure in (3.4) or is correctly specified with $c_n = 0$ and parametric estimation, for each value $x_1$, the asymptotic distributions are identical:

$$\sqrt{nh_1^k [\hat{\tau}(x_1) - \tau(x_1)]} \xrightarrow{d} N(0, V_3(x_1)).$$
4.2 The Cases With Misspecified Models

where

\[ V_3(x_1) = \frac{\sigma_3^2(x_1) \int K_1^2(u) du}{f(x_1)} , \quad \sigma_3^2(x_1) = E \left\{ [\Psi_3(X, Y, D) - \tau(x_1)]^2 \mid X_1 = x_1 \right\} . \]

Remark 1. By some calculations, we can obtain in Proposition 4 below in Section 4.4 that \( \sigma_1^2(x_1) \leq \sigma_3^2(x_1) \), while the analogy does not hold between \( \sigma_2^2(x_1) \) and \( \sigma_1^2(x_1) \). That is, the asymptotic variance of the proposed estimator inflates when the outcome regression models are misspecified, and the propensity score model is parametrically estimated (correctly specified) or semiparametrically estimated. However, whether the asymptotic variance gets larger with a misspecified propensity score model is model-dependent. We show the following example. Suppose that the outcome regression models are correctly specified, while the propensity score model is globally misspecified. Consider a situation that \( p(x) = p_1, \tilde{p}(x; \beta^*) = p_2 \), where \( p_1, p_2 \) are free of \( x \), and \( p_1 \neq p_2 \). We have

\[ \sigma_2^2(x_1) - \sigma_1^2(x_1) = E \left\{ \frac{p^2(X) - \tilde{p}^2(X; \beta^*)}{p^2(X; \beta^*)} Var(Y \mid X, D = 1) \mid X_1 = x_1 \right\} + E \left\{ \frac{[1 - p(X)]^2 - [1 - \tilde{p}^2(X; \beta^*)]^2}{[1 - \tilde{p}(X; \beta^*)]^2[1 - p(X)]} Var(Y \mid X, D = 0) \mid X_1 = x_1 \right\} \]

\[ = \frac{p_2^2 - p_1^2}{p_1 p_2^2} E [Var(Y \mid X, D = 1) \mid X_1 = x_1] + \frac{(1 - p_1)^2 - (1 - p_2)^2}{(1 - p_1)(1 - p_2)^2} E [Var(Y \mid X, D = 0) \mid X_1 = x_1] . \]

To give a clear picture, we further assume that the outcome regression models are homoscedastic that \( Var(Y \mid X, D = 1) = Var(Y \mid X, D = 0) = \xi^2 \),
4.2 The Cases With Misspecified Models

which is free of \(X\). Then we have, 
\[
\sigma^2(x_1) - \sigma^2_1(x_1) = \xi^2 \left( \frac{p_1^2 - p_2^2}{p_1 p_2^2} + \frac{(1-p_1)^2 - (1-p_2)^2}{(1-p_1)(1-p_2)^2} \right).
\]
Define the function 
\[
v_d(p_1, p_2) = \frac{p_1^2 - p_2^2}{p_1 p_2^2} + \frac{(1-p_1)^2 - (1-p_2)^2}{(1-p_1)(1-p_2)^2}.
\]
A negative \(v_d(p_1, p_2)\) implies the variance shrinkage. Consider three true propensity score values 
\(p(x) = p_1 = 0.3, 0.5, 0.7\). The following three curves of \(v_d(p_1, p_2)\) show how the variance inflation or shrinkage occurs.

Figure 1: Curves of \(v_d(p_1, p_2)\) with different \(p_1\)

When \(p_1 = 0.3\) or 0.7, appropriately overestimated propensity score may result in an asymptotic variance shrinking in some cases. When \(p_1 = 0.5\), which means that every individual have an 0.5 probability to be treated regardless of any covariates, misspecification leads to the asymptotic variance augmentation.

We can in effect obtain some more examples since \(Var(Y|X, D = 1)\) and \(Var(Y|X, D = 0)\) are not necessarily equal. Such simple examples show that when only propensity score is misspecified, augmenting or shrinking asymptotic variances are all possible.
Remark 2. Another interesting phenomenon is that once propensity score model is misspecified and outcome regressions are nonparametrically estimated, or vice versa, the asymptotic performance of the proposed estimator is identical to that when all models are correctly specified. As nonparametric estimation takes no risk of misspecification, such an estimation procedure “absorbs” the influence brought by model misspecification due to the doubly robust property. But it is clear that in high-dimensional scenarios, a purely nonparametric estimation is not worthwhile to recommend. Thus, this property mainly serves as an investigation with theoretical interest unless the dimension of the covariates is small.

The results with local misspecification are stated in the following.

Theorem 4. Assume that the propensity score is locally specified with \( c_n \to 0 \). Suppose conditions (C1) – (C6), (A1), (A2) and (B1) are satisfied for

\[
s^* \geq s_3 \geq d, \quad s^* \geq s_4 \geq d, \quad s^* \geq s_6 \geq d_1, \quad s^* \geq s_7 \geq d_0, \quad s_6 < (2s_6+k)(d-d_1),
\]

\[
s_7 < (2s_7+k)(d-d_0).
\]

Then, for each value \( x_1 \), we have

\[
\sqrt{n} h_1^k [\tilde{\tau}(x_1) - \tau(x_1)] \overset{d}{\to} N(0, V_1(x_1)).
\]

Similarly, assume that the outcome regression functions are locally misspecified with \( d_{1n} \to 0 \) and \( d_{0n} \to 0 \). Under the same conditions as those in Theorem 4 for \( s^* \geq s_2 \geq d, \quad s^* \geq s_5 \geq d(2), \quad s_5 < (2s_5+k)(p-p(2)) \). For
each value $x_1$, the asymptotic distribution of $\hat{\tau}(x_1)$ is identical to the above.

4.3 A Further Study: All Models are Misspecified

We study this case as it then has a non-ignorable bias in general and goes
to zero unless the rate of convergence of local misspecification is sufficiently
fast. Recall the definitions of $\gamma_0^*$, $\gamma_1^*$ and $\beta^*$ below (3).

**Theorem 5.** Suppose that all models are globally misspecified with nonzero
constants $c_n$, $d_{1n}$ and $d_{0n}$. Assume that conditions (C1) – (C6) are satisfied.
Then, for each value $x_1$, we have

$$\sqrt{nh_1} [\hat{\tau}(x_1) - \tau(x_1) - \text{bias}(x_1)] \xrightarrow{d} N(0, V_4(x_1)),$$

where

$$\text{bias}(x_1) = E \left\{ \frac{[m_1(X) - \tilde{m}_1(X; \gamma_1^*)][p(X) - \tilde{p}(X; \beta^*)]}{\tilde{p}(X; \beta^*)} - \frac{[m_0(X) - \tilde{m}_0(X; \gamma_0^*)][\tilde{p}(X; \beta^*) - p(X)]}{1 - \tilde{p}(X; \beta^*)} \bigg| X_1 = x_1 \right\},$$

$$V_4(x_1) = \frac{\sigma_4^2(x_1) \int K^2(u)du}{f(x_1)}, \quad \sigma_4^2(x_1) = E \left\{ [\Psi_4(X,Y,D) - \hat{\tau}(x_1)]^2 \bigg| X_1 = x_1 \right\},$$

and

$$\hat{\tau}(x_1) = E \left\{ \left[ \frac{D}{\tilde{p}(X; \beta^*)}[Y - \tilde{m}_1(X; \gamma_1^*)] \right. \right.$$

$$\left. - \frac{1 - D}{1 - \tilde{p}(X; \beta^*)}[Y - \tilde{m}_0(X; \gamma_0^*) + \tilde{m}_1(X; \gamma_1^*) - \tilde{m}_0(X; \gamma_0^*)] \bigg| X_1 = x_1 \right\}.$$
4.3 A Further Study: All Models are Misspecified

The following results show the importance of the convergence rates of $c_n$, $d_{1n}$ and $d_{0n}$ to zero for bias reduction and variance change.

**Theorem 6.** Under the conditions in Theorem 5, when

$$c_n d_{1n} = o \left( \frac{1}{\sqrt{nh_1^k}} \right), \quad c_n d_{0n} = o \left( \frac{1}{\sqrt{nh_1^k}} \right),$$

then, for each $x_1$, we have

$$\sqrt{nh_1^k} \left[ \hat{\tau}(x_1) - \tau(x_1) \right] \overset{d}{\to} N(0, V_1(x_1)).$$

**Remark 3.** This theorem show that to make the bias vanished, $c_n d_{1n}$ and $c_n d_{0n}$ need to tend to zero at the rates faster than the nonparametric convergence rate, $O(1/\sqrt{nh_1^k})$. Recall that Theorems 2 and 3 show that when $c_n = o(1)$, then the variance is $V_3(x_1)$; when $d_{1n} = o(1)$ and $d_{0n} = o(1)$ the variance is $V_2(x_1)$. Altogether, when all misspecifications are local, the asymptotic variances reduces to $V_1(x_1)$. We can then further discuss four cases:

1) All nuisance models are globally misspecified; 2) All nuisance models are locally misspecified; 3) The propensity score function is globally misspecified, and the outcome regression functions are locally misspecified; 4) The propensity score function is locally misspecified, and the outcome regression functions are globally misspecified.
4.4 A summary on the comparison among the asymptotic Variances

The first is the case exactly described in Theorem 5, the second shows that if \( c_n d_{1n} = o(1/\sqrt{nh_1^k}) \) and \( c_n d_{0n} = o(1/\sqrt{nh_1^k}) \), the bias term is negligible, which is the situation in Theorem 6. Otherwise, the estimator is biased. Cases 3 and 4 can be regarded as a combination of those in Theorems 5 and 6. In case 3, once \( d_{1n} = o(1/\sqrt{nh_1^k}) \) and \( d_{0n} = o(1/\sqrt{nh_1^k}) \), the bias goes to 0, and the variance goes to \( ||K_1||^2_2 \sigma_2^2(x_1)/f(x_1) \). In other words, if \( d_{1n} \) and \( d_{0n} \) go to 0 at a rate faster than \( O(1/\sqrt{nh_1^k}) \), Case 3 turns to the case in Theorem 2. We can then also derive that if \( c_n = o(1/\sqrt{nh_1^k}) \), Case 4 is similar to that in Theorem 3.

4.4 A summary on the comparison among the asymptotic Variances

We summarize the comparison among the 4 variances \( V_j(x_1) \) for \( j = 1, 2, 3, 4 \) as listed in Section 1. Note that the variances are \( V_j(x_1) = ||K_1||^2_2 \sigma_j^2(x_1)/f(x_1) \) for \( j = 1, 2, 3, 4 \) and thus the comparison among them is equivalent to the comparison among \( \sigma_j^2(x_1) \) for \( j = 1, 2, 3, 4 \).

**Remark 4.** For any \( x_1 \),

1. \( \sigma_1^2(x_1) \) is not necessarily smaller than \( \sigma_2^2(x_1) \) and as shown in the example in Remark 1, \( \sigma_1^2(x_1) \) can be larger than \( \sigma_2^2(x_1) \) for some \( x_1 \);

2. \( \sigma_1^2(x_1) \leq \sigma_3^2(x_1) \);
3). We have no definitive answer to say whether $\sigma_1^2(x_1)$ is necessarily smaller than $\sigma_4^2(x_1)$.

5. Numerical Study

In this section, we present some Monte Carlo simulations to examine the finite sample performances of the estimators.

5.1 Data-Generating Process

Consider two data-generating processes (DGPs) similarly as those in Abrevaya et al. (2015), the case of $d = 2$ and $d = 4$. Here we only consider that the conditioning covariate $X_1$ is univariate, i.e. $k = 1$. So in the simulations, $\tau(x_1) = E[Y(1) - Y(0)|X_1 = x_1]$.

**Model 1.** It is featured by a 2-dimensional unconfounded covariate, $X = (X_1, X_2)^\top$. In other words, $d = 2$. For further information,

$$X_1 = \rho_1, \quad X_2 = (1 + 2X_1)^2(-1 + X_1)^2 + \rho_2,$$

where $\rho_1, \rho_2$ are independently identically $U(-0.5, 0.5)$ distributed. The potential outcomes and the propensity score function are given as:

$$Y(1) = X_1 X_2 + \epsilon, \quad Y(0) = 0,$$

$$p(X) = \frac{exp(X_1 + X_2)}{1 + exp(X_1 + X_2)}.$$
where $\epsilon \sim N(0, 0.25^2)$. The true CATE conditioning on $X_1$ can be derived as $\tau(x_1) = x_1(1 + 2x_1)^2(-1 + x_1)^2$. Since the misspecification effect is a concern, we use the misspecified parametric model respectively:

$$\tilde{m}_1(X; \gamma_1) = (1, X^\top)\gamma_1, \quad \tilde{p}(X; \beta) = \frac{\exp((1, X_1)\beta)}{1 + \exp((1, X_1)\beta)}.$$

where $\gamma_1 \in \mathbb{R}^3$, $\beta \in \mathbb{R}^2$.

**Model 2.** Another DGP is featured by a 4-dimensional unconfounded covariate for the purpose of a further investigation on higher dimension cases. Write $X = (X_1, X_2, X_3, X_4)^\top$ and

$$X_1 = \rho_1, \quad X_2 = 1 + 2X_1 + \rho_2, \quad X_3 = 1 + 2X_1 + \rho_3, \quad X_4 = (-1 + X_1)^2 + \rho_4,$$

where $\rho_1, \rho_2, \rho_3, \rho_4$ are independently identically $U(-0.5, 0.5)$ distributed. The potential outcomes and the propensity score function are defined as:

$$Y(1) = X_1X_2X_3X_4 + \epsilon, \quad Y(0) = 0,$$

$$p(X) = \frac{\exp\left[\frac{1}{2}(X_1 + X_2 + X_3 + X_4)\right]}{1 + \exp\left[\frac{1}{2}(X_1 + X_2 + X_3 + X_4)\right]}.$$

where $\epsilon \sim \mathcal{N}(0, 0.25^2)$. The true CATE conditioning on $X_1$ remains as $\tau(x_1) = x_1(1 + 2x_1)^2(-1 + x_1)^2$. Still we use the misspecified parametric model respectively:

$$\tilde{m}_1(X; \gamma_1) = (1, X^\top)\gamma_1, \quad \tilde{p}(X; \beta) = \frac{\exp((1, X_1)\beta)}{1 + \exp((1, X_1)\beta)}.$$
where $\gamma_1 \in \mathbb{R}^5$, $\beta \in \mathbb{R}^2$.

### 5.2 Kernel Functions and Bandwidths

As the selections of kernel functions and bandwidths (listed in the supplementary material) have great influence on the asymptotic property when the nuisance models are nonparametrically or semiparametrically estimated, we first discuss this issue.

Let $h = an^{-\eta}$ for $\eta > 0$. Together with condition (A2), how to determine the value $\eta$ goes to a linear programming problem.

For **model 1** ($d = 2$), we consider a kernel function of order 4 ($s_1 = 4$) as the kernel in the second step of N-W estimation, $K_1$. Write $h_1 = a_1h^{-\eta_1}$. For the other bandwidths, take $h_2$ as an example. The results in Section 4 requires that $s^* \geq s_2 \geq d$, we then choose $s_2 = 2$. Also let $h_2 = a_2n^{-\eta_2}$. Then let $(\eta_1, \eta_2) = (\frac{1}{9}, \frac{1}{7})$. The other bandwidths can also be determined similarly as $h_j = a_jn^{-\frac{1}{4}}, (j = 2, 3, 5, 6)$, when $s_j = 2, (j = 1, 2, 3, 5, 6)$. Also, these convergence rates of $h_i$ to meet condition (A16). To choose $a_j, (j = 2, 3, 5, 6)$, we, by the rule of thumb, choose $a_1 = 0.1, a_2 = 0.7, a_3 = 1.5, a_5 = 0.5$ and $a_6 = 1$. For **model 2** ($d = 4$), consider $s_1 = 6$ and $s_j = 4, (j = 2, 3, 5, 6)$ and $h_1 = a_1n^{-\frac{1}{5}}$ and $h_j = a_jn^{-\frac{1}{8}}, (j = 2, 3, 5, 6)$. Further, let $a_1 = 0.1, a_2 = 2, a_3 = 2.5, a_5 = 2.8$ and $a_6 = 1$. In simulations, we chose
many other values and found that the above values are recommendable as the values around them can make the estimators relatively stable.

Consider the Gaussian kernel $K_1$ of order $s_1$ under condition (A1)(i). For other kernel functions, use Epanechnikov kernels of the corresponding orders under conditions (A1)(ii) and (iii).

5.3 Simulation Results

As there are many estimators $\hat{\tau}(x_1)$ with different estimated nuisance models, we then, in Table 2, list them and the corresponding notations for convenience.

To guarantee the regularity conditions and the estimation stability, all estimated propensity scores are trimmed within $[0.005, 0.995]$ as many authors did.

In the simulations, we estimate $\tau(x_1)$ for $x_1 \in \{-0.4, -0.2, 0, 0.2, 0.4\}$. The sample sizes are $n = 500$ and $n = 5,000$ respectively to see their asymptotic behaviours. The experiments are repeated 2,500. Denote $T(x_1) = \sqrt{n h_1} (\hat{\tau}(x_1) - \tau(x_1)).$ we evaluate the estimators based on following criteria: bias of $\hat{\tau}(x_1)$; sample standard deviation (sam-SD) of $T(x_1)$; mean square error (MSE) of $T(x_1)$. We also report the proportions ($P_{0.05}$, $P_{0.95}$) of the standardized $T(x_1)$ below the 5% quantile and above the 95% quan-
### 5.3 Simulation Results

#### Table 2: Estimators involved in simulation

| DRCATE     | $p(x)$                           | $m_1(x)$                           |
|------------|----------------------------------|------------------------------------|
| (O, O)     | oracle                           | oracle                             |
| (cP, cP)   | parametrically estimated (correctly specified) | parametrically estimated (correctly specified) |
| (N, N)     | nonparametrically estimated      | nonparametrically estimated        |
| (S, S)     | semiparametrically estimated     | semiparametrically estimated        |
| (mP, cP)   | parametrically estimated (misspecified) | parametrically estimated (correctly specified) |
| (mP, N)    | parametrically estimated (misspecified) | nonparametrically estimated        |
| (mP, S)    | parametrically estimated (misspecified) | semiparametrically estimated        |
| (cP, mP)   | parametrically estimated (correctly specified) | parametrically estimated (misspecified) |
| (N, mP)    | nonparametrically estimated      | parametrically estimated (misspecified) |
| (S, mP)    | semiparametrically estimated     | parametrically estimated (misspecified) |
5.3 Simulation Results

Table 3: The simulation results under model 1 (part 1)

| DRCATE | $x_1$ | bias  | sam-SD  | MSE  | $P_{0.05}$ | $P_{0.95}$ | bias  | sam-SD  | MSE  | $P_{0.05}$ | $P_{0.95}$ |
|--------|-------|-------|---------|------|------------|------------|-------|---------|------|------------|------------|
| (O,O)  | -0.4  | 0.0001| 0.2776  | 0.0770| 0.052      | 0.046      | -0.0004| 0.2724  | 0.0742| 0.044      | 0.052      |
|        | -0.2  | -0.0023| 0.2378  | 0.0567| 0.056      | 0.044      | -0.0005| 0.2333  | 0.0544| 0.049      | 0.050      |
|        | 0     | -0.0002| 0.2088  | 0.0436| 0.049      | 0.050      | 0.0003 | 0.2014  | 0.0405| 0.047      | 0.048      |
|        | 0.2   | 0.0003 | 0.1997  | 0.0399| 0.052      | 0.047      | 0.0002 | 0.1999  | 0.0400| 0.050      | 0.054      |
|        | 0.4   | 0.0027 | 0.2003  | 0.0403| 0.045      | 0.058      | 0.0004 | 0.2006  | 0.0403| 0.048      | 0.054      |
| (cP,cP) | -0.4  | 0.0000 | 0.2797  | 0.0782| 0.053      | 0.048      | 0.0004 | 0.2725  | 0.0743| 0.044      | 0.052      |
|        | -0.2  | -0.0023| 0.2378  | 0.0567| 0.056      | 0.042      | -0.0005| 0.2333  | 0.0544| 0.051      | 0.048      |
|        | 0     | -0.0002| 0.2089  | 0.0436| 0.048      | 0.050      | 0.0003 | 0.2014  | 0.0405| 0.047      | 0.047      |
|        | 0.2   | 0.0003 | 0.1994  | 0.0397| 0.051      | 0.048      | 0.0002 | 0.2001  | 0.0400| 0.051      | 0.054      |
|        | 0.4   | 0.0027 | 0.2003  | 0.0403| 0.044      | 0.058      | 0.0004 | 0.2007  | 0.0403| 0.047      | 0.054      |
| (N,N)  | -0.4  | 0.0008 | 0.2716  | 0.0738| 0.050      | 0.053      | 0.0001 | 0.2845  | 0.0809| 0.050      | 0.049      |
|        | -0.2  | 0.0015 | 0.2366  | 0.0560| 0.042      | 0.058      | -0.0001| 0.2344  | 0.0549| 0.050      | 0.050      |
|        | 0     | 0.0002 | 0.2046  | 0.0419| 0.043      | 0.052      | -0.0005| 0.1996  | 0.0399| 0.057      | 0.041      |
|        | 0.2   | 0.0010 | 0.2000  | 0.0400| 0.044      | 0.051      | -0.0001| 0.1941  | 0.0377| 0.052      | 0.056      |
|        | 0.4   | 0.0014 | 0.2081  | 0.0433| 0.045      | 0.054      | 0.0009 | 0.2012  | 0.0406| 0.045      | 0.056      |
| (S,S)  | -0.4  | -0.0022| 0.2815  | 0.0794| 0.051      | 0.044      | 0.0002 | 0.2862  | 0.0819| 0.045      | 0.050      |
|        | -0.2  | 0.0004 | 0.2365  | 0.0559| 0.046      | 0.052      | -0.0004| 0.2302  | 0.0530| 0.046      | 0.048      |
|        | 0     | 0.0005 | 0.2082  | 0.0433| 0.053      | 0.052      | 0.0003 | 0.2059  | 0.0424| 0.052      | 0.052      |
|        | 0.2   | -0.0015| 0.1992  | 0.0397| 0.061      | 0.041      | -0.0002| 0.2011  | 0.0404| 0.053      | 0.051      |
|        | 0.4   | 0.0002 | 0.2021  | 0.0408| 0.050      | 0.046      | 0.0012 | 0.2048  | 0.0422| 0.043      | 0.050      |

tile of $N'(0, 1)$ to verify the asymptotic Normality. We display the efficiency comparisons among different estimators under models 1 and 2 in Figures [2]
5.3 Simulation Results

Table 4: The simulation results under **model 1** (part 2)

| DRCATE | $x_1$ | bias | sam-SD | MSE | $P_{0.05}$ | $P_{0.05}$ | bias | sam-SD | MSE | $P_{0.05}$ | $P_{0.05}$ |
|--------|------|------|--------|-----|------------|------------|------|--------|-----|------------|------------|
| (O,O)  | -0.4 | 0.0001 | 0.2776 | 0.0770 | 0.052 | 0.046 | 0.0004 | 0.2724 | 0.0742 | 0.044 | 0.052 |
|        | -0.2 | -0.0023 | 0.2378 | 0.0567 | 0.056 | 0.044 | -0.0005 | 0.2333 | 0.0544 | 0.049 | 0.050 |
|        | 0    | -0.0002 | 0.2088 | 0.0436 | 0.049 | 0.050 | 0.0003 | 0.2014 | 0.0405 | 0.047 | 0.048 |
|        | 0.2  | 0.0003 | 0.1997 | 0.0399 | 0.052 | 0.047 | 0.0002 | 0.1999 | 0.0400 | 0.050 | 0.054 |
|        | 0.4  | 0.0027 | 0.2003 | 0.0403 | 0.045 | 0.058 | 0.0004 | 0.2006 | 0.0403 | 0.048 | 0.054 |
| (mP,P) | -0.4 | 0.0000 | 0.2599 | 0.0675 | 0.052 | 0.049 | 0.0004 | 0.2530 | 0.0640 | 0.044 | 0.052 |
|        | -0.2 | -0.0022 | 0.2363 | 0.0559 | 0.056 | 0.041 | -0.0005 | 0.2323 | 0.0540 | 0.050 | 0.050 |
|        | 0    | -0.0002 | 0.2203 | 0.0485 | 0.049 | 0.048 | 0.0003 | 0.2116 | 0.0448 | 0.047 | 0.052 |
|        | 0.2  | 0.0003 | 0.2041 | 0.0417 | 0.051 | 0.046 | 0.0002 | 0.2048 | 0.0419 | 0.050 | 0.053 |
|        | 0.4  | 0.0027 | 0.1953 | 0.0383 | 0.044 | 0.058 | 0.0004 | 0.1955 | 0.0382 | 0.046 | 0.054 |
| (mP,N) | -0.4 | -0.0046 | 0.2666 | 0.0716 | 0.064 | 0.040 | -0.0011 | 0.2629 | 0.0693 | 0.054 | 0.044 |
|        | -0.2 | -0.0035 | 0.2373 | 0.0566 | 0.059 | 0.044 | -0.0029 | 0.2383 | 0.0584 | 0.074 | 0.037 |
|        | 0    | -0.0068 | 0.2152 | 0.0474 | 0.072 | 0.032 | -0.0027 | 0.2107 | 0.0458 | 0.072 | 0.034 |
|        | 0.2  | -0.0011 | 0.2041 | 0.0417 | 0.052 | 0.047 | -0.0004 | 0.1952 | 0.0381 | 0.050 | 0.045 |
|        | 0.4  | -0.0008 | 0.2003 | 0.0401 | 0.049 | 0.049 | 0.0007 | 0.2002 | 0.0402 | 0.043 | 0.056 |
| (mP,S) | -0.4 | -0.0143 | 0.2701 | 0.0781 | 0.082 | 0.029 | -0.0115 | 0.2722 | 0.0996 | 0.146 | 0.010 |
|        | -0.2 | -0.0094 | 0.2453 | 0.0624 | 0.070 | 0.032 | -0.0073 | 0.2302 | 0.0634 | 0.114 | 0.016 |
|        | 0    | -0.0046 | 0.2116 | 0.0453 | 0.064 | 0.043 | -0.0038 | 0.2099 | 0.0469 | 0.083 | 0.032 |
|        | 0.2  | -0.0019 | 0.2041 | 0.0417 | 0.050 | 0.046 | -0.0006 | 0.1970 | 0.0388 | 0.054 | 0.047 |
|        | 0.4  | 0.0022 | 0.2002 | 0.0402 | 0.046 | 0.058 | 0.0017 | 0.1968 | 0.0393 | 0.037 | 0.062 |

and the detailed results under **model 1** are displayed in Tables 6, 7 and 8. To save space, the other simulation results about **model 2** are
### Table 5: The simulation results under **model 1** (part 3)

| DRCATE | \(x_1\) | \(P_{0.05}\) | \(P_{0.95}\) | bias | sam-SD | MSE | \(P_{0.05}\) | \(P_{0.95}\) | bias | sam-SD | MSE | \(P_{0.05}\) | \(P_{0.95}\) |
|--------|--------|-------------|-------------|------|--------|-----|-------------|-------------|------|--------|-----|-------------|-------------|
| \((O, O)\) | -0.4 | 0.0001 | 0.2776 | 0.0770 | 0.052 | 0.046 | 0.0004 | 0.2724 | 0.0742 | 0.044 | 0.052 |
| | -0.2 | -0.0023 | 0.2378 | 0.0567 | 0.056 | 0.044 | -0.0005 | 0.2333 | 0.0544 | 0.049 | 0.050 |
| | 0 | -0.0002 | 0.2088 | 0.0436 | 0.049 | 0.050 | 0.0003 | 0.2014 | 0.0405 | 0.047 | 0.048 |
| | 0.2 | 0.0003 | 0.1997 | 0.0399 | 0.052 | 0.047 | 0.0002 | 0.1999 | 0.0400 | 0.050 | 0.054 |
| | 0.4 | 0.0027 | 0.2003 | 0.0403 | 0.045 | 0.058 | 0.0004 | 0.2006 | 0.0403 | 0.048 | 0.054 |
| \((cP, mP)\) | -0.4 | -0.0012 | 0.3230 | 0.1044 | 0.051 | 0.048 | 0.0001 | 0.3201 | 0.1024 | 0.050 | 0.049 |
| | -0.2 | -0.0021 | 0.2400 | 0.0577 | 0.052 | 0.042 | -0.0005 | 0.2362 | 0.0558 | 0.054 | 0.044 |
| | 0 | 0.0004 | 0.2147 | 0.0461 | 0.052 | 0.049 | 0.0003 | 0.2050 | 0.0420 | 0.049 | 0.049 |
| | 0.2 | 0.0004 | 0.2012 | 0.0405 | 0.054 | 0.046 | 0.0001 | 0.2016 | 0.0406 | 0.048 | 0.049 |
| | 0.4 | 0.0028 | 0.2059 | 0.0426 | 0.043 | 0.061 | 0.0004 | 0.2039 | 0.0416 | 0.045 | 0.053 |
| \((N, mP)\) | -0.4 | -0.0105 | 0.2840 | 0.0834 | 0.075 | 0.040 | -0.0013 | 0.2970 | 0.0885 | 0.060 | 0.045 |
| | -0.2 | 0.0014 | 0.2353 | 0.0554 | 0.047 | 0.050 | 0.0007 | 0.2288 | 0.0525 | 0.040 | 0.053 |
| | 0 | 0.0013 | 0.2104 | 0.0443 | 0.048 | 0.054 | 0.0002 | 0.2065 | 0.0426 | 0.047 | 0.044 |
| | 0.2 | -0.0014 | 0.1995 | 0.0398 | 0.056 | 0.048 | -0.0004 | 0.2022 | 0.0409 | 0.052 | 0.044 |
| | 0.4 | 0.0008 | 0.2034 | 0.0414 | 0.046 | 0.046 | 0.0000 | 0.2077 | 0.0431 | 0.048 | 0.050 |
| \((S, mP)\) | -0.4 | -0.0051 | 0.2964 | 0.0884 | 0.055 | 0.046 | -0.0005 | 0.3089 | 0.0955 | 0.050 | 0.045 |
| | -0.2 | -0.0002 | 0.2421 | 0.0586 | 0.049 | 0.050 | 0.0001 | 0.2394 | 0.0573 | 0.048 | 0.051 |
| | 0 | 0.0005 | 0.2076 | 0.0431 | 0.050 | 0.050 | -0.0001 | 0.2051 | 0.0421 | 0.048 | 0.049 |
| | 0.2 | -0.0008 | 0.2082 | 0.0433 | 0.049 | 0.049 | -0.0001 | 0.1966 | 0.0386 | 0.054 | 0.048 |
| | 0.4 | 0.0005 | 0.2104 | 0.0443 | 0.044 | 0.052 | 0.0006 | 0.2085 | 0.0435 | 0.048 | 0.054 |

reported in the supplementary material.

Here we present some observations from the simulation results.
5.3 Simulation Results

Figure 2: Relative variance against DRCATE(O,O) in model 1

First, with the sample size growth, the bias and the standard deviation of $\hat{\tau}(x_1)$ reasonably tend to be smaller due to the estimation consistency. The reported proportions $P_{0.05}$ and $P_{0.95}$ can be controlled around 0.05, which implies that the normal approximation of the proposed estimator is valid.

Second, from Figures 2 and 3, the efficiency comparisons among the estimators (O,O), (cP,cP), (N,N) and (S,S) show that the distributions are close to each other. When only the propensity score function is misspecified, variance inflation and shrinkage are both possible. With misspecified outcome regression function, only variance inflation is possible.
Figure 3: Relative variance against DRCATE(O,O) in model 2

Third, the bias and standard deviation of $\hat{\tau}(x_1)$ increase when the covariate dimension grows, see the comparisons in Figures 2 and 3. Possible explanation to this phenomenon would be that the standard deviations of nuisance models’ estimations increase with higher dimension of covariate.

6. Conclusion

In this paper, we investigate the asymptotic behaviours of nine doubly robust estimators (DR), under different combinations of model structures, to provide a relatively complete picture of this methodology.

When all models are correctly specified, the asymptotic equivalence
among all defined estimators does not surprisingly hold. When models are mispecified, we consider local and global misspecifications and some interesting phenomena have been discovered such as asymptotic variance shrinking in some cases due to misspecification. Further, we would recommend semiparametric estimation under dimension reduction structure. This is because nonparametric estimation severely suffers from the curse of dimensionality whereas parametric estimation may not be sufficiently robust against model structure.

Acknowledgements

The research described herein was supported by a NNSF grant of China and a grant from the University Grants Council of Hong Kong, Hong Kong, China.
7. Supplementary Material

The supplementary material contains the detailed proofs of the theorems and propositions, and the additional simulation results.

7.1 Technical Conditions

Here we present some conditions to derive the theoretical results. Together with (C1) and (C2) in the main context, the following conditions in the (C) group are regularity conditions to guarantee the asymptotic properties regardless of the different ways to estimate nuisance models.

(C3) Density functions involved in this article satisfy the following conditions:

(i) For any $x \in \mathcal{X}$, the density function of $X$, $\theta(x)$ is bounded away from 0.

(ii) For any $x_1$, the density function of $X_1$, $f(x_1)$ is bounded away from zero and $s_1$ times continuously differentiable.

(iii) Denote the density functions of $A^\top X$, $B_1^\top X$ and $B_0^\top X$ as $\theta_A(\cdot)$, $\theta_{B_1}(\cdot)$ and $\theta_{B_0}(\cdot)$. For any $x \in \mathcal{X}$, all these density functions are bounded away from 0.

(C4) Denote $\mathcal{C}$ as the parameter space of $\beta$. For any $x \in \mathcal{X}$ and $\beta \in \mathcal{C}$,
\( \tilde{p}(x; \beta) \) is bounded away from 0 and 1.

(C5) \( \sup_{x_1} E[Y(j)^2|X_1 = x_1] < \infty \) for \( j = 0, 1 \).

(C6) \( E[\Psi_1(X, Y, D) - \tau(x_1)]^{2+\kappa_1} \leq \infty \), \( E[\Psi_2(X, Y, D) - \tau(x_1)]^{2+\kappa_2} \leq \infty \),
\( E[\Psi_3(X, Y, D) - \tau(x_1)]^{2+\kappa_3} \leq \infty \), \( E[\Psi_4(X, Y, D) - \tilde{\tau}(x_1)]^{2+\kappa_4} \leq \infty \),
\( \int |K(u)|^{2+\delta} du \leq \infty \) for some constants \( \kappa_1, \kappa_2, \kappa_3, \kappa_4, \delta \geq 0 \).

(C1) and (C2) in the main context are the basic conditions under Rubin’s potential outcome framework, as stated in Section 2. It is obvious that (C4) is an analogue of (C2)(ii). Bounded propensity scores or specified propensity score models, density functions and corresponding conditional moments are required in these conditions, which are common restrictions in the literature, and play important roles in deriving the asymptotic linear expression of the proposed estimators. (C6) ensures the applicability of Lyapunov’s Central Limit Theorem here.

Assume some conditions on kernel functions and bandwidths in non-parametric estimation:

(A1) The kernel functions satisfy the following conditions:

(i) \( K_1(u) \) is a kernel function of order \( s_1 \), which is symmetric around zero and \( s^* \) times continuously differentiable.
(ii) \( K_2(u), K_3(u) \) and \( K_4(u) \) are kernels of order \( s_2 \geq p, s_3 \geq p \) and \( s_4 \geq p \), which are symmetric around zero and equal to zero outside \( \prod_{i=1}^{p} [-1, 1] \), with continuous \((s_2+1), (s_3+1)\) and \((s_4+1)\) order derivatives respectively.

(iii) \( K_5(u), K_6(u) \) and \( K_7(u) \) are kernels of order \( s_5 \geq p(2), s_6 \geq p(1) \) and \( s_7 \geq p(0) \), which are symmetric around zero and equal to zero outside \( \prod_{i=1}^{p(2)} [-1, 1], \prod_{i=1}^{p(1)} [-1, 1] \) and \( \prod_{i=1}^{p(0)} [-1, 1] \), with continuous \((s_5+1), (s_6+1)\) and \((s_7+1)\) order derivatives respectively.

(A2) As different scenarios require different bandwidths, we put them together in the following. As \( n \to \infty \):

(i) \( h_1 \to 0, nh_1^k \to \infty, nh_1^{2s_1+k} \to 0 \).

(ii) \( h_2 \to 0, (\ln n)/(nh_2^{p+s_2}) \to 0 \).

(iii) \( h_3 \to 0, (\ln n)/(nh_3^{p+s_3}) \to 0 \).

(iv) \( h_4 \to 0, (\ln n)/(nh_4^{p+s_4}) \to 0 \).

(v) \( h_5 \to 0, (\ln n)/(nh_5^{p+s_5}) \to 0 \).

(vi) \( h_6 \to 0, (\ln n)/(nh_6^{p+s_6}) \to 0 \).

(vii) \( h_7 \to 0, (\ln n)/(nh_7^{p+s_7}) \to 0 \).

(viii) \( h_2^{2s_2}h_1^{-2s_2-k} \to 0, nh_1^k h_2^{2s_2} \to 0 \).
\( h_3^{2s_3} h_1^{-2s_3-k} \to 0, \, n h_3^{k} h_3^{2s_3} \to 0. \)

\( h_4^{2s_4} h_1^{-2s_4-k} \to 0, \, n h_4^{k} h_4^{2s_4} \to 0. \)

\( h_5^{2s_5} h_1^{-2s_5-k} \to 0, \, n h_5^{k} h_5^{2s_5} \to 0. \)

\( h_6^{2s_6} h_1^{-2s_6-k} \to 0, \, n h_6^{k} h_6^{2s_6} \to 0. \)

\( h_7^{2s_7} h_1^{-2s_7-k} \to 0, \, n h_7^{k} h_7^{2s_7} \to 0. \)

\( n h_1^{k} h_2^{s_2} h_3^{s_3} \to 0, \, n h_1^{k} h_2^{s_2} h_4^{s_4} \to 0, \, n h_1^{k} h_2^{s_2} h_6^{s_6} \to 0, \, n h_1^{k} h_2^{s_2} h_7^{s_7} \to 0, \)
\( n h_1^{k} h_5^{s_5} h_3^{s_3} \to 0, \, n h_1^{k} h_5^{s_5} h_4^{s_4} \to 0, \, n h_1^{k} h_5^{s_5} h_6^{s_6} \to 0, \, n h_1^{k} h_5^{s_5} h_7^{s_7} \to 0. \)

Remind that \( K_j(u), j = 2, 3, \ldots, 7, \) and \( h_j, j = 2, 3, \ldots, 7 \) are corresponding kernels and bandwidths in nonparametric and semiparametric estimators of nuisance models. When only parametric methods are applied to estimate nuisance models, no conditions above, but (A1)(i) and (A2)(i) are required.

The conditions in (A1)(ii) and (A1)(iii) are required when at least one misspecified model is involved. Epanechnikov kernel of corresponding order can be a candidate of \( K_j(u), j = 2, \ldots, 7. \) [Abrevaya et al. (2015)] stated that this restriction on the bounded support can be relaxed to exponential tails.

(A2)(ii)-(xiv) place restrictions on the convergence rates of different bandwidths to ensure remainders of the linear expression negligible. (A2)(xiv),
involving more than 2 bandwidths, can be regarded as an interaction term, which makes it handleable to determine those convergence rates. Here we provide a naive idea to accomplish this task based on linear programming. Assume the corresponding bandwidths converge to 0 in such a manner, \( h_j = a_j n^{-\eta_j}, j = 1, \ldots, 7, \) where \( \eta_j > 0. \) With predetermined \( s_j, j = 1, \ldots, 7 \) and (A2), the problem goes to a linear programming task to find out the feasible region of \( \eta_j. \) For a more detailed example, reader can refer to Section 5.

Lastly, we give a condition to ensure the desired convergence rates of the estimators under semiparametric dimension reduction structure, which will be a favour when pursuing the asymptotic properties of \( \hat{\tau}(x_1): \)

(B1) \( \hat{A} - A = O_p (n^{-1/2}), \hat{B}_1 - B_1 = O_p (n^{-1/2}), \hat{B}_0 - B_0 = O_p (n^{-1/2}) \)

These can be achieved by standard estimations in the literature, see the relevant references such as Li (1991), ????

In summary, these conditions are rather standard.
Recall that
\[
\hat{\tau}(x_1) = \frac{\sum_{i=1}^{n} \left[ \frac{D_i}{\hat{p}_i} (Y_i - \hat{m}_{1i}) - \frac{1-D_i}{1-\hat{p}_i} (Y_i - \hat{m}_{0i}) + \hat{m}_{1i} - \hat{m}_{0i} \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)}{\sum_{t=1}^{n} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)}
\]
\[
= \frac{\sum_{i=1}^{n} \left[ \frac{D_i}{\hat{p}_i} (Y_i - \hat{m}_{1i}) + \hat{m}_{1i} \right]}{\sum_{t=1}^{n} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)} - \frac{\sum_{i=1}^{n} \left[ \frac{1-D_i}{1-\hat{p}_i} (Y_i - \hat{m}_{0i}) + \hat{m}_{0i} \right]}{\sum_{t=1}^{n} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)}
\]
\[
= \hat{\tau}_1(x_1) - \hat{\tau}_0(x_1).
\]

Let
\[
\tau(x_1) = E \left[ \frac{D_i}{p(X)} [Y - m_1(X)] - \frac{1-D_i}{1-p(X)} [Y - m_0(X)] + m_1(X) - m_0(X) \right| X_1 = x_1
\]
\[
= E \left[ \frac{D_i}{p(X)} [Y - m_1(X)] + m_1(X) \right| X_1 = x_1
\]
\[
- E \left[ \frac{1-D_i}{1-p(X)} [Y - m_0(X)] + m_0(X) \right| X_1 = x_1
\]
\[
= \tau_1(x) - \tau_0(x).
\]

For the very first move, we look for the asymptotic linear expression of
\[
\sqrt{nh_1^k} [\hat{\tau}(x_1) - \tau(x_1)].
\]
7.2 Proof of Theorem 1

\( \hat{f}(x_1) \xrightarrow{p} f(x_1) \) ensures that we can use Slutsky’s Theorem later. So we can first consider the asymptotic linear expression of

\[
J(x_1) = \frac{1}{\sqrt{nh_1}} \sum_{i=1}^{n} \left[ \frac{D_i}{p_i} (Y_i - \hat{m}_{1i}) + \hat{m}_{1i} - \tau_1(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right). \tag{7.1}
\]

Consider several combinations of estimation of nuisance functions. Now we list them as below.

Scenario 1. \( p(x) \) parametrically estimated (correctly specified), \( m_1(x) \) parametrically estimated (correctly specified)

Scenario 2. \( p(x) \) nonparametrically estimated, \( m_1(x) \) nonparametrically estimated

Scenario 3. \( p(x) \) semiparametrically estimated, \( m_1(x) \) semiparametrically estimated

Scenario 4. \( p(x) \) parametrically estimated (correctly specified), \( m_1(x) \) nonparametrically estimated

Scenario 5. \( p(x) \) parametrically estimated (correctly specified), \( m_1(x) \) semiparametrically estimated

Scenario 6. \( p(x) \) nonparametrically estimated, \( m_1(x) \) parametrically estimated (correctly specified)

Scenario 7. \( p(x) \) nonparametrically estimated, \( m_1(x) \) semiparametrically estimated
Scenario 8. $p(x)$ semiparametrically estimated, $m_1(x)$ parametrically estimated (correctly specified)

Scenario 9. $p(x)$ semiparametrically estimated, $m_1(x)$ nonparametrically estimated

**Scenario 1:** $p(x)$ and $m_1(x)$ are parametrically estimated. From standard parametric estimation argument,

$$
\sup_{x \in \mathcal{X}} |\tilde{p}(x; \hat{\beta}) - p(x)| = \sup_{x \in \mathcal{X}} |\tilde{p}(x; \hat{\beta}) - \tilde{p}(x; \beta^*)| = O_p\left(\frac{1}{\sqrt{n}}\right), \\
\sup_{x \in \mathcal{X}} |\tilde{m}_1(x; \hat{\gamma}_1) - m_1(x)| = \sup_{x \in \mathcal{X}} |\tilde{m}_1(x; \hat{\gamma}_1) - \tilde{m}_1(x; \gamma_1^*)| = O_p\left(\frac{1}{\sqrt{n}}\right).
$$

We start from (7.1):

$$
J(x_1) = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i}{\hat{p}(X_i; \hat{\beta})} \left[ Y_i - \tilde{m}_1(X_i; \hat{\gamma}_1) \right] + \tilde{m}_1(X_i; \hat{\gamma}_1) - \tau_1(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
$$

$$
= \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i Y_i}{p(X_i)} - \frac{D_i - p(X_i)}{p(X_i)} m_1(X_i) - \tau_1(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
$$

$$
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i (m_{1i}^+ - Y_i)}{p_i^2} \right] \left[ \tilde{p}(X_i; \hat{\beta}) - p(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
$$

$$
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left( \frac{p_i^+ - D_i}{p_i^+} \right) \left[ \tilde{m}_1(X_i; \hat{\gamma}_1) - m_1(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
$$

$$
=: J_{11}(x_1) + J_{12}(x_1) + J_{13}(x_1)
$$

(7.2)

where $p_i^+$ lies between $p(X_i)$ and $\tilde{p}(X_i; \hat{\beta})$, $m_{1i}^+$ lies between $m_1(X_i)$ and $\tilde{m}_1(X_i; \hat{\gamma}_1)$. 
Bounding $J_{12}(x)$ as

$$|J_{12}(x)| = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left| \frac{D_i (m_{i1}^+ - Y_i)}{p_i^{1/2}} \right| \left| \tilde{p}(X_i; \hat{\beta}) - p(X_i) \right| K_1 \left( \frac{X_{i1} - x_1}{h_1} \right)$$

$$\leq \sqrt{nh_1^k} \sup_{x \in \mathcal{X}} \left| \tilde{p}(x; \hat{\beta}) - p(x) \right| \times \frac{1}{nh_1^k} \sum_{i=1}^{n} \left| K_1 \left( \frac{X_{i1} - x_1}{h_1} \right) \right| \left| \frac{D_i (m_{i1}^+ - Y_i)}{p_i^{1/2}} \right|$$

where $\sup_{x \in \mathcal{X}} \left| \tilde{p}(x; \hat{\beta}) - p(x) \right| = O_p \left( \frac{1}{\sqrt{n}} \right)$, $\left| \frac{D_i (m_{i1}^+ - Y_i)}{p_i^{1/2}} \right|$ is bounded due to condition (C2)(ii), (C4) and (C5), $\frac{1}{nh_1^k} \sum_{i=1}^{n} \left| K_1 \left( \frac{X_{i1} - x_1}{h_1} \right) \right| = O_p(1)$ by the standard nonparametric estimation argument. Thus $|J_{12}(x_1)| \leq o_p(1)$.

$O_p(1) = o_p(1)$. With the similar arguments, we can also bound the last term as $|J_{13}(x_1)| = o_p(1)$. So far, we’ve proved $J_{12}(x_1)$ and $J_{13}(x_1)$ converge to 0 in probability. Hence, according to Slutsky’s Theorem, together with (7.2), we have

$$\sqrt{nh_1^k}[\tilde{\tau}_1(x_1) - \tau_1(x_1)] = \frac{1}{f(x_1)} J(x_1) = \frac{1}{f(x_1)} J_{11}(x_1) + o_p(1)$$

$$= \frac{1}{f(x_1)} \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i}{p(X_i)} [Y_i - m_1(X_i)] + m_1(X_i) - \tau_1(x_1) \right] K_1 \left( \frac{X_{i1} - x_1}{h_1} \right) + o_p(1)$$

(7.3)

**Scenario 2:** $p(x)$ and $m_1(x)$ are nonparametrically estimated.

From the standard nonparametric estimation argument, under conditions
(A1)(i), (ii), (iii), we have

\[
\sup_{x \in \mathcal{X}} |\hat{p}(x) - p(x)| = O_p \left( h_2^s + \frac{\ln n}{nh_2^b} \right) = o_p \left( h_2^s \right)
\]

\[
\sup_{x \in \mathcal{X}} |\hat{m}_1(x) - m_1(x)| = O_p \left( h_3^s + \frac{\ln n}{nh_3^b} \right) = o_p \left( h_3^s \right)
\]

Rewrite (7.1):

\[
J(x_1) = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i}{\hat{p}(X_i)} [Y_i - \hat{m}_1(X_i)] + \hat{m}_1(X_i) - \tau_1(x_1) \right] K_1 \left( \frac{X_{ii} - x_1}{h_1} \right)
\]

\[
= \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i}{p(X_i)} [Y_i - m_1(X_i)] + m_1(X_i) - \tau_1(x_1) \right] K_1 \left( \frac{X_{ii} - x_1}{h_1} \right)
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{D_i [m_1(X_i) - Y_i]}{p^2(X_i)} [\hat{p}(X_i) - p(X_i)] K_1 \left( \frac{X_{ii} - x_1}{h_1} \right)
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{p(X_i) - D_i}{p(X_i)} [\hat{m}_1(X_i) - m_1(X_i)] K_1 \left( \frac{X_{ii} - x_1}{h_1} \right) + 0
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{D_i [Y_i - m^+_1]}{p^{+3}_i} [\hat{p}(X_i) - p(X_i)]^2 K_1 \left( \frac{X_{ii} - x_1}{h_1} \right)
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{2D_i}{p^{+2}_i} [\hat{p}(X_i) - p(X_i)] [\hat{m}_1(X_i) - m_1(X_i)] K_1 \left( \frac{X_{ii} - x_1}{h_1} \right)
\]

\[
=: J_{21}(x_1) + J_{22}(x_1) + J_{23}(x_1) + 0 + J_{24}(x_1) + J_{25}(x_1) \tag{7.4}
\]

where \( p_i^+ \) lies between \( p(X_i) \) and \( \hat{p}(X_i) \), \( m^+_1 \) lies between \( m_1(X_i) \) and \( \hat{m}_1(X_i) \).

Rewrite \( J_{22}(x_1) \) as

\[
J_{22}(x_1) = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{D_i [m_1(X_i) - Y_i]}{p^2(X_i)} [\hat{p}(X_i) - p(X_i)] K_1 \left( \frac{X_{ii} - x_1}{h_1} \right)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{D_i [m_1(X_i) - Y_i]}{p^2(X_i)} \frac{1}{\sqrt{h_1^k}} [\hat{p}(X_i) - p(X_i)] K_1 \left( \frac{X_{ii} - x_1}{h_1} \right)
\]
In which \( E \left[ \frac{D_i m_1(X_i) - Y_i}{p^2(X_i)} \right] X_i \) = 0 and thus \( \frac{D_i m_1(X_i) - Y_i}{p^2(X_i)} \) is independent of \( X_i \) for every \( i \); sup\(_{x \in \mathcal{X}} [\hat{p}(x) - p(x)] = O_p \left( h_2^{s_2} + \sqrt{\frac{\ln n}{nh_2^2}} \right) = o_p \left( h_2^{s_2} \right) \). By condition (A1)(viii) and CLT, \( \frac{1}{\sqrt{n h_1^k}} [\hat{p}(X_i) - p(X_i)] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) = o_p(1) \)

and then \( J_{22}(x_1) = o_p(1) \).

Similarly, \( |J_{23}(x_1)| = o_p(1) \). Deal with \( J_{24}(x_1) \) by using the decomposition as

\[
|J_{24}(x_1)| = \left| \frac{1}{\sqrt{n h_1^k}} \sum_{i=1}^{n} D_i \left[ Y_i - m_{1i}^+ \right] \frac{[\hat{p}(X_i) - p(X_i)]^2}{p_i^{s_3}} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right|
\]

\[
\leq \sqrt{n h_1^k} \sup_{x \in \mathcal{X}} [\hat{p}(x) - p(x)]^2 \times \frac{1}{nh_1^k} \sum_{i=1}^{n} \left| K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right| \frac{D_i [Y_i - m_{1i}^+]}{p_i^{s_3}}
\]

in which \( \sup_{x \in \mathcal{X}} [\hat{p}(x) - p(x)] = o_p(h_2^{s_2}) \). Then under condition (A1)(viii),

\( \sqrt{n h_1^k} \sup_{x \in \mathcal{X}} [\hat{p}(x) - p(x)]^2 = o_p(1) \). Under conditions (C2)(ii) and (C5),

\( \left| \frac{D_i [Y_i - m_{1i}^+]}{p_i^{s_3}} \right| \) is bounded. Again by the standard argument for handling non-parametric estimation, \( \frac{1}{nh_1^k} \sum_{i=1}^{n} \left| K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right| = O_p(1) \). Thus, we can obtain that \( |J_{24}(x_1)| = o_p(1) \cdot O_p(1) = o_p(1) \). In a similar way, \( |J_{25}(x_1)| = o_p(1) \) can also be proved. Here we have derived that \( J_{22}(x_1), J_{23}(x_1), J_{24}(x_1) \) and \( J_{25}(x_1) \) can be bounded as \( o_p(1) \). Together with (7.4), we can obtain that

\[
\sqrt{nh_1^k} \left[ \hat{\tau}_1(x_1) - \tau_1(x_1) \right] = \frac{1}{f(x_1)} J(x_1) = \frac{1}{f(x_1)} J_{21}(x_1) + o_p(1)
\]

\[
= \frac{1}{f(x_1)} \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ D_i \frac{[Y_i - m_1(X_i)] + m_1(X_i) - \tau_1(x_1)]}{p(Z_i)} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right] + o_p(1).
\]
Scenario 3: \( p(x) \) and \( m_1(x) \) are semiparametrically estimated.

Under conditions (A2)(v) and (vi), (vii),

\[
\sup_{x \in \mathcal{X}} \left| \hat{g}(A^T x) - g(A^T z) \right| = O_p \left( h_5^{s_5} + \sqrt{\frac{\ln(n)}{nh_5^{p(2)}}} \right) = o_p \left( h_5^{s_5} \right),
\]

\[
\sup_{x \in \mathcal{X}} \left| \hat{r}_1(B_1^T z) - r_1(B_1^T z) \right| = O_p \left( h_6^{s_6} + \sqrt{\frac{\ln n}{nh_6^{p(1)}}} \right) = o_p \left( h_6^{s_6} \right).
\]

Note that under condition (B1), we can first discuss the asymptotic distribution by assuming that the projection matrices \( A, B_0 \) and \( B_1 \) are given.

Then

\[
J(x_1) = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i}{\widehat{g}(A^T X_i)} \left[ Y_i - \widehat{r}_1(B_1^T X_i) \right] + \widehat{r}_1(B_1^T X_i) - \tau_1(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
\]

\[
- \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i}{p(X_i)} [Y_i - m_1(X_i)] + m_1(X_i) - \tau_1(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{p(X_i) - D_i}{p(X_i)} \left[ \widehat{r}_1(B_1^T X_i) - r_1(B_1^T X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) + 0
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{D_i}{g_i^{+3}} [\widehat{g}(A^T X_i) - g(A^T X_i)]^2 K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{2D_i}{g_i^{+2}} \left[ \widehat{r}_1(B_1^T X_i) - r_1(B_1^T X_i) \right] \left[ \widehat{g}(A^T X_i) - g(A^T X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) =: J_{31}(x_1) + J_{32}(x_1) + J_{33}(x_1) + 0 + J_{34}(x_1) + J_{35}(x_1)
\]

(7.6)

where \( g_i^{+} \) lies between \( g(A^T X_i) \) and \( \widehat{g}(A^T X_i) \), \( m_{1i}^{+} \) lies between \( r_1(B_1^T X_i) \) and \( \widehat{r}_1(B_1^T X_i) \). Then we deal with all terms one by one.
Consider $J_{32}(x_1)$ and $J_{33}(x_1)$. We have

$$J_{32}(x_1) = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{D_i[m_1(X_i) - Y_i]}{p^2(X_i)} \left[ \hat{g}(A^T X_i) - g(A^T X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)$$

Again $E \left[ \frac{D_i[m_1(X_i) - Y_i]}{p^2(X_i)} \right] X_i = 0$. Then, $\frac{D_i[m_1(X_i) - Y_i]}{p^2(X_i)}$ is independent of $X_i$ for every $i$; sup$_{x \in X} |\hat{g}(A^T X_i) - g(A^T X_i)| = O_p \left( h_0^{\alpha} + \frac{\ln n}{nh_1^k} \right) = o_p \left( h_0^{\alpha} \right)$. Condition (A2)(xi) yields that $\frac{1}{\sqrt{nh_1^k}} \left[ \hat{g}(A^T X_i) - g(A^T X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) = o_p(1)$. The application of CLT yields that $J_{32}(x_1) = o_p(1)$. Also, we can prove $J_{33}(x_1) = o_p(1)$ similarly.

Deal with $J_{34}(x_1)$ and $J_{35}(x_1)$. We have

$$|J_{34}(x_1)| = \left| \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{D_i[Y_i - r_i^{+1}]}{g_i^{-3}} \left[ \hat{g}(A^T X_i) - g(A^T X_i) \right]^2 K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right|$$

$$\leq \sqrt{nh_1^k} \sup_{x \in X} \left[ \hat{g}(A^T X_i) - g(A^T X_i) \right]^2 \times \frac{1}{nh_1^k} \sum_{i=1}^{n} \left| K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right| \left| \frac{D_i[Y_i - r_i^{+1}]}{g_i^{-3}} \right| .$$

Also sup$_{x \in X} \left[ \hat{g}(A^T X_i) - g(A^T X_i) \right] = o_p (h_0^{\alpha}).$ Condition (A2)(xi) implies that

$$\sqrt{nh_1^k} \sup_{x \in X} \left[ \hat{g}(A^T X_i) - g(A^T X_i) \right]^2 = o_p(1)$$

Under conditions (C2)(ii) and (C5), $\left| \frac{D_i[Y_i - r_i^{+1}]}{g_i^{-3}} \right|$ is bounded. Again, $\frac{1}{nh_1^k} \sum_{i=1}^{n} \left| K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right| = O_p(1)$. We can then achieve $|J_{34}(x_1)| = o_p(1) · O_p(1) = o_p(1)$. This is also the way to prove $|J_{35}(x_1)| = o_p(1)$. In this way, the asymptotic negligibility
of $J_{32}(x_1)$, $J_{33}(x_1)$, $J_{34}(x_1)$ and $J_{35}(x_1)$ has been proved. Together with (7.6), it can be derived that

$$\sqrt{nh_1^k [\hat{\tau}_1(x_1) - \tau_1(x_1)]} = \frac{1}{f(x_1)} J(x_1) = \frac{1}{f(x_1)} J_{31}(x_1) + o_p(1)$$

$$= \frac{1}{f(x_1)} \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i}{p(X_i)} [Y_i - m_1(X_i)] + m_1(X_i) - \tau_1(x) \right] K_1 \left( \frac{X_{i1} - x_1}{h_1} \right) + o_p(1).$$  \quad (7.7)

Consider equations (7.3), (7.5) and (7.7), which imply that the asymptotic linear expressions of $\sqrt{nh_1^k [\hat{\tau}_1(x_1) - \tau_1(x_1)]}$ are identical among scenarios 1, 2 and 3. It is obvious that under the conditions of Theorem 1, in any scenario mentioned above, the asymptotic linear expression remains the same, which leads to the same asymptotic distribution. With the asymptotic linear expression, we can further derive the asymptotic distribution. First, we have the decomposition as

$$\sqrt{nh_1^k [\hat{\tau}(x_1) - \tau(x_1)]}$$

$$= \frac{1}{f(x_1)} \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} [\Psi_1(X_i, Y_i, D_i) - \tau(x_1)] K_1 \left( \frac{X_{i1} - x_1}{h_1} \right) + o_p(1)$$

$$= \frac{1}{f(x_1)} \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} [\Psi_1(X_i, Y_i, D_i) - \tau(X_{i1})] K_1 \left( \frac{X_{i1} - x_1}{h_1} \right)$$

$$+ \frac{1}{f(x_1)} \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} [\tau(X_{i1}) - \tau(x_1)] K_1 \left( \frac{X_{i1} - x_1}{h_1} \right) + o_p(1)$$

$$= \frac{1}{f(x_1)} I_1(x_1) + \frac{1}{f(x_1)} I_2(x_1) + o_p(1).$$ \quad (7.8)
7.2 Proof of Theorem 1

Consider $I_1(x_1)$ first. Note that $\tau(X_{1i}) = E[\Psi_1(X_i, Y_i, D_i) | X_1 = X_{1i}]$. Then $\Psi_1(X_i, Y_i, D_i) - \tau(X_{1i})$ is independent of $X_{1i}$, $K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)$ only depends on $n$ and $X_{1i}$. Thus

$$E \left\{ [\Psi_1(X, Y, D) - \tau(X_1)]K_1 \left( \frac{X_1 - x_1}{h_1} \right) \right\} = E[\Psi_1(X, Y, D) - \tau(X_1)]E K_1 \left( \frac{X_1 - x_1}{h_1} \right) = 0.$$ 

Also, $$\left\{ [\Psi(X_i, Y_i, D_i) - \tau(X_{1i})]K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right\}_{i=1}^{n}$$ independently and identically distributed. We now check the condition of Lyapunov’s CLT: $\exists \kappa > 0$, s.t.

$$\sum_{i=1}^{n} E \left| \frac{1}{\sqrt{n}} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) [\Psi_1(X_i, Y_i, D_i) - \tau(X_{1i})] \frac{1}{\sqrt{h_1^k}} \right|^{2+\kappa} \rightarrow 0 \quad (n \rightarrow \infty)$$ 

Under condition (C6), letting $C = E|\Psi_1(X_i, Y_i, D_i) - \tau(X_{1i})| < \infty$, we have

$$\sum_{i=1}^{n} E \left| \frac{1}{\sqrt{n}} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) [\Psi_1(X_i, Y_i, D_i) - \tau(X_{1i})] \frac{1}{\sqrt{h_1^k}} \right|^{2+\kappa}$$

$$= \left( \frac{1}{\sqrt{nh_1^k}} \right)^{\kappa} E \left| \Psi_1(X, Y, D) - \tau(X_1) \right|^{2+\kappa} E |K_1 \left( \frac{X_1 - x_1}{h_1} \right)|^{2+\kappa} \frac{1}{h_1^k}$$

$$\leq \left( \frac{1}{\sqrt{nh_1^k}} \right)^{\kappa} CE \left| K_1 \left( \frac{X_1 - x_1}{h_1} \right) \right|^{2+\kappa} \frac{1}{h_1^k}$$

where

$$\frac{1}{h_1^k} E \left| K_1 \left( \frac{X_1 - x_1}{h_1} \right) \right|^{2+\kappa} = \int K_1^{2+\kappa}(u)f(x_1 + h_1 u) du \rightarrow f(x_1) \int K_1^{2+\kappa}(u) du < \infty.$$ 

Thus

$$\left( \frac{1}{\sqrt{nh_1^k}} \right)^{\kappa} E \left| \Psi_1(X, Y, D) - \tau(X) \right|^{2+\kappa} E \left| K_1 \left( \frac{X_1 - x_1}{h_1} \right) \right|^{2+\kappa} \frac{1}{h_1^k} \rightarrow 0 (n \rightarrow \infty).$$
7.2 Proof of Theorem 1

The Lyapunov’s condition is satisfied and then

\[ I_1(x) \overset{d}{\to} \mathcal{N}(0, V) \]  

(7.9)

where \( V = \lim_{n \to \infty} \text{Var} \left\{ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} [\Psi_1(X_i, Y_i, D_i) - \tau(x_1)]K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right\} \).

To compute the variance \( V \), we can see that

\[
\text{Var} \left\{ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} [\Psi_1(X_i, Y_i, D_i) - \tau(x_1)]K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right\} = E \left[ \left( \frac{1}{h_1^k} \sum_{i=1}^{n} [\Psi_1(X_i, Y_i, D_i) - \tau(x_1)]K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right)^2 \right]
\]

\[
= h_1^k E \left\{ E \left[ \left( [\Psi_1(X, Y, D) - \tau(x_1)]\frac{1}{h_1^k} K_1 \left( \frac{X_1 - x_1}{h_1} \right) \right)^2 \right] \right\} X_1
\]

\[
= h_1^k \int \left( \frac{1}{h_1^k} K_1 \left( \frac{X_1 - x_1}{h_1} \right) \right)^2 E \left[ \left( [\Psi(X_i, Y_i, D_i) - \tau(x_1)] \right)^2 \right] dF_{X_i}
\]

\[
= h_1^k \int \left( \frac{1}{h_1^k} K_1 \left( \frac{t - x_1}{h_1} \right) \right)^2 E \left[ \left( \Psi(X_i, Y_i, D_i) - \tau(x_1) \right)^2 \right] \bigg| X_1 = t \ f(t) dt
\]

\[
= h_1^k \int \frac{K_1^2(u) E \left[ \left( \Psi(X, Y, D) - \tau(x_1) \right)^2 \right] \bigg| X = x_1 + h_1 u \right] f(x_1 + h_1 u) du
\]

\[
= \sigma_1^2(x_1) f(x_1) \int K_1^2(u) du + O(h_1^k)
\]

where

\[
\sigma_1^2(x_1) = E \left\{ \left[ \Psi_1(X, Y, D) - \tau(x_1) \right]^2 \bigg| X_1 = x_1 \right\}.
\]

Consider \( I_2(x_1) \). We have

\[
I_2(x_1) = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} [\tau(X_{1i}) - \tau(x_1)]K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
\]
where

\[
E \left\{ \frac{1}{\sqrt{n h_1^k}} \sum_{i=1}^{n} \left[ \tau(X_{1i}) - \tau(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right\} = \sqrt{n h_1^k} \int [\tau(x_1 + h_1u) - \tau(x_1)] K_1(u) f(x_1 + h_1u) du = \sqrt{n h_1^k} O_p(h_1^{s_1}) = o_p(1).
\]

Note that its variance is as

\[
Var \left\{ \frac{1}{\sqrt{n h_1^k}} \sum_{i=1}^{n} \left[ \tau(X_{1i}) - \tau(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right\} = E \left\{ \frac{1}{\sqrt{n h_1^k}} \sum_{i=1}^{n} \left[ \tau(X_{1i}) - \tau(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right\}^2
\]

\[
- E \left\{ \frac{1}{\sqrt{n h_1^k}} \sum_{i=1}^{n} \left[ \tau(X_{1i}) - \tau(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right\}^2
\]

\[
= \int [\tau(x_1 + h_1u) - \tau(x)]^2 K_1^2(u) f(x_1 + h_1u) du
\]

\[
- nh_1^k [\tau(x_1 + h_1u) - \tau(x_1)] K_1(u) f(x_1 + h_1u) du^2 = o_p(1).
\]

so \( I_2(x_1) = o_p(1)\).

Combining with (7.8), (7.9) and \( I_2(x_1) = o_p(1)\), we can obtain that

\[
I_1(x_1) + I_2(x_1) \overset{d}{\rightarrow} N \left( 0, \sigma_1^2(x_1) \int K_1^2(u) du \right)
\]

and

\[
\sqrt{nh_1^k [\hat{\tau}(x_1) - \tau(x_1)]} = \frac{1}{f(x_1)} [I_1(x_1) + I_2(x_1)] \overset{d}{\rightarrow} N \left( 0, \frac{\sigma_1^2(x_1) \int K_1^2(u) du}{f(x_1)} \right).
\]

Now we consider the cases with unknown \( A, B_1 \) and \( B_0 \). Note that under condition (B1), \( \hat{A}, \hat{B}_1 \) and \( \hat{B}_0 \) converge in probability to \( A, B_1 \) and
7.3 Proofs of Theorems 2 and 3

$B_0$ respectively at the rate of $O_p\left(\frac{1}{\sqrt{n}}\right)$. Following the similar arguments in [Hu et al. (2014)], we can see easily that the asymptotic distribution retains.

We then do not give the details for space saving. Now we can conclude that, under the conditions of Theorem 1, regardless of which estimation method (parametric, nonparametric, semiparametric dimension reduction) used to estimate nuisance models,

$$\sqrt{n}h_{l_1}^1[\hat{\tau}(x_1) - \tau(x_1)] \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_1^2(x_1) \int K_1^2(u)du}{f(x_1)}\right).$$

The proof is done. \qed

7.3 Proofs of Theorems 2 and 3

We now consider global misspecification cases. Similar to the proof of Theorem 1, we first consider the asymptotic linear expression of $J(x_1)$.

**Scenario 1: $m_1(x)$ is nonparametrically estimated.** In this case, we have

$$\sup_{x \in X} \left| \tilde{p}(x; \hat{\beta}) - \tilde{p}(x; \beta^*) \right| = O_p\left( \frac{1}{\sqrt{n}} \right),$$

$$\sup_{x \in X} \left| \hat{m}_1(x) - m_1(x) \right| = O_p \left( h_{l_1}^{3.3} + \sqrt{\frac{\ln n}{nh_{l_1}^p}} \right) = o_p \left( h_{l_1}^{3.3} \right).$$
We can further rewrite (7.1) as

\[
J(x_1) = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i}{\hat{p}(X_i; \hat{\beta})} [Y_i - \hat{m}_1(X_i)] + \hat{m}_1(X_i) - \tau_1(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) 
\]

\[
= \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \left[ \frac{D_i}{\hat{p}(X_i; \beta^*)} [Y_i - m_1(X_i)] + m_1(X_i) - \tau_1(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) 
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{\hat{p}(X_i; \beta^*) - D_i}{\hat{p}(X_i; \beta^*)} \left[ \hat{m}_1(X_i) - m_1(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) + 0 
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} D_i \left[ Y_i - m_1^+(X_i) \right] \left[ \hat{p}(X_i; \hat{\beta}) - \hat{p}(X_i; \beta^*) \right]^2 K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) 
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} 2D_i \frac{\hat{p}(X_i; \beta^*) - D_i}{\bar{p}_i^{1/2}} \left[ \hat{p}(X_i; \hat{\beta}) - \hat{p}(X_i; \beta^*) \right] \left[ \hat{m}_1(X_i) - m_1(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) 
\]

\[
= J_{41}(x_1) + J_{42}(x_1) + J_{43}(x_1) + 0 + J_{44}(x_1) + J_{45}(x_1) 
\]

(7.10)

where \( p_i^+ \) lies between \( \hat{p}(X_i; \beta^*) \) and \( \hat{p}(X_i; \hat{\beta}) \), \( m_i^+ \) lies between \( m_1(X_i) \) and \( \hat{m}_1(X_i) \).

As we can prove \( J_{42}(x_1) \) and \( J_{44}(x_1) \) are \( o_p(1) \) in the same way as the proof for \( J_{12}(x_1) = o_p(1) \) in scenario 1 of Subsection 7.2, the details are then omitted. For \( J_{45}(x_1) \), obviously

\[
|J_{45}(x_1)| = \left| \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{2D_i}{\bar{p}_i^{1/2}} \left[ \hat{p}(X_i; \hat{\beta}) - \hat{p}(X_i; \beta^*) \right] \left[ \hat{m}_1(X_i) - m_1(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right| 
\]

\[
= \sqrt{nh_1^k} \sup_{x \in \mathcal{X}} |\hat{p}(x; \hat{\beta}) - \hat{p}(x; \beta^*)| \sup_{x \in \mathcal{X}} |\hat{m}_1(x) - m_1(x)| \frac{1}{nh_1^k} \sum_{i=1}^{n} \left| K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right| \left| \frac{2D_i}{\bar{p}_i^{1/2}} \right| 
\]

\[
= o_p(1). 
\]
Consider \(J_{43}(x_1)\). Denote that

\[
\lambda_1(X_i) = E \left[ \frac{\hat{p}(X_i; \beta^*) - D_i}{\hat{p}(X_i; \beta^*)} \right] X_i, \quad \mu_{i1} = \frac{\hat{p}(X_i; \beta^*) - D_i}{\hat{p}(X_i; \beta^*)} - \lambda_1(X_i),
\]

\[
\epsilon_{1i} = Y_i - m_1(X_i), \quad \rho_{ij} = \frac{K_3 \left( \frac{x_i - x_j}{h_3} \right)}{\sum_{t=1}^n D_t K_3 \left( \frac{x_i - x_t}{h_3} \right)}.
\]

We first give a lemma to show their asymptotics below, which is useful for the proof of the theorem.

**Lemma 1.** Under condition (A1)(ii), the outcome regression estimator satisfies

\[
|\rho_{ij} - \rho_{ji}| \leq \frac{C_n h_3}{n h_3^2} K_3 \left( \frac{X_i - X_j}{h_3} \right)
\]

where \(C_n = O_p(h_3)\) and does not depend on \(i, j\).

**Proof.** Note that \(\rho_{ij} = \rho_{ji} = 0\), if \(||X_i - X_j||_\infty > h_3\). We now consider the event that \(||X_i - X_j||_\infty \leq h_3\). For all \(i,\)

\[
\frac{1}{n h_3^p} \sum_{t=1}^n D_t K_3 \left( \frac{X_i - X_t}{h_3} \right) = \frac{\sum_{t=1}^n D_t K_3 \left( \frac{X_i - X_t}{h_3} \right)}{n h_3} \frac{1}{n h_3^p} \sum_{t=1}^n K_3 \left( \frac{X_i - X_t}{h_3} \right) = \hat{p}(X_i)\hat{\theta}(X_i).
\]

Then,

\[
|\rho_{ij} - \rho_{ji}| = \frac{1}{n h_3^p} \left| \sum_{t=1}^n D_t K_3 \left( \frac{X_i - X_t}{h_3} \right) \right| \left| \frac{1}{\hat{p}(X_i)\hat{\theta}(X_i)} - \frac{1}{\hat{p}(X_i)\theta(X_i)} \right| + \left| \frac{1}{p(X_i)\theta(X_i)} - \frac{1}{p(X_i)\theta(X_i)} \right|
\]

\[
+ \left| \frac{1}{p(X_i)\theta(X_i)} - \frac{1}{p(X_j)\hat{\theta}(X_j)} \right|.
\]
Again by the standard arguments for dealing with nonparametric estimation as we have used before and $s_3 \geq p$,

$$
\sup_{x \in X} |\hat{p}(x) - p(x)| = O_p \left( h_3^{s_3} + \sqrt{\frac{\ln n}{nh_3^p}} \right) = o_p(h_3),
$$

$$
\sup_{x \in X} |\hat{\theta}(x) - \theta(x)| = O_p \left( h_3^{s_3} + \sqrt{\frac{\ln n}{nh_3^p}} \right) = o_p(h_3).
$$

Recall the conditions (C2)(ii) and (C3)(i) that $p(x)$ and $\theta(x)$ are bounded away from 0 and 1. The two equations above implies that $\hat{p}(x)$ and $\hat{\theta}(x)$ uniformly converge to $p(x)$ and $\theta(x)$ respectively. We can also obtain that $\hat{p}(x)$ and $\hat{\theta}(x)$ are bounded away from 0 in probability for $n$ large enough.

Then

$$
\sup_{x \in X} \left| \frac{1}{p(x)\theta(x)} - \frac{1}{\hat{p}(x)\hat{\theta}(x)} \right| = \sup_{x \in X} \left| \frac{\hat{p}(x)\hat{\theta}(x) - p(x)\theta(x)}{p(x)\theta(x)\hat{p}(x)\hat{\theta}(x)} \right|
$$

$$
\leq \sup_{x \in X} \left| \frac{\hat{p}(x)\hat{\theta}(x) - \theta(x)}{p(x)\theta(x)\hat{p}(x)\hat{\theta}(x)} \right| + \left| \frac{\hat{p}(x) - p(x)}{p(x)\theta(x)\hat{p}(x)\hat{\theta}(x)} \right| = o_p(1).
$$

This leads to that $\left| \frac{1}{\hat{p}(X_i)\hat{\theta}(X_i)} - \frac{1}{p(X_i)\theta(X_i)} \right| = o_p(1)$, $\left| \frac{1}{\hat{p}(X_j)\hat{\theta}(X_j)} - \frac{1}{p(X_j)\theta(X_j)} \right| = o_p(1)$ uniformly over all $X_i, X_j$. By the Lipchitz continuity, we have $\left| \frac{1}{p(X_i)\theta(X_i)} - \frac{1}{p(X_j)\theta(X_j)} \right| = O_p(h_3)$ uniformly over all $X_i, X_j$. Altogether, we have that the summation in curly brace is $O_p(h_3)$. Therefore, there exists a $C_n = O_p(h_3)$ such that

$$
|\rho_{ij} - \rho_{ji}| \leq \frac{C_n}{nh_3^p} K_3 \left( \frac{X_i - X_j}{h_3} \right).
$$

The proof is completed. \qed
Now we come back to handle the term $J_{43}(x_1)$ that can be decomposed as

\[
J_{43}(x_1) = \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{\hat{p}(X_i; \beta^*) - D_i}{\hat{p}(X_i; \beta^*)} \left[ \hat{m}_1(X_i) - m_1(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
\]

\[
= \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \lambda_1(X_i) \left[ \hat{m}_1(X_i) - m_1(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) + \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \mu_{1i} \left[ \hat{m}_1(X_i) - m_1(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
\]

\[
= \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \lambda_1(X_i) \left[ \sum_{j=1}^{n} \rho_{ij} D_j [\epsilon_{1j} + m_1(X_j)] - m_1(X_i) \right] + \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \mu_{1i} \left[ \hat{m}_1(X_i) - m_1(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
\]

\[
= \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \lambda_1(X_i) \frac{D_i \epsilon_{1i}}{p(X_i)} + \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \frac{D_i \epsilon_{1i}}{p(X_i)} \left[ p(X_i) \sum_{j=1}^{n} \rho_{ji} K_1 \left( \frac{X_{1j} - x_1}{h_1} \right) \lambda_1(X_j) - K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \lambda_1(X_i) \right]
\]

\[
+ \frac{1}{\sqrt{nh_1^k}} K_1 \left( X_{1i} - x_1 \right) \lambda_1(X_i) \left[ \sum_{j=1}^{n} \rho_{ij} D_j m_1(X_j) - m_1(X_i) \right] + \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^{n} \mu_{1i} \left[ \hat{m}_1(X_i) - m_1(X_i) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)
\]

\[
:= J_{431}(x_1) + J_{432}(x_1) + J_{433}(x_1) + J_{434}(x_1).
\] (7.11)

We first prove that $J_{43k}(x_1) = o_p(1)$ for $k = 2, 3, 4$. Consider $J_{432}(x_1)$
by using the following decomposition:

\[
\begin{align*}
&\frac{1}{\sqrt{h_1}} \left[ p(X_i) \sum_{j=1}^{n} \rho_{ji} K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) - K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \lambda_1(X_i) \right] \\
&= \frac{1}{\sqrt{h_1}} p(X_i) \sum_{j=1}^{n} (\rho_{ij} - \rho_{ji}) K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) \\
&\quad + \frac{1}{\sqrt{h_1}} \left[ p(X_i) \sum_{j=1}^{n} \rho_{ij} K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) - K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \lambda_1(X_i) \right]
\end{align*}
\]

:= L_1 + L_2.

\(L_1\) can be bounded as

\[
L_1 \leq \frac{1}{\sqrt{h_1}} \sup_i \left| p(X_i) \sum_{j=1}^{n} \rho_{ij} - \rho_{ji} \right| K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) \\
\leq \frac{1}{\sqrt{h_1}} \sup_i \sum_{j=1}^{n} |\rho_{ij} - \rho_{ji}| \left| K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) \right| \\
\leq \frac{1}{\sqrt{h_1}} \frac{MC_n}{n h_3} \sum_{j=1}^{n} \left| K_3 \left( \frac{X_i - X_j}{h_3} \right) \right| \left| K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) \right| \\
\leq \frac{MC_n}{h_3} \frac{h_3}{\sqrt{h_1}} \sup_i \frac{1}{n h_3} \sum_{j=1}^{n} \left| K_3 \left( \frac{X_i - X_j}{h_3} \right) \right| \left| K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) \right| \\
= O_p(1) \cdot o_p(1) \cdot O_p(1) \\
= o_p(1).
\]
Then $L_2$ can be handled by noting

$$
\frac{1}{\sqrt{h_i^k}} \left[ p(X_i) \sum_{j=1}^{n} \rho_{ij} K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) - K_1 \left( \frac{X_{ii} - x_1}{h_1} \right) \lambda_1(X_i) \right]
$$

$$
= \frac{1}{\sqrt{h_i^k}} \left[ p(X_i) \sum_{s=1}^{n} K_3 \left( \frac{X_{i} - X_s}{h_3} \right) \sum_{j=1}^{n} D_j K_3 \left( \frac{X_j - X_i}{h_3} \right) K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) \right]
$$

$$
- K_1 \left( \frac{X_{ii} - x_1}{h_1} \right) \lambda_1(X_i)
$$

$$
= \frac{1}{\sqrt{h_i^k}} \left\{ \left[ p(X_i) \sum_{s=1}^{n} K_3 \left( \frac{X_{ii} - X_s}{h_3} \right) K_1 \left( \frac{X_{ii} - x_1}{h_1} \right) \lambda_1(X_i) \right]
$$

$$
+ \left[ p(X_i) \sum_{s=1}^{n} K_3 \left( \frac{X_{ii} - X_s}{h_3} \right) K_1 \left( \frac{X_{ii} - x_1}{h_1} \right) \lambda_1(X_i) \right]
$$

$$
- p(X_i) \frac{1}{p(X_i)} K_1 \left( \frac{X_{ii} - x_1}{h_1} \right) \lambda_1(X_i) \right] \right\}
$$

$$
= \frac{1}{\sqrt{h_i^k}} O_p \left( \frac{h_3^x}{h_1^2} + h_3^x \right) = O_p \left( \frac{h_3^x}{h_1^2 + h_3^2/2} \right).
$$

Then under Condition (A2)(ix), $L_2 = o_p(1)$ and, together with the bound

for $L_1$, we have

$$
\frac{1}{\sqrt{h_i^k}} \left[ p(X_i) \sum_{j=1}^{n} \rho_{ji} K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_1(X_j) - K_1 \left( \frac{X_{ii} - x_1}{h_1} \right) \lambda_1(X_i) \right] = o_p(1).
$$

Since $\{\epsilon_{1i}\}_{i=1}^{n}$ are mutually independent given $\{Z_i\}_{i=1}^{n}$, it follows that $J_{432}(x_1) = o_p(1)$.
Second, bound $J_{433}(x_1)$ by noting that

$$|J_{433}(x_1)| = \left| \frac{1}{\sqrt{nh_1^k}} K_1 (X_{1i} - x_1) \lambda_1 (X_i) \left[ \sum_{j=1}^n \rho_{ij} D_j m_1 (X_j) - m_1 (X_i) \right] \right|$$

$$\leq \sqrt{nh_1^k} \sup_{x \in X} \left| \sum_{j=1}^n \rho_{ij} D_j m_1 (X_j) - m_1 (X_i) \right| \frac{1}{nh_1^k} \sum_{i=1}^n |K_1 (X_{1i} - x_1) \lambda_1 (X_i)| \lambda_1 (X_i)$$

$$= \sqrt{nh_1^k} O_p (h_3 s_1) O_p (1) = o_p (1).$$

A similar argument to bound $J_{23}(x_1)$ can lead to $J_{434}(x_1) = o_p (1)$. Altogether and combining (7.11), we have $J_{43}(x_1) = J_{431}(x_1) + o_p (1)$. Recalling that $J_{42}(x_1), J_{44}(x_1)$ and $J_{45}(x_1)$ have all been proved to be $o_p (1)$, together with (7.10), we can conclude that

$$J(x_1) = J_{41}(x_1) + J_{431}(x_1) + o_p (1)$$

$$= \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n \left[ \frac{D_i}{p(X_i)} [Y_i - m_1 (X_i)] + m_1 (X_i) - \tau (x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)$$

$$+ \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \lambda_1 (X_i) \frac{D_i \epsilon_{1i}}{p(X_i)} + o_p (1)$$

$$= \frac{1}{\sqrt{nh_1^k}} \sum_{i=1}^n \left[ \frac{D_i}{p(X_i)} [Y_i - m_1 (X_i)] + m_1 (X_i) - \tau_1 (x) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) + o_p (1).$$

(7.12)

The proof is finished. □

**Scenario 2:** $m_1(x)$ is semiparametrically estimated. First, we
have
\[
\sup_{x \in \mathcal{X}} |\tilde{p}(x; \beta) - \tilde{p}(x; \beta^*)| = o_p \left( \frac{1}{\sqrt{n}} \right),
\]
\[
\sup_{x \in \mathcal{X}} |\hat{r}_1(B_1^T x) - r_1(B_1^T x)| = o_p \left( h_6^{\frac{5}{6}} + \sqrt{\frac{\ln n}{n h_6^{p(1)}}} \right) = o_p \left( h_6^{\frac{5}{6}} \right),
\]
We can further decompose the term in (7.1) as:
\[
J(x) = \frac{1}{\sqrt{n h_1}} \sum_{i=1}^{n} \left[ \frac{D_i}{\tilde{p}(X_i; \beta)} [Y_i - \hat{r}_1(B_1^T X_i)] + \hat{r}_1(B_1^T X_i) - \tau_1(x_i) \right] K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[
= \frac{1}{\sqrt{n h_1}} \sum_{i=1}^{n} \left[ \frac{D_i}{\tilde{p}(X_i; \beta^*)} [Y_i - m_1(X_i)] + m_1(X_i) - \tau(x_i) \right] K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[
+ \frac{1}{\sqrt{n h_1}} \sum_{i=1}^{n} \frac{D_i [r_1(B_1^T X_i) - Y_i]}{\tilde{p}(X_i; \beta^*)} \left[ \tilde{p}(X_i; \beta) - \tilde{p}(X_i; \beta^*) \right] K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[
+ \frac{1}{\sqrt{n h_1}} \sum_{i=1}^{n} \frac{D_i [Y_i - r_1^+]}{\tilde{p}(X_i; \beta^*)} \left[ \tilde{p}(X_i; \beta) - \tilde{p}(X_i; \beta^*) \right] ^2 K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[
+ \frac{1}{\sqrt{n h_1}} \sum_{i=1}^{n} \frac{2D_i [\tilde{p}(X_i; \beta) - \tilde{p}(X_i; \beta^*)]}{p_i^+} \left[ \tilde{r}_1(B_1^T X_i) - r_1(B_1^T X_i) \right] K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[
:= J_{51}(x) + J_{52}(x) + J_{53}(x) + 0 + J_{54}(x) + J_{55}(x)
\]
(7.13)
where $p_i^+$ lies between $\tilde{p}(X_i; \beta^*)$ and $\tilde{p}(X_i; \beta)$, $r_1^+$ lies between $r_1(B_1^T X_i)$ and $\hat{r}_1(B_1^T X_i)$.

Due to the similarity in the proof as the above, we omit the details for proving $J_{52}(x)$, $J_{54}(x)$ and $J_{55}(x)$ to be $o_p(1)$. Now consider $J_{53}(x)$. 


Denote that
\[
\lambda_2(X_i) = E \left[ \frac{\tilde{p}(X_i; \beta^*) - D_i}{\tilde{p}(X_i; \beta^*)} \right] X_i
\]
\[
\mu_{2i} = \frac{\tilde{p}(X_i; \beta^*) - D_i}{\tilde{p}(X_i; \beta^*)} - \lambda_2(X_i)
\]
\[
\epsilon_{2i} = Y_i - r_1(B_1^\top X_i)
\]
\[
\nu_{ij} = \frac{K_0 \left( \frac{B_1^\top x_i - B_1^\top x_j}{\gamma_0} \right)}{\sum_{t=1}^n D_t K_0 \left( \frac{B_1^\top x_i - B_1^\top x_t}{\gamma_0} \right)}
\]

\(J_{53}(x_1)\) can be rewritten as
\[
J_{53}(x_1) = \frac{1}{\sqrt{nh^4}} \sum_{i=1}^n \frac{\tilde{p}(X_i; \beta^*) - D_i}{\tilde{p}(X_i; \beta^*)} \left[ \hat{r}_1(B_1^\top X_i) - r_1(B_1^\top X_i) \right] K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[
= \frac{1}{\sqrt{nh^4}} \sum_{i=1}^n \lambda_2(X_i) \left[ \hat{r}_1(B_1^\top X_i) - r_1(B_1^\top X_i) \right] K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[
+ \frac{1}{\sqrt{nh^4}} \sum_{i=1}^n \mu_{2i} \left[ \hat{r}_1(B_1^\top X_i) - r_1(B_1^\top X_i) \right] K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[
= \frac{1}{\sqrt{nh^4}} \sum_{i=1}^n K_1 \left( \frac{X_{ii} - x_i}{h_1} \right) \lambda_2(X_i) \left[ \sum_{j=1}^n \rho_{ij} D_j [\epsilon_{2j} + r_1(B_j^\top X_j)] - r_1(B_1^\top X_i) \right]
\]
\[
+ \frac{1}{\sqrt{nh^4}} \sum_{i=1}^n \mu_{2i} \left[ \hat{r}_1(B_1^\top X_i) - r_1(B_1^\top X_i) \right] K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[
= \frac{1}{\sqrt{nh^4}} \sum_{i=1}^n D_i \epsilon_{2i} \left[ p(X_i) \sum_{j=1}^n \nu_{ji} K_1 \left( \frac{X_{ij} - x_i}{h_1} \right) \lambda_2(X_j) \right]
\]
\[
+ \frac{1}{\sqrt{nh^4}} K_1 \left( \frac{X_{ij} - x_i}{h_1} \right) \lambda_2(X_i) \left[ \sum_{j=1}^n \nu_{ij} D_j m_1(X_j) - m_1(X_i) \right]
\]
\[
+ \frac{1}{\sqrt{nh^4}} \sum_{i=1}^n \mu_{2i} \left[ \hat{m}_1(X_i) - m_1(X_i) \right] K_1 \left( \frac{X_{ii} - x_i}{h_1} \right)
\]
\[= J_{531}(x_1) + J_{532}(x_1) + J_{533}(x_1). \tag{7.14}\]

It is obvious that \(J_{532}(x_1) = o_p(1)\) and \(J_{533}(x_1) = o_p(1)\). To derive that
7.3 Proofs of Theorems 2 and 3

\( J_{331}(x_1) = o_p(1) \), we start by writing

\[
\frac{1}{\sqrt{h_k}} \left[ p(X_i) \sum_{j=1}^n \nu_{ji} K_1 \left( \frac{X_{1j} - x_1}{h_1} \right) \lambda_2(X_j) \right]
\]

\[
= \frac{1}{\sqrt{h_k}} p(X_i) \sum_{j=1}^n (\nu_{ij} - \nu_{ji}) K_1 \left( \frac{X_{1j} - x_1}{h_1} \right) \lambda_2(X_j)
\]

\[
+ \frac{1}{\sqrt{h_k}} \left[ p(X_i) \sum_{j=1}^n \nu_{ij} K_1 \left( \frac{X_{1j} - x_1}{h_1} \right) \lambda_2(X_j) \right]
\]

:= L_3 + L_4.

Similarly for proving \( L_1 = o_p(1) \) above, we can show that \( L_3 = o_p(1) \), and thus omit the details. To handle \( L_4 = o_p(1) \), we denote \( q_{B_1}(z) \) as the density of \( B_1^\top X \), and

\[
\theta_{B_1}(B_1^\top X) = E[Y(1)|B_1^\top X],
\]

\[
\tilde{\theta}_{B_1}(B_1^\top x) = \frac{\sum_{j=1}^n D_j Y_j K_6 \left( \frac{B_1^\top x_i - B_1^\top x_j}{h_6} \right)}{\sum_{t=1}^n D_t K_6 \left( \frac{B_1^\top x_i - B_1^\top x_t}{h_6} \right)},
\]

\[
\tilde{q}_{B_1}(B_1^\top x) = \frac{\sum_{j=1}^n D_j K_6 \left( \frac{B_1^\top x_i - B_1^\top x_j}{h_6} \right)}{\sum_{t=1}^n K_6 \left( \frac{B_1^\top x_i - B_1^\top x_t}{h_6} \right)}.
\]

Let \( T_1 = \frac{B_1^\top x - B_1^\top x_i}{h_6} \), \( T_1 = \frac{X_i - X_1}{h_6} \), \( T_3 = \frac{X - X_i}{h_6} \). To deal with \( L_4 \), consider
the conditional expectation that can be derived as:

$$E \left\{ p(X_i) \sum_{j=1}^{n} \nu_{ij} K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_2(X_j) \ \bigg| \ X_i \right\}$$

$$= E \left\{ p(X_i) \frac{1}{nh_6^{k(n)}} \sum_{j=1}^{n} K_6 \left( \frac{B_1 X_j - B_1 X_i}{h_6} \right) K_1 \left( \frac{X_{ij} - x_1}{h_1} \right) \lambda_2(X_j) \ \bigg| \ X_i \right\}$$

$$= \frac{[1 + o_p(1)]p(X_i)}{h_6^{k(n)} \theta_{B_1}(B_1 X_i) q_{B_1}(B_1 X_i)} \int K_6 \left( \frac{B_1 u - B_1 X_i}{h_6} \right) K_1 \left( \frac{u - x_1}{h_1} \right) \lambda_2(u) \theta(u) \, du$$

$$= \frac{h_6^{-p(1)} p(X_i)}{q_{B_1}(B_1 X_i) \theta_{B_1}(B_1 X_i)} K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \lambda_2(X_i) + O_p \left( \frac{h_6^{-p(1) + s_6}}{h_1^{s_6}} \right)$$

$$= o_p \left( \frac{h_6^{-p(1)} + \frac{h_6^{-p(1) + s_6}}{h_1^{s_6}}} \right).$$

Then, when $s_6 < (2s_6 + k)(p - p(1))$, $L_4 = O_p \left( \frac{h_6^{-p(1)} + \frac{h_6^{-p(1) + s_6}}{h_1^{s_6}}} \right) = o_p(1)$, $J_{531}(x_1) = o_p(1)$. Together with (7.14), $J_{53}(x_1) = o_p(1)$. Recall that we have proven that $J_{52}(x_1), J_{54}(x_1)$ and $J_{55}(x_1)$ can be bounded by $o_p(1)$. With (7.5), we can eventually derive the asymptotically linear representation as

$$J(x_1) = J_{51}(x_1) + o_p(1)$$

$$= \frac{1}{\sqrt{nh_1}} \sum_{i=1}^{n} \left[ \frac{D_i}{\hat{p}(X_i; \beta^*)} [Y_i - m_1(X_i)] + m_1(X_i) - \tau(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)$$

$$+ o_p(1). \quad (7.15)$$

The proof is completed. \(\Box\)
Scenario 3: \( m_1(x) \) is parametrically estimated (correctly specified). With the similar argument for proving scenario 1 in Theorem 1, we can easily derive that

\[
J(x_1) = \frac{1}{\sqrt{nh_1^2}} \sum_{i=1}^{n} \left[ \frac{D_i}{\hat{p}(X_i; \beta^*)} [Y_i - m_1(X_i)] + m_1(X_i) - \tau(x_1) \right] K_1 \left( \frac{X_{i1} - x_1}{h_1} \right) + o_p(1).
\]

(7.16)

With (7.12), it will be easy to further deduct the asymptotically linear expression of the proposed estimator. When the outcome regression functions are nonparametrically estimated, recalling the relation between the following and \( \sqrt{nh_1^2} [\hat{\tau}(x_1) - \tau(x_1)] \) and \( J(x_1) \) defined in (7.1), we can derive that

\[
\sqrt{nh_1^2}[\hat{\tau}(x_1) - \tau(x_1)] = \frac{1}{f(x_1)} \frac{1}{\sqrt{nh_1^2}} \sum_{i=1}^{n} [\Psi_1(X_i, Y_i, D_i) - \tau(x_1)] K_1 \left( \frac{X_{i1} - x_1}{h_1} \right).
\]

According to (7.15) and (7.16), when the outcome regression functions are semiparametrically or parametrically estimated, we have a similar representation as

\[
\sqrt{nh_1^2}[\hat{\tau}(x_1) - \tau(x_1)] = \frac{1}{f(x_1)} \frac{1}{\sqrt{nh_1^2}} \sum_{i=1}^{n} [\Psi_2(X_i, Y_i, D_i) - \tau(x_1)] K_1 \left( \frac{X_{i1} - x_1}{h_1} \right).
\]

Similarly as the proof for Theorem 1, we can derive that under the conditions of Theorem 2, when the outcome regression functions are estimated nonparametrically, we have

\[
\sqrt{nh_1^2}[\hat{\tau}(x_1) - \tau(x_1)] \overset{d}{\to} \mathcal{N} \left( 0, \frac{\sigma^2_1(x_1) \int K_1^2(u)du}{f(x_1)} \right),
\]
and when the outcome regression functions are estimated semiparametrically or parametrically, we have

\[ \sqrt{n h_1^k} [\hat{\tau}(x_1) - \tau(x_1)] \overset{d}{\to} \mathcal{N} \left( 0, \frac{\sigma_2^2(x_1) \int K_2^2(u) du}{f(x_1)} \right). \]

The proof of Theorem 2 is concluded. \( \square \)

As for Theorem 3, the proof can be very similar to the proof for Theorem 2. Here we only give a crucial lemma in this proof and omit the details of the proof.

**Lemma 2.** Under condition (A2)(viii), the propensity score estimator satisfies

\[ |\omega_{ij} - \omega_{ji}| \leq \frac{E_n}{nh_2^2} K_2 \left( \frac{X_i - X_j}{h_2} \right) \]

where \( E_n = O_p(h_2) \) free of \( i \) and \( j \).

### 7.4 Proof of Theorem 4

This is the case with local misspecification. To check the asymptotic efficiency through the variance comparison, we now compute the difference
7.4 Proof of Theorem 4

between $\sigma_1^2(x_1)$ and $\sigma_2^2(x_1)$:

$$\sigma_2^2(x_1) - \sigma_1^2(x_1)$$

$$= E \left\{ \frac{p(X) - \tilde{p}(X; \beta^*)}{[\tilde{p}(X; \beta^*)]^2} \text{Var}(Y | D = 1, X) + \frac{\tilde{p}(X; \beta^*) - p(X)}{[1 - \tilde{p}(X; \beta^*)]^2} \text{Var}(Y | D = 0, X) \right| X_1 = x_1 \right\}$$

$$= E \left\{ \frac{p(X) - \tilde{p}(X; \beta_0) + \tilde{p}(X; \beta_0) - \tilde{p}(X; \beta^*)}{[\tilde{p}(X; \beta^*)]^2} \text{Var}(Y | D = 1, X) + \frac{\tilde{p}(X; \beta^*) - \tilde{p}(X; \beta_0) + \tilde{p}(X; \beta_0) - p(X)}{[1 - \tilde{p}(X; \beta^*)]^2} \text{Var}(Y | D = 0, X) \right| X_1 = x_1 \right\}$$

(7.17)

and the difference between $\sigma_1^2(x_1)$ and $\sigma_3^2(x_1)$:

$$\sigma_3^2(x_1) - \sigma_1^2(x_1)$$

$$= E \left\{ \left[ \left( 1 - \frac{D}{p(X)} \right) [\tilde{m}_1(X; \gamma_1^*) - m_1(X)] - \left( 1 - \frac{1 - D}{1 - p(X)} \right) [\tilde{m}_0(X; \gamma_0^*) - m_0(X)] \right]^2 \right| X_1 = x_1 \right\}$$

$$= E \left\{ \frac{1 - p(X)}{p(X)} [\tilde{m}_1(X; \gamma_1^*) - \tilde{m}_1(X; \gamma_{10}) + \tilde{m}_1(X; \gamma_{10}) - m_1(X)]^2 
+ \frac{p(X)}{1 - p(X)} [\tilde{m}_0(X; \gamma_0^*) - \tilde{m}_0(X; \gamma_{00}) + \tilde{m}_0(X; \gamma_{00}) - m_0(X)]^2 
+ [\tilde{m}_1(X; \gamma_1^*) - \tilde{m}_1(X; \gamma_{10}) + \tilde{m}_1(X; \gamma_{10}) - m_1(X)] 
\times [\tilde{m}_0(X; \gamma_0^*) - \tilde{m}_0(X; \gamma_{00}) + \tilde{m}_0(X; \gamma_{00}) - m_0(X)] \right| X_1 = x_1 \right\}.$$  

(7.18)
Recall that as the definitions, for all $x \in \mathcal{X}$, there exists $\beta_0, \gamma_{10}, \gamma_{00}$, such that

\[
p(x) = \tilde{p}(x; \beta_0)[1 + c_n a(x)],
\]

\[
m_1(x) = \tilde{m}_1(x; \gamma_{10}) + d_{1n} b_1(x),
\]

\[
m_0(x) = \tilde{m}_0(x; \gamma_{00}) + d_{0n} b_0(x).
\]

That is, $p(x) - \tilde{p}(x; \beta_0) = O(c_n)$, $m_1(x) - \tilde{m}_1(x; \gamma_{10}) = O(d_{1n})$, and $m_0(x) - \tilde{m}_0(x; \gamma_{00}) = O(d_{0n})$. So now we only need to consider $\tilde{p}(x; \beta_0) - \tilde{p}(x; \beta^*)$, $\tilde{m}_1(x; \gamma_{10}) - \tilde{m}_1(x; \gamma^*_1)$ and $\tilde{m}_0(x; \gamma_{00}) - \tilde{m}_0(x; \gamma^*_0)$. Note that $\beta^*, \gamma^*_1, \gamma^*_0$ are the limits of the maximum likelihood estimators $\hat{\beta}, \hat{\gamma}_1, \hat{\gamma}_0$ respectively. Discuss $\beta^*$ first. Given the propensity score function, $D$ is bernoulli distributed.

We can respectively obtain, as the propensity score function would be misspecified, the quasi-likelihood function and the quasi-log likelihood function of the unknown parameter $\beta$:

\[
\tilde{L}(\beta) = \prod_{i=1}^n \tilde{p}(X_i; \beta)^{D_i}[1 - \tilde{p}(X_i; \beta)]^{1-D_i} f(X_i),
\]

and

\[
\tilde{l}(\beta) = \sum_{i=1}^n D_i \ln \tilde{p}(X_i; \beta) + (1 - D_i) \ln[1 - \tilde{p}(X_i; \beta)] + \ln f(X_i).
\]

Then, $\hat{\beta}$ and $\beta^*$ satisfy that

\[
\hat{\beta} = \arg\max_{\beta} \frac{1}{n} \tilde{l}(\beta), \quad \beta^* = \arg\max_{\beta} E[g(W; \beta)].
\]
where \( g(W; \beta) = D \ln \tilde{p}(X; \beta) + (1 - D) \ln[1 - \tilde{p}(X; \beta)] + \ln f(X) \). By the mean value theorem,

\[
E \left[ \frac{\partial g(W, \beta)}{\partial \beta} \bigg|_{\beta = \beta_0} \right] - E \left[ \frac{\partial g(W, \beta)}{\partial \beta} \bigg|_{\beta = \beta^*} \right] = E \left[ \frac{\partial^2 g(W, \beta)}{\partial \beta \partial \beta^\top} \bigg|_{\beta = \tilde{\beta}} \right] (\beta^* - \beta_0),
\]

and

\[
E \left[ \frac{\partial^2 g(W, \beta)}{\partial \beta \partial \beta^\top} \bigg|_{\beta = \tilde{\beta}} \right] (\beta^* - \beta_0) = E \left[ \frac{\partial g(W, \beta)}{\partial \beta} \bigg|_{\beta = \beta_0} \right]
\]

where \( \tilde{\beta} \) takes the value between \( \beta_0 \) and \( \beta^* \). Note that

\[
E \left[ \frac{\partial^2 \ln g(U, \beta)}{\partial \beta \partial \beta^\top} \bigg|_{\beta_0} \right] = O(c_n).
\]

Assume that \( E \left[ \frac{\partial^2 \ln g(U, \beta)}{\partial \beta \partial \beta^\top} \bigg|_{\beta} \right] \) is non-singular for any \( \beta \). We have

\[
\beta^* - \beta_0 = \left\{ E \left[ \frac{\partial^2 g(W, \beta)}{\partial \beta \partial \beta^\top} \bigg|_{\beta = \tilde{\beta}} \right] \right\}^{-1} E \left[ \frac{\partial g(W, \beta)}{\partial \beta} \bigg|_{\beta = \beta_0} \right] = O(c_n) = O(c_n). \]
7.5 Proofs of Theorems 5 and 6

The application of Taylor expansion yields that \( \tilde{p}(x; \beta_0) - \tilde{p}(x; \beta^*) = O(c_n) \).

Similar argument is devoted to deriving that \( \tilde{m}_1(x; \gamma_{10}) - \tilde{m}_1(x; \gamma_{1}^*) = O(d_{1n}) \) and \( \tilde{m}_0(x; \gamma_{00}) - \tilde{m}_0(x; \gamma_{0}^*) = O(d_{0n}) \). Together with these results, we continue to calculate the quantities in \( (7.17) \) and \( (7.18) \) to derive that

\[
\sigma_2^2(x_1) - \sigma_1^2(x_1) = O(c_n),
\]

\[
\sigma_3^2(x_1) - \sigma_1^2(x_1) = O(d_{1n}^2) + O(d_{0n}^2) + O(d_{1n}d_{0n}).
\]

These differences show that when only the propensity score function is or only the outcome regression functions are locally misspecified, the asymptotic distribution remains the same as that without misspecification. \( \Box \)

7.5 Proofs of Theorems 5 and 6

Consider the cases with all models misspecified. The proof of Theorem 5 will be very similar to the proof of scenario 1 in Theorem 6 except that the asymptotic linear expression can be as

\[
\sqrt{nh_1^k} [\hat{\tau}(x_1) - \tau(x_1)] = \frac{1}{\sqrt{nh_1^k} f(x_1)} \sum_{i=1}^n \left[ \Psi_4(X_i, Y_i, D_i) - \tau(x_1) \right] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) + o_p(1).
\]
As the unbiasedness no longer holds, we then compute the bias term. A decomposition is as follows:

\[
E \left\{ \sqrt{n h_1^k} [\hat{\tau}(x_1) - \tau(x_1)] \right\} = \sqrt{n h_1^k} E \left\{ \left( \frac{D}{\hat{p}(X; \beta^*)} - \frac{D}{p(X)} \right) [Y - m_1(X)] + \left( 1 - \frac{D}{\hat{p}(X; \beta^*)} \right) [\hat{m}_1(X; \gamma_1^*) - m_1(X)] - \left( \frac{D}{p(\hat{X}; \beta^*)} - \frac{D}{p(X)} \right) [Y - m_0(X)] + \left( 1 - \frac{D}{\hat{p}(X; \beta^*)} \right) [\hat{m}_0(X; \gamma_0^*) - m_0(X)] \mid X_1 = x_1 \right\} \\
= \sqrt{n h_1^k} E \left\{ \frac{[m_1(X) - \hat{m}_1(X; \gamma_1^*)] [p(X) - \hat{p}(X; \beta^*)]}{\hat{p}(X; \beta^*)} - \frac{[m_0(X) - \hat{m}_0(X; \gamma_0^*)] [\hat{p}(X; \beta^*) - p(X)]}{1 - \hat{p}(X; \beta^*)} \mid X_1 = x_1 \right\} := \sqrt{n h_1^k} bias(x_1).
\]

Let

\[
\bar{\tau}(x_1) = E \left\{ \left[ \frac{D}{\hat{p}(X; \beta^*)} [Y - \hat{m}_1(X; \gamma_1^*)] - \frac{1 - D}{1 - \hat{p}(X; \beta^*)} [Y - \hat{m}_0(X; \gamma_0^*)] + \hat{m}_1(X; \gamma_1^*) - \hat{m}_0(X; \gamma_0^*) \right] \mid X_1 = x_1 \right\}.
\]
The variance term of $\sqrt{nh_1^k} [\hat{\tau}(x_1) - \tau(x_1) - bias(x_1)]$ can be derived as:

$$Var \left\{ \frac{1}{\sqrt{nh_1^k}} f(x_1) \sum_{i=1}^{n} [\Psi_4(X_i, Y_i, D_i) - \tau(x_1) - bias(x_1)] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right) \right\}$$

$$= E \left\{ \frac{1}{\sqrt{nh_1^k}} f(x_1) \sum_{i=1}^{n} [\Psi_4(X_i, Y_i, D_i) - \bar{\tau}(x_1)] K_1 \left( \frac{X_{1i} - x_1}{h_1} \right)^2 \right\}$$

$$= - \frac{h_1^k}{f^2(x_1)} E \left\{ E \left[ \left[ \Psi_4(X,Y,D) - \bar{\tau}(x_1) \right] \frac{1}{h_1^k} K_1 \left( \frac{X_1 - x_1}{h_1} \right) \right]^2 \left| X_1 \right] \right\}$$

$$= - \frac{h_1^k}{f^2(x_1)} \int K_1^2(u) E \left[ \left[ \Psi_4(X,Y,D) - \bar{\tau}(x_1) \right]^2 \left| X_1 = x_1 + h_1 u \right] f(x_1 + h_1 u) du \right.$$  

$$= \frac{\sigma_4^2(x_1)}{f(x_1)} + O(h_1^k).$$

where

$$\sigma_4^2(x_1) = E \left[ \left[ \Psi_4(X,Y,D) - \bar{\tau}(x_1) \right]^2 \mid X_1 = x_1 \right].$$

With the same argument to derive the asymptotic distribution in Theorem 1, we can obtain that

$$\sqrt{nh_1^k} [\hat{\tau}(x_1) - \tau(x_1) - bias(x_1)] \overset{d}{\to} N \left( 0, \frac{\sigma_4^2(x_1)}{f(x_1)} \int K_1^2(u) du \right). \quad (7.19)$$

The proof of Theorem 5 is completed. \qed

Note that Theorem 6 is a variant of Theorem 5. To derive the asymptotic distribution, we only need to consider the bias term and the variance term based on (7.19) when all nuisance models are locally misspecified.

From the definitions of misspecified models before, the bias term can be
bounded by

\[ \text{bias}(x_1) = O_p(c_n d_{1n}) + O_p(c_n d_{0n}). \]

This result implies that if the convergence rates of \(c_n d_{1n}\) and \(c_n d_{0n}\) are faster than \(O\left( \frac{1}{\sqrt{nh_1^k}} \right)\), the bias term vanishes asymptotically. By the central limit theorem, we can also derive the asymptotic normality with the variance term. By (7.19) and (7.19) when \(c_n, d_{1n}\) and \(d_{0n}\) all converge to 0, we have

\[
\frac{\sigma^2_1(x_1) \int K^2_1(u) du}{f(x_1)} = \frac{\sigma^2_1(x_1) \int K^2_1(u) du}{f(x_1)} + o(1).
\]

With Slutsky’s Theorem, we can conclude that, when all nuisance models are locally misspecified, \(c_n d_{1n} = o\left( \frac{1}{\sqrt{nh_1^k}} \right)\), and \(c_n d_{0n} = o\left( \frac{1}{\sqrt{nh_1^k}} \right)\), we then have

\[
\sqrt{nh_1^k}[\hat{\tau}(x_1) - \tau(x_1)] \rightarrow_d N\left( 0, \frac{\sigma^2_1(x_1) \int K^2_1(u) du}{f(x_1)} \right).
\]

Then the proof is completed. \[\square\]

7.6 A simple justification for Remark 5

As we showed in the proof of Theorem 4,

\[
\sigma^2_1(x_1) - \sigma^2_1(x_1) = E \left\{ \frac{p(X) - \tilde{p}(X; \beta^*)}{[\tilde{p}(X; \beta^*)]^2} \right\} Var(Y|D = 1, X) \left| X_1 = x_1 \right\} + E \left\{ \frac{\tilde{p}(X; \beta^*) - p(X)}{[1 - \tilde{p}(X; \beta^*)]^2} \right\} Var(Y|D = 0, X) \left| X_1 = x_1 \right\}.
\]
This difference cannot be showed either positive or negative for all $x_1$. The example in Remark 2 confirms this. For $\sigma_3^2(x_1)$ we have

$$\sigma_3^2(x_1) - \sigma_1^2(x_1) = E \left\{ \left[ \left( 1 - \frac{D}{p(X)} \right) \right] \left[ \tilde{m}_1(X; \gamma_1^*) - m_1(X) \right] - \left( 1 - \frac{1 - D}{1 - p(X)} \right) \left[ \tilde{m}_0(X; \gamma_0^*) - m_0(X) \right] \right\}^2 \bigg| X_1 = x_1 \right\} \geq 0.$$ 

In other words, the variance with $\sigma_3^2(x_1)$ can be larger than that of the estimators with all models correctly specified. Further,

$$\sigma_4^2(x_1) - \sigma_1^2(x_1) = Var(\Psi_4(X,Y,D)|X_1 = x_1) - Var(\Psi_1(X,Y,D)|X_1 = x_1)$$

$$= E \left\{ \left( \frac{p(X)}{\tilde{p}(X; \beta^*)} - \frac{1}{p(X)} \right) Var(Y|X,D = 1) \bigg| X_1 = x_1 \right\}$$

$$- E \left\{ \left( \frac{1 - p(X)}{[1 - \tilde{p}(X; \beta^*)]^2} - \frac{1}{1 - p(X)} \right) Var(Y|X,D = 0) \bigg| X_1 = x_1 \right\}$$

$$+ E \left\{ \frac{p(X)}{[\tilde{p}(X; \beta^*)]^2} \left[ m_1(X) - \tilde{m}_1(X; \gamma_1^*) \right]^2 + \frac{1 - p(X)}{[1 - \tilde{p}(X; \beta^*)]^2} \left[ m_0(X) - \tilde{m}_0(X; \gamma_0^*) \right]^2 \right\} \bigg| X_1 = x_1 \right\}$$

$$+ 2 E \left\{ \left[ \tilde{m}_1(X; \gamma_1^*) - \tilde{m}_0(X; \gamma_0^*) \right][m_1(X) - m_0(X) - \tilde{m}_1(X; \gamma_1^*) + \tilde{m}_0(X; \gamma_0^*)] \right\} \bigg| X_1 = x_1 \right\}$$

$$+ \tau^2(x_1) - \tilde{\tau}^2(x_1).$$

Again $\sigma_4^2(x_1)$ cannot be easily judged whether it is larger than $\sigma_1^2(x_1)$ or not. \[\square\]
### Table 6: The simulation results under model 2 (part 1)

| DRCATE | $x_1$ | bias | sam-SD | MSE | $P_{0.05}$ | $P_{0.95}$ | bias | sam-SD | MSE | $P_{0.05}$ | $P_{0.95}$ |
|--------|------|------|--------|-----|-----------|-----------|------|--------|-----|-----------|-----------|
| **(O,O)** | -0.4 | 0.0012 | 0.2077 | 0.0431 | 0.058 | 0.045 | 0.0006 | 0.2034 | 0.0414 | 0.045 | 0.047 |
|        | -0.2 | 0.0010 | 0.2139 | 0.0457 | 0.040 | 0.052 | 0.0001 | 0.1988 | 0.0395 | 0.050 | 0.050 |
|        | 0    | -0.0005 | 0.2046 | 0.0418 | 0.048 | 0.059 | 0.0007 | 0.1846 | 0.0341 | 0.050 | 0.050 |
|        | 0.2  | -0.0001 | 0.2226 | 0.0495 | 0.045 | 0.050 | 0.0008 | 0.2034 | 0.0415 | 0.038 | 0.057 |
|        | 0.4  | 0.0024 | 0.3312 | 0.1097 | 0.048 | 0.049 | 0.0010 | 0.3114 | 0.0971 | 0.044 | 0.053 |
| **(cP,cP)** | -0.4 | 0.0011 | 0.2077 | 0.0431 | 0.056 | 0.048 | 0.0006 | 0.2035 | 0.0415 | 0.046 | 0.046 |
|        | -0.2 | 0.0009 | 0.2137 | 0.0456 | 0.045 | 0.053 | 0.0001 | 0.1988 | 0.0395 | 0.049 | 0.050 |
|        | 0    | -0.0006 | 0.2044 | 0.0417 | 0.047 | 0.055 | 0.0007 | 0.1846 | 0.0341 | 0.048 | 0.052 |
|        | 0.2  | -0.0001 | 0.2228 | 0.0496 | 0.046 | 0.049 | 0.0007 | 0.2035 | 0.0415 | 0.038 | 0.057 |
|        | 0.4  | 0.0024 | 0.3316 | 0.1100 | 0.047 | 0.047 | 0.0010 | 0.3114 | 0.0971 | 0.044 | 0.053 |
| **(N,N)** | -0.4 | 0.0002 | 0.2653 | 0.0703 | 0.017 | 0.029 | 0.0004 | 0.2136 | 0.0456 | 0.057 | 0.052 |
|        | -0.2 | 0.0011 | 0.2300 | 0.0529 | 0.042 | 0.048 | 0.0004 | 0.1990 | 0.0396 | 0.041 | 0.045 |
|        | 0    | 0.0007 | 0.1962 | 0.0385 | 0.048 | 0.051 | 0.0003 | 0.1917 | 0.0367 | 0.041 | 0.052 |
|        | 0.2  | 0.0011 | 0.2299 | 0.0528 | 0.043 | 0.058 | 0.0006 | 0.2122 | 0.0451 | 0.046 | 0.052 |
|        | 0.4  | 0.0041 | 0.3373 | 0.1141 | 0.054 | 0.057 | 0.0003 | 0.3125 | 0.0976 | 0.050 | 0.052 |
| **(S,S)** | -0.4 | -0.0018 | 0.2058 | 0.0424 | 0.051 | 0.046 | 0.0002 | 0.2501 | 0.0625 | 0.028 | 0.040 |
|        | -0.2 | -0.0021 | 0.2093 | 0.0439 | 0.056 | 0.039 | -0.0008 | 0.2087 | 0.0436 | 0.046 | 0.047 |
|        | 0    | 0.0000 | 0.2040 | 0.0416 | 0.055 | 0.051 | 0.0011 | 0.1868 | 0.0351 | 0.044 | 0.056 |
|        | 0.2  | 0.0060 | 0.2257 | 0.0518 | 0.031 | 0.068 | 0.0014 | 0.2093 | 0.0441 | 0.047 | 0.059 |
|        | 0.4  | 0.0089 | 0.3409 | 0.1181 | 0.039 | 0.064 | 0.0010 | 0.3298 | 0.1089 | 0.043 | 0.062 |
Table 7: The simulation results under model 2 (part 2)

| DRCATE  | $x_1$ | n=500 bias | n=5000 bias | n=500 sam-SD | n=5000 sam-SD | n=500 MSE | n=5000 MSE | P_{0.05} | P_{0.95} | P_{0.05} | P_{0.95} |
|---------|-------|-------------|-------------|--------------|--------------|-----------|-----------|---------|---------|---------|---------|
| (O,O)   | -0.4  | 0.0012      | 0.0006      | 0.2077       | 0.2034       | 0.0431    | 0.0414    | 0.045   | 0.047   | 0.058   | 0.045   |
|         | -0.2  | 0.0010      | 0.0001      | 0.2139       | 0.1988       | 0.0457    | 0.0395    | 0.050   | 0.050   | 0.040   | 0.052   |
|         | 0     | -0.0005     | 0.0007      | 0.2046       | 0.1846       | 0.0418    | 0.0341    | 0.050   | 0.050   | 0.048   | 0.050   |
|         | 0.2   | -0.0001     | 0.0008      | 0.2226       | 0.2034       | 0.0495    | 0.0415    | 0.038   | 0.057   | 0.045   | 0.050   |
|         | 0.4   | 0.0024      | 0.0010      | 0.3312       | 0.3114       | 0.1097    | 0.0971    | 0.044   | 0.053   | 0.048   | 0.044   |
| (mP,cP)| -0.4  | 0.0011      | 0.0006      | 0.2082       | 0.2042       | 0.0433    | 0.0417    | 0.050   | 0.045   | 0.058   | 0.045   |
|         | -0.2  | 0.0009      | 0.0001      | 0.2123       | 0.1974       | 0.0451    | 0.0389    | 0.051   | 0.053   | 0.044   | 0.053   |
|         | 0     | -0.0005     | 0.0006      | 0.2025       | 0.1834       | 0.0410    | 0.0337    | 0.050   | 0.052   | 0.051   | 0.050   |
|         | 0.2   | -0.0002     | 0.0007      | 0.2222       | 0.2030       | 0.0493    | 0.0413    | 0.037   | 0.056   | 0.051   | 0.050   |
|         | 0.4   | 0.0025      | 0.0011      | 0.3315       | 0.3116       | 0.1099    | 0.0972    | 0.043   | 0.053   | 0.048   | 0.043   |
| (mP,N)  | -0.4  | -0.0011     | -0.0005     | 0.2156       | 0.2082       | 0.0464    | 0.0434    | 0.048   | 0.043   | 0.056   | 0.048   |
|         | -0.2  | -0.0019     | -0.0013     | 0.2086       | 0.2062       | 0.0436    | 0.0428    | 0.058   | 0.036   | 0.061   | 0.058   |
|         | 0     | -0.0028     | -0.0011     | 0.2003       | 0.1888       | 0.0403    | 0.0358    | 0.058   | 0.040   | 0.057   | 0.058   |
|         | 0.2   | 0.0021      | -0.0004     | 0.2108       | 0.2099       | 0.0445    | 0.0440    | 0.054   | 0.044   | 0.052   | 0.054   |
|         | 0.4   | 0.0060      | 0.0033      | 0.3258       | 0.3276       | 0.1069    | 0.1093    | 0.033   | 0.069   | 0.045   | 0.044   |
| (mP,S)  | -0.4  | -0.0034     | -0.0010     | 0.2215       | 0.2119       | 0.0493    | 0.0451    | 0.053   | 0.043   | 0.054   | 0.043   |
|         | -0.2  | -0.0055     | -0.0034     | 0.2235       | 0.2115       | 0.0507    | 0.0469    | 0.073   | 0.029   | 0.060   | 0.073   |
|         | 0     | -0.0023     | -0.0025     | 0.2049       | 0.1895       | 0.0421    | 0.0371    | 0.061   | 0.032   | 0.051   | 0.061   |
|         | 0.2   | -0.0003     | -0.0003     | 0.2149       | 0.1982       | 0.0462    | 0.0393    | 0.052   | 0.043   | 0.043   | 0.052   |
|         | 0.4   | 0.0102      | 0.0034      | 0.3351       | 0.3122       | 0.1148    | 0.0997    | 0.039   | 0.058   | 0.032   | 0.039   |
Table 8: The simulation results under model 2 (part 3)

| DRCATE   | $x_1$ | bias  | sam-SD | MSE | $P_{0.05}$ | $P_{0.95}$ | bias  | sam-SD | MSE | $P_{0.05}$ | $P_{0.95}$ |
|----------|-------|-------|--------|-----|------------|------------|-------|--------|-----|------------|------------|
|          | n=500 |       |        |     |            |            | n=5000 |        |     |            |            |
| (O,O)    |       |       |        |     |            |            |       |        |     |            |            |
| -0.4     | 0.0012 | 0.2077 | 0.0431 | 0.058 | 0.045      | 0.0006     | 0.2034 | 0.0414 | 0.045 | 0.047      |
| -0.2     | 0.0010 | 0.2139 | 0.0457 | 0.040 | 0.052      | 0.0001     | 0.1988 | 0.0395 | 0.050 | 0.050      |
| 0        | -0.0005 | 0.2046 | 0.0418 | 0.048 | 0.059      | 0.0007     | 0.1846 | 0.0341 | 0.050 | 0.050      |
| 0.2      | -0.0001 | 0.2226 | 0.0495 | 0.045 | 0.050      | 0.0008     | 0.2034 | 0.0415 | 0.038 | 0.057      |
| 0.4      | 0.0024 | 0.3312 | 0.1097 | 0.048 | 0.049      | 0.0010     | 0.3114 | 0.0971 | 0.044 | 0.053      |
| (cP,mP)  |       |       |        |     |            |            |       |        |     |            |            |
| -0.4     | 0.0008 | 0.2179 | 0.0474 | 0.051 | 0.043      | 0.0004     | 0.2204 | 0.0485 | 0.048 | 0.045      |
| -0.2     | 0.0012 | 0.2233 | 0.0498 | 0.049 | 0.052      | 0.0003     | 0.2069 | 0.0428 | 0.049 | 0.054      |
| 0        | -0.0004 | 0.2104 | 0.0442 | 0.051 | 0.060      | 0.0008     | 0.1890 | 0.0358 | 0.050 | 0.053      |
| 0.2      | -0.0002 | 0.2226 | 0.0495 | 0.048 | 0.049      | 0.0007     | 0.2039 | 0.0416 | 0.036 | 0.054      |
| 0.4      | 0.0028 | 0.3407 | 0.1162 | 0.047 | 0.050      | 0.0010     | 0.3170 | 0.1006 | 0.043 | 0.053      |
| (N,mP)   |       |       |        |     |            |            |       |        |     |            |            |
| -0.4     | -0.0050 | 0.2225 | 0.0501 | 0.060 | 0.036      | -0.0006    | 0.2227 | 0.0496 | 0.051 | 0.050      |
| -0.2     | -0.0015 | 0.2185 | 0.0477 | 0.056 | 0.043      | -0.0011    | 0.1931 | 0.0375 | 0.054 | 0.039      |
| 0        | -0.0020 | 0.2039 | 0.0416 | 0.072 | 0.032      | -0.0013    | 0.1857 | 0.0348 | 0.056 | 0.038      |
| 0.2      | 0.0024 | 0.2178 | 0.0475 | 0.042 | 0.051      | 0.0005     | 0.2064 | 0.0426 | 0.039 | 0.056      |
| 0.4      | 0.0046 | 0.3259 | 0.1066 | 0.035 | 0.064      | 0.0023     | 0.3324 | 0.1115 | 0.044 | 0.050      |
| (S,mP)   |       |       |        |     |            |            |       |        |     |            |            |
| -0.4     | -0.0115 | 0.2117 | 0.0481 | 0.075 | 0.027      | -0.0024    | 0.3260 | 0.1073 | 0.020 | 0.017      |
| -0.2     | -0.0021 | 0.2083 | 0.0434 | 0.051 | 0.051      | -0.0021    | 0.2010 | 0.0412 | 0.065 | 0.033      |
| 0        | -0.0018 | 0.2002 | 0.0401 | 0.044 | 0.053      | -0.0005    | 0.2045 | 0.0418 | 0.045 | 0.038      |
| 0.2      | 0.0035 | 0.2290 | 0.0527 | 0.044 | 0.071      | 0.0001     | 0.2155 | 0.0464 | 0.054 | 0.054      |
| 0.4      | 0.0017 | 0.3460 | 0.1196 | 0.040 | 0.064      | 0.0015     | 0.3519 | 0.1241 | 0.031 | 0.054      |
Abrevaya, J., Y.-C. Hsu, and R. P. Lieli (2015). Estimating conditional average treatment effects. *Journal of Business & Economic Statistics* 33(4), 485–505.

Fan, Q., Y.-C. Hsu, R. P. Lieli, and Y. Zhang (2019). Estimation of conditional average treatment effects with high-dimensional data. *arXiv preprint arXiv:1908.02399*.

Hu, Z., D. A. Follmann, and N. Wang (2014). Estimation of mean response via the effective balancing score. *Biometrika* 101(3), 613–624.

Lee, S., R. Okui, and Y.-J. Whang (2017). Doubly robust uniform confidence band for the conditional average treatment effect function. *Journal of Applied Econometrics* 32(7), 1207–1225.

Li, L., N. Zhou, and L. Zhu (2020). Outcome regression-based estimation of conditional average treatment effect. *Submitted*.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association* 89(427), 846–866.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–55.

Scharfstein, D. O., A. Rotnitzky, and J. M. Robins (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*.
Seaman, S. R. and S. Vansteelandt (2018). Introduction to double robust methods for incomplete data. *Statistical science: a review journal of the Institute of Mathematical Statistics* 33(2), 184.

Shi, C., W. Lu, and R. Song (2019, 04). A sparse random projection-based test for overall qualitative treatment effects. *Journal of the American Statistical Association*, 1–41.

Zhou, N. and L. Zhu (2020). On ipw-based estimation of conditional average treatment effect. Submitted.

Zimmert, M. and M. Lechner (2019). Nonparametric estimation of causal heterogeneity under high-dimensional confounding. *arXiv preprint arXiv:1908.08779*. 