Annals of the Institute of Statistical Mathematics (2022) 74:987–1041
https://doi.org/10.1007/s10463-022-00821-x

Outcome regression-based estimation of conditional average treatment effect

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

Received: 2 December 2020 / Revised: 17 January 2022 / Accepted: 27 January 2022 /
Published online: 29 April 2022

Abstract

The research is about a systematic investigation on the following issues. First, we construct different outcome regression-based estimators for conditional average treatment effect under, respectively, true, parametric, nonparametric and semiparametric dimension reduction structure. Second, according to the corresponding asymptotic variance functions when supposing the models are correctly specified, we answer the following questions: what is the asymptotic efficiency ranking about the four estimators in general? how is the efficiency related to the affiliation of the given covariates in the set of arguments of the regression functions? what do the roles of bandwidth and kernel function selections play for the estimation efficiency; and in which scenarios should the estimator under semiparametric dimension reduction regression structure be used in practice? Meanwhile, the results show that any outcome regression-based estimation should be asymptotically more efficient than any inverse probability weighting-based estimation. Several simulation studies are conducted to examine the finite sample performances of these estimators, and a real dataset is analyzed for illustration.

Keywords Asymptotic variance · Conditional average treatment effect · Regression causal effect · Sufficient dimension reduction

The first two authors are co-first authors. The research was supported by a grant from the University Grants Council of Hong Kong (HKBU12302720) and a grant from the National Natural Science Foundation of China (NSFC12131006).

✉ Lixing Zhu
lzhu@hkbu.edu.hk

[1] School of Mathematical Sciences, Shanghai Jiao Tong University, 800 Dongchuan Rd, Shanghai 200240, China

[2] Center for Statistics and Data Science, Beijing Normal University, 18 Jinfeng Rd, Zhuhai 519087, China

[3] Department of Mathematics, Hong Kong Baptist University, 224 Waterloo Road, Hong Kong, China

1 Introduction

Causal inference has been widely applied for decades to analyse treatment effect based on observational studies, in which treatments are assigned to observations in a non-random fashion. In this paper, we consider causal inference under the potential outcome framework (Rubin 1974; Rosenbaum and Rubin 1983) where the treatment is binary and the outcome variable in the hypothetical complete data set has two components $(Y_{(1)}, Y_{(0)})$. In which $Y_{(1)}$ is the potential outcome if the individual receives treatment and $Y_{(0)}$ is the corresponding potential outcome without treatment. As we can only observe one of $Y_{(1)}$ and $Y_{(0)}$, a commonly used method is to impute a reasonable value in the lieu of the missing one such as linear regression imputation Healy and Westmacott (1956), kernel regression imputation Cheng (1994) and ratio imputation Rao (1996).

In this paper, we consider average treatment effect (ATE, see Rosenbaum and Rubin 1983, 1985) conditional on some covariates to explore the heterogeneity of ATE. Let $X \in \mathbb{R}^p$ be a set of covariates that collects individual's personal information and $X_1 \in \mathbb{R}^k$ be a subvector of X, $1 \le k < p$. Conditional average treatment effect (CATE, hereafter) is defined as $E(Y_{(1)} - Y_{(0)}|X_1)$. To estimate this function, Abrevaya et al. (2015) proposed estimators that are based on inverse probability weighting (IPW, hereafter) method and concluded that, according to the asymptotic variance functions, the estimator with nonparametrically estimated inverse probability (IPW-N) is asymptotically more efficient than the one with parametrically estimated inverse probability (IPW-P). The relevant conclusion is similar to that in Hahn (1998) and Hirano et al. (2003) for the IPW estimators of ATE. But, IPW-P is proved to be asymptotically equivalent to the oracle estimator with true propensity score (IPW-O). This is very different from the unconditional ATE. Zhou and Zhu (2021) proposed an estimator with semiparametically estimated propensity score (IPW-S) and gave some more detailed analysis on the asymptotic efficiency on IPW-N and IPW-S.

As well known, for ATE, outcome regression-based estimation is also a popularly used methodology. Thus, methodologically, the research in this aspect is not new. However, for CATE, the problem becomes more complicated as it involves double conditional expectations on the full set X, or subset $\boldsymbol{\beta}^\top X$ of covariates, if the curse of dimensionality is concerned within dimension reduction framework, and the subset X_1 where $\boldsymbol{\beta}$ is a projection matrix. Three relevant references are Luo et al. (2017), Zhang et al. (2018), Luo et al. (2019) and Ma et al. (2019). To focus on the estimation efficiency issue, we in this paper do not give more details about how to work on dimension reduction and feature selection, while only consider the general setting supposing that a dimension reduction structure already exists. We then consider a systematic investigation on their asymptotic properties to answer the following questions when the model is correctly specified in parametric case.

Q1. When CATE is estimated under nonparametric, semiparametric, parametric and true (oracle) regression structure, what ranking of the asymptotic efficiency can be achieved for these estimators?

Q2. Note that CATE is a function of X_1 and the set of arguments of the regression function, say $\tilde{X}$ that is not necessary to be the full X, and thus X_1 is not necessary to be a strict subset of $\tilde{X}$. Then could the affiliation of X_1 to $\tilde{X}$ affect the asymptotic efficiency of different estimators? This issue is unique for CATE and particularly important under semiparametic dimension reduction framework as the regression function would be a function of $\tilde{X} = \boldsymbol{\beta}^{\top} X$ where $\boldsymbol{\beta}$ is a $p \times r$ matrix with $r \ll p$ in high dimensional scenarios.

Q3. As all estimators use nonparametric estimations for the involved conditional expectations, how could the bandwidth and kernel function affect the efficiency? This study is particularly necessary.

Q4. Comparing with the IPW-based estimation, what efficiency ranking should be concluded?

We will have a very brief discussion in Sect. 5 about the misspecified cases, globally or locally, that will be investigated in the near future, but not be touched in this paper.

Note that CATE is

$$\tau(x_1) = E[(Y_{(1)} - Y_{(0)})|X_1 = x_1] = E[E(Y_{(1)} - Y_{(0)}|X)|X_1 = x_1], \tag{1}$$

where $E(Y_{(1)} - Y_{(0)}|X)$ is the treatment effect heterogeneity. We are interested in, under unconfoundedness assumption, estimating $\tau(x_1)$ in this paper. To well answer the above four questions, we suggest/propose four outcome regression-based estimators (OR, hereafter) when assuming that $m_1(X) - m_0(X) = E(Y_{(1)} - Y_{(0)}|X)$ is completely known function (written as OR-O), parametric function (written as OR-P) ($m_1(X) = m_1(X, \theta_1)$ and $m_0(X) = m_0(X, \theta_0)$), semiparametric function with dimension reduction structure (written as OR-S) ($m_1(X) = m_1(\beta_1^{\top} X)$ and $m_0(X) = m_0(\beta_0^{\top} X)$), and nonparametric function (written as OR-N). The details will be in Sect. 2. When the corresponding nonparametric functions are estimated by, say, kernel estimation, we derive the asymptotically linear representations and asymptotic normality of these estimators in various scenarios and, according to the asymptotic variance functions and using the estimators with true regression/propensity score as the benchmark, we obtain the following results to give a relatively complete picture for the asymptotic efficiencies of the four estimation methods. The following newly derived results show that the estimated CATEs have rather different asymptotic behaviors from the estimated ATEs. Let $A \preceq B$ mean that method A has smaller asymptotic variance function than method B, and $A \cong B$ stand for the asymptotic equivalence of them when the asymptotic variance functions are equal. The results are summarised as follows.

A1. This is the answer for Q1 and Q4. In general, the ranking for the asymptotic efficiencies of the estimators is, together with the results about the IPW-based estimators respectively in Abrevaya et al. (2015) and Zhou and Zhu (2021):

$$\overbrace{\text{OR-O} \cong \text{OR-P} \preceq \text{OR-S} \preceq \text{OR-N}}^{\text{regression-based CATE estimators}} = \overbrace{\text{IPW-N} \preceq \text{IPW-S} \preceq \text{IPW-P} \cong \text{IPW-O}}^{\text{IPW-based CATE estimators}}.$$

A2. For Q2, under semiparametric dimension reduction structure, the affiliation of X_1 to X plays an important role. For Q3, when the CATE functions are smooth sufficiently, and the bandwidth and kernel function are delicately selected, the asymptotic properties are also different. The results are summarized in Table 1. Some more results are included in Sect. 2. Also some similar results about OR-N and more detailed comparisons are described in Sect. 2.

A3. In high-dimensional scenarios, we will see that high order kernel functions are in need and bandwidths must be very delicately selected, to have good estimation efficiency that are very sensitive to the selections. Thus, OR-N is not recommendable. Semiparametric structure-based estimation OR-S can be often preferable due to its advantages of greatly alleviating the curse of dimensionality and avoiding model misspecification. Some more detailed studies and comparisons for the asymptotic efficiency are contained in Sect. 2. The numerical studies in Sect. 3 support this observation.

The rest of this article is organized as follows. Section 2 introduces the CATE function and give the estimators respectively under the true, parametric, nonparametic and semiparametric framework. The asymptotic properties of the proposed estimators are systematically investigated in this section. Section 3 presents some simulation studies to examine the performances of the estimators. Section 4 is devoted to the analysis for a real data example. Conclusions and some further research problems are briefly discussed in Sect. 5. Due to the space limitation, all the technical proofs are relegated to the supplementary material.

2 Estimations and their asymptotic properties

Let D be a dummy variable indicating treatment status with $D = 1$ if an individual receives treatment and $D = 0$ otherwise. We only observe D, X and $Y \equiv D \cdot Y_{(1)} + (1 - D) \cdot Y_{(0)}$ in the real situation. The propensity score $p(D = 1|X)$

Table 1 The efficiency rank among four regression-based CATE estimators

Scenario	Efficiency rank
S1	$OR-O \cong OR-P \preceq OR-S \preceq OR-N$
S2	$OR-O \cong OR-P \cong OR-S \preceq OR-N$
S3	$OR-O \cong OR-P \cong OR-S \cong OR-N$

S1: $X_1 \subseteq \beta_1^\top X \bigcup \beta_0^\top X$;
S2: $X_1 \nsubseteq \beta_1^\top X \bigcup \beta_0^\top X$;
S3: CATE function is smooth enough and kernels and bandwidths are chosen delicately

is denoted by $p(X)$. Let $\{X_i, Y_i, D_i\}$, $i = 1, \ldots, n$ be n independent copies of (X, Y, D). To estimate $\tau(x_1)$, we suggest a two-step estimation procedure when both g_1 and g_0 are unknown. Four estimators are proposed in this paper when the regression causal effect under true (oracle), parametric, nonparametric, and semiparametric dimension reduction structure (OR-O, OR-P, OR-N, and OR-S) respectively.

To clearly state the estimation procedures, recall that the function $m_t(X)$ is defined as

$$m_t(X) = E(Y_{(t)}|X), \quad t = 0, 1.$$

Under the unconfoundedness assumption that is the conditional independence as

$$(Y_{(0)}, Y_{(1)}) \perp\!\!\!\perp D|X,$$

we then first estimate $m_1(X) - m_0(X)$ and then its conditional expectation $\tau(x_1) = E(m_1(X) - m_0(X)|X_1)$. But in semiparametric dimension reduction structure, this unconfoundedness assumption will have a different formula that will be specified in Sect. 2. However, directly estimating $\tau(X_1)$ in terms of $Y_{(1)} - Y_{(0)}$ is not feasible as it is never observed. It is naturally to use $Y_{(1)}$ and $Y_{(0)}$ to estimate $m_1(X)$ and $m_0(X)$ separately. Afterwards $\tau(x_1)$ can be estimated by a nonparametric method such as the N-W estimation (Nadaraya 1964; Watson 1964).

As for OR-S and OR-N, we will have to use high order kernel functions, we give the notation here. A function K_1: $\mathbb{R}^k \to \mathbb{R}$ is a kernel of order s_1 if it integrates to one over $\mathbb{R}^k$, and

$$\int u_1^{p_1} \cdots u_k^{p_k} K_1(u) du = 0$$

for all nonnegative integers $p_1, \cdots, p_k$ such that $1 \leq \sum_{i=1}^k p_i < s_1$, and it is nonzero when $\sum_{i=1}^k p_i = s_1$. Some regularity conditions are listed below.

(C1). (Strong ignorability)

(a) (Unconfoundedness) $(Y_{(0)}, Y_{(1)}) \perp\!\!\!\perp D|X$.

(b) (Common support) For some very small $c > 0$, the propensity score function $p(\cdot)$ satisfies that $c < p(X) < 1 - c$.

(C2). (Distribution of X) The support $\mathcal{X}$ of the p-dimensional covariate X is a Cartesian product of compact intervals, and the density of X, $f(x)$, is bounded away from 0 on $\mathcal{X}$.

(C3). (Kernel functions) $K_1(u)$ is a kernel of order s_1 that is symmetric around zero and s^* times continuously differentiable.

(C4). (Distribution of X_1) The density function of X_1, $f(x_1)$, is bounded away from zero and infinity and $s_1 \geq 2$ times continuously differentiable.

Part (a) of condition (C1) is a commonly used condition on the treatment effect, see e.g., Rosenbaum and Rubin (1983); Abrevaya et al. (2015); Luo et al. (2017).

Moreover, part (a) of condition (C1) is a quite strong but standard assumption in the causal inference literature. Part (b) of condition (C1) implies that there exists overlap between the treated and control observations. Conditions (C2) and (C4) are traditional conditions for nonparametric estimation in the literature (Pagan and Ullah 1999; Yin et al. 2010). Specially, condition (C3) is for high order kernel (Abrevaya et al. 2015). It is noted that Gaussian kernel satisfies this assumption when $k = 1$ and $s_1 = 2$. Furthermore, the value s^* relies on the smoothness of the regression function. More specifically, $s^* \geq 2$ in parametric situation, while $s^* \geq s_2$ and $s^* \geq s_4$ in nonparametric and semiparametric situation, respectively.

In the following, we study the four estimations in separate subsections and give some further analysis for OR-S and OR-N in another subsection.

2.1 OR-O

This estimator will serve as a benchmark to examine the performance of other estimators developed and investigated later. Assume that $m_1(X) - m_0(X)$ is completely known with no need of estimation. Then OR-O can be written as

$$\widehat{\tau}(x_1) = \frac{\frac{1}{nh_1^k}\sum_{i=1}^{n} K_1\left(\frac{X_{1i}-x_1}{h_1}\right)\{m_1(X_i) - m_0(X_i)\}}{\frac{1}{nh_1^k}\sum_{i=1}^{n} K_1\left(\frac{X_{1i}-x_1}{h_1}\right)}. \tag{2}$$

The asymptotically linear representation and asymptotic normality are stated below.

Theorem 1 *Suppose that assumptions* (C1) *through* (C4) *are satisfied. Then, when regression causal effect is given without estimation, for each point x_1 in the support of X_1, we have*

$$\begin{aligned}&\sqrt{nh_1^k}\{\hat{\tau}(x_1) - \tau(x_1)\}\\ &= \frac{1}{\sqrt{nh_1^k}}\frac{1}{f(x_1)}\sum_{i=1}^{n}\{m_1(X_i) - m_0(X_i) - \tau(x_1)\}K_1\left(\frac{X_{1i}-x_1}{h_1}\right) + o_p(1),\end{aligned}$$

and then

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

where $||K_1||_2 = \{\int K_1(u)^2 du\}^{1/2}$, *and*

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

2.2 OR-P

Suppose that both $m_1(X)$ and $m_0(X)$ have parametric structures with unknown parameters $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_0$ respectively. That is, $m_t(X, \boldsymbol{\alpha}_t)$ are parametric functions for $t = 0, 1$. Since each response can only be observed in a subpopulation, to get unbiased estimators of parameters $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_0$, we use a similar method to that of Wang et al. (2004). Write, for $i = 1, \ldots, n$,

$$D_i Y_i = D_i m_1(X_i, \boldsymbol{\alpha}_1) + D_i \epsilon_{1i}, \quad (1 - D_i) Y_i = (1 - D_i) m_0(X_i, \boldsymbol{\alpha}_0) + (1 - D_i) \epsilon_{0i},$$

where ϵ_{ti}, $t = 0, 1$, are random error terms, and independent of X_i, $i = 1, \ldots, n$. Use weighted least squares (Matloff 1981) to estimate $\boldsymbol{\alpha}_t$ for $t = 0, 1$, and write the estimator of $\boldsymbol{\alpha}_t$ and $m_1(X)$ as $\hat{\boldsymbol{\alpha}}_t$ and $\hat{m}_1(X)$. OR-P is then defined as:

$$\hat{\tau}(x_1) = \frac{\frac{1}{nh_1^k} \sum_{i=1}^{n} K_1\left(\frac{X_{1i} - x_1}{h_1}\right) \{\hat{m}_1(X_i) - \hat{m}_0(X_i)\}}{\frac{1}{nh_1^k} \sum_{i=1}^{n} K_1\left(\frac{X_{1i} - x_1}{h_1}\right)}, \tag{3}$$

where

$$\hat{m}_1(X_i) = m_1(X, \hat{\boldsymbol{\alpha}}_1), \quad \hat{m}_0(X_i) = m_0(X, \hat{\boldsymbol{\alpha}}_0), \quad i = 1, \ldots, n.$$

Assume the following additional condition:

(A1). (Bandwidths) $h_1 \to 0$, $nh_1^k \to \infty$, $nh_1^{2s_1 + k} \to 0$.

The following theorem states the asymptotic properties of $\hat{\tau}(x_1)$.

Theorem 2 *Suppose that conditions* (C1) *through* (C4) *and* (A1) *are satisfied for* $s_1 = s^* + 2$. *Then, for each point* x_1 *in the support of* X_1, *we have*

$$\begin{aligned}
&\sqrt{nh_1^k} \{\hat{\tau}(x_1) - \tau(x_1)\} \\
&= \frac{1}{\sqrt{nh_1^k}} \frac{1}{f(x_1)} \sum_{i=1}^{n} \{m_1(X_i) - m_0(X_i) - \tau(x_1)\} K_1\left(\frac{X_{1i} - x_1}{h_1}\right) + o_p(1) \\
&\xrightarrow{d} N\left(0, \frac{||K_1||_2^2 \sigma_P^2(x_1)}{f(x_1)}\right),
\end{aligned}$$

where

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

Remark 1 This theorem states the asymptotic equivalence between OR-P and OR-O in the sense that their asymptotic variance functions are identical.

2.3 OR-N

If we do not have prior information on the structures of $m_1(X)$ and $m_0(X)$ or we try to avoid model misspecification, a nonparametric estimation is feasible. Similarly, we estimate $m_1(X)$ and $m_0(X)$ separately. Therefore, OR-N is written as

$$\widehat{\tau}(x_1) = \frac{\frac{1}{nh_1^k}\sum_{i=1}^{n} K_1\left(\frac{X_{1i}-x_1}{h_1}\right)\{\widehat{m}_1(X_i) - \widehat{m}_0(X_i)\}}{\frac{1}{nh_1^k}\sum_{i=1}^{n} K_1\left(\frac{X_{1i}-x_1}{h_1}\right)}, \tag{4}$$

where

$$\widehat{m}_1(X_i) = \frac{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_2\left(\frac{X_j-X_i}{h_2}\right)Y_{1j}\mathbb{1}(D_j=1)}{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_2\left(\frac{X_j-X_i}{h_2}\right)\mathbb{1}(D_j=1)}, \widehat{m}_0(X_i) = \frac{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_2\left(\frac{X_j-X_i}{h_2}\right)Y_{0j}\mathbb{1}(D_j=0)}{\frac{1}{nh_2^p}\sum_{j=1}^{n} K_2\left(\frac{X_j-X_i}{h_2}\right)\mathbb{1}(D_j=0)}.$$

To study the asymptotic properties of $\widehat{\tau}(x_1)$, we give some more conditions on the kernel function and bandwidths.

(A2). $K_2(u)$ is a kernel of order $s_2 \geq p$, symmetric around zero and equal to zero outside $\prod_{i=1}^{p}[-1,1]$ with continuous (s_2+1) order derivatives.

(A3). $h_2 \to 0, \frac{\log n}{nh_2^{p+s_2}} \to 0$.

(A4). $h_2^{2s_2}h_1^{-2s_2-k} \to 0, nh_1^k h_2^{2s_2} \to 0$.

Conditions (A2), (A3) and (A4) are used to affiliate with the high order derivatives of m_1 and m_0 to ensure the asymptotic normality. The following theorem states the main theoretical results of OR-N. For convenience, define the following function:

$$\Psi_1(X,Y,D) := \frac{D\{Y-m_1(X)\}}{p(X)} - \frac{(1-D)\{Y-m_0(X)\}}{1-p(X)} + m_1(X) - m_0(X).$$

Theorem 3 *Suppose that conditions* (C1) *through* (C4) *and* (A1) *through* (A4) *are satisfied for* $s^* \geq s_2 \geq p$. *Then, for each point* x_1, *we have*

$$\begin{aligned}
&\sqrt{nh_1^k}(\widehat{\tau}(x_1) - \tau(x_1)) \\
&= \frac{1}{\sqrt{nh_1^k}}\frac{1}{f(x_1)}\sum_{i=1}^{n}[\Psi_1(X_i,Y_i,D_i) - \tau(x_1)]K_1\left(\frac{X_{1i}-x_1}{h_1}\right) + o_p(1) \\
&\xrightarrow{d} N\left(0, \frac{||K_1||_2^2\sigma_N^2(x_1)}{f(x_1)}\right),
\end{aligned}$$

where

$$\begin{aligned}\sigma_N^2(x_1) &= E[\{\Psi_1(X,Y,D)-\tau(x_1)\}^2|X_1=x_1]\\ &= \sigma_P^2(x_1)+E\left\{\frac{\text{var}(Y_{(1)}|X)}{p(X)}+\frac{\text{var}(Y_{(0)}|X)}{1-p(X)}\middle|X_1=x_1\right\}\\ &\geq \sigma_P^2(x_1)=\sigma_O^2(x_1),\end{aligned}$$

the equality holds if and only if $\frac{\text{var}(Y_{(1)}|X)}{p(X)}=0$ *and* $\frac{\text{var}(Y_{(0)}|X)}{1-p(X)}=0$, *which rarely happen. Thus, the inequality shows that* $OR-N$ *is asymptotically less efficient than* $OR-P$ *and* $OR-O$.

2.4 OR-S

An obvious limitation of OR-N is its incapability of handling models with high-dimensional covariates X in practice. Therefore, how to alleviate the curse of dimensionality is an important issue. To this end, reducing dimensionality is a natural idea. But we restrict ourselves to the sufficient dimension reduction framework below and use existing methods to estimate the projection directions as the focus of this paper is on asymptotics of the estimations assuming the dimension reduction structure is specified in a semiparametric manner. See the relevant references such as Luo et al. (2017) and Ma et al. (2019) that even considered ultra high-dimensional scenarios under the sufficient dimension reduction framework. A relevant reference is Fan et al. (2020) who proposed nonparametric doubly robust estimators for *CATE* allowing the number of covariates divergent with the sample size. In terms of machine learning to select significant covariates, the dimension reduction is achieved. Thus, we may also classify their method as a semiparametric approach.

We first give a very brief review on sufficient dimension reduction both Luo et al. (2017) and Ma et al. (2019) discussed. For given $\boldsymbol{\beta}^\top X$ where $\boldsymbol{\beta}$ is a $p\times r$ orthonormal matrix with an unknown number $r\ll p$ of columns, suppose that the regression of a response variable W is independent of X, which is written as $E(W|X)\perp\!\!\!\perp X|\boldsymbol{\beta}^\top X,$, where $\perp\!\!\!\perp$ stands for independence. It is generally known that $E(W|X)$ is an unspecified function of $\boldsymbol{\beta}^\top X$, which allows full freedom in the regression with $\boldsymbol{\beta}^\top X$ being the sufficiently reduced covariates (from p to r). This structure has a dimension reduction structure with unknown parameter $\boldsymbol{\beta}$ and also is very much flexible with a nonparametric nature. To identify the projection directions $\boldsymbol{\beta}$, Cook and Li (2002) defined the notion of central mean subspace that is the intersection of all subspaces spanned by any $\boldsymbol{\beta}$ such that the above conditional independence holds. To be specific, without notational confusion, write $\mathcal{S}_{E(Y_{(1)}|X)}$ and $\mathcal{S}_{E(Y_{(0)}|X)}$ respectively spanned by $\boldsymbol{\beta}_1\in\mathbb{R}^{p\times r(1)}$ and $\boldsymbol{\beta}_0\in\mathbb{R}^{p\times r(0)}$ where $r(t)<p$ for $t=0,1$ as the central mean subspaces such that

$$m_1(X) \perp\!\!\!\perp X|\boldsymbol{\beta}_1^\top X, \quad m_0(X) \perp\!\!\!\perp X|\boldsymbol{\beta}_0^\top X. \tag{5}$$

There are a lot of approaches available in the literature to identify $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$, including sliced inverse regression (Li 1991), sliced average variance estimator (Cook and Weisberg 1991), minimum average variance estimation (Xia et al. 2002), directional regression (Li and Wang 2007), the semiparametric methods (Ma and Zhu 2012), and the partial support vector machine (Shin et al. 2017). Then let us introduce how to estimate $\boldsymbol{\beta}_t$, $t = 0, 1$, in detail. Suppose we have a kernel matrix $\mathbf{M}$ which is derived from a certain sufficient dimension reduction method, for example, $\mathbf{M}_{\mathrm{SIR}} = \mathrm{cov}\{E(\mathbf{Z}|Y)\}$ for sliced inverse regression, $\mathbf{M}_{\mathrm{SAVE}} = E\{\mathbf{I}_p - \mathrm{cov}(\mathbf{Z}|Y)\}^2$ for sliced average variance estimation, and $\mathbf{M}_{\mathrm{DR}} = E\{2\mathbf{I}_p - E[(X - X')(X - X')^\top|Y, Y']\}^2$ for directional regression, where $\mathbf{Z} = \mathrm{cov}(X)^{-1/2}(X - EX)$ and (X', Y') is an independent copy of (X, Y), then we can use eigenvalue decomposition of $\mathbf{M}$. Finally, the first $r(t)$ eigenvectors $\boldsymbol{\eta}_t$ of $\mathbf{M}$ are standardized efficient dimension reduction directions under some suitable conditions. Note that $\boldsymbol{\beta}_t = \mathrm{cov}(X)^{-1/2}\boldsymbol{\eta}_t$, $t = 0, 1$, then $\widehat{\boldsymbol{\beta}}_t = \widehat{\mathrm{cov}(X)}^{-1/2}\widehat{\boldsymbol{\eta}}_t$, which is estimated dimension reduction matrices, where $\widehat{\mathrm{cov}(X)} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\top$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

Note that under this dimension reduction structure, we have $m_t(X) = E(Y_{(t)}|X) = E(Y_{(t)}|\boldsymbol{\beta}_t^\top X) = m_t(\boldsymbol{\beta}_t^\top X)$ for $t = 0, 1$. Define a OR-S as

$$\widehat{\tau}(x_1) = \frac{\frac{1}{nh_1^k}\sum_{i=1}^n K_1\left(\frac{X_{1i} - x_1}{h_1}\right)\{\widehat{m}_1(\widehat{\boldsymbol{\beta}}_1^\top X_i) - \widehat{m}_0(\widehat{\boldsymbol{\beta}}_0^\top X_i)\}}{\frac{1}{nh_1^k}\sum_{i=1}^n K_1\left(\frac{X_{1i} - x_1}{h_1}\right)}, \tag{6}$$

where

$$\widehat{m}_1(\widehat{\boldsymbol{\beta}}_1^\top X_i) = \frac{\frac{1}{nh_4^{r(1)}}\sum_{j=1}^n K_4\left(\frac{\widehat{\mathbf{Z}}_j^1 - \widehat{\mathbf{Z}}_i^1}{h_4}\right) Y_{1j}\mathbb{1}(D_j = 1)}{\frac{1}{nh_4^{r(1)}}\sum_{j=1}^n K_4\left(\frac{\widehat{\mathbf{Z}}_j^1 - \widehat{\mathbf{Z}}_i^1}{h_4}\right)\mathbb{1}(D_j = 1)}, \quad \widehat{\mathbf{Z}}^1 = \widehat{\boldsymbol{\beta}}_1^\top X,$$

$$\widehat{m}_0(\widehat{\boldsymbol{\beta}}_0^\top X_i) = \frac{\frac{1}{nh_4^{r(0)}}\sum_{j=1}^n K_4\left(\frac{\widehat{\mathbf{Z}}_j^0 - \widehat{\mathbf{Z}}_i^0}{h_4}\right) Y_{0j}\mathbb{1}(D_j = 0)}{\frac{1}{nh_4^{r(0)}}\sum_{j=1}^n K_4\left(\frac{\widehat{\mathbf{Z}}_j^0 - \widehat{\mathbf{Z}}_i^0}{h_4}\right)\mathbb{1}(D_j = 0)}, \quad \widehat{\mathbf{Z}}^0 = \widehat{\boldsymbol{\beta}}_0^\top X.$$

In order to derive theoretical results, give the following conditions.

(A5). $K_4(u)$ is a kernel of order s_4, is symmetric around zero, is equal to zero outside $\prod_{i=1}^p [-1, 1]$, and is continuously differentiable. The density function of $\boldsymbol{\beta}_t^\top X$, $f_t(\boldsymbol{\beta}_t^\top X)$ is s_4 times continuously differentiable for $t = 0, 1$. For $t = 0, 1$, $p(\boldsymbol{\beta}_t^\top X) \in (c^*, 1 - c^*)$ almost surely for some $c^* \in (0, 0.5)$.

(A6). $h_4 \to 0$, $\frac{\log n}{nh_4^{\max\{r(0), r(1)\}+s_4}} \to 0$.

(A7). $h_4^{2s_4} h_1^{-2s_4-k} \to 0$, $nh_1^k h_4^{2s_4} \to 0$.

(A8). $\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 = O_p(n^{-\frac{1}{2}})$ and $\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0 = O_p(n^{-\frac{1}{2}})$.

Since the treatment effect heterogeneity under the semiparametric structure is based on $\boldsymbol{\beta}_t^\top X$ for $t = 0, 1$, Assumptions (A5) through (A7) play the same role as Assumptions (A2) through (A4). Condition (A8) often holds (Luo et al. 2017).

Define three functions as

$$\begin{aligned}
\Psi_2(X, Y, D) &= \frac{D\{Y - m_1(X)\}}{p(\boldsymbol{\beta}_1^\top X)} + m_1(X) - m_0(X), \\
\Psi_3(X, Y, D) &= -\frac{(1-D)\{Y - m_0(X)\}}{1 - p(\boldsymbol{\beta}_0^\top X)} + m_1(X) - m_0(X), \\
\Psi_4(X, Y, D) &= \frac{D\{Y - m_1(X)\}}{p(\boldsymbol{\beta}_1^\top X)} - \frac{(1-D)\{Y - m_0(X)\}}{1 - p(\boldsymbol{\beta}_0^\top X)} + m_1(X) - m_0(X).
\end{aligned} \tag{7}$$

Next, for ease of explanation of our theoretical results, we introduce some notations. Write A and B as two sets of elements. Without confusion, write $card(A)$ as the cardinality of the set A.

(F1) $A \subset B$ stands for $A \cap B = A$. In other words, elements of A are all in B and $card(B) \geq card(A)$.

(F2) $A \subset^{k-q} B$ stands for $A \cap B = C$ with $card(C) = k - q$, that is, $k - q$ elements of A belong to B. When $k = q$, it means that A and B do not share the same elements, i.e. $A \cap B = \emptyset$, written as $A \nsubseteq B$.

The following theorem states some very detailed investigation on the asymptotic efficiency of OR-S.

Theorem 4 *Suppose that assumptions* (C1) *through* (C4), (A1) *and* (A5) *through* (A8) *are satisfied for* $s^* \geq s_4 \geq \max\{r(0), r(1)\}$. *Then, for each point* x_1 *in the support of* X_1, *noting the definitions of* Ψ_i *for* $i = 2, 3, 4$ *in* (7),

(1) *when* $X_1 \subset^{k-q} \boldsymbol{\beta}_1^\top X$ *and* $X_1 \subset^{k-q} \boldsymbol{\beta}_0^\top X$ *with* $s_4(2 - k/q) + k > 0$ *and* $0 < q \leq k$, *the asymptotically linear representation of* $\widehat{\tau}(x_1)$ *is*

$$\begin{aligned}
&\sqrt{nh_1^k}\{\widehat{\tau}(x_1) - \tau(x_1)\} \\
&= \frac{1}{\sqrt{nh_1^k}} \frac{1}{f(x_1)} \sum_{i=1}^{n} \{m_1(X_i) - m_0(X_i) - \tau(x_1)\} K_1\left(\frac{X_{1i} - x_1}{h_1}\right) + o_p(1),
\end{aligned}$$

and the asymptotic distribution of $\widehat{\tau}(x_1)$ *is*

$$\sqrt{nh_1^k}(\widehat{\tau}(x_1)-\tau(x_1)) \xrightarrow{d} N\left(0, \frac{||K_1||_2^2\sigma_{S,1}^2(x_1)}{f(x_1)}\right);$$

(2) *when* $X_1 \subset \boldsymbol{\beta}_1^\top X$ *and* $X_1 \subset^{k-q} \boldsymbol{\beta}_0^\top X$ *with* $s_4(2-k/q)+k>0$ *and* $0<q\le k$, *the asymptotically linear representation of* $\widehat{\tau}(x_1)$ *is*

$$\begin{aligned}
&\sqrt{nh_1^k}\{\widehat{\tau}(x_1)-\tau(x_1)\}\\
&=\frac{1}{\sqrt{nh_1^k}}\frac{1}{f(x_1)}\sum_{i=1}^{n}\{\Psi_2(X_i,Y_i,D_i)-\tau(x_1)\}K_1\left(\frac{X_{1i}-x_1}{h_1}\right)+o_p(1)\\
&\xrightarrow{d} N\left(0, \frac{||K_1||_2^2\sigma_{S,2}^2(x_1)}{f(x_1)}\right);
\end{aligned}$$

(3) *when* $X_1 \subset^{k-q} \boldsymbol{\beta}_1^\top X$ *and* $X_1 \subset \boldsymbol{\beta}_0^\top X$ *with* $s_4(2-k/q)+k>0$ *and* $0<q\le k$, *the asymptotically linear representation of* $\widehat{\tau}(x_1)$ *is*

$$\begin{aligned}
&\sqrt{nh_1^k}\{\widehat{\tau}(x_1)-\tau(x_1)\}\\
&=\frac{1}{\sqrt{nh_1^k}}\frac{1}{f(x_1)}\sum_{i=1}^{n}\{\Psi_3(X_i,Y_i,D_i)-\tau(x_1)\}K_1\left(\frac{X_{1i}-x_1}{h_1}\right)+o_p(1)\\
&\xrightarrow{d} N\left(0, \frac{||K_1||_2^2\sigma_{S,3}^2(x_1)}{f(x_1)}\right);
\end{aligned}$$

(4) *when* $X_1 \subset \boldsymbol{\beta}_1^\top X$ *and* $X_1 \subset \boldsymbol{\beta}_0^\top X$, *the asymptotically linear representation of* $\widehat{\tau}(x_1)$ *is*

$$\begin{aligned}
&\sqrt{nh_1^k}\{\widehat{\tau}(x_1)-\tau(x_1)\}\\
&=\frac{1}{\sqrt{nh_1^k}}\frac{1}{f(x_1)}\sum_{i=1}^{n}\{\Psi_4(X_i,Y_i,D_i)-\tau(x_1)\}K_1\left(\frac{X_{1i}-x_1}{h_1}\right)+o_p(1)\\
&\xrightarrow{d} N\left(0, \frac{||K_1||_2^2\sigma_{S,4}^2(x_1)}{f(x_1)}\right);
\end{aligned}$$

where

$$\begin{aligned}
\sigma_{S,1}^2(x_1)&=\sigma_O^2(x_1)=E[\{m_1(X)-m_0(X)-\tau(x_1)\}^2|X_1=x_1],\\
\sigma_{S,2}^2(x_1)&=E[\{\Psi_2(X,Y,D)-\tau(x_1)\}^2|X_1=x_1],\\
\sigma_{S,3}^2(x_1)&=E[\{\Psi_3(X,Y,D)-\tau(x_1)\}^2|X_1=x_1],\\
\sigma_{S,4}^2(x_1)&=E[\{\Psi_4(X,Y,D)-\tau(x_1)\}^2|X_1=x_1].
\end{aligned} \tag{8}$$

Remark 2 These results imply that the asymptotic behaviours of OR-S rely on whether X_1 is a subset of $\boldsymbol{\beta}_t^\top X$ for $t = 0, 1$. If $X_1 \subseteq \boldsymbol{\beta}_1^\top X \bigcup \boldsymbol{\beta}_0^\top X$, then the asymptotic variance of OR-S is different from OR-O, OR-P and OR-N. If not in the above cases, then OR-S enjoys same asymptotic variance as OR-O and OR-P. It is also worthwhile to note that even if $X_1 \nsubseteq \boldsymbol{\beta}_1^\top X$ and $X_1 \nsubseteq \boldsymbol{\beta}_0^\top X$, we can still utilize $\widehat{m}_1(\boldsymbol{\beta}_1^\top X)$ and $\widehat{m}_0(\boldsymbol{\beta}_0^\top X)$ to estimate $\tau(X_1)$, since $\boldsymbol{\beta}_1^\top X$ and $\boldsymbol{\beta}_0^\top X$ are sufficient to model $Y_{(1)}$ and $Y_{(0)}$, respectively.

Remark 3 Note that $X_1 \subset^{k-q} \boldsymbol{\beta}_t^\top X$ implies that only $k-q$ elements of X_1 are also the $k-q$ linear combinations of $\boldsymbol{\beta}_t^\top X$ for $t = 0, 1$. In this case, write $\boldsymbol{\beta}_t^\top X$ as $\boldsymbol{\beta}_t^\top X = (X_{1(1)}, \ldots, X_{1(k-q)}, (\widetilde{\boldsymbol{\beta}}_t^\top X)^\top)^\top$ for $t = 0, 1$. Therefore, when $X_1 \subset^{k-q} \boldsymbol{\beta}_t^\top X$ with $s_4(2 - k/q) + k > 0$ and $0 < q \leq k$, we should determine the intersection between X_1 and $\boldsymbol{\beta}_t^\top X$, and then estimate $\boldsymbol{\beta}_t$ through estimating $\widetilde{\boldsymbol{\beta}}_t$ for $t = 0, 1$. It could be done by using partial sufficient dimension reduction (e.g. Feng et al. (2013)). As this is not the focus of this paper, we then assume that $\boldsymbol{\beta}_t$ can be estimated at the rate $1/\sqrt{n}$ of convergence. Obviously, the assumption $s_4(2 - k/q) + k > 0$ is satisfied for $k = 1$.

Corollary 1 *We have*

$$\begin{aligned}
&\sigma_{S,1}^2(x_1) = \sigma_P^2(x_1) = \sigma_O^2(x_1),\\
&\sigma_{S,2}^2(x_1) = \sigma_P^2(x_1) + E\left\{ \frac{\mathrm{var}(Y_{(1)}|X)}{p(\boldsymbol{\beta}_1^\top X)} \middle| X_1 = x_1 \right\} \geq \sigma_P^2(x_1) = \sigma_O^2(x_1),\\
&\sigma_{S,3}^2(x_1) = \sigma_P^2(x_1) + E\left\{ \frac{\mathrm{var}(Y_{(0)}|X)}{1 - p(\boldsymbol{\beta}_0^\top X)} \middle| X_1 = x_1 \right\} \geq \sigma_P^2(x_1) = \sigma_O^2(x_1),\\
&\sigma_{S,4}^2(x_1) = \sigma_P^2(x_1) + E\left\{ \left[\frac{\mathrm{var}(Y_{(1)}|X)}{p(\boldsymbol{\beta}_1^\top X)} + \frac{\mathrm{var}(Y_{(0)}|X)}{1 - p(\boldsymbol{\beta}_0^\top X)} \right] \middle| X_1 = x_1 \right\} \geq \sigma_P^2(x_1) = \sigma_O^2(x_1).
\end{aligned}$$

Assume that $\mathrm{var}(Y_{(t)}|X)$ *is a measurable function with respect to* $\boldsymbol{\beta}_t^\top X$ *for* $t = 0, 1$. *Then*

$$E\left\{ \frac{\mathrm{var}(Y_{(1)}|X)}{p(\boldsymbol{\beta}_1^\top X)} \right\} \leq E\left\{ \frac{\mathrm{var}(Y_{(1)}|X)}{p(X)} \right\}, \quad \textit{and} \quad E\left\{ \frac{\mathrm{var}(Y_{(0)}|X)}{1 - p(\boldsymbol{\beta}_0^\top X)} \right\} \leq E\left\{ \frac{\mathrm{var}(Y_{(0)}|X)}{1 - p(X)} \right\}.$$

Then

$$\begin{aligned}
\sigma_O^2(x_1) = \sigma_P^2(x_1) \leq \sigma_{S,2}^2(x_1) \leq \sigma_{S,4}^2(x_1) \leq \sigma_N^2(x_1),\\
\sigma_O^2(x_1) = \sigma_P^2(x_1) \leq \sigma_{S,3}^2(x_1) \leq \sigma_{S,4}^2(x_1) \leq \sigma_N^2(x_1).
\end{aligned} \tag{9}$$

Remark 4 The results in the above corollary are based on some elementary calculations and the application of Theorem 3 of Luo et al. (2017). We then omit the detailed calculations. Based on these facts, OR-S is more efficient than OR-N in all cases, and less efficient than OR-P and OR-O in cases (2) to (4). in particular, OR-S

shares the same asymptotic distribution as OR-P and OR-O in case (1). Furthermore, OR-S in case (4) is less efficient than cases (2) and (3).

2.5 Further studies on OR-N and OR-S

Inspired by Theorem 4 about the importance of affiliation of X_1 to the set of arguments of the regression functions, we further investigate $OR-S$ and $OR-N$ in more general settings. The results are stated in the following.

Corollary 2 *Suppose that conditions* (C1) *through* (C4) *and* (A1) *through* (A8) *are satisfied. Assume that there is a given* $\widetilde{X}$ *such that* $(Y_{(0)}, Y_{(1)}) \perp\!\!\!\perp X|\widetilde{X}$ *with* $\widetilde{X} \subset X$ *and* $X_1 \notin \widetilde{X}$, *then* $OR-O \cong OR-P \cong OR-N$. *If we further assume* $X_1 \subset^{k-q} \boldsymbol{\beta}_1^\top X$ *and* $X_1 \subset^{k-q} \boldsymbol{\beta}_0^\top X$ *with* $s_4(2-k/q)+k>0$ *and* $0<q\leq k$, *then the four outcome regression-based CATE estimators share the same asymptotic distribution, i.e.,* $OR-O \cong OR-P \cong OR-S \cong OR-N$.

$$\textit{Here, } \widetilde{\sigma}_N^2(x_1) \equiv E[\{m_1(X)-m_0(X)-\tau(x_1)\}^2|X_1=x_1] = \sigma_P^2(x_1) = \sigma_O^2(x_1).$$

Remark 5 Much to our surprise, OR-N can be asymptotically more efficient in this special case to share the same asymptotic variance of OR-P. This shows the importance of covariate affiliation to the set of arguments of the regression function. This is a unique property for CATE as for ATE, this does not happen.

Corollary 3 *In Theorem 3 and Theorem 4, if commonly used constraints on the bandwidths* h_1, h_2 *and* h_4 *are replaced with* $\sqrt{nh_1^k}\left(h_2^s+\sqrt{\log(n)/nh_2^p}\right)=o(1)$ *and* $\sqrt{nh_1^k}\left(h_4^s+\sqrt{\frac{\log(n)}{nh_4^{\max\{r(0),r(1)\}}}}\right)=o(1)$ *for some order* s, *OR-N and OR-S have the same asymptotic distribution as OR-P and OR-O.*

Remark 6 As mentioned above, if we choose the bandwidth to satisfy the above conditions, OR-N and OR-S will share the same asymptotic efficiencies as OR-P and OR-O. It is obvious that the condition $\sqrt{nh_1^k}\left(h_2^s+\sqrt{\log(n)/nh_2^p}\right)=o(1)$ and $\sqrt{nh_1^k}\left(h_4^s+\sqrt{\frac{\log(n)}{nh_4^{\max\{r(0),r(1)\}}}}\right)=o(1)$ are much stronger than the assumptions in Theorem 3 and Theorem 4. However, it is possible to choose such bandwidths if the regression causal effect function is sufficiently smooth such that high order kernel can be used. For details, see Li and Racine (2007) and Zhou and Zhu (2021). Therefore, we obtain that the ranking for the asymptotic efficiencies of four regression-based CATE estimators and four propensity score-based CATE estimators under the condition that $\sqrt{nh_1^k}\left(h_2^s+\sqrt{\log(n)/nh_2^p}\right)=o(1)$ and $\sqrt{nh_1^k}\left(h_4^s+\sqrt{\frac{\log(n)}{nh_4^{\max\{r(0),r(1)\}}}}\right)=o(1)$,

$$\overbrace{\text{OR-O} = \text{OR-P} = \text{OR-S} = \text{OR-N}}^{\text{regression-based CATE estimators}} \leq \overbrace{\text{IPW-N} = \text{IPW-S} = \text{IPW-P} = \text{IPW-O}}^{\text{IPW-based CATE estimators}}. \tag{10}$$

The equality occurs if and only if

$$E\left\{\left[\frac{\text{var}(Y_{(1)}|X)}{p(X)} + \frac{\text{var}(Y_{(0)}|X)}{1-p(X)} + p(X)(1-p(X))\left(\frac{m_1(X)}{p(X)} + \frac{m_0(X)}{1-p(X)}\right)^2\right]\middle| X_1 = x_1\right\} = 0.$$

In other words, regression based estimators are always more efficient than IPW-type estimators in this general setting.

On the other hand, the above investigations are mainly for theoretical studies, and in practice, we may avoid to choose those bandwidths as they are often very difficult to properly select otherwise, the estimators would perform worse.

2.6 Estimation for asymptotic variance

We also very briefly describe the issue of estimating the asymptotic variance functions. In the following, we take *OR-P* as an example to briefly describe an estimation procedure, while the variance functions of the other *CATE* estimators can be similarly estimated.

Recall that the asymptotic variance of *OR-P* in Theorem 2, we then construct its consistent estimator as

$$\hat{\sigma}_P^2(x_1) = \frac{1}{nh_1^k}\sum_{i=1}^{n}\left[\{\hat{m}_1(X_i) - \hat{m}_0(X_i) - \hat{\tau}(x_1)\}K_1\left(\frac{X_{1i}-x_1}{h_1}\right)\right]^2/\hat{f}(x_1).$$

Here $\hat{\tau}(x_1)$ is the corresponding *CATE* estimator $OR-P$, $\hat{f}(x_1)$ is a nonparametric kernel estimation, which can be obtained as $\hat{f}(x_1) = \frac{1}{nh_1^k}\sum_{i=1}^{n} K_1\left(\frac{X_{1i}-x_1}{h_1}\right)$, $\hat{m}_1(X)$ and $\hat{m}_0(X)$ are kernel regressions of Y on X in the treated and control subpopulations respectively. As all are related to nonparametric kernel estimations, the consistency can also be expected. Similarly, we can get the estimator of asymptotic variance of *OR-N* and *OR-S*.

An alternative is the nonparametric bootstrap approximation (Eforn 1979), which is often useful in practice. The procedure can be described by the following steps: given $X_1 = x_1 \in \mathbf{\Omega}$,

- Step 1: Given original random sample $\{(Y_i, X_i, D_i) : i = 1, \cdots, n\}$, obtain the $OR-P$, $\hat{\tau}(x_1)$ as described before;
- Step 2: Generating the b-th bootstrapped sample $\{(Y_i^b, X_i^b, D_i^b) : i = 1, \cdots, n\}$, $b = 1, \cdots, B$ with replacement from $\{(Y_i, X_i, D_i) : i = 1, \cdots, n\}$. For each bootstrapped sample, compute $\hat{\tau}_b(x_1)$;
- Step 3: The estimator of the asymptotic variance of $\hat{\tau}(x_1)$ can be obtained by the empirical variance of $(\hat{\tau}_1(x_1), \cdots, \hat{\tau}_B(x_1))$:

$$\widehat{\sigma}^2(x_1) = \frac{1}{B-1}\sum_{b=1}^{B}\left[\hat{\tau}_b(x_1) - \hat{\tau}(x_1)\right]^2. \tag{11}$$

Similarly, we can get the boostrap-based asymptotic variance estimator of other CATE estimators by replacing the role of $\hat{\tau}(x_1)$. Furthermore, it is standard to obtain confidence intervals for the CATE estimator based on normal approximations, that is $[\hat{\tau}(x_1) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\sigma}^2(x_1)}{nh_1}}, \hat{\tau}(x_1) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\sigma}^2(x_1)}{nh_1}}]$, where $z_{1-\frac{\alpha}{2}}$ is $1-\frac{\alpha}{2}$ critical value of the standard normal distribution and α is a pre-specified confidence level. As this is not the focus of this paper, we then do not give more details about their asymptotic properties.

3 Simulations

To verify our theoretical results, we in this section conduct simulation studies to compare the regression-based OR-O, OR-P, OR-S, OR-N estimators with IPW-based IPW-O, IPW-P, IPW-S, IPW-N estimators (Abrevaya et al. 2015). Set $p = \dim(X) \in \{2, 4\}$ to avoid the curse of dimensionality under nonparametric estimation. Based on our experience and the theoretical results, when p is large, OR-N is very hard to implement. As well known, bandwidth selection plays an important role in the NW estimation. Hence, we first discuss this issue.

3.1 Bandwidth and kernel function selection

Note that OR-O and OR-P only involve one bandwidth h_1 used in the second step of the estimation procedure. We first check how to choose bandwidth sequences and kernel functions satisfying the conditions A1–A7. To this end, consider

$$\begin{aligned} h_1 &= a_1 \cdot n^{-\frac{1}{k+2s_1-\delta_1}}, \quad a_1 0, > \quad \delta_1 0, \\ h_2 7 &= a_2 \cdot n^{-\frac{1}{p+s_2+\delta_2}}, \quad a_2 0, > \quad \delta_2 0, \\ h_4 &= a_3 \cdot n^{-\frac{1}{\max\{r(0),r(1)\}+s_4+\delta_3}}, \quad a_3 > 0, \quad \delta_3 0, \end{aligned} \tag{12}$$

where δ_1, δ_2 and δ_3 can be selected as small as necessary or desired. It is clear that h_1, h_2 and h_4 satisfy conditions A1, A2, A3, A5 and A6. To satisfy condition A4, we set the kernel orders as $s_2 = p$ and $p+1$ for even and odd p respectively; and $s_1 = s_2 + 2$. To satisfy condition 7, under semiparametric dimension reduction structure, set $s_4 = \max\{r(0), r(1)\}$ and $= \max\{r(0), r(1)\} + 1$ respectively for even and odd $\max\{r(0), r(1)\}$. Based on the above values of s_1, s_2 and s_4, we verify th first parts of conditions A4 and A7. Next, consider the second parts of these two conditions. Note that when $s_2 \geq p$ and $s_4 \geq \max\{r(0), r(1)\}$,

$$-\frac{2s_2}{p+s_2} \leq -1, \quad \frac{2s_2+k}{2s_2+4+k} < 1, -\frac{2s_4}{\max\{r(0),r(1)\}+s_4} \leq -1, \quad \frac{2s_4+k}{2s_1+k} < 1.$$

Then

$$-\frac{2s_2}{p+s_2}+\frac{2s_2+k}{2s_2+4+k}<0,\quad -\frac{2s_4}{\max\{r(0),r(1)\}+s_4}+\frac{2s_4+k}{2s_1+k}<0.$$

Therefore, $h_2^{2s_2}h_1^{-2s_2-k}\to 0$ and $h_4^{2s_4}h_1^{-2s_4-k}\to 0$. Invoking condition A3, $nh_1^kh_2^{2s_2}=nh_1^{2s_1+k}h_2^{2s_2}h_1^{-2s_1}\to 0$ when $h_2^{2s_2}h_1^{-2s_1}\to 0$. Since δ_1, δ_2 and δ_3 can be arbitrarily small, we get, because $-s_2/(s_2+p)\le -1/2$ and $(s_2+2)/(2s_2+4+k)<1/2$,

$$-\frac{s_2}{s_2+p}+\frac{s_2+2}{2s_2+4+k}<0.$$

Thus, condition A4 is satisfied. Similarly, together with condition A6, condition A7 can also be satisfied, which has $nh_1^kh_4^{2s_4}\to 0$ by

$$-\frac{s_4}{\max\{r(0),r(1)\}+s_4}+\frac{s_4}{2s_4+k}<0.$$

3.2 Model setting

To examine the finite sample performances of the CATE estimators, consider the following three models:

Model 1: $Y_{(0)}=0,\quad Y_{(1)}=X_1^2+X_2+\epsilon_1,\quad p_1(X)=\frac{\exp(0.2(X_1+X_2))}{1+\exp(0.2(X_1+X_2))}$.

Model 2: $Y_{(0)}=0,\quad Y_{(1)}=X_1+X_2+X_3+X_4+\epsilon_2,\quad p_2(X)=\frac{\exp\{0.2(X_1+X_2+X_3+X_4)\}}{1+\exp\{0.2(X_1+X_2+X_3+X_4)\}}$.

Model 3: $Y_{(0)}=0,\quad Y_{(1)}=X_2+X_3+\epsilon_3,\quad p_3(X)=\frac{\exp\{0.2(X_2+X_3)\}}{1+\exp\{0.2(X_2+X_3))\}}$.

Model 1 is a model with the dimensions 2 and 0 of the central mean subspaces for the treatment and control groups; Model 2 is used to verify Theorem 4. Model 3 is set to justify the theory in Corollary 2. The dimensions of central mean subspaces for the treatment and control group are 1 and 0 in Models 2 and 3. For Model 1, $X=(X_1,X_2)^\top$ is generated by

$$X_1\sim U(-0.5,0.5),\quad X_2=1+2X_1+\zeta,$$

where $\zeta\sim U(-0.5,0.5)$, $\epsilon_1\sim N(0,0.1^2)$. For Model 2, we generate $X=(X_1,X_2,X_3,X_4)^\top$ by

$$X_1\sim U(-0.5,0.5),\quad X_2=1+X_1^2+\zeta_1,$$
$$X_3=(1+X_1)^2+\zeta_2,\quad X_4=(-1+X_1)^2+\zeta_3,$$

where $\zeta_j\overset{iid}{\sim}U(-0.5,0.5)$, $\epsilon_2\sim N(0,0.1^2)$, $j=1,2,3$. In Model 3, $X=(X_1,X_2,X_3)^\top$ are given by

$$X_1\sim U(-0.5,0.5),\quad X_2=X_1+\vartheta_1,\quad X_3=(1+X_1)^2+\vartheta_2,$$

where $\vartheta_j \overset{iid}{\sim} U(-0.5, 0.5)$, $\epsilon_3 \sim N(0, 0.1^2)$, $j = 1, 2$.

Although we introduce how to estimate the asymptotic variances in Sect. 2.6, we should note that its estimation procedure is very complex, since we need to estimate many unknown functions. Hence, we utilize a bootstrap-based method to calculate the asymptotic variance. Furthermore, Let $T(x_1) = \sqrt{(nh_1)}[\widehat{\tau}(x_1) - \tau(x_1)]$, we use the following indices to evaluate the performances of the involved estimators: Standard deviation (SD), the bootstrap-based estimated standard deviation (ESD), Bias, MSE and 95% confidence interval coverage probability based on bootstrap-based estimated standard deviation of $T(x_1)$ (CP). The number of bootstrap time is 200 in this simulation study. The sample size is taken to be respectively $n = 500$ and $n = 1000$. Moreover, the replication time is 500. For the bandwidth selection described in Subsection 3.1, we have the following selections.

a). For Model 1 as $p = 2$, equation (12) gives $s_1 = 4$, $s_2 = 2$, and $s_4 = 2$. We then choose $h_1 = a_1 \cdot n^{-\frac{1}{9}}$ for $a_1 \in \{0.03, 0.05\}$, $h_2 = a_2 \cdot n^{-\frac{1}{4}}$ for $a_2 \in \{0.15, 0.16, 0.17\}$, $h_4 = a_3 \cdot n^{-\frac{1}{4}}$ for $a_3 \in \{0.07, 0.08, 0.1\}$. Here, a_1, a_2 and a_3 are called baselines.

b). For Model 2, as $p = 4$, $h_1 = a_1 \cdot n^{-\frac{1}{13}}$ for $a_1 = 0.1$, $h_2 = a_2 \cdot n^{-\frac{1}{8}}$ for $a_2 = 0.6$, $h_4 = a_3 \cdot n^{-\frac{1}{3}}$ for $a_3 \in \{0.1, 0.12, 0.14, 0.16, 0.18\}$.

c). For Model 3, as $p = 3$, then $h_1 = a_1 \cdot n^{-\frac{1}{13}}$ for $a_1 = 0.05$, $h_2 = a_2 \cdot n^{-\frac{1}{8}}$ for $a_2 \in \{0.4, 0.5, 0.6, 0.8\}$, $h_4 = a_3 \cdot n^{-\frac{1}{3}}$ for $a_3 \in \{0.1, 0.15, 0.2\}$.

3.3 Simulation results

We tubulate the results in Tables 2, 3, 4 below and have some observations.

First, to show the estimation consistency, we can see that larger sample size reasonably results in smaller SD and MSE . The dimension of X also effects the estimation performance. When p increases to 4 from 2, both SD and MSE obviously increase particularly when $n = 1000$.

Second, the comparisons show the significant advantage of outcome regression-based estimation over IPW-based estimation. Even though in theory, OR-N is asymptotically equivalent to IPW-N, the difference on the estimation efficiency is still very significant. All results in the tables obviously indicate this: all IPW-based estimators have much larger SD than all regression-based estimators.

Third, as discussed before, the performances of OR-N and OR-S are highly associated with the affiliation of the given covariates to the set of arguments of the outcome regression. This finding can also be confirmed in Tables 3 and 4 In Model 2, $X_1 \subset^{k-q} \boldsymbol{\beta}_1^\top X$ and $X_1 \subset^{k-q} \boldsymbol{\beta}_0^\top X$ with $k = 1$ and $q = 0$, thus in theory, OR-S shares the same asymptotic variance as OR-P and OR-O and is more efficient than OR-N. From Table 3 we can see that the SDs of OR-S are similar to those of OR-P and OR-O, which are smaller than that of OR-N. In Model 3, $X_1 \not\subset \widetilde{X} = (X_2, X_3)^\top$. the asymptotic efficiencies are equivalent in theory and its SDs in Table 4 are similar

Table 2 The distribution of $\sqrt{nh_1}[\hat{\tau}(x_1) - \tau(x_1)]$ for model 1

*Group*1: $\{a_1 = 0.03, a_4 = 0.08, a_2 = 0.16\}$											
n	x_1	OR-O							OR-P		
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n = 500$	Bias	−0.004	−0.012	0.007	0.003	0.013	−0.002	−0.010	0.008	0.004	0.013
	SD	0.201	0.195	0.204	0.208	0.197	0.201	0.194	0.207	0.210	0.203
	ESD	0.209	0.206	0.210	0.211	0.209	0.209	0.207	0.211	0.211	0.209
	MSE	0.041	0.038	0.042	0.043	0.039	0.040	0.038	0.043	0.044	0.041
	CP	0.932	0.932	0.922	0.928	0.946	0.938	0.936	0.926	0.924	0.920
$n = 1000$	Bias	0.002	0.004	−0.001	0.005	−0.003	0.001	0.005	0.001	0.007	−0.003
	SD	0.206	0.202	0.197	0.203	0.202	0.209	0.205	0.198	0.202	0.203
	ESD	0.201	0.202	0.203	0.203	0.202	0.200	0.202	0.203	0.203	0.202
	MSE	0.043	0.041	0.039	0.041	0.041	0.044	0.042	0.039	0.041	0.041
	CP	0.918	0.936	0.948	0.948	0.956	0.914	0.936	0.948	0.952	0.946
		IPW- N					IPW-S				
$n = 500$	Bias	−0.076	−0.012	−0.004	0.006	0.233	−0.086	−0.027	0.016	0.049	0.317
	SD	0.348	0.469	0.612	0.792	0.879	0.336	0.448	0.610	0.782	0.865
	ESD	0.369	0.509	0.688	0.905	1.110	0.368	0.512	0.697	0.915	1.127
	MSE	0.127	0.220	0.374	0.627	0.827	0.120	0.202	0.372	0.614	0.849
	CP	0.956	0.958	0.976	0.970	0.982	0.962	0.970	0.974	0.982	0.990

Table 2 (continued)

		IPW-N					IPW-S				
$n=1000$	Bias	−0.070	0.000	−0.027	0.031	0.232	−0.087	−0.019	−0.008	0.094	0.305
	SD	0.346	0.417	0.604	0.759	0.876	0.337	0.406	0.574	0.730	0.899
	ESD	0.373	0.510	0.681	0.895	1.103	0.372	0.510	0.688	0.905	1.113
	MSE	0.125	0.174	0.365	0.578	0.822	0.121	0.165	0.330	0.542	0.901
	CP	0.952	0.982	0.974	0.976	0.984	0.960	0.988	0.974	0.980	0.974
*Group*1: $\{a_1=0.03, a_4=0.08, a_2=0.16\}$											
n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.050	0.059	0.112	0.043	−0.118	−0.003	−0.005	0.013	0.012	0.016
	SD	0.202	0.195	0.233	0.242	0.244	0.206	0.193	0.211	0.213	0.206
	ESD	0.202	0.209	0.235	0.247	0.247	0.207	0.208	0.211	0.213	0.213
	MSE	0.043	0.042	0.067	0.060	0.074	0.042	0.037	0.045	0.045	0.043
	CP	0.930	0.930	0.926	0.932	0.926	0.928	0.932	0.928	0.936	0.922
$n=1000$	Bias	−0.041	0.059	0.085	0.041	−0.119	−0.001	0.006	0.004	0.013	−0.003
	SD	0.218	0.215	0.236	0.227	0.245	0.212	0.208	0.206	0.209	0.211
	ESD	0.200	0.207	0.218	0.227	0.227	0.202	0.203	0.203	0.205	0.205
	MSE	0.049	0.050	0.063	0.053	0.074	0.045	0.043	0.042	0.044	0.045
	CP	0.906	0.940	0.930	0.938	0.918	0.908	0.932	0.938	0.936	0.930

Table 2 (continued)

		IPW-P					IPW-O				
$n=500$	Bias	−0.041	0.018	0.008	−0.027	0.068	−0.047	0.011	0.009	−0.010	0.111
	SD	0.348	0.494	0.653	0.843	0.922	0.381	0.515	0.680	0.917	1.101
	ESD	0.377	0.515	0.690	0.900	1.091	0.373	0.511	0.687	0.898	1.086
	MSE	0.123	0.244	0.426	0.712	0.854	0.147	0.265	0.463	0.842	1.224
	CP	0.958	0.954	0.960	0.970	0.978	0.938	0.948	0.942	0.936	0.936
$n=1000$	Bias	−0.034	0.020	−0.024	0.006	0.080	−0.032	0.022	−0.025	0.004	0.068
	SD	0.359	0.439	0.655	0.793	0.928	0.387	0.469	0.687	0.860	1.082
	ESD	0.379	0.512	0.682	0.893	1.090	0.378	0.511	0.680	0.890	1.084
	MSE	0.130	0.193	0.430	0.629	0.868	0.151	0.221	0.473	0.740	1.176
	CP	0.946	0.974	0.950	0.968	0.980	0.940	0.966	0.948	0.940	0.952
*Group*2: $\{a_1=0.05, a_4=0.07, a_2=0.17\}$											
n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.004	0.000	0.006	0.000	−0.013	−0.004	−0.001	0.004	−0.001	−0.014
	SD	0.212	0.212	0.202	0.203	0.211	0.219	0.214	0.205	0.204	0.214
	ESD	0.203	0.202	0.204	0.205	0.205	0.203	0.202	0.204	0.205	0.205
	MSE	0.045	0.045	0.041	0.041	0.045	0.048	0.046	0.042	0.042	0.046
	CP	0.922	0.934	0.930	0.924	0.938	0.910	0.934	0.928	0.928	0.924

Table 2 (continued)

*Group*2: $\{a_1 = 0.05, a_4 = 0.07, a_2 = 0.17\}$

n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=1000$	Bias	−0.003	−0.005	0.007	0.006	0.000	−0.005	−0.009	0.003	0.003	0.000
	SD	0.191	0.203	0.193	0.195	0.204	0.195	0.207	0.193	0.195	0.208
	ESD	0.204	0.204	0.202	0.203	0.203	0.204	0.203	0.202	0.203	0.203
	MSE	0.036	0.041	0.037	0.038	0.042	0.038	0.043	0.037	0.038	0.043
	CP	0.962	0.948	0.958	0.942	0.946	0.964	0.940	0.952	0.942	0.944
		IPW-N					IPW-S				
$n=500$	Bias	−0.059	−0.019	0.022	0.003	0.242	−0.081	−0.034	0.025	0.047	0.303
	SD	0.337	0.476	0.629	0.803	0.904	0.321	0.455	0.607	0.802	0.926
	ESD	0.372	0.511	0.685	0.896	1.103	0.370	0.512	0.691	0.907	1.116
	MSE	0.117	0.227	0.396	0.645	0.876	0.109	0.208	0.369	0.645	0.949
	CP	0.964	0.964	0.960	0.964	0.982	0.958	0.972	0.960	0.962	0.990
$n=1000$	Bias	−0.099	−0.053	0.024	0.046	0.262	−0.090	−0.060	0.030	0.071	0.264
	SD	0.338	0.455	0.614	0.756	0.880	0.317	0.439	0.589	0.711	0.850
	ESD	0.377	0.509	0.683	0.891	1.103	0.380	0.512	0.689	0.898	1.109
	MSE	0.124	0.210	0.378	0.573	0.844	0.108	0.196	0.348	0.510	0.793
	CP	0.968	0.966	0.972	0.972	0.988	0.982	0.976	0.974	0.984	0.990

Table 2 (continued)

*Group*2: $\{a_1=0.05, a_4=0.07, a_2=0.17\}$											
n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.061	0.077	0.134	0.044	−0.175	0.004	0.005	0.009	0.002	−0.007
	SD	0.220	0.223	0.234	0.240	0.260	0.225	0.221	0.212	0.211	0.221
	ESD	0.198	0.206	0.226	0.240	0.239	0.202	0.202	0.205	0.209	0.208
	MSE	0.052	0.056	0.073	0.060	0.098	0.051	0.049	0.045	0.045	0.049
	CP	0.896	0.908	0.934	0.918	0.920	0.902	0.914	0.922	0.924	0.926
$n=1000$	Bias	−0.067	0.064	0.135	0.053	−0.170	−0.004	−0.012	0.013	0.009	0.004
	SD	0.206	0.226	0.231	0.230	0.276	0.202	0.216	0.202	0.207	0.216
	ESD	0.202	0.207	0.220	0.230	0.230	0.203	0.203	0.203	0.206	0.205
	MSE	0.047	0.055	0.072	0.056	0.105	0.041	0.047	0.041	0.043	0.046
	CP	0.932	0.920	0.926	0.942	0.890	0.942	0.922	0.950	0.936	0.938
		IPW-P					IPW-O				
$n=500$	Bias	−0.026	0.004	0.032	−0.030	0.078	−0.016	0.025	0.061	0.003	0.123
	SD	0.343	0.496	0.672	0.826	0.920	0.367	0.529	0.717	0.891	1.108
	ESD	0.380	0.515	0.686	0.891	1.084	0.380	0.516	0.687	0.891	1.079
	MSE	0.119	0.246	0.453	0.683	0.852	0.135	0.280	0.518	0.795	1.243
	CP	0.956	0.958	0.934	0.954	0.968	0.944	0.946	0.932	0.936	0.928

Table 2 (continued)

		IPW-P					IPW-O				
$n=1000$	Bias	−0.058	−0.024	0.038	0.019	0.069	−0.071	−0.041	0.029	0.022	0.096
	SD	0.341	0.480	0.666	0.803	0.900	0.377	0.493	0.695	0.834	1.073
	ESD	0.384	0.514	0.685	0.889	1.087	0.380	0.510	0.683	0.887	1.084
	MSE	0.120	0.231	0.445	0.646	0.814	0.147	0.245	0.484	0.695	1.160
	CP	0.972	0.958	0.962	0.966	0.974	0.956	0.950	0.938	0.950	0.954
*Group*3: $\{a_1=0.03, a_4=0.08, a_2=0.16\}$											
n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.014	−0.002	0.003	−0.001	−0.003	−0.013	−0.002	0.003	−0.002	−0.005
	SD	0.210	0.194	0.201	0.207	0.204	0.215	0.195	0.203	0.210	0.204
	ESD	0.209	0.211	0.205	0.231	0.208	0.209	0.211	0.205	0.231	0.208
	MSE	0.044	0.038	0.041	0.043	0.041	0.046	0.038	0.041	0.044	0.042
	CP	0.916	0.952	0.932	0.930	0.940	0.904	0.952	0.930	0.926	0.940
$n=1000$	Bias	−0.003	0.005	0.007	−0.004	0.000	−0.004	0.006	0.009	−0.003	−0.001
	SD	0.186	0.195	0.205	0.198	0.211	0.190	0.196	0.207	0.198	0.212
	ESD	0.203	0.203	0.204	0.204	0.205	0.203	0.203	0.204	0.204	0.205
	MSE	0.035	0.038	0.042	0.039	0.044	0.036	0.038	0.043	0.039	0.045
	CP	0.962	0.938	0.944	0.946	0.918	0.958	0.948	0.940	0.936	0.912

Table 2 (continued)

		IPW-N					IPW-S				
$n=500$	Bias	−0.054	−0.047	−0.004	0.011	0.160	−0.076	−0.069	0.002	0.083	0.242
	SD	0.348	0.470	0.585	0.759	0.859	0.338	0.464	0.575	0.747	0.875
	ESD	0.372	0.512	0.688	0.903	1.111	0.369	0.512	0.695	0.917	1.126
	MSE	0.124	0.223	0.342	0.575	0.763	0.120	0.220	0.330	0.565	0.824
	CP	0.952	0.958	0.968	0.980	0.982	0.950	0.962	0.80	0.982	0.990
$n=1000$	Bias	−0.070	−0.030	−0.003	−0.018	0.279	−0.090	−0.043	−0.002	0.050	0.370
	SD	0.329	0.452	0.599	0.794	0.874	0.319	0.428	0.583	0.780	0.886
	ESD	0.375	0.512	0.685	0.893	1.109	0.372	0.514	0.690	0.904	1.121
	MSE	0.113	0.205	0.359	0.631	0.842	0.110	0.185	0.340	0.612	0.921
	CP	0.976	0.968	0.976	0.976	0.986	0.982	0.976	0.980	0.988	0.982
*Group*3: $\{a_1=0.03, a_4=0.08, a_2=0.16\}$											
n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.059	0.061	0.108	0.034	−0.136	−0.010	−0.002	0.008	0.000	0.000
	SD	0.214	0.201	0.230	0.243	0.246	0.221	0.203	0.207	0.221	0.210
	ESD	0.203	0.213	0.228	0.264	0.245	0.209	0.209	0.207	0.235	0.213
	MSE	0.049	0.044	0.064	0.060	0.079	0.049	0.041	0.043	0.049	0.044
	CP	0.900	0.940	0.924	0.928	0.938	0.896	0.940	0.926	0.904	0.936

Table 2 (continued)

*Group*3: $\{a_1 = 0.03, a_4 = 0.08, a_2 = 0.16\}$

n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n = 1000$	Bias	−0.049	0.068	0.104	0.034	−0.120	−0.002	0.012	0.013	0.000	0.005
	SD	0.189	0.210	0.243	0.225	0.256	0.194	0.205	0.217	0.205	0.213
	ESD	0.200	0.206	0.222	0.230	0.231	0.201	0.203	0.206	0.206	0.207
	MSE	0.038	0.049	0.070	0.052	0.080	0.038	0.042	0.047	0.042	0.045
	CP	0.952	0.930	0.910	0.936	0.908	0.944	0.930	0.924	0.934	0.916
		IPW-P					IPW-O				
$n = 500$	Bias	−0.022	−0.029	0.003	−0.006	0.018	−0.016	−0.022	0.009	−0.010	0.013
	SD	0.357	0.486	0.624	0.790	0.887	0.389	0.521	0.663	0.823	1.079
	ESD	0.379	0.515	0.691	0.900	1.094	0.377	0.513	0.688	0.895	1.084
	MSE	0.128	0.237	0.389	0.624	0.786	0.152	0.272	0.439	0.678	1.164
	CP	0.950	0.952	0.966	0.970	0.986	0.926	0.932	0.948	0.966	0.948
$n = 1000$	Bias	−0.035	−0.010	0.008	−0.036	0.115	−0.034	−0.010	0.006	−0.042	0.098
	SD	0.333	0.472	0.648	0.845	0.931	0.372	0.504	0.677	0.927	1.037
	ESD	0.380	0.515	0.686	0.891	1.094	0.379	0.513	0.685	0.888	1.088
	MSE	0.112	0.223	0.420	0.715	0.879	0.139	0.254	0.459	0.860	1.085
	CP	0.972	0.958	0.962	0.974	0.972	0.950	0.944	0.946	0.932	0.946

Table 2 (continued)

*Group*4: $\{a_1=0.03, a_4=0.1, a_2=0.15\}$											
n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.012	0.002	0.006	−0.009	0.008	0.010	0.003	0.008	−0.008	0.006
	SD	0.209	0.198	0.211	0.212	0.202	0.213	0.200	0.215	0.213	0.206
	ESD	0.232	0.211	0.208	0.212	0.209	0.233	0.211	0.208	0.212	02.09
	MSE	0.044	0.039	0.044	0.045	0.041	0.046	0.040	0.046	0.046	0.042
	CP	0.926	0.946	0.922	0.932	0.942	0.916	0.946	0.922	0.924	0.942
$n=1000$	Bias	−0.004	0.021	0.002	0.016	−0.006	−0.005	0.022	0.005	0.018	−0.005
	SD	0.204	0.203	0.214	0.191	0.206	0.209	0.203	0.214	0.192	0.209
	ESD	0.205	0.203	0.203	0.202	0.204	0.206	0.203	0.203	0.202	0.204
	MSE	0.042	0.042	0.046	0.037	0.043	0.044	0.042	0.046	0.037	0.044
	CP	0.936	0.940	0.922	0.946	0.930	0.944	0.944	0.928	0.946	0.920
		IPW-N					IPW-S				
$n=500$	Bias	−0.063	0.032	−0.056	−0.035	0.203	−0.091	0.006	−0.039	0.037	0.400
	SD	0.336	0.472	0.596	0.773	0.865	0.327	0.474	0.601	0.779	0.993
	ESD	0.371	0.518	0.692	0.893	1.106	0.367	0.517	0.698	0.906	1.132
	MSE	0.117	0.224	0.358	0.598	0.790	0.115	0.225	0.363	0.609	1.145
	CP	0.956	0.962	0.972	0.982	0.982	0.962	0.964	0.978	0.976	0.978

Table 2 (continued)

		IPW-N					IPW-S				
$n=1000$	Bias	−0.051	−0.057	−0.014	0.033	0.195	−0.071	−0.073	0.012	0.127	0.338
	SD	0.336	0.441	0.602	0.754	0.811	0.335	0.442	0.595	0.762	0.879
	ESD	0.377	0.509	0.680	0.897	1.110	0.375	0.511	0.687	0.909	1.124
	MSE	0.115	0.198	0.363	0.570	0.695	0.117	0.201	0.354	0.597	0.888
	CP	0.974	0.978	0.970	0.970	0.992	0.964	0.976	0.964	0.972	0.986
*Group*4: $\{a_1=0.03, a_4=0.1, a_2=0.15\}$											
n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.040	0.072	0.115	0.027	−0.128	0.010	0.004	0.008	−0.007	0.014
	SD	0.210	0.204	0.242	0.245	0.247	0.215	0.204	0.218	0.223	0.216
	ESD	0.216	0.214	0.231	0.250	0.246	0.235	0.212	0.210	0.215	0.213
	MSE	0.046	0.047	0.072	0.061	0.078	0.046	0.042	0.048	0.050	0.047
	CP	0.920	0.948	0.928	0.928	0.928	0.914	0.936	0.932	0.922	0.920
$n=1000$	Bias	−0.051	0.084	0.095	0.054	−0.130	−0.004	0.027	0.005	0.018	−0.003
	SD	0.212	0.224	0.252	0.224	0.254	0.215	0.216	0.223	0.199	0.218
	ESD	0.202	0.206	0.219	0.226	0.230	0.206	0.204	0.206	0.204	0.207
	MSE	0.048	0.057	0.073	0.053	0.082	0.046	0.047	0.050	0.040	0.047
	CP	0.938	0.926	0.890	0.944	0.908	0.932	0.928	0.920	0.946	0.922

Table 2 (continued)

		IPW-P					IPW-O				
$n=500$	Bias	−0.032	0.050	−0.051	−0.062	0.086	−0.035	0.045	−0.053	−0.066	0.085
	SD	0.344	0.500	0.633	0.814	0.913	0.371	0.509	0.674	0.864	1.070
	ESD	0.378	0.521	0.693	0.890	1.092	0.375	0.518	0.690	0.885	1.083
	MSE	0.119	0.253	0.403	0.667	0.840	0.139	0.261	0.457	0.751	1.152
	CP	0.960	0.954	0.970	0.974	0.976	0.942	0.944	0.946	0.962	0.946
$n=1000$	Bias	−0.017	−0.025	0.015	0.016	0.029	−0.021	−0.035	0.003	−0.001	0.004
	SD	0.346	0.483	0.654	0.806	0.888	0.379	0.510	0.679	0.862	1.012
	ESD	0.383	0.514	0.684	0.895	1.094	0.381	0.511	0.680	0.890	1.087
	MSE	0.120	0.234	0.428	0.651	0.790	0.144	0.261	0.461	0.744	1.024
	CP	0.968	0.960	0.956	0.958	0.986	0.942	0.962	0.948	0.958	0.958
Group5: $\{a_1=0.05, a_4=0.07, a_2=0.15\}$											
n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.005	−0.002	−0.013	−0.003	0.006	−0.002	0.000	−0.013	−0.004	0.003
	SD	0.207	0.199	0.202	0.189	0.205	0.212	0.204	0.204	0.191	0.208
	ESD	0.202	0.205	0.204	0.204	0.204	0.202	0.205	0.204	0.204	0.205
	MSE	0.043	0.040	0.041	0.036	0.042	0.045	0.042	0.042	0.036	0.043
	CP	0.940	0.940	0.940	0.958	0.930	0.936	0.942	0.950	0.958	0.928

Table 2 (continued)

Group5: $\{a_1=0.05, a_4=0.07, a_2=0.15\}$											
n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=1000$	Bias	0.000	0.016	−0.002	0.007	−0.014	0.000	0.014	−0.004	0.007	−0.011
	SD	0.197	0.202	0.197	0.198	0.203	0.200	0.204	0.197	0.200	0.204
	ESD	0.201	0.200	0.202	0.202	0.201	0.200	0.200	0.202	0.202	0.201
	MSE	0.039	0.041	0.039	0.039	0.042	0.040	0.042	0.039	0.040	0.042
	CP	0.950	0.936	0.948	0.942	0.944	0.954	0.932	0.952	0.948	0.948
		IPW-N					IPW-S				
$n=500$	Bias	−0.033	−0.032	−0.064	−0.025	0.208	−0.052	−0.051	−0.050	0.024	0.291
	SD	0.362	0.444	0.597	0.743	0.824	0.344	0.437	0.594	0.742	0.853
	ESD	0.373	0.511	0.684	0.899	1.111	0.371	0.513	0.693	0.910	1.128
	MSE	0.132	0.198	0.360	0.553	0.723	0.121	0.194	0.355	0.550	0.812
	CP	0.946	0.974	0.970	0.980	0.984	0.952	0.978	0.976	0.982	0.988
$n=1000$	Bias	−0.070	−0.001	0.012	0.011	0.196	−0.083	−0.021	0.027	0.068	0.259
	SD	0.333	0.465	0.606	0.782	0.838	0.325	0.438	0.590	0.776	0.869
	ESD	0.372	0.510	0.687	0.897	1.108	0.371	0.511	0.693	0.907	1.117
	MSE	0.116	0.216	0.368	0.611	0.740	0.113	0.192	0.349	0.608	0.822
	CP	0.968	0.964	0.962	0.974	0.994	0.970	0.976	0.970	0.974	0.990

Table 2 (continued)

*Group*5: $\{a_1=0.05, a_4=0.07, a_2=0.15\}$											
n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.058	0.079	0.108	0.036	−0.152	0.005	0.003	−0.012	0.000	0.006
	SD	0.218	0.216	0.236	0.218	0.249	0.222	0.217	0.214	0.199	0.216
	ESD	0.197	0.208	0.225	0.238	0.239	0.202	0.205	0.206	0.208	0.209
	MSE	0.051	0.053	0.067	0.049	0.085	0.049	0.047	0.046	0.039	0.047
	CP	0.926	0.926	0.920	0.956	0.924	0.912	0.922	0.932	0.942	0.934
$n=1000$	Bias	−0.063	0.096	0.127	0.049	−0.184	−0.001	0.020	0.003	0.011	−0.013
	SD	0.212	0.226	0.238	0.231	0.265	0.209	0.218	0.207	0.212	0.212
	ESD	0.198	0.204	0.220	0.229	0.229	0.200	0.201	0.204	0.204	0.204
	MSE	0.049	0.060	0.073	0.056	0.104	0.044	0.048	0.043	0.045	0.045
	CP	0.942	0.914	0.908	0.938	0.896	0.942	0.914	0.940	0.930	0.942
		IPW-P					IPW-O				
$n=500$	Bias	−0.001	−0.011	−0.059	−0.045	0.099	0.006	−0.010	−0.061	−0.057	0.072
	SD	0.365	0.467	0.634	0.798	0.871	0.408	0.492	0.671	0.856	1.090
	ESD	0.381	0.515	0.686	0.896	1.098	0.379	0.512	0.682	0.890	1.085
	MSE	0.133	0.218	0.405	0.638	0.768	0.167	0.242	0.454	0.736	1.192
	CP	0.946	0.956	0.960	0.978	0.984	0.932	0.950	0.948	0.950	0.934
$n=1000$	Bias	−0.035	0.019	0.021	−0.008	0.062	−0.035	0.016	0.014	−0.027	0.036
	SD	0.349	0.487	0.655	0.835	0.905	0.382	0.515	0.688	0.886	1.069
	ESD	0.377	0.512	0.689	0.895	1.095	0.376	0.511	0.687	0.891	1.088
	MSE	0.123	0.238	0.430	0.696	0.822	0.147	0.266	0.474	0.786	1.143
	CP	0.958	0.958	0.948	0.968	0.984	0.954	0.946	0.938	0.950	0.944

Table 3 The distribution of $\sqrt{nh_1}[\hat{\tau}(x_1) - \tau(x_1)]$ for model 2

*Group*1: $\{a_1=0.1, a_4=0.1, a_2=0.6\}$											
n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.048	0.015	0.015	0.014	0.109	0.048	0.016	0.018	0.018	0.117
	SD	0.339	0.351	0.327	0.352	0.357	0.342	0.355	0.331	0.353	0.365
	ESD	0.337	0.350	0.347	0.354	0.341	0.337	0.350	0.347	0.354	0.341
	MSE	0.117	0.124	0.107	0.124	0.139	0.119	0.126	0.110	0.125	0.147
	CP	0.942	0.938	0.970	0.944	0.926	0.938	0.936	0.966	0.948	0.920
$n=1000$	Bias	0.076	0.012	−0.003	0.006	0.190	0.076	0.008	−0.008	0.001	0.187
	SD	0.338	0.354	0.332	0.337	0.340	0.345	0.356	0.335	0.336	0.346
	ESD	0.338	0.346	0.349	0.349	0.341	0.338	0.346	0.349	0.349	0.341
	MSE	0.120	0.126	0.110	0.114	0.152	0.125	0.127	0.112	0.113	0.155
	CP	0.940	0.940	0.958	0.944	0.934	0.940	0.942	0.952	0.952	0.928
		IPW-N					IPW-S				
$n=500$	Bias	−0.159	−0.274	−0.203	0.500	1.196	−0.398	−0.410	−0.391	0.109	0.412
	SD	1.474	1.267	1.272	1.406	1.866	1.503	1.385	1.406	1.550	2.003
	ESD	1.751	1.533	1.595	1.712	2.129	1.754	1.536	1.592	1.698	2.099
	MSE	2.198	1.680	1.658	2.226	4.912	2.419	2.087	2.130	2.414	4.183
	CP	0.982	0.980	0.986	0.976	0.974	0.972	0.966	0.966	0.960	0.960
$n=1000$	Bias	−0.144	−0.431	−0.428	0.590	1.800	−0.417	−0.491	−0.547	0.153	0.884
	SD	1.522	1.402	1.405	1.509	1.878	1.720	1.571	1.580	1.613	2.123
	ESD	1.725	1.542	1.597	1.715	2.086	1.720	1.545	1.598	1.700	2.057
	MSE	2.338	2.150	2.157	2.626	6.767	3.133	2.708	2.794	2.625	5.289
	CP	0.962	0.976	0.968	0.958	0.964	0.954	0.956	0.954	0.958	0.956

Table 3 (continued)

*Group*1: $\{a_1=0.1, a_4=0.1, a_2=0.6\}$											
n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.052	0.023	0.021	0.018	0.108	0.220	0.075	0.068	0.036	−0.101
	SD	0.340	0.356	0.329	0.353	0.360	0.334	0.348	0.340	0.349	0.399
	ESD	0.337	0.349	0.347	0.354	0.341	0.299	0.307	0.331	0.332	0.349
	MSE	0.118	0.127	0.109	0.125	0.141	0.160	0.127	0.120	0.123	0.169
	CP	0.936	0.936	0.966	0.948	0.924	0.920	0.914	0.938	0.934	0.898
$n=1000$	Bias	0.078	0.012	−0.004	0.000	0.180	0.315	0.077	0.035	0.028	−0.079
	SD	0.342	0.356	0.334	0.339	0.346	0.416	0.335	0.320	0.350	0.350
	ESD	0.338	0.346	0.349	0.349	0.342	0.381	0.305	0.319	0.337	0.324
	MSE	0.123	0.127	0.111	0.115	0.152	0.273	0.118	0.103	0.123	0.128
	CP	0.942	0.944	0.956	0.950	0.928	0.918	0.920	0.942	0.934	0.916
		IPW-P					IPW-O				
$n=500$	Bias	−0.027	0.078	0.051	0.181	−0.153	−0.101	0.013	−0.008	0.115	−0.265
	SD	0.971	1.204	1.261	1.289	1.163	1.685	1.469	1.501	1.624	1.893
	ESD	1.755	1.544	1.596	1.678	2.032	1.728	1.528	1.581	1.661	2.002
	MSE	0.944	1.456	1.593	1.695	1.375	2.849	2.160	2.254	2.651	3.654
	CP	0.998	0.988	0.988	0.978	1.000	0.960	0.958	0.956	0.954	0.954
$n=1000$	Bias	−0.011	0.041	−0.072	0.190	−0.053	−0.049	0.002	−0.120	0.125	−0.179
	SD	1.076	1.316	1.398	1.337	1.185	1.681	1.604	1.660	1.718	1.973
	ESD	1.717	1.548	1.595	1.679	1.983	1.705	1.540	1.588	1.670	1.966
	MSE	1.158	1.733	1.960	1.825	1.407	2.827	2.572	2.770	2.968	3.923
	CP	0.996	0.984	0.972	0.984	1.000	0.966	0.934	0.944	0.928	0.936

Table 3 (continued)

*Group*2: $\{a_1=0.1, a_4=0.12, a_2=0.6\}$

n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.070	−0.004	0.006	−0.004	0.135	0.068	−0.003	0.007	−0.002	0.139
	SD	0.341	0.356	0.356	0.338	0.339	0.345	0.361	0.361	0.338	0.346
	ESD	0.339	0.347	0.350	0.350	0.344	0.339	0.348	0.350	0.351	0.344
	MSE	0.121	0.127	0.127	0.114	0.133	0.123	0.130	0.130	0.114	0.139
	CP	0.942	0.946	0.946	0.966	0.956	0.930	0.948	0.946	0.958	0.950
$n=1000$	Bias	0.097	−0.006	0.002	−0.016	0.156	0.097	−0.010	−0.003	−0.018	0.160
	SD	0.332	0.338	0.356	0.346	0.337	0.337	0.341	0.360	0.351	0.347
	ESD	0.336	0.348	0.350	0.352	0.343	0.336	0.348	0.350	0.352	0.343
	MSE	0.120	0.114	0.127	0.120	0.138	0.123	0.116	0.130	0.124	0.146
	CP	0.944	0.946	0.940	0.954	0.962	0.950	0.948	0.936	0.950	0.952
		IPW-N					IPW-S				
$n=500$	Bias	−0.225	−0.302	−0.219	0.446	1.230	−0.416	−0.329	−0.265	0.165	0.527
	SD	1.524	1.310	1.338	1.415	1.756	1.415	1.255	1.292	1.370	1.636
	ESD	1.753	1.523	1.590	1.702	2.101	1.750	1.529	1.595	1.695	2.069
	MSE	1.540	1.345	1.356	1.483	2.144	1.475	1.298	1.319	1.380	1.719
	CP	0.964	0.974	0.976	0.984	0.976	0.970	0.976	0.984	0.982	0.978
$n=1000$	Bias	−0.149	−0.429	−0.428	0.577	1.824	−0.441	−0.459	−0.510	0.155	0.810
	SD	1.504	1.279	1.294	1.532	1.856	1.579	1.385	1.424	1.619	2.129
	ESD	1.729	1.535	1.596	1.715	2.082	1.720	1.538	1.596	1.700	2.044
	MSE	2.284	1.821	1.857	2.679	6.773	2.688	2.128	2.287	2.645	5.186
	CP	0.976	0.986	0.980	0.974	0.974	0.962	0.964	0.966	0.950	0.944

Table 3 (continued)

*Group*2: $\{a_1=0.1, a_4=0.12, a_2=0.6\}$

n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.073	0.003	0.011	−0.003	0.128	0.233	0.065	0.045	0.022	−0.089
	SD	0.342	0.362	0.359	0.339	0.345	0.356	0.339	0.389	0.415	0.388
	ESD	0.339	0.347	0.350	0.350	0.344	0.335	0.302	0.353	0.396	0.365
	MSE	0.122	0.131	0.129	0.115	0.135	0.181	0.119	0.153	0.173	0.158
	CP	0.940	0.944	0.946	0.958	0.948	0.894	0.912	0.924	0.944	0.928
$n=1000$	Bias	0.099	−0.003	0.002	−0.020	0.150	0.329	0.060	0.049	0.005	−0.103
	SD	0.335	0.341	0.357	0.352	0.346	0.325	0.321	0.345	0.352	0.346
	ESD	0.335	0.348	0.349	0.352	0.344	0.299	0.307	0.320	0.334	0.326
	MSE	0.122	0.116	0.128	0.125	0.142	0.214	0.106	0.121	0.124	0.130
	CP	0.948	0.946	0.940	0.948	0.952	0.934	0.940	0.926	0.928	0.936
		IPW-P					IPW-O				
$n=500$	Bias	−0.054	0.081	0.033	0.104	−0.174	−0.088	0.080	0.072	0.176	−0.099
	SD	1.005	1.194	1.319	1.276	1.124	1.680	1.561	1.535	1.626	1.914
	ESD	1.758	1.536	1.591	1.668	2.000	1.737	1.524	1.583	1.660	1.983
	MSE	1.006	1.197	1.319	1.280	1.137	1.682	1.563	1.537	1.635	1.917
	CP	1.000	0.986	0.982	0.988	1.000	0.950	0.930	0.946	0.960	0.954
$n=1000$	Bias	0.006	0.063	−0.068	0.158	−0.104	−0.095	−0.033	−0.157	0.069	−0.196
	SD	1.005	1.184	1.279	1.407	1.241	1.728	1.478	1.497	1.630	2.037
	ESD	1.722	1.541	1.594	1.678	1.976	1.706	1.531	1.585	1.669	1.959
	MSE	1.010	1.407	1.642	2.005	1.551	2.997	2.186	2.264	2.660	4.188
	CP	0.998	0.992	0.982	0.972	0.998	0.952	0.958	0.962	0.950	0.928

Table 3 (continued)

*Group*3: $\{a_1=0.1, a_4=0.14, a_2=0.6\}$

n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.064	0.006	0.003	−0.017	0.102	0.064	0.005	0.002	−0.020	0.101
	SD	0.328	0.350	0.359	0.352	0.347	0.329	0.351	0.361	0.354	0.354
	ESD	0.336	0.350	0.347	0.353	0.343	0.336	0.350	0.347	0.354	0.343
	MSE	0.111	0.123	0.129	0.124	0.131	0.112	0.123	0.131	0.126	0.136
	CP	0.950	0.940	0.934	0.944	0.932	0.952	0.940	0.938	0.940	0.932
$n=1000$	Bias	0.052	0.006	−0.023	−0.023	0.124	0.054	0.006	−0.023	−0.023	0.128
	SD	0.322	0.350	0.330	0.357	0.363	0.332	0.353	0.333	0.363	0.365
	ESD	0.336	0.347	0.349	0.350	0.343	0.336	0.347	0.349	0.350	0.344
	MSE	0.106	0.123	0.109	0.128	0.147	0.113	0.125	0.112	0.132	0.149
	CP	0.950	0.946	0.962	0.932	0.918	0.948	0.948	0.964	0.936	0.928
		IPW-N					IPW-S				
$n=500$	Bias	−0.132	−0.256	−0.378	0.429	1.287	−0.312	−0.270	−0.426	0.172	0.641
	SD	1.518	1.294	1.420	1.493	1.904	1.351	1.251	1.336	1.424	1.736
	ESD	1.746	1.533	1.595	1.704	2.105	1.746	1.540	1.598	1.696	2.077
	MSE	2.322	1.741	2.159	2.412	5.281	1.923	1.638	1.966	2.057	3.425
	CP	0.974	0.970	0.956	0.974	0.968	0.984	0.974	0.964	0.980	0.972
$n=1000$	Bias	−0.133	−0.472	−0.398	0.642	1.677	−0.307	−0.473	−0.398	0.367	0.874
	SD	1.451	1.299	1.352	1.478	1.841	1.387	1.285	1.335	1.454	1.859
	ESD	1.713	1.529	1.589	1.700	2.073	1.707	1.530	1.590	1.690	2.041
	MSE	2.123	1.911	1.987	2.596	6.200	2.019	1.874	1.940	2.249	4.217
	CP	0.986	0.982	0.972	0.978	0.960	0.986	0.974	0.974	0.970	0.954

Table 3 (continued)

*Group*3: $\{a_1=0.1, a_4=0.14, a_2=0.6\}$											
n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	-0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.070	0.014	0.008	-0.023	0.087	0.260	0.071	0.049	0.004	-0.116
	SD	0.325	0.348	0.360	0.353	0.353	0.463	0.368	0.366	0.347	0.362
	ESD	0.335	0.349	0.347	0.353	0.343	0.451	0.353	0.324	0.331	0.327
	MSE	0.110	0.122	0.130	0.125	0.132	0.283	0.140	0.137	0.120	0.144
	CP	0.956	0.942	0.936	0.938	0.930	0.932	0.924	0.924	0.924	0.910
$n=1000$	Bias	0.058	0.014	-0.019	-0.025	0.116	0.285	0.075	0.022	-0.004	-0.136
	SD	0.331	0.355	0.331	0.360	0.363	0.318	0.339	0.327	0.359	0.363
	ESD	0.336	0.346	0.348	0.350	0.344	0.298	0.306	0.320	0.332	0.325
	MSE	0.113	0.126	0.110	0.130	0.145	0.182	0.120	0.107	0.129	0.150
	CP	0.948	0.946	0.966	0.934	0.928	0.930	0.922	0.940	0.928	0.910
		IPW-P					IPW-O				
$n=500$	Bias	−0.020	0.092	-0.123	0.115	−0.077	-0.035	0.078	−0.135	0.101	−0.113
	SD	0.990	1.196	1.410	1.361	1.200	1.742	1.546	1.622	1.687	1.948
	ESD	1.746	1.542	1.595	1.671	2.007	1.725	1.531	1.585	1.658	1.984
	MSE	0.980	1.439	2.004	1.864	1.445	3.034	2.397	2.648	2.857	3.808
	CP	0.998	0.986	0.956	0.982	1.000	0.936	0.934	0.946	0.940	0.956
$n=1000$	Bias	-0.018	0.000	-0.038	0.242	-0.219	0.051	0.035	-0.011	0.275	-0.172
	SD	0.938	1.230	1.347	1.321	1.183	1.708	1.496	1.551	1.673	2.022
	ESD	1.704	1.535	1.587	1.663	1.968	1.697	1.532	1.584	1.659	1.958
	MSE	0.880	1.513	1.815	1.805	1.447	2.920	2.238	2.405	2.876	4.116
	CP	1.000	0.990	0.980	0.988	1.000	0.948	0.956	0.950	0.938	0.934

Table 3 (continued)

*Group*4: $\{a_1=0.1,\ a_4=0.16,\ a_2=0.6\}$

n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.071	−0.004	−0.018	0.031	0.105	0.068	−0.008	−0.022	0.028	0.105
	SD	0.331	0.328	0.351	0.352	0.334	0.339	0.331	0.352	0.356	0.337
	ESD	0.335	0.347	0.351	0.348	0.341	0.335	0.347	0.351	0.348	0.341
	MSE	0.114	0.107	0.124	0.125	0.123	0.119	0.110	0.125	0.128	0.125
	CP	0.952	0.952	0.928	0.948	0.954	0.946	0.954	0.932	0.940	0.944
$n=1000$	Bias	0.079	−0.015	−0.020	0.005	0.179	0.079	−0.020	−0.026	0.002	0.176
	SD	0.332	0.350	0.327	0.340	0.340	0.344	0.353	0.327	0.345	0.342
	ESD	0.337	0.348	0.346	0.349	0.342	0.337	0.348	0.346	0.349	0.342
	MSE	0.117	0.123	0.107	0.116	0.148	0.125	0.125	0.108	0.119	0.148
	CP	0.952	0.950	0.956	0.958	0.946	0.946	0.944	0.956	0.944	0.942
		IPW-N					IPW-S				
$n=500$	Bias	−0.270	−0.302	−0.222	0.567	1.289	−0.382	−0.286	−0.227	0.339	0.672
	SD	1.475	1.238	1.284	1.470	1.835	1.292	1.154	1.202	1.379	1.649
	ESD	1.742	1.528	1.594	1.697	2.097	1.745	1.534	1.597	1.687	2.067
	MSE	2.250	1.623	1.697	2.483	5.028	1.815	1.414	1.496	2.015	3.171
	CP	0.978	0.978	0.982	0.980	0.966	0.992	0.988	0.990	0.984	0.982
$n=1000$	Bias	−0.247	−0.383	−0.516	0.661	1.863	−0.319	−0.324	−0.484	0.397	1.106
	SD	1.594	1.361	1.358	1.466	1.925	1.376	1.258	1.278	1.385	1.831
	ESD	1.729	1.528	1.585	1.714	2.079	1.730	1.532	1.586	1.704	2.049
	MSE	2.603	1.998	2.110	2.587	7.176	1.996	1.687	1.869	2.076	4.574
	CP	0.972	0.970	0.984	0.976	0.954	0.988	0.974	0.986	0.990	0.966

Table 3 (continued)

*Group*4: $\{a_1=0.1,\ a_4=0.16,\ a_2=0.6\}$

n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.073	0.004	−0.015	0.025	0.089	0.235	0.075	0.029	0.043	−0.222
	SD	0.337	0.332	0.352	0.354	0.334	0.335	0.527	0.336	0.369	2.347
	ESD	0.334	0.345	0.349	0.347	0.341	0.303	0.518	0.315	0.347	2.643
	MSE	0.119	0.110	0.124	0.126	0.119	0.167	0.284	0.114	0.138	5.557
	CP	0.944	0.952	0.932	0.942	0.948	0.914	0.932	0.916	0.916	0.914
$n=1000$	Bias	0.083	−0.009	−0.019	−0.002	0.161	0.308	0.056	0.058	0.026	−0.098
	SD	0.339	0.351	0.326	0.345	0.340	0.326	0.334	0.869	0.348	0.344
	ESD	0.336	0.347	0.346	0.348	0.342	0.303	0.307	0.861	0.335	0.330
	MSE	0.122	0.123	0.107	0.119	0.141	0.201	0.115	0.758	0.122	0.128
	CP	0.952	0.946	0.958	0.956	0.944	0.932	0.928	0.946	0.936	0.924
		IPW-P					IPW-O				
$n=500$	Bias	−0.054	0.088	0.028	0.184	−0.158	−0.144	0.055	0.053	0.261	−0.061
	SD	1.026	1.122	1.272	1.300	1.172	1.712	1.445	1.498	1.629	1.944
	ESD	1.752	1.542	1.595	1.660	1.995	1.724	1.528	1.585	1.653	1.980
	MSE	1.055	1.268	1.619	1.723	1.399	2.952	2.090	2.247	2.722	3.784
	CP	0.998	0.992	0.980	0.992	1.000	0.952	0.950	0.964	0.948	0.934
$n=1000$	Bias	0.005	0.121	−0.187	0.213	−0.057	−0.103	0.099	−0.165	0.253	−0.039
	SD	1.056	1.265	1.341	1.319	1.245	1.797	1.614	1.556	1.657	2.098
	ESD	1.729	1.536	1.583	1.678	1.975	1.711	1.528	1.577	1.672	1.962
	MSE	1.116	1.614	1.833	1.786	1.554	3.239	2.614	2.447	2.811	4.405
	CP	0.998	0.984	0.980	0.988	0.998	0.940	0.942	0.952	0.950	0.932

Table 3 (continued)

Group5: $\{a_1=0.1,\ a_4=0.18,\ a_2=0.6\}$

n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.032	−0.010	−0.013	−0.003	0.111	0.030	−0.012	−0.016	−0.007	0.105
	SD	0.345	0.330	0.344	0.363	0.343	0.352	0.334	0.344	0.369	0.346
	ESD	0.333	0.347	0.350	0.350	0.342	0.333	0.347	0.350	0.349	0.342
	MSE	0.120	0.109	0.119	0.132	0.130	0.125	0.112	0.119	0.136	0.131
	CP	0.934	0.956	0.940	0.934	0.946	0.940	0.954	0.940	0.938	0.938
$n=1000$	Bias	0.068	−0.012	−0.008	0.011	0.157	0.066	−0.013	−0.010	0.009	0.152
	SD	0.328	0.352	0.335	0.352	0.324	0.338	0.353	0.340	0.354	0.330
	ESD	0.335	0.346	0.347	0.351	0.343	0.335	0.346	0.347	0.351	0.343
	MSE	0.112	0.124	0.112	0.124	0.130	0.119	0.125	0.116	0.125	0.132
	CP	0.944	0.948	0.958	0.938	0.960	0.944	0.948	0.948	0.948	0.960
		IPW-N					IPW-S				
$n=500$	Bias	−0.181	−0.300	−0.296	0.335	1.354	−0.304	−0.298	−0.285	0.159	0.837
	SD	1.570	1.178	1.453	1.467	1.907	1.365	1.108	1.330	1.370	1.723
	ESD	1.734	1.529	1.594	1.705	2.102	1.732	1.534	1.599	1.698	2.078
	MSE	2.498	1.478	2.198	2.263	5.470	1.954	1.317	1.851	1.901	3.671
	CP	0.964	0.990	0.962	0.978	0.968	0.978	0.994	0.968	0.988	0.976
$n=1000$	Bias	−0.217	−0.433	−0.384	0.599	1.797	−0.316	−0.402	−0.368	0.377	1.126
	SD	1.542	1.286	1.353	1.414	1.747	1.405	1.201	1.274	1.358	1.670
	ESD	1.712	1.522	1.576	1.708	2.075	1.709	1.524	1.576	1.697	2.047
	MSE	2.426	1.841	1.977	2.358	6.282	2.074	1.603	1.759	1.988	4.055
	CP	0.964	0.988	0.976	0.984	0.966	0.982	0.986	0.980	0.990	0.976

Table 3 (continued)

*Group*5: $\{a_1=0.1,\ a_4=0.16,\ a_2=0.6\}$											
n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.037	−0.001	−0.007	−0.010	0.087	0.201	0.047	0.035	0.009	−0.102
	SD	0.351	0.331	0.344	0.368	0.348	0.354	0.322	0.337	0.385	0.363
	ESD	0.331	0.345	0.348	0.348	0.340	0.303	0.303	0.316	0.336	0.347
	MSE	0.125	0.110	0.118	0.136	0.128	0.166	0.106	0.115	0.148	0.142
	CP	0.936	0.956	0.940	0.938	0.938	0.914	0.928	0.920	0.914	0.916
$n=1000$	Bias	0.073	−0.003	−0.004	0.005	0.135	0.303	0.056	0.040	0.026	−0.153
	SD	0.336	0.354	0.337	0.353	0.329	0.328	0.338	0.332	0.354	1.124
	ESD	0.334	0.345	0.346	0.350	0.342	0.302	0.305	0.317	0.332	1.129
	MSE	0.118	0.126	0.114	0.125	0.127	0.200	0.117	0.112	0.126	1.286
	CP	0.950	0.946	0.950	0.942	0.966	0.916	0.916	0.938	0.918	0.954
		IPW-P					IPW-				
$n=500$	Bias	−0.024	0.058	−0.039	0.010	−0.045	−0.061	0.053	−0.038	0.031	−0.021
	SD	1.050	1.094	1.447	1.306	1.228	1.723	1.386	1.594	1.683	2.053
	ESD	1.737	1.540	1.596	1.673	2.005	1.715	1.529	1.586	1.662	1.982
	MSE	1.102	1.201	2.095	1.706	1.510	2.973	1.923	2.542	2.834	4.214
	CP	0.998	0.994	0.958	0.984	0.998	0.942	0.968	0.950	0.938	0.938
$n=1000$	Bias	−0.008	0.055	−0.052	0.170	−0.098	−0.045	0.076	−0.006	0.218	−0.075
	SD	1.035	1.168	1.336	1.275	1.110	1.714	1.535	1.536	1.582	1.853
	ESD	1.709	1.529	1.574	1.672	1.971	1.696	1.524	1.570	1.667	1.961
	MSE	1.071	1.367	1.787	1.653	1.242	2.940	2.361	2.359	2.549	3.440
	CP	0.996	0.996	0.974	0.986	1.000	0.950	0.946	0.946	0.966	0.958

Table 4 The distribution of $\sqrt{nh_1}[\hat{\tau}(x_1) - \tau(x_1)]$ for model 3

*Group*1: $\{a_1=0.1, a_4=0.1, a_2=0.1\}$											
n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.005	−0.024	0.010	−0.001	−0.009	−0.005	−0.023	0.010	−0.002	−0.013
	SD	0.320	0.334	0.334	0.346	0.336	0.327	0.336	0.335	0.350	0.338
	ESD	0.336	0.334	0.336	0.332	0.337	0.336	0.334	0.335	0.332	0.338
	MSE	0.102	0.112	0.112	0.119	0.113	0.107	0.114	0.112	0.123	0.114
	CP	0.944	0.936	0.926	0.938	0.946	0.944	0.934	0.924	0.932	0.944
$n=1000$	Bias	0.013	−0.008	0.018	0.019	−0.016	0.015	−0.009	0.016	0.018	−0.012
	SD	0.331	0.344	0.347	0.319	0.327	0.332	0.345	0.346	0.318	0.330
	ESD	0.328	0.326	0.328	0.329	0.331	0.327	0.326	0.328	0.329	0.331
	MSE	0.109	0.119	0.121	0.102	0.107	0.110	0.119	0.120	0.101	0.109
	CP	0.942	0.922	0.938	0.956	0.946	0.944	0.920	0.934	0.960	0.940
		IPW-N					IPW-S				
$n=500$	Bias	0.020	0.024	0.022	−0.064	0.185	0.072	−0.080	−0.091	0.048	0.674
	SD	0.437	0.528	0.675	0.895	0.851	0.432	0.523	0.779	1.168	1.677
	ESD	0.488	0.585	0.861	1.194	1.576	0.458	0.576	0.871	1.231	1.645
	MSE	0.191	0.279	0.456	0.805	0.759	0.192	0.280	0.615	1.366	3.266
	CP	0.966	0.974	0.986	0.990	1.000	0.954	0.968	0.964	0.958	0.932
$n=1000$	Bias	−0.027	0.041	0.038	0.022	0.223	0.056	−0.133	−0.144	0.126	0.859
	SD	0.419	0.522	0.696	0.876	0.830	0.419	0.538	0.854	1.337	1.997
	ESD	0.486	0.577	0.853	1.193	1.557	0.459	0.564	0.858	1.228	1.629
	MSE	0.176	0.274	0.485	0.767	0.738	0.179	0.307	0.750	1.805	4.724
	CP	0.964	0.966	0.986	0.992	1.000	0.958	0.958	0.942	0.908	0.878

Table 4 (continued)

*Group*1: $\{a_1=0.1, a_4=0.1, a_2=0.1\}$

n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.001	−0.022	0.009	−0.001	−0.011	0.058	−0.019	0.012	−0.004	−0.042
	SD	0.328	0.338	0.338	0.346	0.339	0.316	0.322	0.331	0.341	0.328
	ESD	0.337	0.336	0.337	0.333	0.340	0.319	0.320	0.325	0.321	0.325
	MSE	0.108	0.115	0.114	0.120	0.115	0.103	0.104	0.110	0.116	0.109
	CP	0.942	0.936	0.930	0.940	0.940	0.928	0.938	0.916	0.928	0.934
$n=1000$	Bias	0.014	−0.009	0.018	0.017	−0.013	0.077	−0.004	0.018	0.023	−0.045
	SD	0.334	0.347	0.348	0.318	0.334	0.321	0.339	0.339	0.313	0.322
	ESD	0.329	0.327	0.329	0.330	0.333	0.312	0.315	0.319	0.320	0.320
	MSE	0.112	0.120	0.121	0.102	0.112	0.109	0.115	0.115	0.098	0.106
	CP	0.944	0.920	0.936	0.960	0.936	0.930	0.920	0.920	0.950	0.936
		IPW-P					IPW-O				
$n=500$	Bias	−0.044	0.009	0.016	−0.040	0.090	−0.037	0.015	0.023	−0.035	0.101
	SD	0.476	0.545	0.757	1.041	0.989	0.503	0.595	0.838	1.175	1.538
	ESD	0.500	0.583	0.857	1.191	1.556	0.496	0.582	0.854	1.185	1.541
	MSE	0.229	0.297	0.573	1.086	0.985	0.255	0.354	0.704	1.381	2.375
	CP	0.960	0.972	0.966	0.974	1.000	0.952	0.950	0.958	0.956	0.952
$n=1000$	Bias	−0.095	0.031	0.052	0.044	0.078	−0.096	0.024	0.050	0.040	0.061
	SD	0.472	0.542	0.815	1.049	1.054	0.488	0.580	0.877	1.207	1.577
	ESD	0.494	0.576	0.851	1.190	1.540	0.491	0.574	0.849	1.186	1.530
	MSE	0.232	0.295	0.667	1.102	1.114	0.247	0.337	0.772	1.457	2.491
	CP	0.954	0.954	0.946	0.968	0.996	0.936	0.950	0.936	0.942	0.936

Table 4 (continued)

*Group*2: $\{a_1=0.1, a_4=0.12, a_2=0.1\}$

n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.023	−0.001	−0.002	−0.004	−0.007	−0.027	−0.004	−0.003	−0.002	−0.003
	SD	0.329	0.343	0.328	0.313	0.324	0.336	0.342	0.330	0.313	0.328
	ESD	0.329	0.338	0.332	0.335	0.340	0.329	0.338	0.332	0.335	0.340
	MSE	0.109	0.118	0.107	0.098	0.105	0.114	0.117	0.109	0.098	0.107
	CP	0.928	0.928	0.934	0.954	0.950	0.936	0.930	0.930	0.952	0.944
$n=1000$	Bias	−0.005	−0.009	−0.006	−0.006	−0.010	−0.004	−0.010	−0.007	−0.006	−0.008
	SD	0.330	0.307	0.324	0.344	0.330	0.334	0.310	0.328	0.348	0.334
	ESD	0.331	0.326	0.329	0.331	0.329	0.331	0.325	0.329	0.331	0.329
	MSE	0.109	0.094	0.105	0.118	0.109	0.112	0.096	0.108	0.121	0.112
	CP	0.942	0.956	0.946	0.946	0.950	0.932	0.952	0.934	0.946	0.946
		IPW-N					IPW-S				
$n=500$	Bias	0.001	−0.012	−0.010	−0.117	0.052	0.065	−0.113	−0.142	−0.035	0.668
	SD	0.453	0.523	0.662	0.835	0.856	0.454	0.535	0.803	1.161	1.729
	ESD	0.486	0.576	0.855	1.205	1.572	0.459	0.566	0.858	1.241	1.655
	MSE	0.205	0.273	0.438	0.711	0.736	0.210	0.299	0.665	1.350	3.436
	CP	0.958	0.962	0.980	0.992	1.000	0.938	0.950	0.962	0.954	0.910
$n=1000$	Bias	−0.009	0.041	−0.040	−0.096	0.122	0.091	−0.091	−0.153	0.107	1.051
	SD	0.413	0.485	0.689	0.879	0.825	0.439	0.515	0.808	1.293	1.952
	ESD	0.489	0.578	0.851	1.189	1.541	0.460	0.568	0.860	1.228	1.625
	MSE	0.170	0.237	0.477	0.781	0.695	0.201	0.273	0.676	1.683	4.915
	CP	0.982	0.974	0.972	0.984	1.000	0.948	0.966	0.960	0.928	0.890

Table 4 (continued)

*Group*2: $\{a_1=0.1, a_4=0.12, a_2=0.1\}$

n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
n = 500	Bias	−0.024	−0.005	−0.004	−0.003	−0.005	0.006	−0.005	−0.018	−0.002	−0.012
	SD	0.340	0.348	0.330	0.316	0.329	0.338	0.340	0.408	0.316	0.327
	ESD	0.331	0.340	0.334	0.338	0.344	0.324	0.333	0.411	0.333	0.337
	MSE	0.116	0.121	0.109	0.100	0.108	0.115	0.115	0.167	0.100	0.107
	CP	0.932	0.928	0.934	0.958	0.942	0.926	0.930	0.936	0.950	0.940
n = 1000	Bias	−0.005	−0.009	−0.007	−0.004	−0.009	0.025	−0.008	−0.008	−0.006	−0.018
	SD	0.336	0.311	0.331	0.349	0.336	0.329	0.309	0.327	0.345	0.331
	ESD	0.332	0.327	0.331	0.333	0.331	0.324	0.321	0.325	0.328	0.325
	MSE	0.113	0.097	0.110	0.122	0.113	0.109	0.096	0.107	0.119	0.110
	CP	0.932	0.950	0.934	0.942	0.942	0.932	0.944	0.938	0.934	0.948
		IPW-P					IPW-O				
n = 500	Bias	−0.041	0.011	0.045	−0.073	0.136	−0.049	0.017	0.044	−0.071	0.090
	SD	0.510	0.569	0.799	1.009	1.041	0.537	0.605	0.819	1.187	1.496
	ESD	0.495	0.580	0.856	1.203	1.564	0.494	0.578	0.851	1.194	1.545
	MSE	0.262	0.323	0.641	1.024	1.102	0.291	0.366	0.672	1.414	2.247
	CP	0.946	0.940	0.962	0.980	0.996	0.926	0.940	0.964	0.940	0.956
n = 1000	Bias	−0.056	0.042	−0.016	−0.045	0.146	−0.040	0.049	−0.015	−0.062	0.156
	SD	0.464	0.541	0.808	1.072	1.073	0.492	0.581	0.864	1.151	1.563
	ESD	0.494	0.578	0.848	1.186	1.531	0.493	0.577	0.847	1.183	1.525
	MSE	0.218	0.295	0.653	1.152	1.173	0.244	0.340	0.746	1.329	2.468
	CP	0.958	0.946	0.956	0.954	0.990	0.948	0.948	0.942	0.950	0.944

Table 4 (continued)

*Group*3: $\{a_1=0.1, a_4=0.14, a_2=0.13\}$

n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.020	−0.001	−0.005	−0.025	−0.022	0.019	−0.001	−0.004	−0.024	−0.022
	SD	0.330	0.331	0.336	0.327	0.311	0.335	0.333	0.340	0.329	0.315
	ESD	0.339	0.336	0.335	0.333	0.335	0.339	0.336	0.335	0.333	0.335
	MSE	0.110	0.109	0.113	0.108	0.097	0.113	0.111	0.115	0.109	0.100
	CP	0.928	0.950	0.934	0.930	0.952	0.930	0.946	0.934	0.942	0.938
$n=1000$	Bias	0.008	−0.022	−0.008	−0.020	0.026	0.011	−0.021	−0.009	−0.020	0.030
	SD	0.310	0.322	0.336	0.331	0.316	0.316	0.325	0.337	0.333	0.321
	ESD	0.330	0.329	0.328	0.333	0.328	0.331	0.329	0.328	0.333	0.329
	MSE	0.096	0.104	0.113	0.110	0.100	0.100	0.106	0.113	0.111	0.104
	CP	0.956	0.936	0.928	0.936	0.946	0.950	0.924	0.926	0.942	0.942
		IPW-N					IPW-S				
$n=500$	Bias	0.043	0.032	0.041	0.042	0.559	0.071	−0.026	0.003	0.170	1.068
	SD	0.473	0.546	0.768	1.020	1.024	0.443	0.528	0.722	1.105	1.420
	ESD	0.482	0.584	0.863	1.211	1.602	0.468	0.578	0.873	1.241	1.666
	MSE	0.226	0.300	0.592	1.042	1.361	0.201	0.279	0.521	1.250	3.156
	CP	0.954	0.954	0.964	0.982	0.998	0.954	0.960	0.976	0.974	0.970
$n=1000$	Bias	0.038	0.057	0.044	0.130	0.665	0.067	−0.029	−0.019	0.320	1.331
	SD	0.457	0.534	0.746	1.021	0.934	0.429	0.529	0.735	1.055	1.541
	ESD	0.479	0.578	0.852	1.206	1.572	0.469	0.571	0.856	1.233	1.627
	MSE	0.210	0.288	0.559	1.059	1.314	0.189	0.281	0.541	1.216	4.147
	CP	0.958	0.960	0.976	0.972	1.000	0.962	0.958	0.972	0.974	0.982

Table 4 (continued)

*Group*3: $\{a_1=0.1, a_4=0.14, a_2=0.13\}$											
n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	0.027	0.001	−0.004	−0.028	−0.028	0.284	0.061	−0.008	−0.041	−0.196
	SD	0.334	0.335	0.339	0.332	0.312	0.287	0.299	0.316	0.308	0.290
	ESD	0.339	0.336	0.336	0.334	0.336	0.269	0.285	0.300	0.300	0.281
	MSE	0.112	0.112	0.115	0.111	0.098	0.163	0.093	0.100	0.097	0.122
	CP	0.938	0.946	0.934	0.936	0.942	0.920	0.928	0.924	0.920	0.916
$n=1000$	Bias	0.019	−0.018	−0.007	−0.019	0.023	0.297	0.041	−0.010	−0.033	−0.166
	SD	0.318	0.327	0.335	0.335	0.322	0.276	0.296	0.311	0.316	0.292
	ESD	0.330	0.329	0.329	0.334	0.329	0.273	0.286	0.298	0.304	0.284
	MSE	0.102	0.107	0.112	0.112	0.105	0.165	0.089	0.097	0.101	0.113
	CP	0.950	0.920	0.928	0.942	0.948	0.936	0.918	0.926	0.936	0.926
		IPW-P					IPW-O				
$n=500$	Bias	−0.035	0.027	0.002	−0.103	0.133	−0.038	0.009	−0.015	−0.110	0.108
	SD	0.495	0.548	0.821	1.102	1.069	0.505	0.584	0.856	1.263	1.497
	ESD	0.504	0.584	0.856	1.193	1.563	0.497	0.579	0.849	1.185	1.545
	MSE	0.246	0.301	0.674	1.225	1.160	0.257	0.342	0.732	1.606	2.253
	CP	0.950	0.960	0.952	0.970	0.998	0.936	0.940	0.940	0.930	0.954
$n=1000$	Bias	−0.065	0.040	0.001	−0.014	0.135	−0.068	0.028	−0.010	−0.012	0.135
	SD	0.476	0.539	0.806	1.089	1.036	0.502	0.565	0.855	1.229	1.488
	ESD	0.496	0.577	0.845	1.192	1.534	0.493	0.575	0.842	1.187	1.526
	MSE	0.231	0.292	0.650	1.187	1.092	0.256	0.320	0.730	1.511	2.233
	CP	0.958	0.972	0.966	0.964	0.996	0.942	0.950	0.944	0.924	0.962

Table 4 (continued)

*Group*4: $\{a_1=0.1, a_4=0.12, a_2=0.13\}$

n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.006	0.021	−0.015	−0.016	−0.014	−0.006	0.020	−0.017	−0.018	−0.013
	SD	0.337	0.324	0.344	0.319	0.337	0.343	0.326	0.346	0.323	0.338
	ESD	0.331	0.330	0.335	0.330	0.338	0.331	0.329	0.335	0.330	0.339
	MSE	0.114	0.105	0.119	0.102	0.114	0.118	0.107	0.120	0.105	0.115
	CP	0.932	0.948	0.938	0.954	0.942	0.932	0.940	0.932	0.954	0.936
$n=1000$	Bias	−0.009	0.031	0.003	−0.007	−0.007	−0.008	0.031	0.002	−0.007	−0.006
	SD	0.326	0.335	0.328	0.315	0.335	0.331	0.336	0.334	0.318	0.339
	ESD	0.328	0.330	0.330	0.327	0.333	0.328	0.330	0.330	0.328	0.334
	MSE	0.106	0.113	0.108	0.099	0.113	0.110	0.114	0.111	0.101	0.115
	CP	0.954	0.944	0.940	0.950	0.944	0.952	0.950	0.934	0.946	0.946
		IPW-N					IPW-S				
$n=200$	Bias	0.048	0.021	0.037	−0.020	0.281	0.086	−0.122	−0.145	−0.057	0.573
	SD	0.446	0.508	0.706	0.910	0.880	0.422	0.510	0.821	1.183	1.816
	ESD	0.486	0.582	0.858	1.205	1.583	0.459	0.568	0.862	1.237	1.647
	MSE	0.201	0.258	0.500	0.829	0.853	0.186	0.275	0.696	1.403	3.626
	CP	0.964	0.974	0.986	0.996	0.996	0.960	0.970	0.954	0.950	0.894
$n=1000$	Bias	−0.006	0.102	0.068	−0.122	0.445	0.047	−0.044	−0.098	−0.005	1.057
	SD	0.454	0.523	0.690	0.897	0.902	0.441	0.530	0.777	1.233	1.902
	ESD	0.480	0.584	0.851	1.195	1.550	0.457	0.574	0.856	1.227	1.617
	MSE	0.206	0.284	0.480	0.820	1.012	0.197	0.283	0.614	1.519	4.733
	CP	0.962	0.960	0.976	0.984	0.996	0.954	0.962	0.962	0.946	0.890

Table 4 (continued)

*Group*4: $\{a_1=0.1, a_4=0.12, a_2=0.13\}$

n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.006	0.021	−0.015	−0.016	−0.014	0.118	0.036	−0.013	−0.017	−0.084
	SD	0.337	0.324	0.344	0.319	0.337	0.318	0.310	0.334	0.315	0.325
	ESD	0.331	0.330	0.335	0.330	0.338	0.298	0.305	0.316	0.313	0.314
	MSE	0.114	0.105	0.119	0.102	0.114	0.115	0.097	0.112	0.100	0.113
	CP	0.932	0.948	0.938	0.954	0.942	0.920	0.940	0.934	0.928	0.924
$n=1000$	Bias	−0.009	0.031	0.003	−0.007	−0.007	0.122	0.041	0.005	−0.007	−0.080
	SD	0.326	0.335	0.328	0.315	0.335	0.310	0.323	0.321	0.310	0.329
	ESD	0.328	0.330	0.330	0.327	0.333	0.301	0.310	0.313	0.313	0.313
	MSE	0.106	0.113	0.108	0.099	0.113	0.111	0.106	0.103	0.096	0.115
	CP	0.954	0.944	0.940	0.950	0.944	0.938	0.932	0.936	0.940	0.932
		IPW-N					IPW-O				
$n=200$	Bias	0.048	0.021	0.037	−0.020	0.281	−0.025	0.002	0.013	−0.055	0.138
	SD	0.446	0.508	0.706	0.910	0.880	0.499	0.554	0.861	1.181	1.557
	ESD	0.486	0.582	0.858	1.205	1.583	0.498	0.578	0.849	1.190	1.549
	MSE	0.201	0.258	0.500	0.829	0.853	0.249	0.306	0.742	1.398	2.445
	CP	0.964	0.974	0.986	0.996	0.996	0.950	0.960	0.942	0.946	0.936
$n=1000$	Bias	−0.006	0.102	0.068	−0.122	0.445	−0.082	0.088	0.055	−0.154	0.252
	SD	0.454	0.523	0.690	0.897	0.902	0.513	0.571	0.842	1.199	1.552
	ESD	0.480	0.584	0.851	1.195	1.550	0.490	0.582	0.845	1.184	1.523
	MSE	0.206	0.284	0.480	0.820	1.012	0.270	0.334	0.712	1.461	2.471
	CP	0.962	0.960	0.976	0.984	0.996	0.934	0.936	0.936	0.934	0.952

Table 4 (continued)

*Group*5: $\{a_1=0.1, a_4=0.1, a_2=0.15\}$

n	x_1	OR-O					OR-P				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.011	−0.014	−0.002	−0.029	−0.018	−0.010	−0.011	0.001	−0.028	−0.017
	SD	0.325	0.330	0.332	0.335	0.322	0.328	0.330	0.337	0.335	0.324
	ESD	0.331	0.330	0.335	0.332	0.330	0.331	0.330	0.336	0.333	0.330
	MSE	0.105	0.109	0.110	0.113	0.104	0.108	0.109	0.113	0.113	0.105
	CP	0.946	0.932	0.946	0.938	0.952	0.954	0.932	0.942	0.940	0.950
$n=1000$	Bias	0.009	0.020	−0.002	0.037	−0.029	0.005	0.022	0.002	0.040	−0.032
	SD	0.328	0.324	0.318	0.336	0.333	0.333	0.326	0.320	0.336	0.334
	ESD	0.329	0.330	0.328	0.330	0.332	0.329	0.330	0.328	0.330	0.332
	MSE	0.108	0.106	0.101	0.114	0.111	0.111	0.107	0.102	0.115	0.113
	CP	0.940	0.952	0.960	0.936	0.948	0.950	0.952	0.960	0.934	0.930
		IPW-N					IPW-S				
$n=500$	Bias	0.021	0.020	0.006	0.084	0.530	0.033	−0.063	−0.095	0.166	0.830
	SD	0.464	0.505	0.700	0.888	0.958	0.440	0.474	0.731	0.962	1.424
	ESD	0.483	0.582	0.860	1.202	1.581	0.474	0.576	0.865	1.231	1.634
	MSE	0.216	0.255	0.490	0.796	1.199	0.195	0.228	0.543	0.952	2.716
	CP	0.952	0.968	0.984	0.994	0.996	0.972	0.974	0.974	0.986	0.978
$n=1000$	Bias	0.029	0.082	0.065	0.113	0.648	0.048	−0.032	−0.040	0.267	1.294
	SD	0.441	0.564	0.731	0.939	0.996	0.414	0.526	0.709	1.041	1.688
	ESD	0.481	0.583	0.854	1.208	1.564	0.469	0.575	0.858	1.235	1.626
	MSE	0.196	0.325	0.539	0.894	1.411	0.174	0.278	0.504	1.155	4.521
	CP	0.972	0.956	0.976	0.980	0.998	0.972	0.954	0.984	0.966	0.958

Table 4 (continued)

*Group*5: $\{a_1=0.1, a_4=0.1, a_2=0.15\}$

n	x_1	OR-S					OR-N				
		−0.4	−0.2	0	0.2	0.4	−0.4	−0.2	0	0.2	0.4
$n=500$	Bias	−0.004	−0.011	0.001	−0.030	−0.020	0.303	0.057	−0.007	−0.045	−0.229
	SD	0.331	0.332	0.335	0.337	0.325	0.290	0.296	0.318	0.319	0.301
	ESD	0.333	0.331	0.336	0.334	0.331	0.264	0.280	0.301	0.300	0.276
	MSE	0.110	0.110	0.112	0.115	0.106	0.176	0.091	0.101	0.104	0.143
	CP	0.954	0.932	0.938	0.938	0.954	0.910	0.914	0.918	0.922	0.928
$n=1000$	Bias	0.010	0.021	−0.001	0.039	−0.032	0.345	0.088	−0.001	0.021	−0.249
	SD	0.332	0.329	0.320	0.337	0.336	0.297	0.302	0.304	0.321	0.312
	ESD	0.329	0.331	0.329	0.330	0.332	0.272	0.288	0.298	0.301	0.287
	MSE	0.110	0.109	0.103	0.115	0.114	0.207	0.099	0.092	0.104	0.159
	CP	0.946	0.956	0.956	0.940	0.934	0.926	0.942	0.946	0.924	0.912
		IPW-P					IPW-O				
$n=500$	Bias	−0.060	0.013	−0.045	−0.073	0.119	−0.071	0.010	−0.029	−0.044	0.192
	SD	0.491	0.506	0.763	0.947	1.010	0.505	0.547	0.822	1.068	1.529
	ESD	0.505	0.582	0.852	1.184	1.545	0.503	0.579	0.847	1.179	1.534
	MSE	0.244	0.256	0.584	0.903	1.034	0.260	0.299	0.676	1.142	2.375
	CP	0.952	0.970	0.956	0.988	0.996	0.946	0.946	0.944	0.968	0.950
$n=1000$	Bias	−0.087	0.053	−0.003	−0.064	0.130	−0.079	0.059	0.007	−0.044	0.158
	SD	0.469	0.583	0.792	1.029	1.073	0.483	0.607	0.847	1.173	1.624
	ESD	0.498	0.581	0.845	1.190	1.528	0.496	0.580	0.844	1.188	1.522
	MSE	0.228	0.342	0.627	1.062	1.167	0.240	0.372	0.718	1.377	2.662
	CP	0.970	0.950	0.962	0.972	0.998	0.962	0.938	0.952	0.944	0.932

to, even slightly smaller than, the others. In this case, all outcome regression-based estimations have smaller SDs than all IPW-based estimations.

Last, as we can see from Table 2, 3, 4, the difference between standard deviation and the bootstrap-based estimated standard deviation is very small. Furthermore, with n goes larger, the difference becomes smaller and smaller, even zero in many cases, which implies the bootstrap-based method performs well. Furthermore, the values of 95% confidence interval coverage probability (CP) are closer to the nominal level 0.95 (Table 2, 3, 4), which indicates that the normal approximation works well.

4 Empirical applications

In this section, we apply OR-S, as the dimensionality ($p = 15$) of X is high, to analyse the ACTG 175 data set that can be obtained from the R package speff2trial. This data set was collected from a randomized clinical trial that evaluated treatment effect when either one or two therapies were used for HIV-infected adults; see Hammer et al. (1996); Song and Ma (2008) for more details. As discussed before, our goal is to explore the heterogeneity of this treatment effect across subpopulations. Take *age* as X_1 to check how the expected pesticide effect changes with *age*.

A very brief description about the data set is as follows. The outcome here is CD4 T cell count at baseline and the treatment indicator variable D is a binary variable. $D = 0$ means receiving zidovudine only and $D = 1$ means receiving two therapies simultaneously. As documented by a number of authors, we take $Y = \log_{10}(\text{CD4})$ and delete some infinite value after logarithmic transformation, then the number of observations is $n = 2136$. Further, to guarantee the unconfoundedness assumption, X consists of the following 15 covariates: the pidnum (patient's ID number); age (age in years at baseline); wtkg (weight in kg at baseline); hemo (hemophilia); homo (homosexual activity); drugs (history of intravenous drug use); karnof (Karnofsky score); oprior (non-zidovudine antiretroviral therapy prior to initiation of study treatment); zprior (zidovudine use prior to treatment initiation); preanti (number of days of previously received antiretroviral therapy); race; gender; str2 (antiretroviral history); offtrt (indicator of off-treatment before 96pm5 weeks); days (number of days until the first occurrence of: (i) a decline in CD4 T cell count of at least 50 (ii) an event indicating progression to AIDS, or (iii) death).

We now estimate CATE in the interval between 20 and 57 to avoid the boundary effect when nonparametric estimation method is involved. This range is about from 0.025 quantile to 0.975 quantile of the data. To apply OR-S, we use the sufficient dimension reduction developed by Xia et al. (2002), which is now known to be MAVE to estimate the projection matrices $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$, and the associated dimensions. The results are $r(1) = 2$ and $r(0) = 3$. From these, we then have $s_4 = \max\{r(1), r(0)\} + 1 = 4$ and $h_4 = \widehat{\sigma}_r n^{-1/7}$ and $h = \widehat{\sigma}_1 n^{-1/31}$, where $\widehat{\sigma}_r = \sqrt{\text{var}(\boldsymbol{\beta}_0^\top X)}$, $\widehat{\boldsymbol{\beta}}_0$ is the estimated projection and $\widehat{\sigma}_1 = 2\sqrt{\text{var}(X_1)}$. Similar to the simulation studies, Gaussian kernel is used.

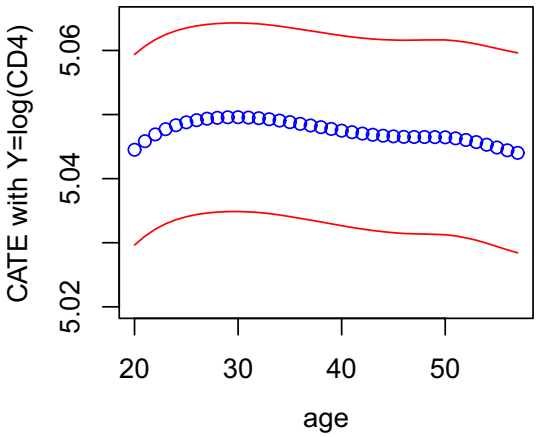

Fig. 1 The curves of conditional average treatment effects over age with the 95% pointwise confidence band if $Y = \log_{10}(\text{CD4})$

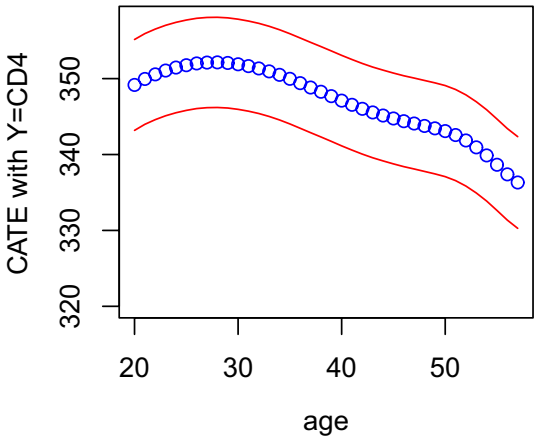

Fig. 2 The curves of conditional average treatment effects over age with the 95% pointwise confidence band if $Y = \text{CD4}$

Figure 1 shows, as a function of *age*, the curve of the estimated CATE and the pointwise 95% confidence band. Furthermore, to show the results more intuitively, we also provide the estimated CATE and the corresponding 95% confidence band with original $Y = \text{CD4}$ in Fig. 2. Note that the curve is much above zero. In other words, receiving two therapies simultaneously has a much better treatment effect than receiving only one (zidovudine). Song and Ma (2008) also obtained this conclusion. But the investigation on the heterogeneity shows that the treatment effect is influenced by *age*. As shown in Fig. 1, before the age of 30, receiving two therapies leads to the immunity rise. After that, the advantage of this treatment is gradually weakened. Thus, such a treatment seems more useful for patients whose ages are around 30.

5 Conclusion

In this paper, we propose four regression-based estimators of CATE, aimed to capture the heterogeneity of a treatment effect across subpopulations. The systematic investigation shows the important factors that affect the asymptotic behaviours of the estimators: the convergence rates of the outcome regression functions and the affiliation of the given covariates to the set of arguments of the outcome regression functions. Further, any regression-based estimation can be asymptotically more efficient than any propensity score-based estimation, and can at most achieve the asymptotic efficiency of nonparametric regression-based estimation in some cases. These results can give a relatively complete profile of propensity score-based and regression-based estimation for CATE. From the research, semiparametric regression-based estimation (OR-S) is worth of recommendation as it can avoid model misspecification as well as the curse of dimensionality when some dimension reduction and feature selection approaches are combined. see Luo et al. (2017) and Ma et al. (2019). In this paper, we only discuss the cases with correctly specified models. When the model is misspecified globally, further topics are about the asymptotic bias. Here global misspecification means that the assumed model is not convergent to the underlying model. If it is convergent, we call it local misspecification. Thus, we will check at which rate of convergence, the asymptotic bias vanishes and then also study its asymptotic efficiency. Another topic is about double robust estimation as it can greatly avoid model misspecification. As we have known, the uniform confidence band can provide a lot of useful information for us. However, the theoretical work of uniform band needs more theoretical support and more skillful technical requirements, which are left to further research. The research is ongoing.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00821-x.

References

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

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

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

Cook, R. D., & Weisberg, S. (1991). Sliced inverse regression for dimension reduction: Comment. *Journal of the American Statistical Association, 86,* 328–332.

Eforn, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7,* 1–26.

Fan, Q., Hsu, Y. C., Lieli, R. P., & Zhang, Y. (2020). Estimation of conditional average treatment effects with high-dimensional data. *Journal of Business and Economic Statistics, 40*(1), 313–327.

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

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

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine. Econometrica, 335,* 1081–1090.

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

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71,* 1161–1189.

Li, B., & Wang, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association, 102,* 997–1008.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association, 86,* 316–327.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Luo, W., Zhu, Y., & Ghosh, D. (2017). On estimating regression-based causal effects using sufficient dimension reduction. *Biometrika, 104,* 51–65.

Luo, W., Wu, W., & Zhu, Y. (2019). Learning heterogeneity in causal inference using sufficient dimension reduction. Published online.

Ma, S., Zhu, L., Zhang, Z., Tsai, C. L., & Carroll, R. J. (2019). A robust and efficient approach to causal inference based on sparse sufficient dimension reduction. *The Annals of Statistics, 47,* 1505–1535.

Ma, Y., & Zhu, L. (2012). A semiparametric approach to dimension reduction. *Journal of the American Statistical Association, 107,* 168–179.

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

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications, 9,* 141–142.

Pagan, A., & Ullah, A. (1999). *Nonparametric econometrics*. Cambridge University Press.

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

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

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

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

Shin, S. J., Wu, Y., Zhang, H. H., & Liu, Y. (2017). Principal weighted support vector machines for sufficient dimension reduction in binary classification. *Biometrika, 104,* 67–81.

Song, X., & Ma, S. (2008). Multiple augmentation for interval censored data with measurement error. *Statistics in Medicine, 27,* 3178–3190.

Wang, Q., Linton, O., & Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association, 99,* 334–345.

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

Yin, J., Geng, Z., Li, R., & Wang, H. (2010). Nonparametric covariance model. *Statistica Sinica, 20,* 469–479.

Xia, Y., Li, W. K., Tong, H., & Zhu, L. X. (2002). An adaptive estimation of dimension reduction space. *Journal of Royal Statistical Society, Series B, 64,* 363–410.

Zhang, Y., Shao, J., Yu, M., & Wang, L. (2018). Impact of sufficient dimension reduction in nonparametric estimation of causal effect. *Statistical Theory and Related Fields, 2*(1), 89–95.

Zhou, N. W., & Zhu, L. X. (2021). On IPW-based estimation of conditional average treatment effects. *Journal of Statistical Planning and Inference, 215,* 1–22.