Estimation of Causal Effects with a Binary Endogenous Variable: A Unified M-estimation Framework

S. Derya Uysal[*]

University of Munich

November 23, 2020

Abstract

In this paper, we review several estimators of the average treatment effect (ATE) that belong to three main groups: regression, weighting and doubly robust methods. We unify the exposition of these estimators within an M-estimation framework and we derive the corresponding sandwich form variance estimators. We compare the finite sample properties of the estimators by a Monte Carlo study, where the emphasis lies on the effect of various types of misspecifications and degree of overlap. Additionally, we re-estimate the causal return to higher education on earnings by the reviewed methods using the rich dataset provided by the British National Child Development Study (NCDS) as an empirical illustration.

JEL classification: $C21, C31$

Keywords: ATE; M-estimation; Treatment Effects; Double Robustness

[*] Department of Economics, University of Munich, Munich, Germany. Phone +49 89 2180 2224, email: derya.uysal@econ.lmu.de. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged.

1 Introduction

In this article we will review several estimators of average treatment effect (ATE) under unconfounded treatment assignment. We concentrate on methods that belong to three main groups: regression, weighting and doubly robust methods. We then unify the exposition of these estimators within the M-estimation framework. We believe that a unified estimation approach will help to ease our understanding of how these methods work and also how they relate to each other. Moreover, studying these methods within M-estimation framework can facilitate the programming of these estimators. It can also facilitate the study of their extensions for treatment effects in the case of multivalued treatment or local average treatment effect with instruments.

The identification and estimation of a causal effect is one of the central questions in empirical economics. Not surprisingly, there are many different identification approaches and corresponding estimators. Imbens (2004), Imbens and Wooldridge (2009), Athey and Imbens (2017), Abadie and Cattaneo (2018) provide excellent reviews on the existing methods and the new directions that the literature is moving towards. This study does not aim to review all of the existing method but instead concentrates on the regression and weighting methods, as well as certain combinations of these two approaches. The goal here is to provide an M-estimation representation of several estimators and derive the corresponding sandwich form variance estimators.

The hybrid methods that we consider posses a so-called double robustness property (for further discussion on double robustness see Robins, Rotnitzky, and Zhao, 1995, Robins and Ritov, 1997, Scharfstein, Rotnitzky, and Robins, 1999, Hirano and Imbens, 2001, Wooldridge, 2007, Bang and Robins, 2005, Kang and Schafer, 2007, among others). Even though the doubly robust methods have been known for a while by statisticians, they only gained in popularity in econometrics and economics after 2010. Some of the most recent examples include Graham, de Xavier Pinto, and Egel (2012, 2016), Lee, Okui, and Whang (2017), Słoczyński and Wooldridge (2018), Rothe and Firpo (2019), Muris (2020), Sant'Anna and Zhao (2020), Heiler and Kazak (2020).

We also compare the finite sample properties of the methods reviewed. Because the estimation of ATE has been a popular topic for a while now, several studies have

focused on finite sample properties. Busso, DiNardo, and McCrary (2014), Frölich, Huber, and Wiesenfarth (2017), Bodory, Camponovo, Huber, and Lechner (2020) are among the most important. The simulation setup here shares some similarities with previous studies but the emphasis of this study is different. Here, the emphasis lies on the finite sample properties of the parametric estimators under misspecification of the propensity score or the outcome equation. Therefore, the properties of these estimation methods are examined for various misspecification designs. The overlap problem, which is an important issue for propensity score based methods, is also investigated in the simulation studies for the reviewed methods. Additionally, the treatment indicator is simulated in several ways to examine different treated-control ratios. This extension of the Monte Carlo design allows us to evaluate the sensitivity of these methods to the distribution of the propensity score. We also compare the small sample performance of these methods in terms of the Monte Carlo mean square error (MCMSE) and also in terms of average of variance estimates based on asymptotic results.
Finally, we provide an applications of the considered methods to illustrate the practical aspects. This application is an estimation of the causal returns of higher education using the rich dataset provided by the British National Child Development Study (NCDS). Although the measurement of the individual returns to education has been an important research question for the last few decades (for a review see Card, 1999), due to the strong data requirements there are very few papers where returns to schooling are estimated under CIA (see, for example Blundell, Dearden, and Sianesi, 2005, Flossmann, 2010, Pohlmeier and Pfeiffer, 2004). The NCDS dataset is used by Blundell et al. (2005) to estimate the returns to higher education by regression and matching methods under unconfounded treatment assignment. Given that they do not use the weighting or the doubly robust methods, we estimate the causal effects by means of the methods reviewed here.

The organization of this paper is as follows. Section 2 first introduces the potential outcome framework. This section will also explain the treatment effects of interest and the identifying assumptions. We will then review several existing methods within the M-estimation framework. To investigate the finite sample properties of the given estimators, a Monte Carlo study is carried out in Section 3. In Section 4, the average causal effect of higher education on earnings is estimated. Finally, Section 5 summarizes the main results and concludes the paper.

2 Econometric Method

Consider N units, which are drawn from a large population. For each individual i in the sample, where $i = 1, ..., N$, the triple (Y_i, D_i, X_i) is observed. D_i shows the binary treatment status for individual i:

$$D_i = \begin{cases} 1, & \text{if the } i^{th} \text{ individual is treated} \\ 0, & \text{otherwise} \end{cases}$$

X_i denotes the characteristics of the individual i. There are two potential outcomes (Y_{0i}, Y_{1i}) for each individual. Y_{id} denotes the outcome for individual i, for which $D_i = d$ where $d \in \{0, 1\}$. Thus, Y_{1i} is the outcome that the individual would receive if they get the treatment. Meanwhile, Y_{0i} is the outcome that the individual would get without receiving the treatment. However, it is not possible to observe both of the outcomes for one individual. Either they would receive the treatment and Y_{1i} would be observed or they would not receive the treatment and Y_{0i} would be observed. The observed outcome (Y_i) can be written in terms of treatment indicator (D_i) and the potential outcomes (Y_{id}):

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$$

The advantage of the potential outcome framework is that it provides a definition of the causal effects without any functional form or distributional assumptions. Using the potential outcome framework, several treatment effects can be defined. One of these effects is the ATE, which measures the mean effect of treatment over the entire population:

$$\tau \equiv \mathrm{E}\left[Y_{1i} - Y_{0i}\right] = \mu_1 - \mu_0.$$

Because only one of the potential outcomes is observed, the previously defined ATE cannot be identified without further assumptions. In this paper, we discuss the identification and estimation of the ATE under following assumptions:

Assumption 2.1. *Conditional Independence Assumption (CIA)*
$Y_{0i}, Y_{1i} \perp D_i | X_i$,

where $\perp$ stands for independence. This assumption implies that after controlling for the effect of covariates, treatment and potential outcomes are independent. This requires that all of the confounders are observed and there is no selection into treatment due to the unobservables. Obviously, this assumption puts strong requirements on the data.

Assumption 2.2. *Overlap Assumption*
$0 < \Pr\left[D_i = 1 \,|X_i\right] < 1$

Assumption 2.2 implies that for all X_i there is a positive probability of either participating ($D_i = 1$) or not participating ($D_i = 0$). In other words, for each value of covariates there are both treated and untreated cases. Khan and Tamer (2010) show that standard overlap assumption (Assumption 2.2) is not sufficient to guarantee $\sqrt{N}$-consistency of the semiparametric treatment effect estimators. However, the strict overlap assumption—that is, $\xi < \Pr\left[D_i = 1 \,|X_i\right] < 1 - \xi$ for some $\xi > 0$—is sufficient for the $\sqrt{N}$-consistency.

Rosenbaum and Rubin (1983) show that under CIA, identification can be achieved by conditioning on a function of X_i instead of a high dimensional X_i. The propensity score is the most commonly used function in the evaluation literature. It is simply the conditional probability of assignment to the treatment given the covariates:

$$p(x) = \Pr\left[D_i = 1 \,|X_i = x\right] = \mathrm{E}\left[\left. D_i \right| X_i = x\right].$$

Lemma 2.1. *Unconfoundedness Given the Propensity Score*
Given the CIA (2.1) and Common Support (2.2) assumptions, outcomes Y_{0i} and Y_{1i} are independent of treatment given the propensity score.
$Y_{0i}, Y_{1i} \perp D_i | p(X_i)$

If these assumptions are satisfied, then several methods can be used to estimate the ATE. In this paper, we investigate several parametric estimation methods for the ATE, which can be classified into three groups: regression, propensity score weighting and doubly robust methods (which are a combination of the first two approaches). All of these methods require estimation of the unconditional means μ_1 and μ_0. The estimators are then used to estimate the ATE. Independent of which method is used to estimate unconditional means, the estimator for the ATE has the following simple form

$$\hat{\tau} = \hat{\mu}_1 - \hat{\mu}_0.$$

where $\hat{\mu}_1$ and $\hat{\mu}_0$ are consistent and asymptotically normal estimators of the unconditional means μ_1 and μ_0. Thus, the asymptotic distribution of $\hat{\tau}$ is given by

$$\sqrt{N}(\hat{\tau} - \tau) \xrightarrow{d} N\left(0, \mathrm{AV}[\hat{\mu}_1] + \mathrm{AV}[\hat{\mu}_0] - 2\mathrm{ACov}[\hat{\mu}_1, \hat{\mu}_0]\right), \tag{1}$$

where AV and ACov refer to the asymptotic variance and covariance, respectively. Thus, the task of deriving asymptotic distribution of different ATE estimators is reduced to derivation of variance-covariance matrix of $(\hat{\mu}_1, \hat{\mu}_0)$ based on different estimation methods.

In the following subsections we will discuss several estimation methods and theoretical properties of the resulting estimators in a unified framework of M-estimation. The M-estimator, $\hat{\theta}$, can be derived as a solution to the sample moment equation

$$\frac{1}{N}\sum_{i=1}^{N}\psi(Z_i, \hat{\theta}) = 0,$$

where Z_i is the observed data (See, for example, Huber, 1964, Stefanski and Boos, 2002, Wooldridge, 2010, for more on M-estimation.). Thus, $\hat{\theta}$ is the estimator of θ, $k\times$ 1 unknown parameter vector, which satisfies the population relation $\mathrm{E}\left[\psi(Z_i, \theta)\right] = 0$. Under standard regularity conditions, the asymptotic distribution of an M-estimator is given by

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, A^{-1}VA^{-1\prime}\right) \tag{2}$$

with

$$\begin{aligned} A &= \mathrm{E}\left[\frac{\partial\psi(Z_i, \theta)}{\partial\theta'}\right] \\ V &= \mathrm{E}\left[\psi(Z_i, \theta)\psi(Z_i, \theta)'\right]. \end{aligned}$$

The derivations of asymptotic distributions of different ATE estimators in the following sections rely on (2) and, if necessary, (1).

2.1 Regression

The regression method is one of several estimation methods. This requires estimation of the conditional means $\mathrm{E}\left[Y_{id}|\, X_i\right]$, from which one can estimate the unconditional means $\mathrm{E}\left[Y_{id}\right] = \mu_d$ for $d \in 0, 1$ using treated and untreated samples separately.

Unless CIA (Assumption 2.1) is satisfied, one cannot estimate the population parameters of the conditional means based on these two subsamples. CIA assumption guarantees that conditional on observable characteristics, X_i's, selection into the treatment can be treated as random; therefore, the observed subsamples can identify the population parameters.

For the conditional mean functions of the potential outcomes, we consider generalized linear models (GLM) with a link function η. The conditional mean functions of the potential outcomes are specified as follows:

$$\mathrm{E}\left[Y_{id} \middle| X_i \right] = \eta[X_i'\beta_d], \quad \text{for } d \in \{0, 1\},$$

where β_d are vectors of parameters and η is the link function.[1] As mentioned earlier, β_1 and β_0 can be consistently estimated using the treated and untreated samples separately. Thus, $\hat{\beta}_1$ and $\hat{\beta}_0$ can be represented as solutions to the following minimization problems:

$$\begin{aligned} \{\hat{\beta}_1\} &= \underset{\beta_1}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} D_i q(Y_i, X_i; \beta_1) \\ \{\hat{\beta}_0\} &= \underset{\beta_0}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} (1 - D_i) q(Y_i, X_i; \beta_0), \end{aligned}$$

where $q(\cdot)$ is the objective function. If, for example, η is the identity link function, then the objective function, $q(\cdot)$, is simply the sum of squared residuals. The identity link function is suitable for continuous dependent variables. For binary or count dependent variables, the logit or Poisson link functions should be preferred. In this case, the objective function is the negative of the log-likelihood function. Given the estimators, $\hat{\beta}_d$, unconditional means can be consistently estimated by taking the average of the predicted Y_{id} over the distribution of X_i:

$$\hat{\mu}_{d,reg} = \frac{1}{N} \sum_{i=1}^{N} \eta[X_i'\hat{\beta}_d].$$

The consistency follows from weak law of large numbers—that is, $\frac{1}{N}\sum_{i=1}^{N} \eta[X_i'\hat{\beta}_d] \xrightarrow{p} \mathrm{E}\left[\eta[X_i'\beta_d]\right]$—and by law of iterated expectation—that is, $\mathrm{E}\left[\eta[X_i'\beta_d]\right] = \mathrm{E}\left[\mathrm{E}\left[Y_{id} \middle| X_i \right]\right] =$

[1] X_i should be considered more as a function of covariates. Without loss of generality, we use X_i instead of a function of X_i, $g(X_i)$, for the sake of notational simplicity.

$\mathrm{E}\left[Y_{id}\right] = \mu_d$. Thus, one can estimate the ATE using the regression coefficients, as follows:

$$\begin{aligned}
\hat{\tau}_{reg} &= \frac{1}{N}\sum_{i=1}^{N}\left(\eta[X_i'\hat{\beta}_1] - \eta[X_i'\hat{\beta}_0]\right) \\
&= \hat{\mu}_{1,reg} - \hat{\mu}_{0,reg}.
\end{aligned}$$

The regression estimator $\hat{\tau}_{reg}$ is consistent as long as the conditional means are correctly specified. The estimators $\hat{\beta}_d$ and $\hat{\tau}_{reg}$ can also be written as a solution of the sample moment equation:

$$\frac{1}{N}\sum_{i=1}^{N}\psi(Z_i, \hat{\theta}_{reg}) = 0, \tag{3}$$

where $\hat{\theta}_{reg}$ is the estimator of the parameter vector $\theta_{reg} = (\beta_1, \beta_0, \tau)$ and $Z_i = (Y_i, X_i, D_i)$. Using the moment functions related to each parameter vector in θ_{reg}, one can explicitly rewrite the moment function in (3) as follows:

$$\psi(Z_i, \theta_{reg}) = \begin{pmatrix} \psi_1(Z_i, \theta_{reg}) \\ \psi_2(Z_i, \theta_{reg}) \\ \psi_3(Z_i, \theta_{reg}) \end{pmatrix} = \begin{pmatrix} D_i \frac{\partial q(Y_i, X_i; \beta_1)}{\partial \beta_1} \\ (1 - D_i)\frac{\partial q(Y_i, X_i; \beta_0)}{\partial \beta_0} \\ \eta[X_i'\beta_1] - \eta[X_i'\beta_0] - \tau \end{pmatrix}.$$

Writing the estimation problem in M-estimation framework makes it easier to derive the asymptotic distribution of the resulting estimator. By standard results and under regularity conditions for M-estimation, the asymptotic distribution of the estimator is given by:

$$\sqrt{N}(\hat{\theta}_{reg} - \theta) \xrightarrow{d} N\left(0, A_{reg}^{-1} V_{reg} A_{reg}^{-1\,\prime}\right), \tag{4}$$

where

$$\begin{aligned}
A_{reg} &\equiv E\left[\frac{\partial \psi(Z_i, \theta)}{\partial \theta'}\right] \\
V_{reg} &\equiv V\left[\psi(Z_i, \theta)\right] = E\left[\psi(Z_i, \theta)\psi(Z_i, \theta)'\right].
\end{aligned}$$

Hence, depending on the regression model chosen for the outcome model, A_{reg} and V_{reg} can be derived. To estimate the variance-covariance matrix, we can replace the expectations with the sample means and the true parameter vector with its estimate. Furthermore, the asymptotic distribution of the regression estimator can be isolated from the variance-covariance matrix $A_{reg}^{-1} V_{reg} A_{reg}^{-1\prime}$ as[2]

$$\sqrt{N}(\hat{\tau}_{reg} - \tau) \xrightarrow{d} N\left(0, \mathrm{AV}_{\hat{\tau},reg}\right), \tag{5}$$

where $\mathrm{AV}_{\hat{\tau},reg}$ is given by

$$\begin{aligned} \mathrm{AV}_{\hat{\tau},reg} &= \mathrm{E}\left[(\eta[X_i'\beta_1] - \eta[X_i'\beta_0] - \tau)^2\right] + \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_1]}{\partial\beta_1'}\right] \mathrm{AV}_{\hat{\beta}_1} \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_1]}{\partial\beta_1'}\right]' \\ &\quad + \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_0]}{\partial\beta_0'}\right] \mathrm{AV}_{\hat{\beta}_0} \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_0]}{\partial\beta_0'}\right]'. \end{aligned} \tag{6}$$

$\mathrm{AV}_{\hat{\beta}_1}$ and $\mathrm{AV}_{\hat{\beta}_0}$ are the asymptotic variances of $\hat{\beta}_1$ and $\hat{\beta}_0$, respectively. The explicit forms for the asymptotic variances can be written as follows:

$$\begin{aligned} \mathrm{AV}_{\hat{\beta}_1} &= \mathrm{E}\left[D_i H_1(\beta_1)\right]^{-1} \mathrm{E}\left[D_i S_1(\beta_1) S_1(\beta_1)'\right] \mathrm{E}\left[D_i H_1(\beta_1)\right]^{-1} \\ \mathrm{AV}_{\hat{\beta}_0} &= \mathrm{E}\left[(1-D_i) H_0(\beta_0)\right]^{-1} \mathrm{E}\left[(1-D_i) S_0(\beta_0) S_1(\beta_0)'\right] \mathrm{E}\left[(1-D_i) H_0(\beta_0)\right]^{-1}. \end{aligned}$$

where $H_d(\beta_d)$ stands for the Hessian (second derivative of the objective function, $\frac{\partial^2 q(Y_i, X_i; \beta_d)}{\partial\beta_d \partial\beta_d'}$) and $S_d(\beta_d)$ stands for the score (first derivative of the objective function, $\frac{\partial q(Y_i, X_i; \beta_d)}{\partial\beta_d}$) of the regression for β_d.

2.2 Weighting by Propensity Score

The second group of estimation methods relies on another identification result. The mean outcomes for the treated and control groups can be identified by weighting the observations with the inverse of the propensity score:

$$\mathrm{E}\left[Y_{1i}\right] = \mathrm{E}\left[\frac{D_i Y_i}{p(X_i)}\right] \tag{7}$$

$$\mathrm{E}\left[Y_{0i}\right] = \mathrm{E}\left[\frac{(1-D_i) Y_i}{(1-p(X_i))}\right]. \tag{8}$$

Thus, the ATE is equal to:

$$\tau = \mathrm{E}\left[\frac{D_i Y_i}{p(X_i)} - \frac{(1-D_i) Y_i}{(1-p(X_i))}\right]. \tag{9}$$

[2] Simple application of the Delta Method can also be used for the derivation. See Appendix B.1 for further details.

Because the probabilities are usually unknown, one has to estimate them first. Consider the following regression function for the propensity score:

$$\Pr\left[D_i = 1 \,|X_i\right] = \pi[X_i'\alpha] = \pi_i,$$

where π is the link function and α is the unknown parameter vector.

The obvious way to estimate μ_1 and μ_0 is to replace the expectations with sample means and unknown probabilities with the estimated ones in (7) and (8), respectively. However, we represent the weighting type estimators in a more general way by using a weighting function $\omega_{di}(\alpha)$ for $d = 0, 1$, which is a function of the propensity score, and we then examine possible forms of ω_{di}. Let $\frac{1}{N}\sum_{i=1}^{N}\omega_{1i}(\hat{\alpha})Y_i$ and $\frac{1}{N}\sum_{i=1}^{N}\omega_{0i}(\hat{\alpha})Y_0$ denote the general form of the weighting estimators of μ_1 and μ_0, respectively. Thus, the general weighting estimator of the ATE is given by:

$$\hat{\tau}_{ps} = \frac{1}{N}\sum_{i=1}^{N}\left(\omega_{1i}(\hat{\alpha})Y_i - \omega_{0i}(\hat{\alpha})Y_i\right),$$

where $\hat{\alpha}$ is estimated by a regression of D_i on X_i with the link function π and $\omega_{di}(\hat{\alpha})$ is the weighting function at the estimated propensity score.

Several possible weighting functions are proposed in the literature. We consider here three commonly used weighting functions. The first propensity score weighting estimator of the ATE is the sample counterpart of the population expectations in (9), where the true probability of getting the treatment is replaced by its estimate:[3]

$$\begin{aligned} \hat{\tau}_{ps1} &= \frac{1}{N}\sum_{i=1}^{N}[\omega_{1i}^{(1)}(\hat{\alpha})Y_i - \omega_{0i}^{(1)}(\hat{\alpha})Y_i] \\ &= \frac{1}{N}\sum_{i=1}^{N}[\frac{D_iY_i}{\pi[X_i'\hat{\alpha}]} - \frac{(1-D_i)Y_i}{(1-\pi[X_i'\hat{\alpha}])}]. \end{aligned} \tag{10}$$

$\hat{\tau}_{ps}$ is consistent as long as the propensity score is correctly specified (see for further discussion Horvitz and Thompson, 1952, Rosenbaum, 1987, Bang and Robins, 2005).[4] As in the previous subsection, we present the weighting estimators in the

[3] This estimator is identical to an estimator from Horvitz and Thompson (1952) for handling nonrandom sampling.

[4] Hirano, Imbens, and Ridder (2003) examine the estimator in (10), where $\pi[X_i'\hat{\alpha}]$ is replaced by nonparametric estimates.

M-estimation framework and we then derive the asymptotic properties. For the first weighting estimator of τ, the estimators $\hat{\alpha}$, $\hat{\mu}_{1,ps1}$ and $\hat{\mu}_{0,ps1}$ solve the following sample moment equation

$$\frac{1}{N}\sum_{i=1}^{N}\psi(Z_i,\hat{\theta}_{ps1})=0$$

with

$$\psi(Z_i,\theta_{ps1}) \;=\; \begin{pmatrix} \psi_1(Z_i,\theta_{ps}) \\ \psi_2(Z_i,\theta_{ps}) \\ \psi_3(Z_i,\theta_{ps}) \end{pmatrix} = \begin{pmatrix} \frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha} \\ \frac{D_iY_i}{\pi[X_i'\alpha]}-\mu_1 \\ \frac{(1-D_i)Y_i}{1-\pi[X_i'\alpha]}-\mu_0 \end{pmatrix},$$

where $\theta_{ps1}=(\alpha,\mu_1,\mu_0)$. The first moment condition ψ_1 corresponds to the score of the maximum likelihood estimation of α. The general result in Equation (2) applies here with the following A and V:

$$\begin{aligned} A_{ps1} &\equiv E\left[\frac{\partial\psi(Z_i,\theta_{ps1})}{\partial\theta_{ps1}'}\right] \\ V_{ps1} &\equiv V\left[\psi(Z_i,\theta_{ps1})\right]=E\left[\psi(Z_i,\theta_{ps1})\psi(Z_i,\theta_{ps1})'\right], \end{aligned}$$

where the asymptotic distribution of $\hat{\tau}_{ps1}$ can be derived from the joint distribution. The explicit forms of A and V are given in the Appendix B.1.

$$\sqrt{N}(\hat{\tau}_{ps1}-\tau)\xrightarrow{d}N\left(0,\mathrm{AV}_{\hat{\tau},ps1}\right),$$

with

$$\begin{aligned} \mathrm{AV}_{\hat{\tau},ps1} = \mathrm{E}\left[\tfrac{Y_{1i}^2}{\pi[X_i'\alpha]}+\tfrac{Y_{0i}^2}{1-\pi[X_i'\alpha]}\right]-\tau^2 \qquad (11) \\ -\mathrm{E}\left[\left(\tfrac{Y_{1i}}{\pi[X_i'\alpha]}+\tfrac{Y_{0i}}{(1-\pi[X_i'\alpha])}\right)\tfrac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\left(-\mathrm{E}\left[H(Z_i,\alpha)\right]^{-1}\right)\mathrm{E}\left[\left(\tfrac{Y_{1i}}{\pi[X_i'\alpha]}+\tfrac{Y_{0i}}{(1-\pi[X_i'\alpha])}\right)\tfrac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]'. \end{aligned}$$

The first line in (11) corresponds to the asymptotic variance of $\hat{\tau}_{ps1}$ for known α.[5] Note that the last term has a quadratic form with the term in the middle, $-\mathrm{E}\left[H(Z_i,\alpha)\right]^{-1}$, which is the asymptotic variance of $\hat{\alpha}$. Hence, the last term is positive semidefinite and the variance of $\hat{\tau}$ with estimated probabilities is smaller than the variance with known probabilities. This is a well-known fact and was established by Hirano et al. (2003). The variance in (11) can be estimated by replacing the expectations with the sample counterparts and the unknown quantities with

[5] See Appendix B.1 for the proof.

their estimates.

Although several papers use the first weighting function to estimate the treatment effects (see, for example, Dehejia and Wahba, 1999, Hirano et al., 2003, among others), there are other weighting functions. The problem with the first type of weighting estimators is that the estimated weights do not necessarily add up to one. Therefore, an adjusted version of the estimator is proposed in the literature, which is given by:

$$\begin{aligned}\hat{\tau}_{ps2} &= \frac{1}{N}\sum_{i=1}^{N}\left(\omega_{1i}^{(2)}(\hat{\alpha})Y_i - \omega_{0i}^{(2)}(\hat{\alpha})Y_i\right)\\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\left(\frac{D_i/\pi[X_i'\hat{\alpha}]}{\frac{1}{N}\sum_{i=1}^{N} D_i/\pi[X_i'\hat{\alpha}]}\right)Y_i - \left(\frac{(1-D_i)/(1-\pi[X_i'\hat{\alpha}])}{\frac{1}{N}\sum_{i=1}^{N}(1-D_i)/(1-\pi[X_i'\hat{\alpha}])}\right)Y_i\right),\end{aligned}$$

where the weights are adjusted such that they sum up to one. In the ATE estimation framework, this adjustment is advocated by Johnston and DiNardo (1996) and Imbens (2004). Because $\mathrm{E}\left[\frac{D_i}{p(X_i)}\right] = \mathrm{E}\left[\frac{\mathrm{E}[D_i|X_i]}{p(X_i)}\right] = 1$ and $\mathrm{E}\left[\frac{1-D_i}{1-p(X_i)}\right] = \mathrm{E}\left[\mathrm{E}\left[\frac{1-D_i}{1-p(X_i)}\middle| X_i\right]\right] = 1$, the denominators converge to one and (as long as the propensity score is correctly specified) the ATE will be consistently estimated by the second weighting function. By rescaling the weights, one also avoids the problem that too small probabilities cause huge weights for some observations. Although the sign of the theoretical variance difference between first two weighting estimators is ambiguous, the small sample studies indicate that the adjusted weighting estimator is more efficient (see, for example, Lunceford and Davidian, 2004, and simulation results in Section 3).

The moment conditions for the second propensity score weighting estimator can be written as follows

$$\psi(Z_i,\theta_{ps2}) = \begin{pmatrix}\psi_1(Z_i,\theta_{ps2})\\ \psi_2(Z_i,\theta_{ps2})\\ \psi_3(Z_i,\theta_{ps2})\end{pmatrix} = \begin{pmatrix}\frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha}\\ \frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\\ \frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}\end{pmatrix},$$

where $\theta_{ps2} = (\alpha,\mu_1,\mu_0)$. The corresponding A and V matrices in the variance term in Equation (2) are given in the Appendix B.1. As earlier, the asymptotic distribution of $\hat{\tau}_{ps2}$ can be derived from the joint distribution.

$$\sqrt{N}(\hat{\tau}_{ps2}-\tau) \xrightarrow{d} N\left(0, \mathrm{AV}_{\hat{\tau},ps2}\right),$$

with

$$
\begin{aligned}
\mathrm{AV}_{\hat{\tau},ps2} = \mathrm{E}\left[\tfrac{(Y_{1i}-\mu_1)^2}{\pi[X_i'\alpha]} + \tfrac{(Y_{0i}-\mu_0)^2}{1-\pi[X_i'\alpha]}\right] \qquad (12)\\
- \mathrm{E}\left[\left(\tfrac{(Y_{1i}-\mu_1)}{\pi[X_i'\alpha]} + \tfrac{(Y_{0i}-\mu_0)}{(1-\pi[X_i'\alpha])}\right)\tfrac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \mathrm{AV}\left[\hat{\alpha}\right] \mathrm{E}\left[\left(\tfrac{(Y_{1i}-\mu_1)}{\pi[X_i'\alpha]} + \tfrac{(Y_{0i}-\mu_0)}{(1-\pi[X_i'\alpha])}\right)\tfrac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]',
\end{aligned}
$$

where $\mathrm{AV}\left[\hat{\alpha}\right]$ stands for the asymptotic variance of $\hat{\alpha}$. The first line in (12) corresponds to the variance of the second weighting estimator with known propensity score. Like the asymptotic variance of the first weighting estimator, a positive semidefinite matrix is subtracted from the first part. Hence, for the second weighting estimator, using the estimated probabilities instead of known probabilities also increases the efficiency.[6]

The third weighting function, which is less known in the literature, is proposed by Lunceford and Davidian (2004). The third estimator is based on an asymptotic variance minimizing linear combination of the first and second weighting estimators. For this estimator, the weighting functions are given by:

$$
\begin{aligned}
\omega_{1i}^{(3)}(\hat{\alpha}) &= \frac{D_i}{\pi[X_i'\hat{\alpha}]}(1-C_i^1)/\left(\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi(X_i'\hat{\alpha})}(1-C_i^1)\right)\\
\omega_{0i}^{(3)}(\hat{\alpha}) &= \frac{1-D_i}{1-\pi[X_i'\hat{\alpha}]}(1-C_i^0)/\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1-D_i}{1-\pi(X_i'\hat{\alpha})}(1-C_i^0)\right),
\end{aligned}
$$

where C_i^1 and C_i^0 are correction factors in the following form:

$$
\begin{aligned}
C_i^1 &= \frac{\frac{1}{\pi[X_i'\hat{\alpha}]}N^{-1}\sum_{i=1}^{N}(A_i(1-\pi[X_i'\hat{\alpha}])-(1-D_i))}{N^{-1}\sum_{i=1}^{N}(A_i(1-\pi[X_i'\hat{\alpha}])-(1-D_i))^2}\\
C_i^0 &= \frac{\frac{1}{1-\pi[X_i'\hat{\alpha}]}N^{-1}\sum_{i=1}^{N}(B_i(1-\pi[X_i'\hat{\alpha}])-D_i)}{N^{-1}\sum_{i=1}^{N}(B_i(1-\pi[X_i'\hat{\alpha}])-D_i)^2}\\
A_i &= \frac{D_i}{\pi[X_i'\hat{\alpha}]}\\
B_i &= \frac{1-D_i}{1-\pi[X_i'\hat{\alpha}]}.
\end{aligned}
$$

Lunceford and Davidian (2004) derive these weighting functions by minimizing the asymptotic variances of $\hat{\mu}_{1,ps3}$ and $\hat{\mu}_{0,ps3}$ with respect to (η_1, η_0). For given (η_1, η_0),

[6] The details are provided in Appendix B.1.

$\hat{\mu}_{1,ps3}$ and $\hat{\mu}_{0,ps3}$ are the solutions to the following sample moment conditions:

$$\sum_{i=1}^{N}\left(\frac{D_i(Y_i-\hat{\mu}_{1,ps3})}{\pi[X_i'\alpha]}+\eta_1\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\right)=0 \quad \text{and} \tag{13}$$

$$\sum_{i=1}^{N}\left(\frac{(1-D_i)(Y_i-\hat{\mu}_{0,ps3})}{1-\pi[X_i'\alpha]}-\eta_0\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right)=0. \tag{14}$$

Note that if $(\eta_0,\eta_1)=(\hat{\mu}_{0,ps3},\hat{\mu}_{1,ps3})$, then Equations (13) and (14) are equivalent to $\psi_2(Z_i,\theta_{ps1})$ and $\psi_3(Z_i,\theta_{ps1})$, respectively. Meanwhile, if $(\eta_0,\eta_1)=(0,0)$, then Equations (13) and (14) yield $\psi_2(Z_i,\theta_{ps2})$ and $\psi_3(Z_i,\theta_{ps2})$. By the central limit theorem, the asymptotic variances of $\hat{\mu}_{1,ps3}$ and $\hat{\mu}_{0,ps3}$ are given by

$$V_{\mu_{1,ps3}}=\mathrm{E}\left[\left(\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}+\eta_1\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\right)^2\right]$$

$$V_{\mu_{0,ps3}}=\mathrm{E}\left[\left(\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}-\eta_0\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right)^2\right].$$

Minimizing these stated variances with respect to η_1 and η_0 and solving for η_1 and η_0 leads to the following:

$$\eta_1 = -\frac{\mathrm{E}\left[D_i(Y_i-\mu_1)/\pi[X_i'\alpha]^2\right]}{\mathrm{E}\left[(D_i-\pi[X_i'\alpha])^2/\pi[X_i'\alpha]^2\right]} \tag{15}$$

$$\eta_0 = -\frac{\mathrm{E}\left[(1-D_i)(Y_i-\mu_0)/(1-\pi[X_i'\alpha])^2\right]}{\mathrm{E}\left[(D_i-\pi[X_i'\alpha])^2/(1-\pi[X_i'\alpha])^2\right]}. \tag{16}$$

By plugging the sample counterparts of (15) and (16) in (13) and (14), the third weighting function can be derived.

The moment conditions for the third propensity score weighting estimator for given η_1 and η_0 can be written as follows

$$\psi(Z_i,\theta_{ps3}) = \begin{pmatrix}\psi_1(Z_i,\theta_{ps3})\\ \psi_2(Z_i,\theta_{ps3})\\ \psi_3(Z_i,\theta_{ps3})\end{pmatrix} = \begin{pmatrix}\frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha}\\ \frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}+\eta_1\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\\ \frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}-\eta_0\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\end{pmatrix}, \tag{17}$$

where $\theta_{ps3}=(\alpha,\mu_1,\mu_0)$. The corresponding A and V matrices in the variance term in Equation (2) are given in Appendix B.1.

The asymptotic distribution of $\hat{\tau}_{p3}$ follows the general results (as before) and is given by

$$\sqrt{N}(\hat{\tau}_{ps3} - \tau) \xrightarrow{d} N\left(0, \mathrm{AV}_{\hat{\tau},ps3}\right),$$

with

$$\begin{aligned}\mathrm{AV}_{\hat{\tau},ps3} = {} & \mathrm{E}\left[\tfrac{(Y_{1i}-\mu_1)^2}{\pi[X_i'\alpha]} + \tfrac{(Y_{0i}-\mu_0)^2}{1-\pi[X_i'\alpha]}\right] + \eta_1\,\mathrm{E}\left[\tfrac{Y_{1i}-\mu_1}{\pi[X_i'\alpha]}\right] + \eta_0\,\mathrm{E}\left[\tfrac{Y_{0i}-\mu_0}{1-\pi[X_i'\alpha]}\right] + 2\eta_1\eta_0 \\ & - \mathrm{E}\left[\left(\tfrac{Y_{1i}-\mu_1+\eta_1}{\pi[X_i'\alpha]} + \tfrac{Y_{0i}-\mu_0+\eta_0}{1-\pi[X_i'\alpha]}\right)\tfrac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \mathrm{AV}\left[\hat{\alpha}\right] \mathrm{E}\left[\left(\tfrac{Y_{1i}-\mu_1+\eta_1}{\pi[X_i'\alpha]} + \tfrac{Y_{0i}-\mu_0+\eta_0}{1-\pi[X_i'\alpha]}\right)\tfrac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]'.\end{aligned}$$

Similar to the first two weighting estimators, the third weighting estimator is more efficient if the probabilities are estimated. This can be seen from the fact that the first line corresponds to the asymptotic variance with the known propensity score and the second line is a positive semidefinite matrix subtracted from the first line. The details of the derivation can be found in Appendix B.1. Note that in practice one will estimate η_1 and eta_0 jointly with other parameters. This can easily be done by adding the moment functions derived from (15) and (16) to (17). It is easy to show that adding these moment conditions does not change the asymptotic variance of the resulting ATE estimator.

In this section, the existing weighting estimators are reviewed and represented in M-estimation framework, including a discussion of their theoretical properties. Henceforth, the weighting estimators are denoted by *IPW* followed by the number of the weighting function (i.e., *IPW1*, *IPW2* and *IPW3*).

2.3 Doubly Robust Methods

Both of the previously mentioned estimation methods, regression and propensity score weighting, can be easily implemented. There are no computational difficulties, or curse of dimensionality problems as in nonparametric methods. However, the consistency of the estimates hinges upon the true specification of the mean of the outcome variable or the propensity score depending on the estimation method used. Over the last decade, a lot of progress has been made on the development of doubly robust estimators in incomplete data analysis(for a review, see Słoczyński and Wooldridge, 2018, Seaman and Vansteelandt, 2018). These estimators can be seen as a combination of two estimation strategies, where each of them alone estimates the same parameter of interest. The benefit of combining these two methods is that the resulting mixed estimation strategy is more robust against misspecification.

When two different estimation approaches are used alone to estimate the parameter of interest, a misspecification of the corresponding model delivers an inconsistent estimate. However, the combination will still consistently estimate the parameter of interest, even if one of the models suffers from misspecification. Given that for a researcher it is difficult to be sure about the correct specification of the model, doubly robust methods provide more protection against misspecification.

Here, we consider two doubly robust estimation methods of the ATE. Both methods combine the regression adjustment and propensity score weighting. The first doubly robust estimator of the ATE uses the following estimators for μ_1 and μ_0:[7]

$$\begin{aligned}\hat{\mu}_{1,dr1} &= \frac{1}{N}\sum_{i=1}^{N}(\omega_{1i}(\hat{\alpha})Y_i - (\omega_{1i}(\hat{\alpha}) - 1)\eta[X_i'\hat{\beta}_1]) \\ &= \frac{1}{N}\sum_{i=1}^{N}(\omega_{1i}(\hat{\alpha})(Y_i - \eta[X_i'\hat{\beta}_1]) + \eta[X_i'\hat{\beta}_1]) \qquad (18) \\ \hat{\mu}_{0,dr1} &= \frac{1}{N}\sum_{i=1}^{N}(\omega_{0i}(\hat{\alpha})Y_i - (\omega_{0i}(\hat{\alpha}) - 1)\eta[X_i'\hat{\beta}_0]) \\ &= \frac{1}{N}\sum_{i=1}^{N}(\omega_{0i}(\hat{\alpha})(Y_i - \eta[X_i'\hat{\beta}_0]) + \eta[X_i'\hat{\beta}_0]) \qquad (19)\end{aligned}$$

Thus, the first doubly robust estimator is given by

$$\hat{\tau}_{dr1} = \hat{\mu}_{1,dr1} - \hat{\mu}_{0,dr1}.$$

If the propensity score model is correctly specified, then the second terms in first equalities of (18) and (19) converge to zero. Because the first terms correspond to weighting estimators for μ_1 and μ_0, the unconditional means are estimated consistently; independent of whether the outcome models, $\eta[X_i'\beta_d]$, are correctly specified or not. Meanwhile, if the outcome model is correctly specified, then it estimates the treatment effect parameter consistently because the first terms in second equalities converge to zero and therefore the wrongly specified propensity scores disappear. It has also been shown that adding the "augmenting" the weighting method with the regression adjustment increases the efficiency with respect to the weighting method (see Robins et al., 1994, Lunceford and Davidian, 2004). The proof of doubly ro-

[7] See Robins, Rotnitzky, and Zhao (1994), Robins et al. (1995) and Lunceford and Davidian (2004) for a detailed discussion of this method.

bustness is given in Appendix B.1.

Alternatively, the estimation procedure can be stated as an M-estimation with following moment functions related to the coefficient vector $\theta_{dr1} = (\alpha, \beta_1, \beta_0, \mu_1, \mu_0)$

$$\psi(Z_i, \theta_{dr1}) = \begin{pmatrix} \psi_1(Z_i, \theta_{dr1}) \\ \psi_2(Z_i, \theta_{dr1}) \\ \psi_3(Z_i, \theta_{dr1}) \\ \psi_4(Z_i, \theta_{dr1}) \\ \psi_5(Z_i, \theta_{dr1}) \end{pmatrix} = \begin{pmatrix} \frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial \pi[X_i'\alpha]}{\partial \alpha} \\ D_i \frac{\partial q(Y_i, X_i; \beta_1)}{\partial \beta_1} \\ (1-D_i)\frac{\partial q(Y_i, X_i; \beta_0)}{\partial \beta_0} \\ \omega_{1i}(\alpha)(Y_i - \eta[X_i'\beta_1]) + \eta[X_i'\beta_1] - \mu_1 \\ \omega_{0i}(\alpha)(Y_i - \eta[X_i'\beta_0]) + \eta[X_i'\beta_0] - \mu_0 \end{pmatrix}.$$

Because the results on M-estimation apply, the asymptotic variance has the standard form. Using the relevant parts of the asymptotic variance and applying (1) will give the asymptotic variance of $\hat{\tau}_{dr1}$.

$$\begin{aligned} \mathrm{AV}_{\hat{\tau},dr1} = {} & \mathrm{E}\left[\omega_{1i}(\alpha)^2 (Y_{1i} - \mu_1)^2\right] + \mathrm{E}\left[\omega_{0i}(\alpha)^2 (Y_{0i} - \mu_0)^2\right] - \mathrm{E}\Big[\left(\omega_{1i}(\alpha)^2 - 1\right)\left(\eta[X_i'\beta_1] - \mu_1\right) \\ & + \left(\omega_{0i}(\alpha)^2 - 1\right)\left(\eta[X_i'\beta_0] - \mu_0\right) + 2\left(\eta[X_i'\beta_1] - \mu_1\right)\left(\eta[X_i'\beta_0] - \mu_0\right)\Big] \end{aligned}$$

By plugging in $\omega_{di} = \omega_{di}^{(1)}$, we get the following asymptotic variance of $\hat{\tau}_{dr1}$ as in Lunceford and Davidian (2004):

$$\begin{aligned} \mathrm{AV}_{\hat{\tau},dr1} = {} & \mathrm{E}\left[\frac{(Y_{1i} - \mu_1)^2}{\pi(X_i'\alpha)} + \frac{(Y_{0i} - \mu_0)^2}{1 - \pi(X_i'\alpha)}\right] \\ & - \mathrm{E}\left[\left(\sqrt{\frac{1 - \pi(X_i'\alpha)}{\pi(X_i'\alpha)}}\left(\eta[X_i'\hat{\beta}_1] - \mu_1\right) + \sqrt{\frac{\pi(X_i'\alpha)}{1 - \pi(X_i'\alpha)}}\left(\eta[X_i'\hat{\beta}_0] - \mu_0\right)\right)^2\right]. \end{aligned}$$

Due to the results by Robins et al. (1994), this doubly robust estimator is more efficient than weighting estimators IPW1, IPW2 and IPW3. Also note that the asymptotic variance does not depend on the estimation error of the propensity score or the asymptotic variance of regression part. Although this method was originally proposed with the weighting function $\omega_{di}^{(1)}$, in the Monte Carlo we consider study all three weighting functions described in the previous subsection. The resulting doubly robust estimators are denoted by DR1a, DR1b and DR1c, respectively.

The second possible way of getting a doubly robust estimator of ATE for certain types of conditional mean functions is to use a weighted version of the regression adjustment (Kang and Schafer, 2007, Wooldridge, 2007). The main idea is to weight the objective function by $(\pi[X_i'\hat{\alpha}])$ for the treated and by $(1-\pi[X_i'\hat{\alpha}])$ for the untreated. In the M-estimation framework, the sample moments of regression method should be replaced by the weighted sample moments and the sample moment for the propensity score should be included in the moment function:

$$\frac{1}{N}\sum_{i=1}^{N}\psi(Z_i,\hat{\theta}_{dr2})=0$$

where $\hat{\theta}_{dr}$ is the estimator of the parameter vector $\theta_{dr2}=(\beta_1,\beta_0,\alpha,\tau)$ and $Z_i = (Y_i, X_i, D_i)$. The sample moments are given by:

$$\psi(Z_i,\theta_{dr2}) \;=\; \begin{pmatrix} \psi_1(Z_i,\theta_{dr2}) \\ \psi_2(Z_i,\theta_{dr2}) \\ \psi_3(Z_i,\theta_{dr2}) \\ \psi_4(Z_i,\theta_{dr2}) \end{pmatrix} = \begin{pmatrix} \frac{D_i}{\pi[X_i'\alpha]}\frac{\partial q(Y_i,X_i;\beta_1)}{\partial\beta_1} \\ (\frac{1-D_i}{1-\pi[X_i'\alpha]})\frac{\partial q(Y_i,X_i;\beta_0)}{\partial\beta_0} \\ \frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha} \\ \eta[X_i'\beta_1]-\eta[X_i'\beta_0]-\tau \end{pmatrix}. \qquad (20)$$

The fourth moment function corresponds to the doubly robust estimator of ATE. Similar to the regression estimator of the ATE, the second doubly robust estimator can be written as follows:

$$\hat{\tau}_{dr2}=\frac{1}{N}\sum_{i=1}^{N}[\eta(X_i'\hat{\beta}_{1,dr})-\eta(X_i'\hat{\beta}_{0,dr})],$$

where $\hat{\beta}_{1,dr}$ and $\hat{\beta}_{0,dr}$ are estimated by a weighted regression. The resulting ATE estimator is doubly robust, if X_i includes a constant and $\eta(\cdot)^{-1}$ is a canonical link function. For a continuous outcome variable, the suitable link function is the identity link. Whereas, for a dichotomous outcome the logit link $(g(a)^{-1}=\ln\left(\frac{a}{1-a}\right)$), $g(a)=\frac{\exp(a)}{1+\exp(a)}$) and for a nonnegative discrete outcome variable the log link $(g(a)^{-1}=\ln(a)$, $g(a)=\exp(a))$ will be suitable. The proof of the doubly robustness can be found in Appendix B.1.

The asymptotic variance of $\hat{\tau}_{dr2}$ has the same form as in (6) with the asymptotic variances of $\hat{\beta}_{1,dr}$ and $\hat{\beta}_{0,dr}$. If we only consider the first three moments in (20), then the variance of $\hat{\beta}_{1,dr}$ and $\hat{\beta}_{0,dr}$ can be derived where the two step nature of the estimation is taken care of. Thus, the asymptotic variance of $\hat{\tau}_{dr2}$ is given by

$$\begin{aligned} \mathrm{AV}_{\hat{\tau},dr2} &= \mathrm{E}\left[(\eta[X_i'\beta_1] - \eta[X_i'\beta_0] - \tau)^2\right] + \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_1]}{\partial\beta_1'}\right] \mathrm{AV}_{\hat{\beta}_{1,dr}}\, \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_1]}{\partial\beta_1'}\right]' \\ &\quad + \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_0]}{\partial\beta_0'}\right] \mathrm{AV}_{\hat{\beta}_{0,dr}}\, \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_0]}{\partial\beta_0'}\right]' \end{aligned}$$

with

$$\begin{aligned} \mathrm{AV}_{\hat{\beta}_{1,dr}} = {} & \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1} \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)S_1(\beta_1)'\right] \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1} \\ & - \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1} \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}S_1(\beta_1)S(\alpha)'\right] \mathrm{AV}\left[\hat{\alpha}\right] \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}S_1(\beta_1)S(\alpha)'\right]' \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1} \end{aligned}$$

and

$$\begin{aligned} \mathrm{AV}_{\hat{\beta}_{0,dr}} &= \mathrm{E}\left[\frac{(1-D_i)}{(1-\pi[X_i'\alpha])}H_0(\beta_0)\right]^{-1} \mathrm{E}\left[\frac{(1-D_i)}{(1-\pi[X_i'\alpha])^2}S_0(\beta_0)S_1(\beta_0)'\right] \mathrm{E}\left[\frac{(1-D_i)}{(1-\pi[X_i'\alpha])}H_0(\beta_0)\right]^{-1} \\ &\quad - \mathrm{E}\left[\frac{(1-D_i)}{(1-\pi[X_i'\alpha])}H_0(\beta_0)\right]^{-1} \mathrm{E}\left[\frac{(1-D_i)}{(1-\pi[X_i'\alpha])}S_0(\beta_0)S(\alpha)'\right] \mathrm{AV}\left[\hat{\alpha}\right] \\ &\quad \times \mathrm{E}\left[\frac{(1-D_i)}{(1-\pi[X_i'\alpha])}S_0(\beta_0)S(\alpha)'\right]' \mathrm{E}\left[\frac{(1-D_i)}{(1-\pi[X_i'\alpha])}H_0(\beta_0)\right]^{-1} \end{aligned}$$

where $H_d(\beta_d)$ stands for the Hessian (second derivative of the objective function) and $S_d(\beta_d)$ stands for the score (first derivative of the objective function) of unweighted regression for β_d. The details of the derivations can be found in Appendix B.1. Wooldridge (2007) shows that for the objective functions, $q(Y_i, X_i; \beta_d)$ for $d = \{0, 1\}$, which satisfy the generalized conditional information matrix equality, the (unweighted) regression estimator is more efficient than any other weighted regression estimator (Wooldridge, 2007, Theorem 4.3). Basically, the weighted regression increases the robustness at the expense of efficiency with respect to the regression estimator. However, the weighted regression type of doubly robust estimator is still more efficient than the weighting estimators (see Robins et al., 1994).

3 Monte Carlo Study

In this section, the finite sample properties of regression, propensity score weighting and doubly robust estimators described earlier are compared. Propensity score weighting and the first doubly robust method explained in Section 2 are applied using all three weighting functions. Thus, in total eight estimators are compared.[8]

[8] All simulations are conducted in GAUSS, the codes are available upon request.

Several recent papers have investigated the finite sample properties of treatment effect estimators (see, for example, Busso et al., 2014, Frölich et al., 2017, Bodory et al., 2020). The simulation setup shares some similarities with previous studies, but the emphasis of our study is different from these studies. Here, the emphasis lies on the finite sample properties of the parametric estimators under misspecification of the propensity score or the outcome equation. Therefore, the properties of these estimation methods are examined for various misspecification designs. The overlap problem, which is an important issue for propensity score based methods, is also investigated in the simulation studies for the reviewed methods.

Table 1: Data Generating Processes

Model	Outcome Equations & Treatment Equation
DGP 1	$Y_{0i} = \beta_{00} + \beta_{01}X_{1i} + \beta_{02}X_{2i} + \varepsilon_{0i}$
	$Y_{1i} = \beta_{10} + \beta_{11}X_{1i} + \beta_{12}X_{2i} + \varepsilon_{1i}$
	$D_i = \mathbb{1}\{\alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} - \nu_i > 0\}$
DGP 2	$Y_{0i} = \beta_{00} + \beta_{01}X_{1i} + \beta_{02}X_{2i} + \varepsilon_{0i}$
	$Y_{1i} = \beta_{10} + \beta_{11}X_{1i} + \beta_{12}X_{2i} + \varepsilon_{1i}$
	$D_i = \mathbb{1}\{\alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_2 X_{3i} - \nu_i > 0\}$
DGP 3	$Y_{0i} = \beta_{00} + \beta_{01}X_{1i} + \beta_{02}X_{2i} + \beta_{03}X_{3i} + \varepsilon_{0i}$
	$Y_{1i} = \beta_{10} + \beta_{11}X_{1i} + \beta_{12}X_{2i} + \beta_{13}X_{3i} + \varepsilon_{1i}$
	$D_i = \mathbb{1}\{\alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} - \nu_i > 0\}$
DGP 4	$Y_{0i} = \beta_{00} + \beta_{01}X_{1i} + \beta_{02}X_{2i} + \beta_{03}X_{3i} + \varepsilon_{0i}$
	$Y_{1i} = \beta_{10} + \beta_{11}X_{1i} + \beta_{12}X_{2i} + \beta_{13}X_{3i} + \varepsilon_{1i}$
	$D_i = \mathbb{1}\{\alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_2 X_{3i} - \nu_i > 0\}$

The data generating processes (DGPs) are given in Table 1. Four different DGPs are considered. The difference comes from the fact that the Xs that determine potential outcomes and treatment indicator are not the same for all DGPs. (Y_{1i}, Y_{0i}, D_i) generated by the first DGP are determined only by X_{1i} and X_{2i}. The second DGP generates (Y_{0i}, Y_{1i}), the same as the DGP1; however, the treatment indicator is additionally affected by X_{3i}. The third DGP for the treatment indicator is the same as the first one, but the potential outcomes are affected by all three covariates.

Finally, the last DGP produces (Y_{1i}, Y_{0i}, D_i) as functions of all three covariates. The error term of the treatment indicator, ν_i, is drawn from a logistic distribution, whereas the error terms of the potential outcomes, $(\varepsilon_{1i}, \varepsilon_{0i})$, are independent vectors of standard normal variables. In the first part of the simulation study, X_{1i}, X_{2i} and X_{3i} are drawn from a uniform distribution over $[-1, 1]$ with the correlation matrix V_X, which is given by

$$V_X = \begin{bmatrix} 1.0 & 0.7 & 0.6 \\ 0.7 & 1.0 & 0.6 \\ 0.6 & 0.6 & 1.0 \end{bmatrix}.$$

According to the results by Khan and Tamer (2010) with bounded covariates, it is guaranteed that the strict overlap assumption is satisfied. Therefore, in the first part we use uniformly distributed variables. Khan and Tamer (2010) note that strict overlap is a sufficient assumption for $\sqrt{N}-$consistency of semiparametric treatment effect estimators. For the second part of the simulation study, we evaluate the properties of the methods discussed here under violation of strict overlap. For this purpose, the X variables are drawn from a multivariate normal distribution with the same correlation matrix V_X.

The coefficients for the potential outcome models and treatment indicators are chosen to create various interesting scenarios. In general, treatment effects can be categorized as being homogeneous and heterogeneous. Homogeneity for treatment effect means that the effect of the treatment does not change with different X characteristics. Meanwhile, heterogeneity implies that the treatment effect varies with different characteristics. We use different parameter combinations of $\beta_1 = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})'$ and $\beta_0 = (\beta_{00}, \beta_{02}, \beta_{01}, \beta_{03})'$ to investigate the differences of the estimators in finite samples with respect to the homogeneity and heterogeneity of the treatment effect. If the parameter vectors β_1 and β_0 differ only in the constant term, then the treatment effect is homogeneous. Otherwise, the treatment effect is heterogeneous. The parameter combinations for the potential outcomes are given in Table 2.

Table 2: Parameter Configurations for the Outcome Equation

Homogeneous Treatment				**Heterogeneous Treatment**			
β_{00}	3	β_{10}	1	β_{00}	3	β_{10}	1
β_{01}	4	β_{11}	4	β_{01}	4	β_{11}	5
β_{02}	2	β_{12}	2	β_{02}	2	β_{12}	-1
β_{03}	1	β_{13}	1	β_{03}	1	β_{13}	2

Note: For DGP1 and DGP2 the coefficients of X_3 are set to zero. See Table 1.

We use different values of the parameter vector $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)'$ to generate samples with different expected treated-control ratios. We investigate three ratios. Table 3 summarizes the parameter vectors for each ratio. The DGPs given in Table 1, together with the coefficient vector α, give the expected treated-control ratios stated at the bottom of Table 3.

Table 3: Parameter Configurations for the Treatment Equation with bounded X

	Ratio 1	**Ratio 2**	**Ratio 3**
α_0	1.5	0	-1.5
α_1	1.5	1.5	1.5
α_2	1	1	1
α_3	0.5	0.5	0.5
$\mathrm{E}\left[D\middle\| X\right]$	0.25	0.5	0.75
Treated-Control	1:3	1:1	3:1

Note: For DGP1 and DGP3 the coefficients of X_3 are set to zero. See Table 1.

For all four DGPs, we use the same regression models for the outcome equation and we used the selection equation to estimate the ATE. Both conditional means are modeled as a linear function of a constant, X_1 and X_2. For the outcome equation, an identity link function is used and for the treatment variable a logit link function is used. In other words, we estimate the following regression equations independent of the true data generating process:

$$\mathrm{E}\left[Y_{1i}\middle| X_{1i}, X_{2i}\right] = \delta_{10} + \delta_{11}X_{1i} + \delta_{12}X_{2i} \tag{21}$$

$$\begin{aligned} \mathrm{E}\left[Y_{0i}\middle|\, X_{1i}, X_{2i}\right] &= \delta_{00} + \delta_{01}X_{1i} + \delta_{02}X_{2i} & (22) \\ \mathrm{E}\left[D_{i}\middle|\, X_{1i}, X_{2i}\right] &= F(\gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i}) \\ &= \frac{\exp\left(\gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i}\right)}{1 + \exp\left(\gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i}\right)} & (23) \end{aligned}$$

Theoretically, all of the methods based on the regression models given by (21)-(23) give consistent estimates for the first GDP. For the second GDP, regression (23) is a misspecified model because X_3 is omitted. Thus, the weighting methods, IPW1-IPW3, estimate the ATE inconsistently, but regression and doubly robust methods are not affected by this misspecification. For the third GDP, the outcome regression models given in (21) and (22) are misspecified because the regression models for Y_{1i} and Y_0 omit X_3, which is a confounder for the outcome variables according to the third DGP. Therefore, the estimator based on the regression method is inconsistent. IPW methods are not supposed to be affected by this misspecification. Because doubly robust methods use both model specifications, if the true data generating process is either the second or the third, then one of the underlying models for doubly robust method will be misspecified. However, the theoretical results suggest that these methods estimate the ATE consistently, even for DGP2 and DGP3 with given regression models. None of these methods estimates the parameter of interest consistently for the fourth DGP because true DGPs of potential outcomes and treatment indicator include X_3, whereas the regression models do not.

We estimate the ATE for all possible scenarios with all eight estimators. Tables B1-B8 summarize the results, where the odd numbered tables give the results for the homogeneous treatment and the even numbered tables give the results for the heterogeneous treatment. Each table has three panels. The panels correspond to the different treatment-control ratios. Within the panels, four summary statistics are presented for each sample size, as follows: (i) BIAS: the average bias over Monte Carlo samples; (ii) MCVAR: Monte Carlo Variance multiplied by 100; (iii) MCMSE: Mean squared error over Monte Carlo samples multiplied by 100; and (iv) AAVAR: average of the variance estimates based on sandwich form over Monte Carlo samples multiplied by 100.[9] We investigate the properties of the methods for the sample sizes N=100, 400 and 1600. The number of replications (R) is chosen to be proportional to the sample sizes by setting $N \times R$ constant. For the sample sizes 100,

[9] Because the last three statistics are very close for some estimators, they are multiplied by 100 to facilitate a comparison.

400 and 1600, we use 16000, 4000 and 1000 replications, respectively. Regression adjustment is denoted by REG. IPW1-IPW3 refer to the weighting methods with the three weighting functions explained in Section 2. DR1a-DR1c stand for the first doubly robust method with three different weighting functions. Lastly, DR2 stands for the second doubly robust method.

Tables B1-B2 give a summary of the simulation results for correctly specified regression models for outcome and treatment variables. The regression and doubly robust estimators of ATE are unbiased, even in small samples under each treated-control ratio. However, the weighting estimators are biased in small samples. Although the biases get smaller as the sample size increases, they are still slightly biased even with a sample size of 1600 if the treated to control ratio is not 1:1. The IPW methods are also very inefficient compared to the regression and doubly robust methods. Among the IPW methods, the third weighting function gives the smallest variance as the theory suggests; however, it has the highest bias among the IPW methods. Interestingly, for each treatment-control ratio the biases of IPW methods are smaller under heterogeneous settings than those under homogeneous settings for the smallest sample size. Furthermore, in almost all cases, the weighting methods are more efficient under heterogeneity. Meanwhile, heterogeneous treatment leads to a decrease in the efficiency of regression and doubly robust methods. The regression estimator is the most efficient among all of the estimators except for the smallest sample size. The second doubly robust estimator is the most efficient if sample size is 100. The regression estimator has the smallest MCMSE. Doubly robust methods are slightly less efficient than the regression estimators and is much more efficient than the weighting estimators. All four doubly robust methods are quite close to each other in terms of efficiency. However, the second doubly robust estimator (DR2) always has the smallest average asymptotic variance and the first doubly robust estimator with second weighting function (DR1b) has the second smallest. The first doubly robust method with the third weighting function (DR1c) has the smallest MCMSE among the doubly robust methods. All of the methods become less efficient if the treated-control ratio deviates from 1:1. The first doubly robust estimator with first weighting function (DR1a) seems to be the least efficient one among all doubly robust estimators.

Tables B3-B4 give the results for correctly specified outcome model along with misspecified treatment model. Thus, the IPW methods are theoretically inconsistent, whereas the regression and doubly robust methods are consistent. For each treated-control ratio, the regression and doubly robust methods are unbiased, whereas the weighting methods results in biased estimators for all sample sizes. The variances of doubly robust methods slightly increase due to the misspecification of the propensity score method. The regression method is still the best with respect to the MCMSE. As earlier, for small samples the average asymptotic variance estimates of the second doubly robust estimator is smaller than of the others. Doubly robust methods do not differ much from each other in terms efficiency. DR2 has lowest asymptotic variance, whereas DR1c has the lowest Monte Carlo Variance. As earlier, heterogeneity in treatment increases the variance of the regression and doubly robust estimators. Deviations from 1:1 treated-control ratio have similar effects.

The results of misspecified outcome model with correctly specified treatment model are given in Tables B5-B6. Interestingly, for small sample sizes the bias of the weighting estimator is even larger than the bias of regression estimator. However, as opposed to the weighting estimator, the bias of the regression estimator does not vanish as the sample size grows. When compared with the case where both outcome and propensity score models are correctly specified (Tables B1-B2), the variances of doubly robust estimators decrease due to misspecified regression model. However, the difference is not very large. Even with the misspecified outcome model, the regression estimator is the most efficient. This observation is also consistent with the theoretical results in Wooldridge (2007). The previous observations on the efficiency rankings of the doubly robust estimators, the effect of the treated-control ratio and the effect of the heterogeneity are also valid for the case where outcome model is misspecified.

Tables B7-B8 summarize the results for the misspecified outcome and treatment regression models. All of the methods give inconsistent estimates, as suggested by the theory. The biases are higher than previous cases. This observation suggests that if the omitted variable is only relevant for the outcome model or the propensity score model, then the resulting treatment effect estimates are less affected by the omission. However, if the omitted variable is relevant for both, then all the estimators are affected significantly by the omission.

To give an overview the efficiency results, the asymptotic variances of all estimator for all possible scenarios are tabulated in Tables A1 and A2 as a factor of the asymptotic variance of the regression estimator. The numbers stand for the asymptotic variance of the given estimator divided by the asymptotic variance of the regression estimator. Numbers smaller than 1 indicate an asymptotic variance that is smaller than the variance of the regression estimator. We see for the smallest sample size that the second doubly robust method and sometimes first doubly robust estimator with second weighting function have smaller asymptotic variance than regression method. However, it should be noted that for a sample size of 100, it is very likely that estimating the variance by its asymptotic variance is not a good approximation. At the bottom of each table, the minimum and maximum factors are reported. Doubly robust methods are closer to the regression method, whereas the asymptotic variance of IPW estimators can be considerably higher than regression estimator. Among the doubly robust estimators, DR2 has the smallest variances and the DR1b has the second smallest variance. If the weighting estimators are considered, then the third weighting estimator is the most efficient. The variances get slightly closer if the treatment effect is heterogeneous, but the ranking does not change.

In the second part of the Monte Carlo study, we use unbounded Xs, which cause the violation of the strict overlap assumption. Here, different parameter combinations are used to fix the treated-control ratio (as in the first part). Figure A2 displays the conditional density graphs of the propensity score for the treated and untreated samples for three different treated-control ratios where Xs are drawn from a normal distribution. Visual inspection on Figures A1 and A2 clarifies the distinction between overlap and strict overlap. The estimated propensity scores are strictly smaller than 0 and strictly greater than 1. However, to sustain the strict overlap assumption, the probabilities have to be $\epsilon > 0$ away from the boundaries. The densities displayed in Figures A1 show that with bounded Xs we guarantee that the probabilities away from the boundaries. However, with unbounded Xs, as displayed in Figure A2, it is possible that we observe probabilities which are almost at the boundaries. Because the theory by Khan and Tamer (2010) suggests that if the strict overlap is violated, as in Figure A2, $\sqrt{N}$ convergence is not guaranteed. The first goal of the second part of the Monte Carlo study is to investigate the small sample properties under the violation of strict overlap assumption. For this part of the analysis, we only consider the first GDP under which all methods give consistent estimates. The second goal of this part is related to the possible solutions to this problem. One possible way to deal with this problem is to estimate the treatment effect over the common support. Several methods are proposed in the literature to estimate the common support (for a review, see Caliendo and Kopeinig, 2008). The second goal is to compare two commonly used rules of determining the common support. According to the first rule, the common support is determined by deleting all observations whose propensity score is smaller than the minimum and larger than the maximum in the opposite group. This rule was proposed by Dehejia and Wahba (1999). The second rule proposed by Crump, Hotz, Imbens, and Mitnik (2009) determines the common support by dropping all units with an estimated propensity score outside the interval $[0.1; 0.9]$.

The results are summarized in Tables B9-B11. IPW methods are affected by violation of strict overlap, especially for the heterogeneous treatment effect for the small sample sizes and if the treated-control ratio is different from 1:1. Among the weighting methods, IPW3 is affected at most by the violation. The doubly robust methods and the regression method are not much affected by violation of strict overlap in terms of bias. All the methods have larger variances than the overlap problem case. As the ratio of treated to control deviates from 1:1, the biases of the IPW methods increase considerably, especially if the treatment is heterogeneous. Meanwhile, there is no increase of the biases of the doubly robust methods or regression methods. It is difficult to generalize the effects of trimming rules. Under some settings they help to decrease the bias, but under some settings they can even harm the estimators. If the treated-control ratio is 1:1, then both trimming rules seem to improve the performance of weighting estimator in terms of bias and variance. The second trimming rule has slightly stronger effect. For a 1:1 ratio, the trimming rules have neither positive nor negative effects on the regression and doubly robust methods. If the ratio is different than 1:1, then the effects of the trimming rules differ considerably for homogeneous and heterogeneous treatment. If the treatment is homogeneous, then trimming rules help to decrease the bias and variance of the weighting methods. The second trimming rule works slightly better than the first trimming rule. Trimming rules do not change much for the regression and doubly robust estimator if the treatment homogeneous. However, if the treatment is heterogeneous and the ratio is different than 1:1, then second trimming rule has a worsening effect on all estimators. The first trimming rule has only a small positive effect on the third weighting method if the sample size is large. However, it has negative effect on the performance of the other estimators.

According to the results of the Monte Carlo study, the doubly robust estimation methods outperform the propensity score weighting methods. The first advantage of the doubly robust methods is that it provides double protection against misspecification. The second advantage over weighting methods is that they are more efficient. The last advantage is the violation of strict overlap assumption is not as harmful for harm the doubly robust estimators as the weighting estimators. Because neither theoretical results nor Monte Carlo evidence suggests any optimal trimming rule that corrects the problems associated with violation of (strict) overlap assumption, using doubly robust methods provides additional protection against violation of the strict overlap assumption.

In summary, the doubly robust estimator based on weighted estimation methods with weights given by the inverse of the propensity score performs better than the simple regression estimator under incorrect outcome model specification. Under correct outcome model specification, both estimates are unbiased. However, the regression estimates are slightly more efficient than doubly robust estimates. The doubly robust estimator performs better than the propensity weighting estimator under both correct and incorrect propensity score specification. The lack of overlap between the treated and control group considerably effects the quality of the weighting estimator. Based on the results of the Monte Carlo study, it can be concluded that the doubly robust estimator provides double protection against misspecification for finite samples at no significant cost.

4 Empirical Example: Returns to Higher Education

As an illustration of the methods studied here, we estimate the causal effects of obtaining a higher education degree using data from the NCDS (National Child Development Survey). The NCDS is a continuing longitudinal study that follows all those living in Great Britain who were born in one particular week in 1958.[10] We

[10] These datasets have been used in several empirical papers (see for example Blundell, Dearden, Goodman, and Reed, 2000, Blundell et al., 2005, Dearden, 1999a,b, Dearden, Ferri, and Meghir, 2002).

construct the sample based on the waves, which were undertaken in 1965, 1969, 1974 and 1991. We only use males to avoid problems related with female labour force participation decision. The dependent variable is log of hourly wages at the time of the 1991 survey. The treatment variable is defined as having a higher education degree versus less than higher education at the age of 33 (in year 1991).[11]

Blundell et al. (2005) estimate the causal effects of education under conditional independence assumption using the same dataset.[12] However, they do not consider doubly robust methods or weighting methods reviewed here. Therefore, we reconsider the estimation of causal effect of higher education by means of the methods reviewed here. Before applying these estimation methods, one should carefully investigate the validity of the assumptions. First, one should include variables related to both treatment status and potential outcomes in the estimation so that the CIA holds approximately. The study by Blundell et al. (2005) is one of the very few examples where the returns to education are estimated under CIA assumption.[13] Because the CIA assumption is not testable, it is important that one uses a rich dataset. Following Blundell et al. (2005), we argue that the rich set of control variables available in the dataset should be enough to satisfy the CIA assumption approximately. The dataset consists of detailed information on parents, school related topics and ability measures. We use variables that measure the individual's mathematical and reading ability. The variables are based on ability tests that were undertaken when the child was 7 and 11. Five dummy variables that indicate to which quintile an individual belongs are constructed for each test to rank the individuals.[14] We control for the school types that the individuals attended at the age of 16. To control for family background, the parents' years of education, parents' ages, father's social class, mother's employment status, number of siblings when the child was 16 are included as covariates. The variables measuring the parents' interest on the child's education are based on the teachers' assessments. Other than these variables, we control for

[11] Higher education group includes the Higher National Certificate or Diploma, the Scottish Higher National Certificate or Scottish Higher National Diploma, Technician Education Council or Business and Technician Education Council Higher or Higher National Certificate or Diploma or Scottish equivalent of those, professional qualifications, nursing qualifications including National Nursery Examining Board, polytechnic qualifications, university certificates or diplomas, first degrees, postgraduate diplomas and higher degrees.

[12] We try to closely follow the data preparation by Blundell et al. (2005) to create the sample used here. We get very similar figures, but the samples are not identical.

[13] Flossmann (2010), Flossmann and Pohlmeier (2006) and Pohlmeier and Pfeiffer (2004) estimate the returns to schooling under CIA using different datasets from Germany.

[14] The quintiles refer to quintiles at the time of the test was taken.

the region the individual used to live at age of 16 and experiencing financial problems in 1969 or 1974. The indicator variable for past financial problems is constructed following Dearden (1999b). This variable identifies individuals who received free school meals in 1969 or 1974 or whose parents were seriously troubled financially in the year prior to the 1969 or 1974 survey. As in Blundell et al. (2005), we only drop the observations if the treatment variable or the outcome variable is missing. For all of the other variables, an indicator variable for missing cases is used and the missing values are set to zero.

The summary statistics are presented in Table A4. All of the variables differ in terms of their means by treatment status. This indicates the need to control for covariates. Another important requirement for the validity of our analysis is to use covariates that are not affected by the treatment or the outcome variable. Because all of the covariates are measured before the treatment and the outcome, they are obviously not affected by them.

The second important assumption is the (strict) overlap assumption. We evaluate the common support assumption by comparing the distributions (histograms) of the estimated propensity scores by the treatment variable, as suggested in Lechner (2010). The propensity score is estimated by logit using all the covariates listed in Table A4. The estimation results for propensity score can be found in Table A5. There are untreated individuals with very low probabilities of getting a higher education. However, there are also individuals in treated group who have low probabilities of getting a higher education (see Figure A3). Thus, the histogram does not indicate overlap problems. Therefore, we estimate the ATE without applying any common support correction.[15] The outcome model is specified as a linear model with identity link function. The estimation results by different methods are summarized in Table 4. All of the reviewed methods estimate the average returns of higher education as around 20%. This means that on average the individuals with higher educational degree earn 20% more than the individuals with any educational degree less than higher education. Blundell et al. (2005) estimate the ATE with several methods and their estimates vary over an interval from 20% to 40%. Because the sample we use here is not a one-to-one match to their sample, the differences are not surprising. Nevertheless, their ATE estimate (where they consider treatment heterogeneity is

[15] The estimation results do not change after applying the first trimming rule discussed in Section 3.

22%) is very close to the results presented in Table 4.

Table 4: Estimated ATE of higher education

REG1	**IPW1**	**IPW2**	**IPW3**	**DR1a**	**DR1b**	**DR1c**	**DR2**
0.21***	0.21***	0.19***	0.20***	0.21***	0.21***	0.21***	0.21***
(0.018)	(0.040)	(0.021)	(0.020)	(0.020)	(0.020)	(0.019)	(0.016)

Note: The methods are explained in Section 2. Standard errors are reported in parentheses. The sample size is 3,092.

The proximity of different estimates might indicate that the model specifications are correct. Unfortunately, it may also indicate that all the specifications are wrong. Because there is no way to determine the correctness of the specifications, one needs to evaluate the estimates according to their plausibility. In our case, the results seem to be quite plausible and moreover they are similar to the previous results in the literature.

5 Conclusion

In this paper we have reviewed the treatment effect estimation methods, which belong to the three main groups: regression, weighting and doubly robust methods, within the M-estimation framework. Although they are not new, the doubly robust methods have only recently come to much deserved attention. The appealing feature of a doubly robust estimator is that it stays consistent, despite certain types of model misspecifications. Given that in the applied work it is not clear whether a model is correctly specified or not, the use of doubly robust methods offers double protection against misspecification. A unified representation of the methods is important to demonstrate the relation between these estimators. A unified approach also helps to study the asymptotic properties of the methods that are generalizations of those studied here. One generalization is an estimation of the average treatment effects in case of a nonbinary treatment variable (see, for example, Cattaneo, 2010, Uysal, 2015). Another extension is related to estimation of the local average treatment effect (LATE) with an instrument. It has been shown that the LATE is the ratio of the ATEs (see, for example, Donald, Hsu, and Lieli, 2014, Uysal, 2011, chap. 2). Thus, the unified M-estimation approach can be directly applied to the estimation of LATE where the denominator and numerator are estimated by any method discussed here.

Besides providing a unified M-estimation representation for the methods, we also compare the finite sample properties of these estimators in a Monte Carlo study. We demonstrate the double robustness property and the performance of double robust estimation methods compared to the regression and propensity weighting methods in small samples. The treatment indicator is simulated in several ways to examine different treated/control ratios. This extension of the Monte Carlo design allows us to evaluate the sensitivity of these methods to the distribution of the propensity score. Another aspect of the Monte Carlo study lies in examining the effect of the overlap assumption. Our results show that the doubly robust estimators outperform the other two methods in finite samples under misspecification of either the mean or the propensity score function. Compared to the weighting methods, the doubly robust estimators are less sensitive to the lack of overlap between treated and control groups.

In the last part of this study, we provided an application example of the considered methods. The goal of the example is to give more insights on the application side. We estimate the causal returns of higher education using the rich NCDS dataset. Due to the data limitations, very few studies apply methods that are valid under CIA to estimate the returns of schooling. Because the dataset used in this paper is very rich in terms of variables, the CIA assumption is more likely to be valid. In fact, the same dataset was used to estimate returns to education under unconfoundedness of treatment assumption by different methods in Blundell et al. (2005). Our estimates indicate that higher education increases the earnings by around 20%, which is in line with the results in Blundell et al. (2005).

References

Abadie, A. and M. D. Cattaneo (2018): "Econometric methods for program evaluation," *Annual Review of Economics*, 10, 465–503, URL `https://doi.org/10.1146/annurev-economics-080217-053402`.

Athey, S. and G. W. Imbens (2017): "The state of applied econometrics: Causality and policy evaluation," *Journal of Economic Perspectives*, 31, 3–32, URL `https://www.aeaweb.org/articles?id=10.1257/jep.31.2.3`.

Bang, H. and J. M. Robins (2005): "Doubly robust estimation in missing data and causal inference models," *Biometrics*, 61, 962–973, URL `http://dx.doi.org/10.1111/j.1541-0420.2005.00377.x`.

Blundell, R., L. Dearden, A. Goodman, and H. Reed (2000): "The returns to higher education in Britain: Evidence from a British cohort," *The Economic Journal*, 110, 82–99.

Blundell, R., L. Dearden, and B. Sianesi (2005): "Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 168, 473–512, URL `http://www.jstor.org/stable/3559836`.

Bodory, H., L. Camponovo, M. Huber, and M. Lechner (2020): "The finite sample performance of inference methods for propensity score matching and weighting estimators," *Journal of Business & Economic Statistics*, 38, 183–200, URL `https://doi.org/10.1080/07350015.2018.1476247`.

Busso, M., J. DiNardo, and J. McCrary (2014): "New evidence on the finite sample properties of propensity score reweighting and matching estimators," *The Review of Economics and Statistics*, 96, 885–897, URL `https://doi.org/10.1162/REST_a_00431`.

Caliendo, M. and S. Kopeinig (2008): "Some practical guidance for the implementation of propensity score matching," *Journal of Economic Surveys*, 22, 31–72.

Card, D. (1999): "Chapter 30 the causal effect of education on earnings," in O. C. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, *Handbook of Labor Economics*, volume 3, Part A, Elsevier, 1801 – 1863, URL `http://www.sciencedirect.com/science/article/pii/S1573446399030114`.

Cattaneo, M. D. (2010): "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, 155, 138 – 154, URL `http://www.sciencedirect.com/science/article/pii/S030440760900236X`.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009): "Dealing with limited overlap in estimation of average treatment effects," *Biometrika*, 96, 187–199.

Dearden, L. (1999a): "The effects of families and ability on men's education and earnings in Britain," *Labour Economics*, 6, 551–567, URL `http://www.sciencedirect.com/science/article/pii/S0927537198000153`.

Dearden, L. (1999b): "Qualifications and earnings in Britain: How reliable are conventional OLS estimates of the returns to education?" Working Paper 99/7, Institute for Fiscal Studies, London, URL `http://www.ifs.org.uk/wps/wp9907.pdf`.

Dearden, L., J. Ferri, and C. Meghir (2002): "The effect of school quality on educational attainment and wages," *The Review of Economics and Statistics*, 84, 1–20, URL `https://doi.org/10.1162/003465302317331883`.

Dehejia, R. and S. Wahba (1999): "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs," *Journal of the American Statistical Association*, 94, 1053–1062.

Donald, S. G., Y.-C. Hsu, and R. P. Lieli (2014): "Testing the unconfoundedness assumption via inverse probability weighted estimators of (l)att," *Journal of Business & Economic Statistics*, 32, 395–415, URL `https://doi.org/10.1080/07350015.2014.888290`.

Flossmann, A. (2010): "Accounting for missing data in M-estimation: A general matching approach," *Empirical Economics*, 38, 85–117.

Flossmann, A. and W. Pohlmeier (2006): "Causal returns to education: A survey in empirical evidence for Germany," *Journal of Economics and Statistics*, 226, 6–23, URL `http://ideas.repec.org/a/jns/jbstat/v226y2006i1p6-23.html`.

Frölich, M., M. Huber, and M. Wiesenfarth (2017): "The finite sample performance of semi- and non-parametric estimators for treatment effects and policy evaluation," *Computational Statistics & Data Analysis*, 115, 91 – 102, URL `http://www.sciencedirect.com/science/article/pii/S0167947317301020`.

Graham, B. S., C. C. de Xavier Pinto, and D. Egel (2012): "Inverse probability tilting for moment condition models with missing data," *The Review of Economic Studies*, 79, 1053–1079, URL `http://www.jstor.org/stable/23261379`.

Graham, B. S., C. C. de Xavier Pinto, and D. Egel (2016): "Efficient estimation of data combination models by the method of auxiliary-to-study tilting (ast)," *Journal of Business & Economic Statistics*, 34, 288–301, URL `https://doi.org/10.1080/07350015.2015.1038544`.

Heiler, P. and E. Kazak (2020): "Valid inference for treatment effect parameters under irregular identification and many extreme propensity scores," *Journal of Econometrics*, URL `http://www.sciencedirect.com/science/article/pii/S0304407620303377`.

Hirano, K. and G. W. Imbens (2001): "Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization," *Health Services and Outcomes Research Methodology*, 2, 259–278, URL http://dx.doi.org/10.1023/A%3A1020371312283.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71.

Horvitz, D. G. and D. J. Thompson (1952): "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, 47, 663–685.

Huber, P. J. (1964): "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, 35, pp. 73–101.

Imbens, G. W. (2004): "Nonparametric estimation of average treatment effects under exogeneity," *Review of Economics and Statistics*, 86, 4–29, URL http://dx.doi.org/10.1162/003465304323023651.

Imbens, G. W. and J. M. Wooldridge (2009): "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, 47, 5–86, URL https://www.aeaweb.org/articles?id=10.1257/jel.47.1.5.

Johnston, J. and J. E. DiNardo (1996): *Econometric Methods*, New York, NY: McGraw-Hill.

Kang, J. D. Y. and J. L. Schafer (2007): "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data," *Statist. Sci.*, 22, 523–539, URL https://doi.org/10.1214/07-STS227.

Khan, S. and E. Tamer (2010): "Irregular identification, support conditions, and inverse weight estimation," *Econometrica*, 78, 2021–2042.

Lechner, M. (2010): "A note on the common support problem in applied evaluation studies," *Annales d'Économie et de Statistique*, 91-92, 217–234.

Lee, S., R. Okui, and Y.-J. Whang (2017): "Doubly robust uniform confidence band for the conditional average treatment effect function," *Journal of Applied Econometrics*, 32, 1207–1225, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2574.

Lunceford, J. and M. Davidian (2004): "Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study," *Statistics in Medicine*, 23, 2937–2960.

Muris, C. (2020): "Efficient gmm estimation with incomplete data," *The Review of Economics and Statistics*, 102, 518–530, URL https://doi.org/10.1162/rest_a_00836.

Pohlmeier, W. and F. Pfeiffer (2004): "Returns to education and individual heterogeneity," ZEW Discussion Papers 04-34, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research, URL http://ideas.repec.org/p/zbw/zewdip/1860.html.

Robins, J. M. and Y. Ritov (1997): "Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models," *Statistics in Medicine*, 16, 285–319.

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

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1995): "Analysis of semiparametric regression models for repeated outcomes in the presence of missing data," *Journal of the American Statistical Association*, 90, 106–121, URL `http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476493`.

Rosenbaum, P. R. (1987): "Model-based direct adjustment," *Journal of the American Statistical Association*, 82, 387–394.

Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55, URL `http://biomet.oxfordjournals.org/content/70/1/41.abstract`.

Rothe, C. and S. Firpo (2019): "Properties of doubly robust estimators when nuisance functions are estimated nonparametrically," *Econometric Theory*, 35, 1048–1087.

Sant'Anna, P. H. and J. Zhao (2020): "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, 219, 101 – 122, URL `http://www.sciencedirect.com/science/article/pii/S0304407620301901`.

Scharfstein, D. O., A. Rotnitzky, and J. M. Robins (1999): "Adjusting for nonignorable drop-out using semiparametric nonresponse models," *Journal of the American Statistical Association*, 94, 1096–1120, URL `https://www.tandfonline.com/doi/abs/10.1080/01621459.1999.10473862`.

Seaman, S. R. and S. Vansteelandt (2018): "Introduction to double robust methods for incomplete data," *Statist. Sci.*, 33, 184–197, URL `https://doi.org/10.1214/18-STS647`.

Słoczyński, T. and J. M. Wooldridge (2018): "A general double robustness result for estimating average treatment effects," *Econometric Theory*, 34, 112–133.

Stefanski, L. A. and D. D. Boos (2002): "The calculus of m-estimation," *The American Statistician*, 56, 29–38.

Uysal, S. D. (2011): *Three Essays on Doubly Robust Estimation Methods*, Ph.D. thesis, University of Konstanz.

Uysal, S. D. (2015): "Doubly robust estimation of causal effects with multivalued treatments: An application to the returns to schooling," *Journal of Applied Econometrics*, 30, 763–786, URL `http://dx.doi.org/10.1002/jae.2386`.

Wooldridge, J. M. (2007): "Inverse probability weighted estimation for general missing data problems," *Journal of Econometrics*, 141, 1281 – 1301, URL `http://www.sciencedirect.com/science/article/pii/S0304407607000437`.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

A Appendix

A.1 Tables: Monte Carlo Study

Table A1: Average variances relative to regression estimator (Homogeneous Treatment)

Ratio	N	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2
Both Models Correct								
1/3	100	4.503	3.434	1.531	1.085	1.066	1.164	0.960
	400	3.281	3.028	1.733	1.126	1.125	1.149	1.100
	1600	3.037	2.833	1.744	1.138	1.138	1.143	1.132
1/1	100	22.607	4.091	1.582	1.494	0.947	1.100	0.744
	400	13.343	8.198	2.565	1.189	1.175	1.269	1.032
	1600	12.370	8.982	3.117	1.248	1.248	1.282	1.212
3/1	100	3.797	4.141	1.581	1.187	0.943	1.098	0.739
	400	3.492	8.307	2.590	1.177	1.166	1.258	1.022
	1600	3.567	9.162	3.114	1.272	1.268	1.299	1.217
Correct Regression Model and Wrong Propensity Score								
1/3	100	86.228	3.967	1.464	1.607	0.954	1.130	0.732
	400	17.213	9.495	2.449	1.284	1.257	1.382	1.043
	1600	16.949	11.756	3.269	1.371	1.364	1.409	1.289
1/1	100	6.558	4.089	1.550	1.126	1.088	1.229	0.946
	400	4.542	4.145	1.955	1.204	1.202	1.243	1.159
	1600	4.131	3.898	2.011	1.214	1.214	1.227	1.207
3/1	100	4.196	4.008	1.451	1.234	0.953	1.130	0.728
	400	4.053	9.513	2.466	1.295	1.263	1.388	1.048
	1600	4.161	11.784	3.268	1.380	1.375	1.428	1.301
Wrong Regression Model and Correct Propensity Score								
1/3	100	22.639	4.238	1.616	1.186	0.941	1.094	0.747
	400	13.383	8.556	2.639	1.155	1.145	1.233	1.008
	1600	12.878	9.547	3.261	1.209	1.206	1.235	1.164
1/1	100	4.748	3.597	1.561	1.061	1.044	1.141	0.946
	400	3.384	3.197	1.794	1.104	1.104	1.126	1.081
	1600	3.119	2.988	1.807	1.112	1.112	1.117	1.107
3/1	100	4.924	4.252	1.610	1.119	0.937	1.089	0.742
	400	4.067	8.527	2.666	1.168	1.156	1.245	1.014
	1600	4.103	9.552	3.229	1.211	1.210	1.240	1.170
Min		3.037	2.833	1.451	1.061	0.937	1.089	0.728
Max		86.228	11.784	3.268	1.607	1.375	1.428	1.301

Note: Average of the estimated variances from Tables B1, B3 and B5 (homogeneous treatment) are summarized relative to the average of the estimated variance of the regression estimator.

Table A2: Average variances relative to regression estimator (Heterogeneous Treatment)

Ratio	N	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2
				Both Models Correct				
1/3	100	12.069	2.424	1.412	1.242	0.958	1.086	0.788
	400	6.716	4.062	1.900	1.155	1.145	1.223	1.023
	1600	6.249	4.350	2.175	1.210	1.209	1.239	1.179
1/1	100	2.613	2.452	1.337	1.063	1.047	1.120	0.972
	400	2.034	2.257	1.480	1.099	1.099	1.117	1.080
	1600	1.909	2.164	1.494	1.107	1.107	1.110	1.103
3/1	100	3.273	3.573	1.420	1.119	0.954	1.081	0.783
	400	3.145	7.312	2.345	1.179	1.161	1.236	1.026
	1600	3.103	7.835	2.785	1.200	1.200	1.228	1.167
				Correct Regression Model and Wrong Propensity Score				
1/3	100	20.677	2.399	1.396	1.452	0.959	1.103	0.773
	400	8.586	4.659	1.940	1.251	1.225	1.338	1.034
	1600	8.144	5.352	2.293	1.305	1.301	1.344	1.248
1/1	100	3.310	2.841	1.364	1.099	1.066	1.170	0.959
	400	2.593	2.948	1.625	1.153	1.152	1.183	1.118
	1600	2.415	2.786	1.667	1.165	1.165	1.175	1.160
3/1	100	4.458	3.495	1.311	1.330	0.961	1.112	0.769
	400	3.612	8.637	2.263	1.242	1.218	1.328	1.035
	1600	3.769	10.289	2.904	1.312	1.304	1.336	1.238
				Wrong Regression Model and Correct Propensity Score				
1/3	100	12.951	2.479	1.292	1.216	0.936	1.082	0.754
	400	6.982	4.504	1.869	1.125	1.117	1.209	0.990
	1600	6.582	4.877	2.172	1.146	1.146	1.175	1.120
1/1	100	2.994	2.575	1.344	1.034	1.022	1.100	0.943
	400	2.295	2.379	1.511	1.067	1.067	1.085	1.048
	1600	2.128	2.231	1.521	1.072	1.072	1.076	1.068
3/1	100	4.151	3.738	1.474	1.100	0.946	1.079	0.772
	400	3.755	7.672	2.422	1.141	1.129	1.208	1.007
	1600	3.714	8.408	2.930	1.176	1.174	1.195	1.141
Min		1.909	2.164	1.292	1.034	0.936	1.076	0.754
Max		20.677	10.289	2.930	1.452	1.304	1.338	1.238

Note: Average of the estimated variances from Tables B2, B4 and B6 (Heterogeneous treatment) are summarized relative to the average of the estimated variance of the regression estimator.

A.2 Tables: Empirical Study

Table A3: Description of the Variables (NCDS)

Variable	**Description**
WHITE	=1 if Euro-Caucasian
MATH7	Mathematics test score at the age of 7
MATH7D5	=1 if math7 is in the top 20th quintile
MATH7D4	=1 if math7 is in the top 20-40th quintile
MATH7D3	=1 if math7 is in the top 40-60th quintile
MATH7D2	=1 if math7 in the top 60-80th quintile
MATH7D1	=1 if math7 in the bottom 20th quintile
MATH7MIS	=1 if math7 is missing
READ7	Reading comprehension test score at the age of 7
READ7D5	=1 if read7 is in the top 20th quintile
READ7D4	=1 if read7 is in the top 20-40th quintile
READ7D3	=1 if read7 is in the top 40-60th quintile
READ7D2	=1 if read7 in the top 60-80th quintile
READ7D1	=1 if read7 in the bottom 20th quintile
READ7MIS	=1 if read7 is missing
MATH11	Mathematics comprehension test score at the age of 11
MATH11D5	=1 if math11 is in the top 20th quintile
MATH11D4	=1 if math11 is in the top 20-40th quintile
MATH11D3	=1 if math11 is in the top 40-60th quintile
MATH11D2	=1 if math11 in the top 60-80th quintile
MATH11D1	=1 if math11 in the bottom 20th quintile
MATH11MISS	=1 if math11 is missing
READ11	Reading comprehension test score at the age of 11
READ11D5	=1 if read11 is in the top 20th quintile
READ11D4	=1 if read11 is in the top 20-40th quintile
READ11D3	=1 if read11 is in the top 40-60th quintile
READ11D2	=1 if read11 in the top 60-80th quintile
READ11D1	=1 if read11 in the bottom 20th quintile
READ11MIS	=1 if read11 is missing
COMPREHEN	=1 if Comprehensive school is attended in 1974
SECOND	=1 if Secondary modern school is attended in 1974
GRAMMAR	=1 if Grammar school is attended in 1974
PRIV	=1 if Private school is attended in 1974
OTHER	=1 if other type of school is attended in 1974

Note: Source: NCDS, own definitions

Table A3 ***(cont'd)*****:** Description of the Variables (NCDS)

Variable	**Description**
EDUFAT	Father's years of education
EDUFATM	=1 if edufat is missing
EDUMOT	Mother's years of education
EDUMOTM	=1 if edumot is missing
AGEF	Father's age when the child was 16
AGEFM	=1 if agef is missing
AGEM	Mother's age when the child was 16
AGEMM	=1 if agem is missing
FATSOCD1	=1 if father's social class is professional
FATSOCD2	=1 if father's social class is intermediate
FATSOCD3	=1 if father's social class is skilled non-manual
FATSOCD4	=1 if father's social class is skilled manual
FATSOCD5	=1 if father's social class is semi-skilled nonmanual
FATSOCD6	=1 if father's social class is semi-skilled manual
FATSOCD7	=1 if father's social class is unskilled
FATSOCD8	=1 if father's social class is missing, or father is unemployed, or no father
MOTEMP	=1 if the mother is employed
FATINTD1	=1 if father expects to much
FATINTD2	=1 if father is very interested
FATINTD3	=1 if father shows some interest
MOTINTD1	=1 if mother expects to much
MOTINTD2	=1 if mother is very interested
MOTINTD3	=1 if mother shows some interest
BADFIN	=1 if the family experienced financial problems in 1969 or 1974
DEGD1	=1 if the family lived in North in 1974
DEGD2	=1 if the family lived in North West in 1974
DEGD3	=1 if the family lived in North in 1974
DEGD4	=1 if the family lived in East and West Riding in 1974
DEGD5	=1 if the family lived in North Midlands in 1974
DEGD6	=1 if the family lived in Midlands in 1974
DEGD7	=1 if the family lived in East in 1974
DEGD8	=1 if the family lived in London and South East in 1974
DEGD9	=1 if the family lived in South in 1974
DEGD10	=1 if the family lived in Wales in 1974
DEGD11	=1 if the family lived in Scotland in 1974
DEGD12	=1 if other region
NOSIB	number of siblings
NOSIBM	=1 if nosib is missing

Note: Source: NCDS, own definitions

Table A4: Summary statistics for entire sample and by treatment status (NCDS)

		Entire Sample		**By Treatment Status**				
				D=1		**D=0**		
		Mean	**Std. Dev**	**Mean**	**Std. Dev**	**Mean**	**Std. Dev**	**p-value**
Dependent Variable	log(hourly wage) 1991	1.633	0.456	1.865	0.406	1.528	0.439	0.00
Treatment Variable	HE	0.312	0.463	1	0	0	0	
Covariates	White	0.978	0.148	0.981	0.136	0.976	0.153	0.33
Mathematics ability	5th quintile(highest)	0.195	0.396	0.302	0.459	0.146	0.354	0.00
at 7 years	4th quintile	0.106	0.308	0.133	0.340	0.094	0.292	0.00
	3rd quintile	0.254	0.436	0.280	0.449	0.243	0.429	0.01
	2nd quintile	0.121	0.326	0.086	0.281	0.137	0.344	0.00
	1st quintile(lowest)	0.213	0.410	0.088	0.283	0.270	0.444	0.00
	Missing	0.110	0.313	0.110	0.313	0.109	0.312	0.95
Reading ability	5th quintile (highest)	0.151	0.358	0.251	0.434	0.106	0.308	0.00
at 7 years	4th quintile	0.129	0.336	0.182	0.386	0.106	0.308	0.00
	3rd quintile	0.241	0.428	0.263	0.440	0.231	0.421	0.03
	2nd quintile	0.187	0.390	0.142	0.349	0.207	0.406	0.00
	1st quintile(lowest)	0.186	0.389	0.053	0.225	0.247	0.431	0.00
	Missing	0.105	0.307	0.108	0.311	0.104	0.305	0.67
Mathematics ability	5th quintile (highest)	0.211	0.408	0.431	0.495	0.111	0.315	0.00
at 11 years	4th quintile	0.186	0.389	0.229	0.421	0.166	0.373	0.00
	3rd quintile	0.171	0.377	0.129	0.335	0.190	0.392	0.00
	2nd quintile	0.164	0.370	0.062	0.241	0.210	0.407	0.00
	1st quintile(lowest)	0.133	0.340	0.018	0.133	0.185	0.389	0.00
	Missing	0.135	0.342	0.131	0.338	0.137	0.344	0.66
Reading ability	5th quintile (highest)	0.209	0.407	0.403	0.491	0.122	0.327	0.00
at 11 years	4th quintile	0.134	0.340	0.179	0.384	0.113	0.317	0.00
	3rd quintile	0.218	0.413	0.186	0.389	0.233	0.423	0.00
	2nd quintile	0.292	0.455	0.186	0.389	0.340	0.474	0.00
	1st quintile(lowest)	0.175	0.380	0.039	0.195	0.236	0.425	0.00
	Missing	0.135	0.342	0.131	0.338	0.137	0.344	0.66
School attended 1974	Comprehensive school	0.470	0.499	0.411	0.492	0.496	0.500	0.00
	Secondary modern school	0.159	0.366	0.114	0.318	0.180	0.384	0.00
	Grammar school	0.094	0.292	0.176	0.381	0.057	0.232	0.00
	Private school	0.050	0.218	0.104	0.306	0.025	0.157	0.00
	Other school	0.009	0.096	0.003	0.057	0.012	0.109	0.01
	Missing school information	0.096	0.294	0.085	0.280	0.100	0.300	0.15
Family background	Father's years of education	7.137	4.692	7.866	5.003	6.806	4.507	0.00
	Father's education missing	0.282	0.450	0.260	0.439	0.292	0.455	0.04
	Mother's years of education	7.252	4.551	7.830	4.707	6.990	4.454	0.00
	Mother's education missing	0.269	0.444	0.247	0.431	0.279	0.449	0.03
	Father's age in 1974	43.215	13.708	43.324	13.677	43.166	13.725	0.74
	Father's age missing	0.074	0.262	0.076	0.264	0.074	0.261	0.84
	Mother's age in 1974	41.523	10.851	41.328	11.502	41.612	10.543	0.45
	Mother's age missing	0.048	0.215	0.058	0.234	0.044	0.205	0.05
	Motheremployedin1974	0.508	0.500	0.520	0.500	0.502	0.500	0.30
	Number of siblings	1.686	1.782	1.482	1.479	1.779	1.897	0.00
	Number of siblings missing	0.268	0.443	0.242	0.429	0.280	0.449	0.01
Father's social class	Professional	0.042	0.200	0.093	0.290	0.018	0.134	0.00
in 1974	Intermediate	0.141	0.348	0.214	0.411	0.108	0.310	0.00
	Skilled non-manual	0.073	0.260	0.085	0.278	0.068	0.251	0.06
	Skilled manual	0.296	0.457	0.242	0.428	0.321	0.467	0.00
	Semi-skilled non-manual	0.010	0.098	0.003	0.057	0.013	0.112	0.01
	Semi-skilled manual	0.096	0.294	0.057	0.231	0.114	0.317	0.00
	Unskilled	0.029	0.169	0.017	0.130	0.035	0.184	0.00
	Missing,or unemployed	0.313	0.464	0.289	0.454	0.324	0.468	0.03
Father's interest	Expects too much	0.012	0.110	0.022	0.147	0.008	0.088	0.00
in education	Very interested	0.242	0.428	0.353	0.478	0.191	0.393	0.00
	Some interest	0.220	0.414	0.210	0.408	0.224	0.417	0.35
Mother's interest	Expects too much	0.032	0.175	0.041	0.199	0.028	0.164	0.03
in education	Very interested	0.332	0.471	0.462	0.499	0.274	0.446	0.00
	Some interest	0.361	0.480	0.298	0.458	0.390	0.488	0.00
	Badfinancesin1969or1974	0.167	0.373	0.089	0.284	0.202	0.402	0.00
Region in 1974	North West	0.102	0.303	0.108	0.311	0.099	0.299	0.39
	North	0.074	0.262	0.070	0.255	0.076	0.265	0.50
	East and West Riding	0.084	0.277	0.073	0.260	0.089	0.284	0.11
	North Midlands	0.073	0.260	0.071	0.256	0.074	0.261	0.73
	East	0.075	0.263	0.084	0.277	0.070	0.256	0.14
	London and South East	0.148	0.356	0.149	0.356	0.148	0.355	0.97
	South	0.060	0.237	0.075	0.263	0.053	0.224	0.01
	South West	0.060	0.237	0.071	0.258	0.054	0.227	0.04
	Midlands	0.091	0.288	0.082	0.275	0.095	0.294	0.19
	Wales	0.053	0.224	0.048	0.213	0.055	0.228	0.33
	Scotland	0.095	0.293	0.088	0.283	0.098	0.297	0.34
	Other	0.086	0.281	0.081	0.273	0.088	0.284	0.48
	No. of Obs.	3,902		1,210		2,685		

Note: Source: NCDS, own definitions. p-value for t-test of mean equality by treatment status.

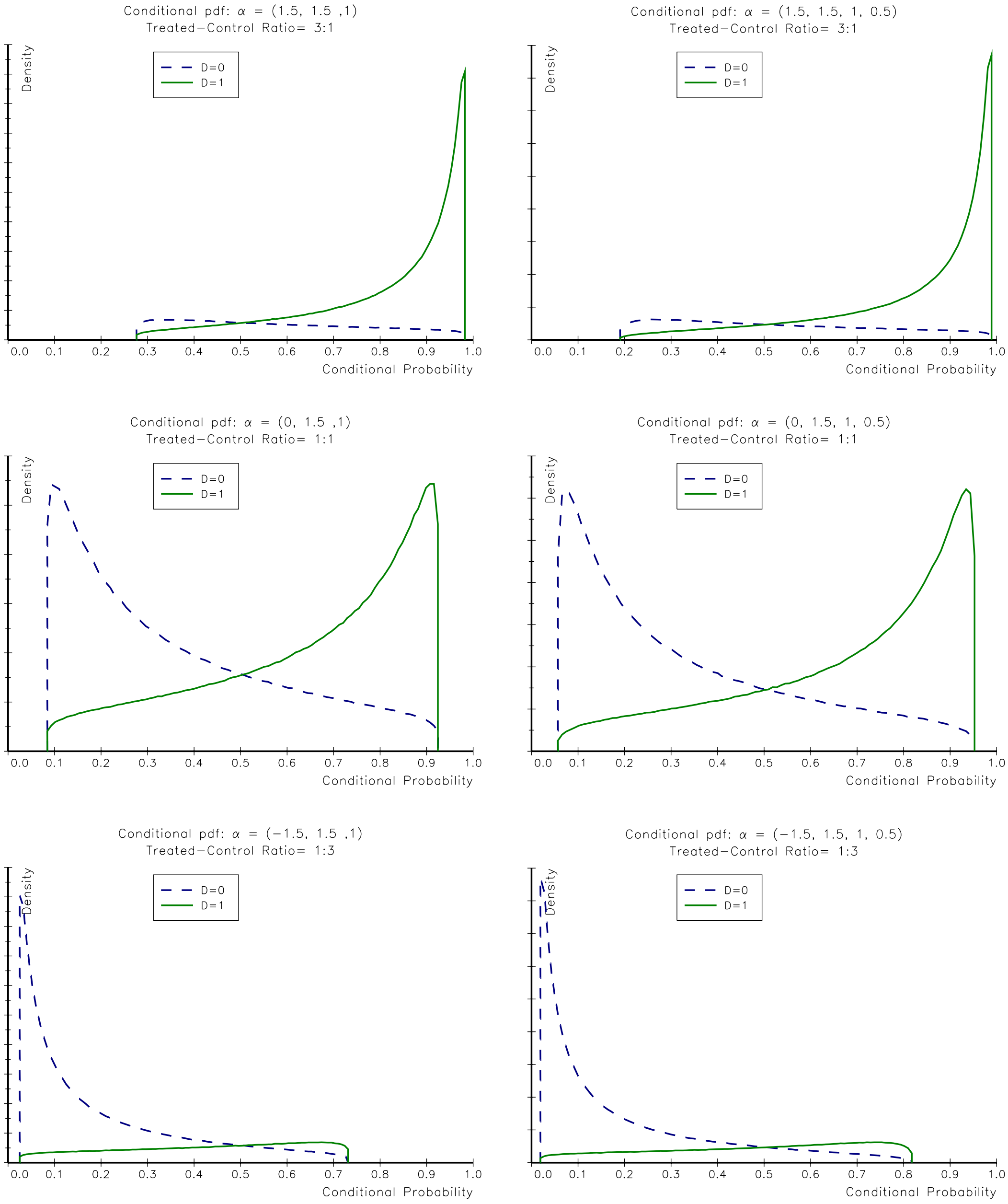
Table A5: Propensity Score Estimation Results (Higher Education)

Variable	**Coeff.**	**(Std. Err.)**	**Variable**	**Coeff.**	**(Std. Err.)**
WHITE	-0.243	(0.307)	AGEM	0.002	(0.012)
MATH7D5	0.334	(0.400)	AGEMM	$1.018^{\dagger}$	(0.608)
MATH7D4	0.373	(0.406)	FATSOCD1	$0.692^{\dagger}$	(0.357)
MATH7D3	0.360	(0.398)	FATSOCD2	0.274	(0.297)
MATH7D2	0.166	(0.412)	FATSOCD3	0.049	(0.309)
MATH7D1	-0.062	(0.408)	FATSOCD4	0.110	(0.281)
READ7D4	-0.070	(0.138)	FATSOCD5	-0.937	(0.625)
READ7D3	-0.069	(0.127)	FATSOCD6	-0.138	(0.306)
READ7D2	0.086	(0.151)	FATSOCD8	0.136	(0.322)
READ7D1	-0.443^{*}	(0.191)	MOTEMP	-0.147	(0.103)
READ7MIS	0.381	(0.410)	FATINTD1	0.866^{*}	(0.409)
MATH11D5	0.890^{**}	(0.271)	FATINTD2	-0.031	(0.137)
MATH11D4	0.238	(0.263)	FATINTD3	0.093	(0.117)
MATH11D3	-0.134	(0.266)	MOTINTD1	0.178	(0.279)
MATH11D2	-0.589^{*}	(0.277)	MOTINTD2	$0.314^{\dagger}$	(0.161)
MATH11D1	-1.297^{**}	(0.341)	MOTINTD3	0.079	(0.142)
READ11D5	$0.463^{\dagger}$	(0.249)	BADFIN	-0.380^{**}	(0.137)
READ11D4	0.118	(0.250)	DEGD1	0.181	(0.233)
READ11D3	0.149	(0.169)	DEGD2	0.359	(0.220)
READ11D2	-0.421^{*}	(0.185)	DEGD3	0.126	(0.230)
READ11D1	-0.581^{*}	(0.284)	DEGD4	0.208	(0.236)
COMPRH	0.044	(0.133)	DEGD5	0.085	(0.228)
SECOND	0.070	(0.158)	DEGD6	0.297	(0.232)
GRAMMER	0.274	(0.168)	DEGD7	-0.001	(0.210)
PRIV	0.545^{*}	(0.212)	DEGD8	0.276	(0.243)
OTHER	-0.832	(0.586)	DEGD9	0.280	(0.245)
EDUFAT	0.124^{**}	(0.037)	DEGD11	0.052	(0.222)
EDUFATM	1.543^{**}	(0.474)	DEGD12	0.400	(0.265)
EDUMOT	0.049	(0.042)	NOSIB	-0.040	(0.033)
EDUMOTM	0.281	(0.523)	NOSIBM	-0.434	(0.273)
AGEF	0.004	(0.011)	Intercept	-3.123^{**}	(0.835)
AGEFM	-0.430	(0.561)			
No. of Obs.	3,902				
Log-likelihood	-1,886.08				
LR chi2(k)	1,071.05				

Note: Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

A.3 Figures

Figure A1: Overlap Plots with uniformly distributed Xs


Conditional pdf: α = (1.5, 1.5 ,1)
Treated−Control Ratio= 3:1

Conditional pdf: α = (1.5, 1.5, 1, 0.5)
Treated−Control Ratio= 3:1

Conditional pdf: α = (0, 1.5 ,1)
Treated−Control Ratio= 1:1

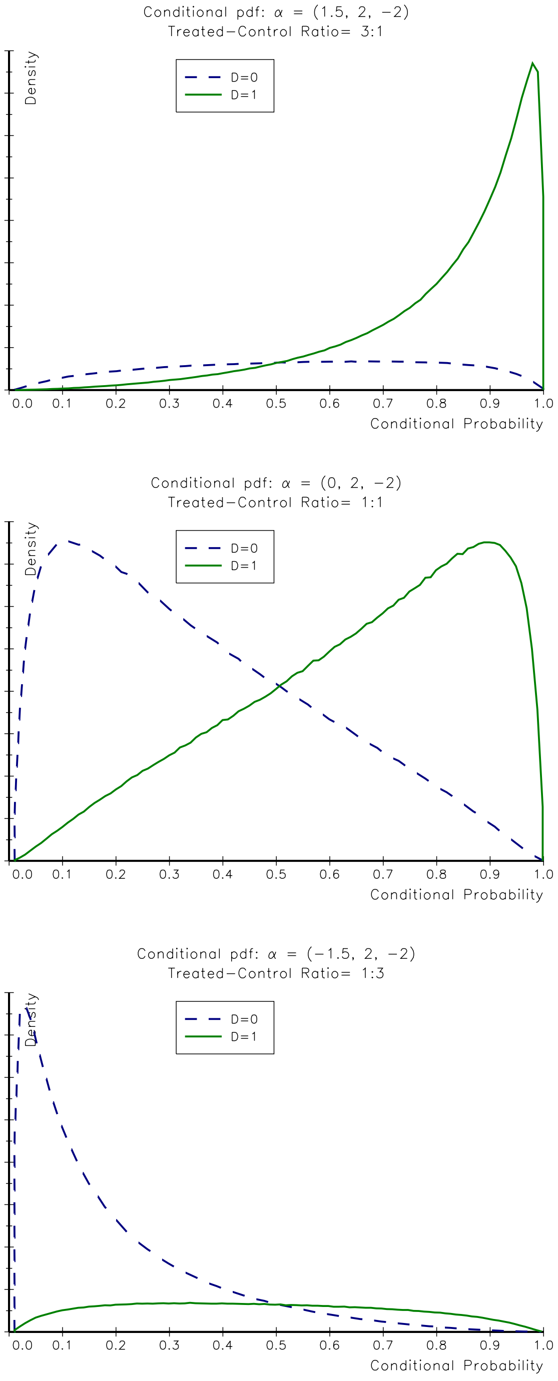
Conditional pdf: α = (0, 1.5, 1, 0.5)
Treated−Control Ratio= 1:1

Conditional pdf: α = (−1.5, 1.5 ,1)
Treated−Control Ratio= 1:3

Conditional pdf: α = (−1.5, 1.5, 1, 0.5)
Treated−Control Ratio= 1:3

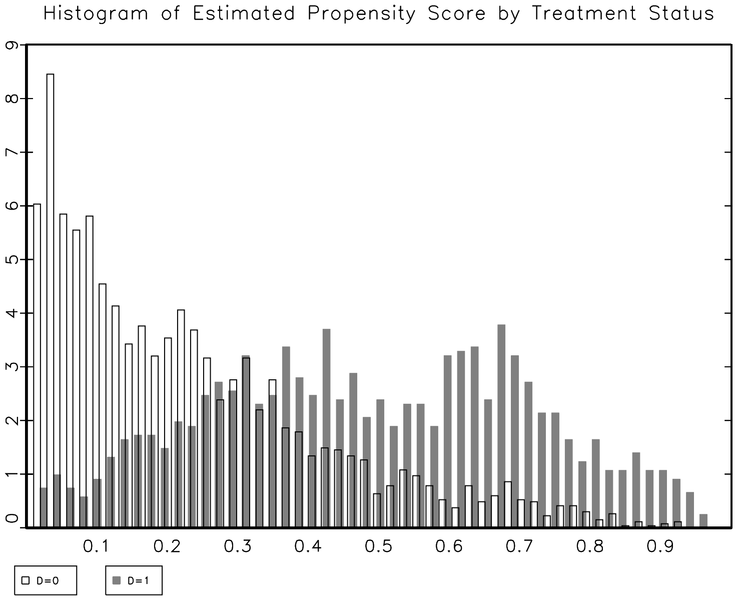
Note: The graphs display estimated densities of conditional probabilities for treated (D=1, solid line) and control (D=0, dashed line) groups where Xs are drawn from a uniform distribution. Each row corresponds to different treated-control ratio. The graphs on the left-hand side are the distribution of the propensity score where the treatment indicator is generated by DGP1 (without X_3) and the graphs on the right-hand side are the distribution of the propensity score where treatment indicator is generated by DGP2 (with X_3). See Table 3 for parameter configurations and Table 1 for DGPs.

Figure A2: Overlap Plots with normally distributed Xs



Note: The graphs display estimated densities of conditional probabilities for treated (D=1, solid line) and control (D=0, dashed line) groups where Xs are drawn from a normal distribution. Each row corresponds to different treated-control ratio. The treatment indicator is generated by DGP1 and parameters in Table 3 are adjusted such that the desired treated-control ratio is generated with normally distributed Xs. Parameter values are given above the graphs.

Figure A3: Histogram of Estimated Propensity Score by Treatment Status (Higher Education or less).



Note: Histogram of the estimated probability of getting higher education by treatment status. Estimation results are reported in Table A5. The empty bars are the estimated probabilities of the individuals who have higher education and the light-gray filled bars are the estimated probabilities of the individuals who have less than higher education

B Web Appendix

B.1 Supplementary Proofs

Regression Estimator

The explicit forms of A_{reg} and V_{reg} in (4) are as follows:

$$\begin{aligned}
A_{reg} &\equiv E\left[\frac{\partial\psi(Z_i,\theta)}{\partial\theta'}\right]\\
&= E\begin{bmatrix} \frac{\partial\psi_1(Z_i,\theta_{reg})}{\partial\beta_1'} & \frac{\partial\psi_1(Z_i,\theta_{reg})}{\partial\beta_0'} & \frac{\partial\psi_1(Z_i,\theta_{reg})}{\partial\tau} \\ \frac{\partial\psi_2(Z_i,\theta_{reg})}{\partial\beta_1'} & \frac{\partial\psi_2(Z_i,\theta_{reg})}{\partial\beta_0'} & \frac{\partial\psi_2(Z_i,\theta_{reg})}{\partial\tau} \\ \frac{\partial\psi_3(Z_i,\theta_{reg})}{\partial\beta_1'} & \frac{\partial\psi_3(Z_i,\theta_{reg})}{\partial\beta_0'} & \frac{\partial\psi_3(Z_i,\theta_{reg})}{\partial\tau} \end{bmatrix} = E\begin{bmatrix} \frac{\partial\psi_1(Z_i,\theta_{reg})}{\partial\beta_1'} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\partial\psi_2(Z_i,\theta_{reg})}{\partial\beta_0'} & \mathbf{0} \\ \frac{\partial\psi_3(Z_i,\theta_{reg})}{\partial\beta_1'} & \frac{\partial\psi_3(Z_i,\theta_{reg})}{\partial\beta_0'} & -1 \end{bmatrix}\\
V_{reg} &\equiv V\left[\psi(Z_i,\theta)\right] = E\left[\psi(Z_i,\theta)\psi(Z_i,\theta)'\right]\\
&= \mathrm{E}\left[\begin{pmatrix} \psi_1(Z_i,\theta_{reg}) \\ \psi_2(Z_i,\theta_{reg}) \\ \psi_3(Z_i,\theta_{reg}) \end{pmatrix}\begin{pmatrix} \psi_1(Z_i,\theta_{reg})' & \psi_2(Z_i,\theta_{reg})' & \psi_3(Z_i,\theta_{reg})' \end{pmatrix}\right].
\end{aligned}$$

Hence, depending on the regression model chosen for the outcome model A_{reg} and V_{reg} can be derived. To estimate the variance-covariance matrix, we can replace the expectations with the sample means and true parameter vector with its estimate, in the following way

$$\begin{aligned}
\hat{A}_{reg} &= \frac{1}{N}\sum_i \frac{\partial\psi(Z_i,\hat{\theta}_{reg})}{\partial\theta'}\\
&= \frac{1}{N}\sum_i \begin{pmatrix} \frac{\partial\psi_1(Z_i,\hat{\theta}_{reg})}{\partial\beta_1'} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\partial\psi_2(Z_i,\hat{\theta}_{reg})}{\partial\beta_0'} & \mathbf{0} \\ \frac{\partial\psi_3(Z_i,\hat{\theta}_{reg})}{\partial\beta_1'} & \frac{\partial\psi_3(Z_i,\hat{\theta}_{reg})}{\partial\beta_0'} & -1 \end{pmatrix}\\
\hat{V}_{reg} &= \frac{1}{N}\sum_i \psi(Z_i,\hat{\theta}_{reg})\psi(Z_i,\hat{\theta}_{reg})'.
\end{aligned}$$

Asymptotic Variance of Regression Estimator

To verify the asymptotic distribution of $\hat{\tau}_{reg}$ given in Equation (5), I use the first order Taylor series approximation of $\sqrt{N}(\tau(\hat{\beta}_1, \hat{\beta}_0) - \tau)$ around the true values (β_1, β_0).

$$
\begin{align}
\sqrt{N}\left(\tau(\hat{\beta}_1, \hat{\beta}_0) - \tau\right) &\approx \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\eta[X_i'\beta_1] - \eta[X_i'\beta_0] - \tau\right)\right) \notag \\
&+ \sqrt{N}(\hat{\beta}_1 - \beta_1)\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\eta[X_i'\beta_1]}{\partial\beta_1'} \tag{W.1} \\
&+ \sqrt{N}(\hat{\beta}_0 - \beta_0)\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\eta[X_i'\beta_0]}{\partial\beta_0'}. \tag{W.2}
\end{align}
$$

The asymptotic distributions of the single terms in Equation (W.1) are easy to verify. The first term has the following asymptotic distribution

$$
\begin{align}
\sqrt{N}\quad \left(\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\left(\eta[X_i'\beta_1] - \eta[X_i'\beta_0] - \tau\right)\right) &\xrightarrow{d} N\left(0, \mathrm{V}\left[\eta[X_i'\beta_1] - \eta[X_i'\beta_0] - \tau\right]\right) \notag \\
&\xrightarrow{d} N\left(0, \mathrm{E}\left[\left(\eta[X_i'\beta_1] - \eta[X_i'\beta_0] - \tau\right)^2\right]\right) \tag{W.3}
\end{align}
$$

by Central Limit Theorem. For $\hat{\beta}_1$ and $\hat{\beta}_0$ with $\sqrt{N}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N\left(0, \mathrm{AV}_{\hat{\beta}_1}\right)$ and $\sqrt{N}(\hat{\beta}_0 - \beta_0) \xrightarrow{d} N\left(0, \mathrm{AV}_{\hat{\beta}_0}\right)$ the asymptotic distributions of the second and third terms are as follows

$$
\begin{align}
\sqrt{N}(\hat{\beta}_1 - \beta_1)\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\eta[X_i'\beta_1]}{\partial\beta_1'} &\xrightarrow{d} N\left(0, \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_1]}{\partial\beta_1'}\right]\mathrm{AV}_{\hat{\beta}_1}\,\mathrm{E}\left[\frac{\partial\eta[X_i'\beta_1]}{\partial\beta_1'}\right]'\right) \tag{W.4} \\
\sqrt{N}(\hat{\beta}_0 - \beta_0)\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\eta[X_i'\beta_0]}{\partial\beta_0'} &\xrightarrow{d} N\left(0, \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_0]}{\partial\beta_0'}\right]\mathrm{AV}_{\hat{\beta}_0}\,\mathrm{E}\left[\frac{\partial\eta[X_i'\beta_0]}{\partial\beta_0'}\right]'\right). \tag{W.5}
\end{align}
$$

by the weak law of large numbers ($\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\eta[X_i'\beta_d]}{\partial\beta_d} \xrightarrow{p} \mathrm{E}\left[\frac{\partial\eta[X_i'\beta_d]}{\partial\beta_d}\right]$) and Slutsky's theorem. Because the covariance between $\hat{\beta}_1$ and $\hat{\beta}_0$ is equal to zero[16], combining single asymptotic distributions leads to Equation (5). $\mathrm{AV}_{\hat{\tau},reg}$ can be estimated by replacing unknown parameter with their estimates and the expectations by sample averages; that

[16] The reason is that $\hat{\beta}_1$ and $\hat{\beta}_0$ are estimated using different subsamples.

is,

$$\begin{aligned}
\hat{\mathrm{V}}\left[\hat{\tau}_{reg}\right] &= \frac{1}{N}\left[\frac{1}{N}\sum_i \left(\eta[X_i'\hat{\beta}_1] - \eta[X_i'\hat{\beta}_0] - \hat{\tau}\right)^2\right. \\
&\quad + \left(\frac{1}{N}\sum_i \frac{\partial\eta[X_i'\hat{\beta}_1]}{\partial\beta_1'}\right)\hat{\mathrm{V}}\left[\hat{\beta}_1\right]\left(\frac{1}{N}\sum_i \frac{\partial\eta[X_i'\hat{\beta}_1]}{\partial\beta_1'}\right)' \\
&\quad \left. + \left(\frac{1}{N}\sum_i \frac{\partial\eta[X_i'\hat{\beta}_0]}{\partial\beta_0'}\right)\hat{\mathrm{V}}\left[\hat{\beta}_0\right]\left(\frac{1}{N}\sum_i \frac{\partial\eta[X_i'\hat{\beta}_0]}{\partial\beta_0'}\right)'\right],
\end{aligned}$$

where $\hat{\mathrm{V}}\left[\hat{\beta}_1\right]$ and $\hat{\mathrm{V}}\left[\hat{\beta}_0\right]$ are the estimated variance-covariance matrices of $\hat{\beta}_1$ and $\hat{\beta}_0$.

Weighting Estimators

Asymptotic Variance of IPW1

The first type of weighting estimator, IPW1, is estimated based on the set of moment conditions given in Equation (2.2). To prove the asymptotic distribution with estimated propensity score, consider the M-estimation framework. By standard results on M-estimators the asymptotic variance of θ_{ps1} is given by $A_{ps1}^{-1}V_{ps1}A_{ps1}^{-1}{}'$ with A_{ps1} and V_{ps1}, as given below.

$$\begin{aligned}
A_{ps1} &\equiv E\left[\frac{\partial\psi(Z_i,\theta_{ps1})}{\partial\theta_{ps1}'}\right] \\
&= E\begin{bmatrix}
\frac{\partial\psi_1(Z_i,\theta_{ps})}{\partial\alpha'} & \frac{\partial\psi_1(Z_i,\theta_{ps1})}{\partial\mu_1} & \frac{\partial\psi_1(Z_i,\theta_{ps1})}{\partial\mu_0} \\
\frac{\partial\psi_2(Z_i,\theta_{ps})}{\partial\alpha'} & \frac{\partial\psi_2(Z_i,\theta_{ps1})}{\partial\mu_1} & \frac{\partial\psi_2(Z_i,\theta_{ps1})}{\partial\mu_0} \\
\frac{\partial\psi_3(Z_i,\theta_{ps})}{\partial\alpha'} & \frac{\partial\psi_3(Z_i,\theta_{ps1})}{\partial\mu_1} & \frac{\partial\psi_3(Z_i,\theta_{ps1})}{\partial\mu_0}
\end{bmatrix} = E\begin{bmatrix}
H(Z_i,\alpha) & \mathbf{0} & \mathbf{0} \\
-\frac{D_iY_i}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'} & -1 & 0 \\
\frac{(1-D_i)Y_i}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'} & 0 & -1
\end{bmatrix} \\
&= \begin{bmatrix}
\mathrm{E}\left[H(Z_i,\alpha)\right] & \mathbf{0} & \mathbf{0} \\
-\mathrm{E}\left[\frac{D_iY_i}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] & -1 & 0 \\
\mathrm{E}\left[\frac{(1-D_i)Y_i}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] & 0 & -1
\end{bmatrix} = \begin{bmatrix}
E_H & \mathbf{0} & \mathbf{0} \\
E_{11} & -1 & 0 \\
E_{10} & 0 & -1
\end{bmatrix}
\end{aligned}$$

where $H(Z_i,\alpha)$ stands for the Hessian. Furthermore, $E_H \equiv \mathrm{E}\left[H(Z_i,\alpha)\right]$,

$$E_{11} \equiv -\mathrm{E}\left[\frac{D_iY_i}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = -\mathrm{E}\left[\mathrm{E}\left[\frac{D_iY_i}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\middle| X_i\right]\right]$$

$$= \quad -\mathrm{E}\left[\frac{\mathrm{E}\left[D_i|\, X_i\right] Y_{1i}}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = -\,\mathrm{E}\left[\frac{Y_{1i}}{\pi[X_i'\alpha]}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]$$

and

$$\begin{aligned} E_{10} &\equiv \mathrm{E}\left[\frac{(1-D_i)Y_i}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = \mathrm{E}\left[\mathrm{E}\left[\frac{(1-D_i)Y_i}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right| X_i\right]\right] \\ &= \mathrm{E}\left[\frac{(1-\mathrm{E}\left[D_i|\, X_i\right])Y_{0i}}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = \mathrm{E}\left[\frac{Y_{0i}}{1-\pi[X_i'\alpha]}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]. \end{aligned}$$

Here, we used the law of iterated expectations (LIE), CIA and the fact that $\mathrm{E}\left[D_i|\, X_i\right] = \Pr\left[D_i = 1\,|X_i\right] = \pi[X_i'\alpha]$. Using the matrix inversion rule for block form matrices[17] the inverse of A_{ps1} can be derived:

$$A_{ps1}^{-1} \quad = \quad \begin{bmatrix} E_H^{-1} & \mathbf{0} & \mathbf{0} \\ E_{11}E_H^{-1} & -1 & 0 \\ E_{10}E_H^{-1} & 0 & -1 \end{bmatrix}$$

The matrix V_{ps1} is as follows

$$\begin{aligned} V_{ps1} &\equiv V\left[\psi(Z_i,\theta_{ps1})\right] = E\left[\psi(Z_i,\theta_{ps1})\psi(Z_i,\theta_{ps1})'\right] \\ &= \mathrm{E}\left[\begin{pmatrix} \psi_1(Z_i,\theta_{ps1}) \\ \psi_2(Z_i,\theta_{ps1}) \\ \psi_3(Z_i,\theta_{ps1}) \end{pmatrix}\begin{pmatrix} \psi_1(Z_i,\theta_{ps1})' & \psi_2(Z_i,\theta_{ps1})' & \psi_3(Z_i,\theta_{ps1})' \end{pmatrix}\right] \\ &= \begin{bmatrix} \mathrm{E}\left[\psi_1(Z_i,\theta_{ps1})\psi_1(Z_i,\theta_{ps1})'\right] & \mathrm{E}\left[\psi_1(Z_i,\theta_{ps1})\psi_2(Z_i,\theta_{ps1})'\right] & \mathrm{E}\left[\psi_1(Z_i,\theta_{ps1})\psi_3(Z_i,\theta_{ps1})'\right] \\ \mathrm{E}\left[\psi_2(Z_i,\theta_{ps1})\psi_1(Z_i,\theta_{ps1})'\right] & \mathrm{E}\left[\psi_2(Z_i,\theta_{ps1})\psi_2(Z_i,\theta_{ps1})'\right] & \mathrm{E}\left[\psi_2(Z_i,\theta_{ps1})\psi_3(Z_i,\theta_{ps1})'\right] \\ \mathrm{E}\left[\psi_3(Z_i,\theta_{ps1})\psi_1(Z_i,\theta_{ps1})'\right] & \mathrm{E}\left[\psi_3(Z_i,\theta_{ps1})\psi_2(Z_i,\theta_{ps1})'\right] & \mathrm{E}\left[\psi_3(Z_i,\theta_{ps1})\psi_3(Z_i,\theta_{ps1})'\right] \end{bmatrix} \end{aligned}$$

[17] Let a matrix be partitioned into a block form:

$$\mathbf{M}_{(m+n)\times(m+n)} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

with the blocks $\mathbf{A}_{m\times m}$ and $\mathbf{D}_{n\times n}$ are invertible. Then,

$$\mathbf{M}^{-1} = \begin{bmatrix} \left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix}.$$

$$= \begin{bmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi_{22} & \psi_{23} \\ \psi_{31} & \psi_{32} & \psi_{33} \end{bmatrix}.$$

Remember that $A_{ps1}^{-1} V_{ps1} A_{ps1}^{-1}{}'$ is the asymptotic variance-covariance matrix of $\sqrt{N}\,(\theta_{ps1} - \theta)$, which is easy to derive by multiplying the given matrices. However, because the goal here is to derive the asymptotic variance of $\sqrt{N}\,(\hat{\tau}_{ps1} - \tau)$, I present only the 2×2 submatrix of $A_{ps1}^{-1} V_{ps1} A_{ps1}^{-1}{}'$, which corresponds to the asymptotic variance of $(\hat{\mu}_{1,ps1}, \hat{\mu}_{0,ps1})$. The asymptotic distribution of $(\hat{\mu}_{1,ps1}, \hat{\mu}_{0,ps1})$ is given by[18]:

$$\sqrt{N}\begin{pmatrix} \hat{\mu}_{1,ps1} - \mu_1 \\ \hat{\mu}_{0,ps1} - \mu_0 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \right. \tag{W.6}$$

$$\left.\begin{pmatrix} E_{11}E_H^{-1}\psi_{11}E_H^{-1}E_{11}^{'} - 2E_{11}E_H^{-1}\psi_{12} + \psi_{22} & E_{11}E_H^{-1}\psi_{11}E_H^{-1}E_{10}^{'} - \psi_{12}E_H^{-1}E_{10}^{'} - \psi_{13}E_H^{-1}E_{11}^{'} + \psi_{23} \\ (*) & E_{10}E_H^{-1}\psi_{11}E_H^{-1}E_{10}^{'} - 2E_{10}E_H^{-1}\psi_{13} + \psi_{33} \end{pmatrix}\right).$$

Now, Equation (1) can be used together with the asymptotic distribution of $(\hat{\mu}_{1,ps1}, \hat{\mu}_{0,ps1})$ to derive the asymptotic distribution of $\hat{\tau}_{ps1}$. A couple of equalities should be noted before proceeding. First, by information equality $-E_H = \psi_{11}$ because E_H is the Hessian and ψ_{11} is the score function of α. Second, it can be shown that $\psi_{21} = -E_{11}$ and $\psi_{31} = -E_{10}$. The following proves the first equality:

$$\begin{aligned}
\psi_{21} &\equiv \mathrm{E}\left[\psi_2\psi_1'\right] = \mathrm{E}\left[\left(\frac{D_iY_i}{\pi[X_i'\alpha]} - \mu_1\right)\frac{(D_i - \pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\left(\frac{D_iY_i - D_i\pi[X_i'\alpha]\mu_1 - \pi[X_i'\alpha]D_iY_i + \pi[X_i'\alpha]^2\mu_1}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\frac{D_iY_i - D_i\pi[X_i'\alpha]\mu_1 - \pi[X_i'\alpha]D_iY_i + \pi[X_i'\alpha]^2\mu_1}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\middle| X_i\right]\right] \\
&= \mathrm{E}\left[\left(\frac{\mathrm{E}\left[D_i|\,X_i\right]Y_{1i} - \mathrm{E}\left[D_i|\,X_i\right]\pi[X_i'\alpha]\mu_1 - \pi[X_i'\alpha]\,\mathrm{E}\left[D_i|\,X_i\right]Y_{1i} + \pi[X_i'\alpha]^2\mu_1}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\frac{Y_{1i}}{\pi[X_i'\alpha]}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = -E_{11} \quad \square
\end{aligned}$$

[18] $(*)$ is used due to the space constraint. Because the matrix is symmetric, it is given by the upper right element of the variance-covariance matrix.

We make use of LIE, CIA and the fact that $\mathrm{E}\left[D_i|\, X_i\right] = \Pr\left[D_i = 1\,|X_i\right] = \pi[X_i'\alpha]$. Similarly, the second equality $\psi_{13} = -E_{10}$ can be shown as follows

$$
\begin{aligned}
&\psi_{31} \equiv \mathrm{E}\left[\psi_3\psi_1'\right] = \mathrm{E}\left[\left(\frac{(1-D_i)Y_i}{1-\pi[X_i'\alpha]} - \mu_1\right)\frac{(D_i - \pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\left(\frac{D_i(1-D_i)Y_i - D_i(1-\pi[X_i'\alpha])\mu_0 - \pi[X_i'\alpha](1-D_i)Y_i + \pi[X_i'\alpha](1-\pi[X_i'\alpha])\mu_0}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])^2}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\frac{-D_i(1-\pi[X_i'\alpha])\mu_0 - \pi[X_i'\alpha](1-D_i)Y_i + \pi[X_i'\alpha](1-\pi[X_i'\alpha])\mu_0}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])^2}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right| X_i\right]\right] \\
&= \mathrm{E}\left[\left(\frac{-\mathrm{E}\left[D_i|\, X_i\right](1-\pi[X_i'\alpha])\mu_0 - \pi[X_i'\alpha](1-\mathrm{E}\left[D_i|\, X_i\right])Y_{0i} + \pi[X_i'\alpha](1-\pi[X_i'\alpha])\mu_0}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])^2}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[-\frac{Y_{0i}}{1-\pi[X_i'\alpha]}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = -E_{10}. \quad \square
\end{aligned}
$$

Here, additionally I use the fact that $D_i(1-D_i) = 0$. By Equation (1), the asymptotic variance of $\hat{\tau}_{ps1}$ is given by

$$
\begin{aligned}
\mathrm{AV}\left[\hat{\tau}_{ps1}\right] &= \mathrm{AV}[\hat{\mu}_{1,ps1}] + \mathrm{AV}[\hat{\mu}_{0,ps1}] - 2\mathrm{ACov}[\hat{\mu}_{1,ps1}, \hat{\mu}_{0,ps1}] \\
&= E_{11}E_H^{-1}\psi_{11}E_H^{-1}E_{11}^{'} - 2E_{11}E_H^{-1}\psi_{12} + \psi_{22} \\
&\quad +E_{10}E_H^{-1}\psi_{11}E_H^{-1}E_{10}^{'} - 2E_{10}E_H^{-1}\psi_{13} + \psi_{33} \\
&\quad -2\left(E_{11}E_H^{-1}\psi_{11}E_H^{-1}E_{10}^{'} - \psi_{12}E_H^{-1}E_{10}^{'} - \psi_{13}E_H^{-1}E_{11}^{'} + \psi_{23}\right) \\
&= E_{11}E_H^{-1}E_{11}' + E_{10}E_H^{-1}E_{10}' - 2E_{11}E_H^{-1}E_{10}' + \psi_{22} + \psi_{33} - 2\psi_{23} \\
&= (E_{11} - E_{10})E_H^{-1}(E_{11} - E_{10})' + \psi_{22} + \psi_{33} - 2\psi_{23} \\
&= -(E_{11} - E_{10})(-E_H^{-1})(E_{11} - E_{10})' + \psi_{22} + \psi_{33} - 2\psi_{23} \qquad \text{(W.7)}
\end{aligned}
$$

Note that ψ_{22} and ψ_{33} correspond to the asymptotic variance of $\hat{\mu}_{1,ps1}$ $\hat{\mu}_{0,ps1}$ with known propensity score, respectively. Furthermore, $\hat{\mu}_{1,ps1}$ corresponds to the asymptotic covariance between $\hat{\mu}_{1,ps}$ $\hat{\mu}_{0,ps}$ with known propensity score. Hence, $\psi_{22} + \psi_{33} - 2\psi_{23}$ is simply the asymptotic variance for $\hat{\tau}_{ps1}$ with known propensity score. For further simplification, first consider $\psi_{22} + \psi_{33} - 2\psi_{23}$ separately:

$$
\psi_{22} + \psi_{33} - 2\psi_{23} = \mathrm{E}\left[\psi_2\psi_2'\right] + \mathrm{E}\left[\psi_3\psi_3'\right] - 2\,\mathrm{E}\left[\psi_2\psi_3'\right]
$$

$$
\begin{aligned}
&= \mathrm{E}\left[\left(\frac{D_iY_i}{\pi[X_i'\alpha]}-\mu_1\right)^2\right]+\mathrm{E}\left[\left(\frac{(1-D_i)Y_i}{1-\pi[X_i'\alpha]}-\mu_0\right)^2\right] \\
&\quad -2\,\mathrm{E}\left[\left(\frac{D_iY_i}{\pi[X_i'\alpha]}-\mu_1\right)\left(\frac{(1-D_i)Y_i}{1-\pi[X_i'\alpha]}-\mu_0\right)\right] \\
&= \mathrm{E}\left[\frac{D_i^2Y_i^2}{\pi[X_i'\alpha]^2}-2\frac{D_iY_i}{\pi[X_i'\alpha]}\mu_1+\mu_1^2\right] \\
&\quad +\mathrm{E}\left[\frac{(1-D_i)^2Y_i^2}{(1-\pi[X_i'\alpha])^2}-2\frac{(1-D_i)Y_i}{(1-\pi[X_i'\alpha])}\mu_0+\mu_0^2\right] \\
&\quad -2\,\mathrm{E}\left[-\frac{D_iY_i}{\pi[X_i'\alpha]}\mu_0-\mu_1\frac{(1-D_i)Y_i}{1-\pi[X_i'\alpha]}+\mu_1\mu_0\right]
\end{aligned}
$$

Using the LIE and Conditional Independence Assumption leads to further simplifications:

$$
\begin{aligned}
\psi_{22}+\psi_{33}-2\psi_{23} &= \mathrm{E}\left[\mathrm{E}\left[\frac{D_iY_i^2}{\pi[X_i'\alpha]^2}-2\frac{D_iY_i}{\pi[X_i'\alpha]}\mu_1+\mu_1^2\middle|\, X_i\right]\right] \quad \text{(Because } D_i(1-D_i)=0\text{)} \\
&\quad +\mathrm{E}\left[\mathrm{E}\left[\frac{(1-D_i)Y_i^2}{(1-\pi[X_i'\alpha])^2}-2\frac{(1-D_i)Y_i}{(1-\pi[X_i'\alpha])}\mu_0+\mu_0^2\middle|\, X_i\right]\right] \quad \text{(LIE)} \\
&\quad -2\,\mathrm{E}\left[\mathrm{E}\left[-\frac{D_iY_i}{\pi[X_i'\alpha]}\mu_0-\mu_1\frac{(1-D_i)Y_i}{1-\pi[X_i'\alpha]}+\mu_1\mu_0\middle|\, X_i\right]\right] \\
&= \mathrm{E}\left[\frac{\mathrm{E}\left[D_i|\,X_i\right]Y_{1i}^2}{\pi[X_i'\alpha]^2}-2\frac{\mathrm{E}\left[D_i|\,X_i\right]Y_{1i}}{\pi[X_i'\alpha]}\mu_1+\mu_1^2\right] \quad \text{(CIA)} \\
&\quad +\mathrm{E}\left[\frac{(1-\mathrm{E}\left[D_i|\,X_i\right])Y_{0i}^2}{(1-\pi[X_i'\alpha])^2}-2\frac{(1-\mathrm{E}\left[D_i|\,X_i\right])Y_{0i}}{(1-\pi[X_i'\alpha])}\mu_0+\mu_0^2\right] \\
&\quad -2\,\mathrm{E}\left[-\frac{\mathrm{E}\left[D_i|\,X_i\right]Y_{1i}}{\pi[X_i'\alpha]}\mu_0-\mu_1\frac{(1-\mathrm{E}\left[D_i|\,X_i\right])Y_{0i}}{1-\pi[X_i'\alpha]}+\mu_1\mu_0\right] \\
&= \mathrm{E}\left[\frac{Y_{1i}^2}{\pi[X_i'\alpha]}+\frac{Y_{0i}^2}{1-\pi[X_i'\alpha]}\right]-\mu_1^2-\mu_0^2+2\mu_1\mu_0 \\
&= \mathrm{E}\left[\frac{Y_{1i}^2}{\pi[X_i'\alpha]}+\frac{Y_{0i}^2}{1-\pi[X_i'\alpha]}\right]-(\mu_1-\mu_0)^2 \\
&= \mathrm{E}\left[\frac{Y_{1i}^2}{\pi[X_i'\alpha]}+\frac{Y_{0i}^2}{1-\pi[X_i'\alpha]}\right]-\tau^2.
\end{aligned}
$$

Putting the last expression back into the equality for $\mathrm{AV}\left[\hat{\tau}_{ps1}\right]$ and writing the explicit form for $(E_{11}-E_{10})$ gives the asymptotic variance as in Equation (11).

Asymptotic Variance of IPW2

For the second weighting estimator, a similar method can be followed to derive AV_{ps2} in Equation (12). First, consider the explicit form for A_{ps2}:

$$\begin{aligned}
A_{ps2} &\equiv E\left[\frac{\partial\psi(Z_i,\theta_{ps2})}{\partial\theta'_{ps2}}\right] \\
&= E\begin{bmatrix} \frac{\partial\psi_1(Z_i,\theta_{ps2})}{\partial\alpha'} & \frac{\partial\psi_1(Z_i,\theta_{ps2})}{\partial\mu_1} & \frac{\partial\psi_1(Z_i,\theta_{ps2})}{\partial\mu_0} \\ \frac{\partial\psi_2(Z_i,\theta_{ps2})}{\partial\alpha'} & \frac{\partial\psi_2(Z_i,\theta_{ps2})}{\partial\mu_1} & \frac{\partial\psi_2(Z_i,\theta_{ps2})}{\partial\mu_0} \\ \frac{\partial\psi_3(Z_i,\theta_{ps2})}{\partial\alpha'} & \frac{\partial\psi_3(Z_i,\theta_{ps2})}{\partial\mu_1} & \frac{\partial\psi_3(Z_i,\theta_{ps2})}{\partial\mu_0} \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{E}\left[H(Z_i,\alpha)\right] & \mathbf{0} & \mathbf{0} \\ \mathrm{E}\left[-\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] & \mathrm{E}\left[-\frac{D_i}{1-\pi[X_i'\alpha]}\right] & 0 \\ \mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] & 0 & \mathrm{E}\left[-\frac{1-D_i}{1-\pi[X_i'\alpha]}\right] \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{E}\left[H(Z_i,\alpha)\right] & \mathbf{0} & \mathbf{0} \\ \mathrm{E}\left[-\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] & -1 & 0 \\ \mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] & 0 & -1 \end{bmatrix} = \begin{bmatrix} E_H & \mathbf{0} & \mathbf{0} \\ E_{21} & -1 & 0 \\ E_{20} & 0 & -1 \end{bmatrix}
\end{aligned}$$

where

$$\begin{aligned}
E_{21} &\equiv \mathrm{E}\left[-\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = \mathrm{E}\left[\mathrm{E}\left[-\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\middle| X_i\right]\right] \\
&= \mathrm{E}\left[-\frac{\mathrm{E}\left[D_i|\, X_i\right](Y_{1i}-\mu_1)}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = -\,\mathrm{E}\left[\frac{(Y_{1i}-\mu_1)}{\pi[X_i'\alpha]}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]
\end{aligned}$$

and

$$\begin{aligned}
E_{20} &\equiv \mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = \mathrm{E}\left[\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\middle| X_i\right]\right] \\
&= \mathrm{E}\left[\frac{(1-\mathrm{E}\left[D_i|\, X_i\right])(Y_{0i}-\mu_0)}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = \mathrm{E}\left[\frac{(Y_{0i}-\mu_0)}{(1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right].
\end{aligned}$$

The asymptotic distribution of $(\hat{\mu}_{1,ps2},\hat{\mu}_{0,ps2})$ has the same form as in Equation (W.6) where E_{10} and E_{11} are replaced by E_{20} and E_{21}. Furthermore, ψ_{kl} now stands for

$\mathrm{E}\left[\psi_k(Z_i,\theta_{ps2})\psi_l(Z_i,\theta_{ps2})'\right]$ for $k,l=1,2,3$. It can easily be shown that $\psi_{21}=-E_{21}$ and $\psi_{31}=-E_{20}$ for this case, too.

$$\begin{aligned}
\psi_{21} &\equiv \mathrm{E}\left[\psi_2\psi_1'\right]=\mathrm{E}\left[\left(\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\right)\frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\\
&= \mathrm{E}\left[\left(\frac{D_iY_i-D_iY_i\pi[X_i'\alpha]-D_i\mu_1+D_i\mu_1\pi[X_i'\alpha]}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\\
&= \mathrm{E}\left[\left(\frac{D_iY_i(1-\pi[X_i'\alpha])-D_i\mu_1(1-\pi[X_i'\alpha])}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\\
&= \mathrm{E}\left[\left(\frac{(1-\pi[X_i'\alpha])D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\frac{(1-\pi[X_i'\alpha])D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\middle| X_i\right]\right]\\
&= \mathrm{E}\left[\left(\frac{(1-\pi[X_i'\alpha])\,\mathrm{E}\left[D_i|\,X_i\right](Y_{1i}-\mu_1)}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\\
&= \mathrm{E}\left[\left(\frac{(Y_{1i}-\mu_1)}{\pi[X_i'\alpha]}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]=-E_{21}\quad\square
\end{aligned}$$

$$\begin{aligned}
\psi_{31} &\equiv \mathrm{E}\left[\psi_3\psi_1'\right]=\mathrm{E}\left[\left(\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}\right)\frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\\
&= \mathrm{E}\left[\left(\frac{\pi[X_i'\alpha](D_i-1)(Y_i-\mu_0)}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])^2}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\frac{\pi[X_i'\alpha](D_i-1)(Y_i-\mu_0)}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])^2}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\middle| X_i\right]\right]\\
&= \mathrm{E}\left[-\left(\frac{\pi[X_i'\alpha](1-\mathrm{E}\left[D_i|\,X_i\right])(Y_{0i}-\mu_0)}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])^2}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\\
&= \mathrm{E}\left[-\left(\frac{(Y_{0i}-\mu_0)}{(1-\pi[X_i'\alpha])}\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]=-E_{20}.\quad\square
\end{aligned}$$

Furthermore,

$$\psi_{23}=\mathrm{E}\left[\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}\right]=0$$

due to the fact that $D_i(1-D_i)=0$. Thus, the asymptotic variance of $\hat{\tau}_{ps2}$ is given by

$$\begin{aligned}
\text{AV}\left[\hat{\tau}_{ps2}\right] &= \text{AV}[\hat{\mu}_{1,ps2}] + \text{AV}[\hat{\mu}_{0,ps2}] - 2\text{ACov}[\hat{\mu}_{1,ps2}, \hat{\mu}_{0,ps2}] \\
&= -(E_{21}-E_{20})(-E_H^{-1})(E_{21}-E_{20})' + \psi_{22} + \psi_{33}
\end{aligned}$$

which corresponds to Equation (12).

Asymptotic Variance of IPW3

The derivation of the variance of the third weighting estimator given in Equation (18) is similar to the previous weighting estimators.

$$\begin{aligned}
A_{ps3} &\equiv E\left[\frac{\partial\psi(Z_i,\theta_{ps3})}{\partial\theta'_{ps3}}\right] \\
&= E\begin{bmatrix}
\frac{\partial\psi_1(Z_i,\theta_{ps3})}{\partial\alpha'} & \frac{\partial\psi_1(Z_i,\theta_{ps3})}{\partial\mu_1} & \frac{\partial\psi_1(Z_i,\theta_{ps3})}{\partial\mu_0} \\
\frac{\partial\psi_2(Z_i,\theta_{ps3})}{\partial\alpha'} & \frac{\partial\psi_2(Z_i,\theta_{ps3})}{\partial\mu_1} & \frac{\partial\psi_2(Z_i,\theta_{ps3})}{\partial\mu_0} \\
\frac{\partial\psi_3(Z_i,\theta_{ps3})}{\partial\alpha'} & \frac{\partial\psi_3(Z_i,\theta_{ps3})}{\partial\mu_1} & \frac{\partial\psi_3(Z_i,\theta_{ps3})}{\partial\mu_0}
\end{bmatrix} \\
&= \begin{bmatrix}
E_H & \mathbf{0} & \mathbf{0} \\
E_{31} & -1 & 0 \\
E_{30} & 0 & -1
\end{bmatrix}
\end{aligned}$$

with

$$\begin{aligned}
E_{31} &\equiv \text{E}\left[-\frac{D_i(Y_i-\mu_1+\eta_1)}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = \text{E}\left[\text{E}\left[-\frac{D_i(Y_i-\mu_1+\eta_1)}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\middle| X_i\right]\right] \\
&= \text{E}\left[-\frac{\text{E}\left[D_i|\, X_i\right](Y_{1i}-\mu_1+\eta_1)}{\pi[X_i'\alpha]^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = -\,\text{E}\left[\frac{(Y_{1i}-\mu_1+\eta_1)}{\pi[X_i'\alpha]}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]
\end{aligned}$$

and

$$E_{30} \equiv \text{E}\left[\frac{(1-D_i)(Y_i-\mu_0+\eta_0)}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = \text{E}\left[\text{E}\left[\frac{(1-D_i)(Y_i-\mu_0+\eta_0)}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\middle| X_i\right]\right]$$

$$= \mathrm{E}\left[\frac{(1-\mathrm{E}\left[D_i|\,X_i\right])(Y_{0i}-\mu_0+\eta_0)}{(1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = \mathrm{E}\left[\frac{(Y_{0i}-\mu_0+\eta_0)}{(1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right].$$

Because the equalities $\psi_{21} = -E_{31}$ and $\psi_{31} = -E_{30}$ are true for the third weighting estimator, the asymptotic variance given in Equation (W.7) also holds for this estimator, except that E_{10} and E_{11} are replaced with E_{30} and E_{31}, respectively.

$$\begin{aligned}
\psi_{21} &\equiv \mathrm{E}\left[\psi_2\psi_1'\right] \\
&= \mathrm{E}\left[\left(\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}+\eta_1\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\right)\frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\frac{D_i(1-\pi[X_i'\alpha])(Y_i-\mu_1)+\eta_1(D_i-\pi[X_i'\alpha])^2}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\frac{D_i(1-\pi[X_i'\alpha])(Y_i-\mu_1)+\eta_1(D_i-2D_i\pi[X_i'\alpha]+\pi[X_i'\alpha]^2)}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right| X_i\right]\right] \\
&= \mathrm{E}\left[\frac{\pi[X_i'\alpha](1-\pi[X_i'\alpha])(Y_{1i}-\mu_1+\eta_1)}{\pi[X_i'\alpha]^2(1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\frac{(Y_{1i}-\mu_1+\eta_1)}{\pi[X_i'\alpha]}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = -E_{31} \quad \square
\end{aligned}$$

$$\begin{aligned}
\psi_{31} &\equiv \mathrm{E}\left[\psi_3\psi_1'\right] \\
&= \mathrm{E}\left[\left(\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}-\eta_0\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right)\frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)(D_i-\pi[X_i'\alpha])-\eta_0(D_i-\pi[X_i'\alpha])^2}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)(D_i-\pi[X_i'\alpha])-\eta_0(D_i-\pi[X_i'\alpha])^2}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])^2}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right| X_i\right]\right] \\
&= \mathrm{E}\left[-\frac{(1-\pi[X_i'\alpha])\pi[X_i'\alpha](Y_{0i}-\mu_0+\eta_0)}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])^2}\right] \\
&= \mathrm{E}\left[-\frac{(Y_{0i}-\mu_0+\eta_0)}{(1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] = -E_{30}. \quad \square
\end{aligned}$$

The last part of the proof requires us to identify the term $\psi_{22}+\psi_{33}-2\psi_{23}$ with the moments of the third weighting estimator. First consider ψ_{22}:

$$\begin{aligned}
\psi_{22} &\equiv \mathrm{E}\left[\psi_2(Z_i,\theta_{ps3})\psi_2(Z_i,\theta_{ps3})'\right] = \mathrm{E}\left[\left(\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}+\eta_1\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\right)^2\right] \\
&= \mathrm{E}\left[\frac{D_i(Y_i-\mu_1)^2}{\pi[X_i'\alpha]^2}+\eta_1^2\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)^2+2\eta_1\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\right] \\
&= \mathrm{E}\left[\frac{D_i(Y_i-\mu_1)^2}{\pi[X_i'\alpha]^2}\right]+\eta_1^2\,\mathrm{E}\left[\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)^2\right]+2\eta_1\,\mathrm{E}\left[\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\right] \\
&= \mathrm{E}\left[\frac{D_i(Y_i-\mu_1)^2}{\pi[X_i'\alpha]^2}\right]-\eta_1\,\mathrm{E}\left[\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2}\right]+2\eta_1\,\mathrm{E}\left[\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\right] \\
&= \mathrm{E}\left[\frac{D_i(Y_i-\mu_1)^2}{\pi[X_i'\alpha]^2}\right]-\eta_1\,\mathrm{E}\left[\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2}-2\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\right] \\
&= \mathrm{E}\left[\frac{D_i(Y_i-\mu_1)^2}{\pi[X_i'\alpha]^2}\right]-\eta_1\,\mathrm{E}\left[\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2}-2\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]^2}+2\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\right] \\
&= \mathrm{E}\left[\frac{(Y_{1i}-\mu_1)^2}{\pi[X_i'\alpha]}\right]+\eta_1\,\mathrm{E}\left[\frac{Y_{1i}-\mu_1}{\pi[X_i'\alpha]}\right]
\end{aligned}$$

To get from the third to the fourth line, Equation (15) is used. Other simplifications are results of CIA, LIE and the fact that $D_i=D_i^2$. Note also that the last term in the second equality from bottom, $\mathrm{E}\left[\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\right]$ has zero expectation. Similarly, ψ_{33} can be derived as follows:

$$\begin{aligned}
&\psi_{33}\equiv \mathrm{E}\left[\psi_3(Z_i,\theta_{ps3})\psi_3(Z_i,\theta_{ps3})'\right]=\mathrm{E}\left[\left(\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}-\eta_0\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right)^2\right] \\
&=\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)^2}{(1-\pi[X_i'\alpha])^2}+\eta_0^2\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)^2-2\eta_0\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])}\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right] \\
&=\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)^2}{(1-\pi[X_i'\alpha])^2}\right]+\eta_0^2\,\mathrm{E}\left[\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)^2\right]-2\eta_0\,\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])}\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right] \\
&=\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)^2}{(1-\pi[X_i'\alpha])^2}\right]-\eta_0\,\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}\right]-2\eta_0\,\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])}\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right] \\
&=\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)^2}{(1-\pi[X_i'\alpha])^2}\right]-\eta_0\,\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}+2\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])}\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right] \\
&=\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)^2}{(1-\pi[X_i'\alpha])^2}\right]-\eta_0\,\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}+2\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])}\left(\frac{-(1-D_i)+1-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right] \\
&=\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)^2}{(1-\pi[X_i'\alpha])^2}\right]-\eta_0\,\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}-2\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}+2\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])}\right] \\
&=\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)^2}{(1-\pi[X_i'\alpha])^2}\right]+\eta_0\,\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{(1-\pi[X_i'\alpha])^2}\right]
\end{aligned}$$

Equation (16) is used to get from the third to the fourth line. As earlier, CIA, LIE and the fact that $(1 - D_i) = (1 - D_i)^2$ are used to simplify the results. Note also that $\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_1)}{(1-\pi[X_i'\alpha])}\right]$ has zero expectation. The last part of the variance term is ψ_{23}.

$$\begin{aligned}
\psi_{23} &\equiv \mathrm{E}\left[\psi_2(Z_i,\theta_{ps3})\psi_3(Z_i,\theta_{ps3})'\right] \\
&= \mathrm{E}\left[\left\{\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]} + \eta_1\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\right\}\left\{\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]} - \eta_0\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right\}\right] \\
&= \mathrm{E}\left[-\eta_0\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right) + \eta_1\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}\right] \\
&\quad - \mathrm{E}\left[\eta_1\eta_0\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\left(\frac{D_i-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right)\right] \\
&= \mathrm{E}\left[-\eta_0\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\left(\frac{D_i-1+1-\pi[X_i'\alpha]}{1-\pi[X_i'\alpha]}\right) + \eta_1\left(\frac{D_i-\pi[X_i'\alpha]}{\pi[X_i'\alpha]}\right)\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}\right] \\
&\quad -\eta_1\eta_0\,\mathrm{E}\left[\frac{D_i-2D_i\pi[X_i'\alpha]+\pi[X_i'\alpha]^2}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\right] \\
&= \mathrm{E}\left[-\eta_0\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\left(\frac{-(1-D_i)}{1-\pi[X_i'\alpha]}+1\right) + \eta_1\left(\frac{D_i}{\pi[X_i'\alpha]}-1\right)\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}\right] \\
&\quad -\eta_1\eta_0\,\mathrm{E}\left[\mathrm{E}\left[\left.\frac{D_i-2D_i\pi[X_i'\alpha]+\pi[X_i'\alpha]^2}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\right| X_i\right]\right] \\
&= -\eta_0\,\mathrm{E}\left[\frac{D_i(Y_i-\mu_1)}{\pi[X_i'\alpha]}\right] + \eta_1\,\mathrm{E}\left[\frac{(1-D_i)(Y_i-\mu_0)}{1-\pi[X_i'\alpha]}\right] - \eta_1\eta_0\,\mathrm{E}\left[\frac{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\right] \\
&= -\eta_1\eta_0.
\end{aligned}$$

Substituting everything in

$$\mathrm{AV}\left[\hat{\tau}_{ps3}\right] = -(E_{31}-E_{30})(-E_H^{-1})(E_{31}-E_{30})' + \psi_{22} + \psi_{33} - 2\psi_{23}$$

gives the asymptotic variance in Equation (18).

Doubly Robust Estimators

First Doubly Robust Estimator (DR1)

In the following, I illustrate the doubly robustness property of the ATE estimator in Equation (20). To show that the ATE estimator is doubly robust, it is sufficient to show that the first two terms estimate μ_1 doubly robustly. By law of large numbers, the first two terms converge to the following population mean:

$$\mathrm{E}\left[\frac{D_iY_i}{\pi(X_i'\alpha^*)} - \frac{D_i - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}\eta[X_i'\beta_1^*]\right],$$

where α^* and β_1^* are the probability limits of $\hat{\alpha}$ and $\hat{\beta}_1$, respectively. Now, using some simple algebra we can rewrite this expectation:

$$\begin{aligned}
\mathrm{E}\left[\frac{D_iY_i}{\pi(X_i'\alpha^*)} - \frac{D_i - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}\eta[X_i'\beta_1^*]\right] &= \mathrm{E}\left[\frac{D_iY_{1i}}{\pi(X_i'\alpha^*)} - \frac{D_i - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}\eta[X_i'\beta_1^*]\right] \qquad \text{(W.8)}\\
&= \mathrm{E}\left[Y_{1i} + \frac{D - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}(Y_{1i} - \eta[X_i'\beta_1^*])\right]\\
&= \mathrm{E}\left[Y_{1i}\right] + \mathrm{E}\left[\frac{D - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}(Y_{1i} - \eta[X_i'\beta_1^*])\right].
\end{aligned}$$

If the second term in the last equality is equal to zero, then μ_1 is estimated consistently. By LIE, I rewrite the second term in Equation (W.8):

$$\begin{aligned}
\mathrm{E}\left[\frac{D_i - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}(Y_{1i} - \eta[X\beta_1^*])\right] &= \mathrm{E}\left[\mathrm{E}\left[\left.\frac{D_i - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}(Y_{1i} - \eta[X_i'\beta_1^*])\right| D_i, X_i\right]\right]\\
&= \mathrm{E}\left[\frac{D_i - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}\mathrm{E}\left[\left.(Y_{1i} - \eta[X_i'\beta_1^*])\right| D_i, X_i\right]\right]\\
&= \mathrm{E}\left[\frac{D_i - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}\mathrm{E}\left[\left.(Y_{1i} - \eta[X_i'\beta_1^*])\right| X_i\right]\right]\\
&= \mathrm{E}\left[\frac{D_i - \pi(X_i'\alpha^*)}{\pi(X_i'\alpha^*)}(\mathrm{E}\left[\left.Y_{1i}\right| X_i\right] - \eta[X_i'\beta_1^*])\right]
\end{aligned}$$

where the CIA assumption is used from second to third equality. We see from the last equality the doubly robustness property of this estimator. If the propensity score is correctly specified (i.e. $\pi(X_i'\alpha^*) = \mathrm{E}[D_i|\, X_i]$) but the outcome equation is misspecified (i.e. $\eta[X_i'\beta_1^*] \neq \mathrm{E}[Y_{1i}|\, X_i]$), then the first term in the expectation will be zero. Thus, as long as the propensity score model is correct, even if the outcome regression model is incorrect, the unconditional mean μ_1 is consistently estimated. Because the arguments are symmetric, μ_0 is also consistently estimated. Hence, the ATE is consistently estimated. If the outcome regression model is correct (i.e. $\eta[X_i'\beta_1^*] = \mathrm{E}[Y_{1i}|\, X_i]$) but the propensity score model is not (i.e. $\pi(X_i'\alpha^*) \neq \mathrm{E}[D_i|\, X_i]$), then the second term in the expectation will be zero. Thus, the unconditional mean μ_1 is still consistently estimated.

Asymptotic Variance of DR1

$$\begin{aligned}
A_{dr1} &\equiv E\left[\frac{\partial\psi(Z_i,\theta_{dr1})}{\partial\theta'_{dr1}}\right] \\
&= E\begin{bmatrix}
\frac{\partial\psi_1(Z_i,\theta_{dr1})}{\partial\alpha'} & \frac{\partial\psi_1(Z_i,\theta_{dr1})}{\partial\beta_1} & \frac{\partial\psi_1(Z_i,\theta_{dr1})}{\partial\beta_0} & \frac{\partial\psi_1(Z_i,\theta_{dr1})}{\partial\mu_1} & \frac{\partial\psi_1(Z_i,\theta_{dr1})}{\partial\mu_0} \\
\frac{\partial\psi_2(Z_i,\theta_{dr1})}{\partial\alpha'} & \frac{\partial\psi_2(Z_i,\theta_{dr1})}{\partial\beta_1} & \frac{\partial\psi_2(Z_i,\theta_{dr1})}{\partial\beta_0} & \frac{\partial\psi_2(Z_i,\theta_{dr1})}{\partial\mu_1} & \frac{\partial\psi_2(Z_i,\theta_{dr1})}{\partial\mu_0} \\
\frac{\partial\psi_3(Z_i,\theta_{dr1})}{\partial\alpha'} & \frac{\partial\psi_3(Z_i,\theta_{dr1})}{\partial\beta_1} & \frac{\partial\psi_3(Z_i,\theta_{dr1})}{\partial\beta_0} & \frac{\partial\psi_3(Z_i,\theta_{dr1})}{\partial\mu_1} & \frac{\partial\psi_3(Z_i,\theta_{dr1})}{\partial\mu_0} \\
\frac{\partial\psi_4(Z_i,\theta_{dr1})}{\partial\alpha'} & \frac{\partial\psi_4(Z_i,\theta_{dr1})}{\partial\beta_1} & \frac{\partial\psi_4(Z_i,\theta_{dr1})}{\partial\beta_0} & \frac{\partial\psi_4(Z_i,\theta_{dr1})}{\partial\mu_1} & \frac{\partial\psi_4(Z_i,\theta_{dr1})}{\partial\mu_0} \\
\frac{\partial\psi_5(Z_i,\theta_{dr1})}{\partial\alpha'} & \frac{\partial\psi_5(Z_i,\theta_{dr1})}{\partial\beta_1} & \frac{\partial\psi_5(Z_i,\theta_{dr1})}{\partial\beta_0} & \frac{\partial\psi_5(Z_i,\theta_{dr1})}{\partial\mu_1} & \frac{\partial\psi_5(Z_i,\theta_{dr1})}{\partial\mu_0}
\end{bmatrix} \\
&= \begin{bmatrix}
\mathrm{E}[H(Z_i,\alpha)] & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
0 & \mathrm{E}[D_i H_1(\beta_1)] & 0 & 0 & 0 \\
0 & 0 & \mathrm{E}[(1-D_i)H_0(\beta_0)] & 0 & 0 \\
0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & -1
\end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned}
V_{dr1} &\equiv V\left[\psi(Z_i,\theta_{dr1})\right] = E\left[\psi(Z_i,\theta_{dr1})\psi(Z_i,\theta_{dr1})'\right] \\
&= \mathrm{E}\left[\begin{pmatrix} \psi_1(Z_i,\theta_{dr1}) \\ \psi_2(Z_i,\theta_{dr1}) \\ \psi_3(Z_i,\theta_{dr1}) \\ \psi_4(Z_i,\theta_{dr1}) \\ \psi_5(Z_i,\theta_{dr1}) \end{pmatrix} \begin{pmatrix} \psi_1(Z_i,\theta_{dr1})' & \dots & \psi_5(Z_i,\theta_{dr1})' \end{pmatrix}\right] \\
&= \begin{bmatrix} \mathrm{E}\left[\psi_1(Z_i,\theta_{dr1})\psi_1(Z_i,\theta_{dr1})'\right] & \dots & \mathrm{E}\left[\psi_1(Z_i,\theta_{dr1})\psi_5(Z_i,\theta_{dr1})'\right] \\ \vdots & \ddots & \vdots \\ \mathrm{E}\left[\psi_5(Z_i,\theta_{dr1})\psi_1(Z_i,\theta_{dr1})'\right] & \dots & \mathrm{E}\left[\psi_5(Z_i,\theta_{dr1})\psi_5(Z_i,\theta_{dr1})'\right] \end{bmatrix}.
\end{aligned}$$

Second Doubly Robust Estimator (DR2)

The doubly robustness of the second estimator relies on the properties of the estimation in the linear exponential family with a canonical link function (Scharfstein et al., 1999, Wooldridge, 2007, see). Without loss of generality, I assume here an identity link for the outcome model. Consistent estimation of the unconditional mean, μ_1, requires that $\mathrm{E}\left[\frac{D_i}{\pi(X\alpha^*)}(Y_i - X_i'\beta_1^*)\right]$ is equal to zero. By the LIE, we can write the following equality:

$$\begin{aligned}
\mathrm{E}\left[\frac{D_i}{\pi(X_i'\alpha^*)}(Y_i - X_i'\beta_1^*)\right] &= \mathrm{E}\left[\frac{D_iY_i - D_i(X_i'\beta_1^*)}{\pi(X\alpha^*)}\right] = \mathrm{E}\left[\frac{D_iY_{1i} - D_i(X_i'\beta_1^*)}{\pi(X_i'\alpha^*)}\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\frac{D_i}{\pi(X_i'\alpha^*)}(Y_{1i} - X_i'\beta_1^*)\middle| X_i\right]\right] \\
&= \mathrm{E}\left[\frac{\mathrm{E}\left[D_i|\, X_i\right]}{\pi(X_i'\alpha^*)}\mathrm{E}\left[(Y_{1i} - X_i'\beta_1^*)|\, X_i\right]\right] \\
&= \mathrm{E}\left[\frac{\mathrm{E}\left[D_i|\, X_i\right]}{\pi(X_i'\alpha^*)}(\mathrm{E}\left[Y_{1i}|\, X_i\right] - X_i'\beta_1^*)\right].
\end{aligned}$$

If $\mathrm{E}\left[Y_{1i}|\, X_i\right] = X_i'\beta_1^*$, then the last equality will be equal to zero even if the propensity score is wrongly specified. If the propensity score is correctly specified, then the first term is equal to one and the equality can be simplified as:

$$\mathrm{E}\left[\mathrm{E}\left[Y_{1i}|\, X_i\right] - X_i'\beta_1^*\right] = \mathrm{E}\left[Y_{1i}\right] - \mathrm{E}\left[X_i'\beta_1^*\right]$$

This term is equal to zero even if $\mathrm{E}\left[Y_{1i}|\, X_i\right] \neq X_i'\beta$, because the special property of the estimation in the linear exponential family with a canonical link function $\mathrm{E}\left[Y_{1i}\right] = \mathrm{E}\left[X_i'\hat{\beta}_1\right]$ holds.

Asymptotic Variance of Weighted Regression Coefficients

In the last part of this section, I provide the details for the derivation of $\mathrm{AV}_{\hat{\beta}_{1,dr}}$.[19] The weighted regression estimator of β_1 can be derived from the solution of the following sample moment equations

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\psi(Z_i,\alpha,\beta_1) &= \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N}\psi_1(Z_i,\alpha,\beta_1) \\ \frac{1}{N}\sum_{i=1}^{N}\psi_2(Z_i,\alpha,\beta_1) \end{pmatrix} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N}\frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha} \\ \frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi[X_i'\alpha]}\frac{\partial q(Y_i,X_i;\beta_1)}{\partial\beta_1} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N}S(\alpha) \\ \frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi[X_i'\alpha]}S_1(\beta_1) \end{pmatrix} = 0.
\end{aligned}$$

The asymptotic variance-covariance matrix of $(\hat{\alpha},\hat{\beta}_{1,dr})$ can be derived by applying the general results on M-estimators. Let $A_{we}^{-1}V_{we}A_{we}^{-1\prime}$ be the asymptotic variance of $(\hat{\alpha},\hat{\beta}_{1,dr})$ with A_{we} and V_{we} as follows:

$$\begin{aligned}
A_{we} &\equiv E\begin{bmatrix} \frac{\partial\psi_1(Z_i,\alpha,\beta_1)}{\partial\alpha'} & \frac{\partial\psi_2(Z_i,\alpha,\beta_1)}{\partial\beta_1} \\ \frac{\partial\psi_2(Z_i,\alpha,\beta_1)}{\partial\alpha'} & \frac{\partial\psi_2(Z_i,\alpha,\beta_1)}{\partial\beta_1} \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{E}\left[H(Z_i,\alpha)\right] & \mathbf{0} \\ \mathrm{E}\left[-\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] & \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right] \end{bmatrix} \\
V_{we} &\equiv E\left[\psi(Z_i,\alpha,\beta_1)\psi(Z_i,\alpha,\beta_1)'\right]
\end{aligned}$$

[19] Because the derivation of $\mathrm{AV}_{\hat{\beta}_{0,dr}}$ is completely symmetric, I skipped the details.

$$
\begin{aligned}
&= \mathrm{E}\left[\begin{pmatrix} \psi_1(Z_i,\alpha,\beta_1) \\ \psi_2(Z_i,\alpha,\beta_1) \end{pmatrix} \begin{pmatrix} \psi_1(Z_i,\alpha,\beta_1)' & \psi_2(Z_i,\alpha,\beta_1)' \end{pmatrix}\right] \\
&= \begin{bmatrix} \mathrm{E}\left[\psi_1\psi_1'\right] & \mathrm{E}\left[\psi_1\psi_2'\right] \\ \mathrm{E}\left[\psi_2\psi_1'\right] & \mathrm{E}\left[\psi_2\psi_2'\right] \end{bmatrix} = \begin{bmatrix} \mathrm{E}\left[S(\alpha)S(\alpha)'\right] & \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}S(\alpha)S_1(\beta_1)'\right] \\ \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}S_1(\beta_1)S(\alpha)'\right] & \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)S_1(\beta_1)'\right] \end{bmatrix}.
\end{aligned}
$$

By matrix inversion rule, A_{we}^{-1} is given by

$$
A_{we}^{-1} = \begin{bmatrix} \mathrm{E}\left[H(Z_i,\alpha)\right]^{-1} & \mathbf{0} \\ \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1}\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\mathrm{E}\left[H(Z_i,\alpha)\right]^{-1} & \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1} \end{bmatrix}.
$$

Multiplication of the matrices yields the following asymptotic variance for $\hat{\beta}_{1,dr}$

$$
\begin{aligned}
\mathrm{AV}_{\hat{\beta}_{1,dr}} &= \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1}\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\mathrm{E}\left[H(Z_i,\alpha)\right]^{-1}\mathrm{E}\left[S(\alpha)S(\alpha)'\right] \\
&\quad\times\mathrm{E}\left[H(Z_i,\alpha)\right]^{-1}\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]'\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1} \\
&\quad+2\,\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1}\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}S_1(\beta_1)S(\alpha)'\right] \\
&\quad\times\mathrm{E}\left[H(Z_i,\alpha)\right]^{-1}\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]'\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1} \\
&\quad+\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1}\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)S_1(\beta_1)'\right]\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1}. \quad \text{(W.9)}
\end{aligned}
$$

Note that

$$
\begin{aligned}
\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}S_1(\beta_1)S(\alpha)'\right] &= \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}S_1(\beta_1)\frac{(D_i-\pi[X_i'\alpha])}{\pi[X_i'\alpha](1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]'}{\partial\alpha}\right] \\
&= \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\frac{((D_i-1)+(1-\pi[X_i'\alpha]))}{(1-\pi[X_i'\alpha])}\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\left(\frac{-(1-D_i)}{(1-\pi[X_i'\alpha])}+1\right)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right] \\
&= \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]
\end{aligned}
$$

where the last equality follows from the fact that $D_i(1-D_i)=0$. Using this fact and the information equality, $\mathrm{E}\left[H(Z_i,\alpha)\right] = -\,\mathrm{E}\left[S(\alpha)S(\alpha)'\right]$, Equation (W.9) can be further simplified:

$$\begin{aligned}
\mathrm{AV}_{\hat{\beta}_{1,dr}} \;=\;& \mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1}\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]\mathrm{E}\left[H(Z_i,\alpha)\right]^{-1}\\
&\times\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)\frac{\partial\pi[X_i'\alpha]}{\partial\alpha'}\right]'\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1}\\
&+\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1}\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]^2}S_1(\beta_1)S_1(\beta_1)'\right]\mathrm{E}\left[\frac{D_i}{\pi[X_i'\alpha]}H_1(\beta_1)\right]^{-1} \qquad \square
\end{aligned}$$

Replacing $\mathrm{E}\left[H(Z_i,\alpha)\right]^{-1}$ with $-\mathrm{AV}\left[\hat{\alpha}\right]$ gives the asymptotic variance in Equation (21).

B.2 Web Appendix Tables: Monte Carlo Study

Table B1: MC Results for correctly specified regression and propensity score models, Homogeneous Treatment, Bounded X

Ratio	**N**		**REG**	**IPW1**	**IPW2**	**IPW3**	**DR1a**	**DR1b**	**DR1c**	**DR2**
1/3	100	BIAS	0.00	0.10	0.30	0.34	0.00	0.00	0.00	0.00
		MCVAR	12.36	248.72	103.23	48.53	16.67	14.50	14.15	14.93
		MCMSE	12.36	249.72	112.34	60.29	16.67	14.50	14.15	14.94
		AAVAR	9.90	223.91	40.52	15.67	14.80	9.38	10.89	7.37
	400	BIAS	0.00	0.01	0.05	0.07	0.00	0.00	0.00	0.00
		MCVAR	2.54	36.65	25.05	9.65	3.27	3.21	3.17	3.23
		MCMSE	2.54	36.65	25.25	10.14	3.28	3.22	3.17	3.23
		AAVAR	2.43	32.39	19.90	6.23	2.89	2.85	3.08	2.51
	1600	BIAS	0.00	0.02	0.03	0.02	0.00	0.00	0.00	0.00
		MCVAR	0.61	8.36	6.23	2.31	0.78	0.78	0.78	0.79
		MCMSE	0.61	8.40	6.31	2.37	0.78	0.78	0.78	0.79
		AAVAR	0.62	7.64	5.54	1.92	0.77	0.77	0.79	0.75
1/1	100	BIAS	0.00	0.01	0.07	0.12	0.00	0.00	0.00	0.00
		MCVAR	5.88	26.55	23.60	12.94	7.12	6.97	6.82	6.88
		MCMSE	5.88	26.57	24.05	14.34	7.12	6.97	6.82	6.88
		AAVAR	5.56	25.02	19.09	8.51	6.03	5.92	6.47	5.33
	400	BIAS	0.00	0.01	0.01	0.02	0.00	0.00	0.00	0.00
		MCVAR	1.41	4.62	4.05	2.47	1.61	1.61	1.60	1.61
		MCMSE	1.41	4.63	4.06	2.52	1.61	1.61	1.60	1.61
		AAVAR	1.39	4.56	4.21	2.41	1.57	1.57	1.60	1.53
	1600	BIAS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		MCVAR	0.34	1.01	0.93	0.58	0.40	0.39	0.39	0.39
		MCMSE	0.34	1.01	0.93	0.59	0.40	0.39	0.39	0.39
		AAVAR	0.35	1.06	0.99	0.61	0.40	0.40	0.40	0.40
3/1	100	BIAS	0.00	0.04	0.30	0.34	0.00	0.00	0.00	0.00
		MCVAR	12.42	48.08	99.56	48.08	15.82	14.42	14.13	14.84
		MCMSE	12.42	48.25	108.33	59.30	15.82	14.42	14.13	14.84
		AAVAR	9.75	37.01	40.37	15.41	11.57	9.19	10.70	7.20
	400	BIAS	0.00	0.00	0.05	0.07	0.00	0.00	0.00	0.00
		MCVAR	2.55	9.53	25.40	9.95	3.36	3.29	3.22	3.26
		MCMSE	2.55	9.53	25.64	10.46	3.36	3.29	3.22	3.26
		AAVAR	2.46	8.58	20.40	6.36	2.89	2.86	3.09	2.51
	1600	BIAS	0.00	-0.01	-0.01	0.01	0.00	0.00	0.00	0.00
		MCVAR	0.61	2.38	6.19	2.23	0.82	0.82	0.82	0.82
		MCMSE	0.61	2.39	6.20	2.23	0.82	0.82	0.82	0.82
		AAVAR	0.61	2.19	5.63	1.91	0.78	0.78	0.80	0.75

Note: MC Results for the DGP1 with homogeneous treatment and bounded Xs. Both regression and propensity score models are correctly specified. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Ratio refers to the treated-control ratio. BIAS: average bias over Monte Carlo simulations. MCVAR: Monte Carlo Variance × 100. MCMSE: Monte Carlo Mean Squared Error × 100. AAVAR: Average of the estimated variances based on the asymptotic variance of the M-estimators × 100.

Table B2: MC Results for correctly specified regression and propensity score models, Heterogeneous Treatment, Bounded X

Ratio	**N**		**REG**	**IPW1**	**IPW2**	**IPW3**	**DR1a**	**DR1b**	**DR1c**	**DR2**
1/3	100	BIAS	0.00	0.07	0.20	0.23	0.00	0.00	0.00	0.00
		MCVAR	14.24	146.02	61.60	37.27	17.93	16.29	16.05	16.84
		MCMSE	14.25	146.55	65.69	42.76	17.93	16.29	16.05	16.84
		AAVAR	11.80	142.40	28.59	16.66	14.65	11.30	12.82	9.30
	400	BIAS	0.00	0.01	0.04	0.05	0.00	0.00	0.00	0.00
		MCVAR	3.07	21.84	14.51	7.60	3.78	3.72	3.67	3.71
		MCMSE	3.07	21.86	14.67	7.88	3.78	3.72	3.67	3.71
		AAVAR	2.91	19.56	11.83	5.53	3.36	3.33	3.56	2.98
	1600	BIAS	0.00	0.01	0.02	0.01	0.00	0.00	0.00	0.00
		MCVAR	0.77	4.84	3.40	1.74	0.92	0.92	0.92	0.92
		MCMSE	0.77	4.86	3.42	1.76	0.92	0.92	0.92	0.92
		AAVAR	0.73	4.57	3.18	1.59	0.89	0.88	0.91	0.86
1/1	100	BIAS	0.00	0.01	0.05	0.10	0.00	0.00	0.00	0.00
		MCVAR	7.68	20.32	22.54	13.74	8.94	8.79	8.66	8.70
		MCMSE	7.68	20.33	22.79	14.65	8.94	8.79	8.66	8.70
		AAVAR	7.44	19.45	18.25	9.95	7.91	7.80	8.33	7.24
	400	BIAS	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00
		MCVAR	1.90	3.81	4.34	2.97	2.12	2.12	2.11	2.11
		MCMSE	1.90	3.81	4.35	3.01	2.12	2.12	2.11	2.11
		AAVAR	1.87	3.80	4.21	2.76	2.05	2.05	2.09	2.02
	1600	BIAS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		MCVAR	0.45	0.90	1.04	0.71	0.49	0.49	0.49	0.49
		MCMSE	0.45	0.90	1.04	0.71	0.49	0.49	0.49	0.49
		AAVAR	0.47	0.89	1.01	0.70	0.52	0.52	0.52	0.52
3/1	100	BIAS	0.00	0.03	0.29	0.33	0.00	0.00	0.00	0.00
		MCVAR	14.23	51.95	107.12	51.06	17.60	16.27	16.07	16.93
		MCMSE	14.23	52.06	115.60	62.13	17.60	16.27	16.08	16.93
		AAVAR	11.77	38.53	42.06	16.71	13.17	11.23	12.73	9.22
	400	BIAS	0.00	0.00	0.04	0.06	0.00	0.00	0.00	0.00
		MCVAR	2.94	10.10	25.90	10.00	3.77	3.67	3.62	3.66
		MCMSE	2.94	10.10	26.06	10.41	3.77	3.67	3.62	3.66
		AAVAR	2.91	9.14	21.26	6.82	3.43	3.37	3.59	2.98
	1600	BIAS	0.00	0.01	0.02	0.02	0.00	0.00	0.00	0.00
		MCVAR	0.73	2.37	5.87	2.21	0.91	0.91	0.92	0.92
		MCMSE	0.73	2.37	5.90	2.25	0.91	0.91	0.92	0.92
		AAVAR	0.73	2.26	5.70	2.03	0.87	0.87	0.89	0.85

Note: MC Results for the DGP1 with heterogeneous treatment and bounded Xs. Both regression and propensity score models are correctly specified. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Ratio refers to the treated-control ratio. BIAS: average bias over Monte Carlo simulations. MCVAR: Monte Carlo Variance × 100. MCMSE: Monte Carlo Mean Squared Error × 100. AAVAR: Average of the estimated variances based on the asymptotic variance of the M-estimators × 100.

Table B3: MC Results for correctly specified regression and misspecified propensity score models, Homogeneous Treatment, Bounded X

Ratio	**N**		**REG**	**IPW1**	**IPW2**	**IPW3**	**DR1a**	**DR1b**	**DR1c**	**DR2**
1/3	100	BIAS	0.00	0.12	0.39	0.42	0.00	0.00	0.00	0.00
		MCVAR	12.46	501.37	116.36	51.82	17.57	14.86	14.61	15.55
		MCMSE	12.46	502.81	131.27	69.64	17.57	14.86	14.61	15.55
		AAVAR	10.32	889.49	40.92	15.10	16.58	9.84	11.66	7.55
	400	BIAS	0.00	0.04	0.09	0.11	0.00	0.00	0.00	0.00
		MCVAR	2.64	50.34	33.15	11.33	3.74	3.64	3.57	3.66
		MCMSE	2.64	50.48	33.98	12.60	3.74	3.64	3.57	3.66
		AAVAR	2.57	44.18	24.37	6.29	3.30	3.23	3.55	2.68
	1600	BIAS	0.00	0.02	0.03	0.03	0.00	0.00	0.00	0.00
		MCVAR	0.63	11.07	7.99	2.50	0.92	0.91	0.91	0.91
		MCMSE	0.63	11.11	8.09	2.60	0.92	0.91	0.91	0.91
		AAVAR	0.64	10.80	7.49	2.08	0.87	0.87	0.90	0.82
1/1	100	BIAS	0.00	0.04	0.13	0.18	0.00	0.00	0.00	0.00
		MCVAR	6.41	43.04	34.26	16.34	8.44	8.13	7.87	7.95
		MCMSE	6.41	43.19	35.83	19.67	8.45	8.13	7.87	7.95
		AAVAR	5.95	39.02	24.33	9.22	6.70	6.47	7.31	5.63
	400	BIAS	0.00	0.02	0.04	0.05	0.00	0.00	0.00	0.00
		MCVAR	1.54	7.20	6.65	3.40	1.93	1.93	1.93	1.93
		MCMSE	1.54	7.25	6.82	3.62	1.93	1.93	1.93	1.93
		AAVAR	1.49	6.77	6.18	2.91	1.79	1.79	1.85	1.73
	1600	BIAS	0.00	0.02	0.03	0.02	0.00	0.00	0.00	0.00
		MCVAR	0.40	1.57	1.50	0.82	0.51	0.51	0.51	0.51
		MCMSE	0.40	1.63	1.58	0.87	0.51	0.51	0.51	0.51
		AAVAR	0.37	1.54	1.45	0.75	0.45	0.45	0.46	0.45
3/1	100	BIAS	0.00	0.05	0.37	0.41	0.00	0.00	0.00	0.00
		MCVAR	12.82	58.32	116.78	52.28	17.36	15.34	15.09	16.05
		MCMSE	12.82	58.54	130.61	69.25	17.36	15.34	15.09	16.05
		AAVAR	10.35	43.42	41.47	15.01	12.76	9.86	11.69	7.53
	400	BIAS	0.00	0.02	0.09	0.11	0.00	0.00	0.00	0.00
		MCVAR	2.75	12.30	33.82	11.67	3.99	3.85	3.77	3.84
		MCMSE	2.75	12.34	34.58	12.92	3.99	3.85	3.77	3.84
		AAVAR	2.57	10.43	24.48	6.35	3.33	3.25	3.57	2.70
	1600	BIAS	0.00	0.01	0.02	0.02	0.00	0.00	0.00	0.00
		MCVAR	0.64	2.67	7.44	2.29	0.96	0.96	0.96	0.96
		MCMSE	0.64	2.67	7.48	2.35	0.96	0.96	0.96	0.96
		AAVAR	0.65	2.69	7.60	2.11	0.89	0.89	0.92	0.84

Note: MC Results for the DGP2 with homogeneous treatment and bounded Xs. The regression model is correctly specified, but the propensity score model is wrongly specified. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Ratio refers to the treated-control ratio. BIAS: average bias over Monte Carlo simulations. MCVAR: Monte Carlo Variance × 100. MCMSE: Monte Carlo Mean Squared Error × 100. AAVAR: Average of the estimated variances based on the asymptotic variance of the M-estimators× 100.

Table B4: MC Results for correctly specified regression and misspecified propensity score models, heterogeneous Treatment, Bounded X

Ratio	**N**		**REG**	**IPW1**	**IPW2**	**IPW3**	**DR1a**	**DR1b**	**DR1c**	**DR2**
1/3	100	BIAS	0.00	0.08	0.26	0.29	0.01	0.01	0.00	0.00
		MCVAR	14.59	222.32	68.83	39.62	19.78	17.10	16.83	17.73
		MCMSE	14.59	222.99	75.66	48.16	19.78	17.10	16.84	17.73
		AAVAR	12.28	253.83	29.45	17.13	17.83	11.77	13.54	9.49
	400	BIAS	0.00	0.03	0.07	0.08	0.00	0.00	0.00	0.00
		MCVAR	3.01	28.96	18.14	8.79	4.24	4.10	4.00	4.08
		MCMSE	3.01	29.07	18.62	9.46	4.24	4.10	4.00	4.08
		AAVAR	3.05	26.20	14.22	5.92	3.82	3.74	4.08	3.16
	1600	BIAS	0.00	0.02	0.02	0.02	0.00	0.00	0.00	0.00
		MCVAR	0.73	6.28	4.24	1.85	0.97	0.96	0.96	0.97
		MCMSE	0.73	6.31	4.30	1.90	0.97	0.96	0.96	0.97
		AAVAR	0.76	6.16	4.05	1.73	0.99	0.98	1.02	0.94
1/1	100	BIAS	0.00	0.04	0.11	0.16	0.00	0.00	0.00	0.00
		MCVAR	8.25	29.60	31.09	16.93	10.29	9.97	9.73	9.81
		MCMSE	8.25	29.76	32.33	19.37	10.29	9.97	9.73	9.81
		AAVAR	7.80	25.83	22.17	10.65	8.57	8.32	9.13	7.48
	400	BIAS	0.00	0.01	0.03	0.03	-0.01	-0.01	-0.01	-0.01
		MCVAR	2.12	5.06	5.89	3.52	2.46	2.45	2.45	2.46
		MCMSE	2.12	5.08	5.96	3.62	2.46	2.46	2.45	2.46
		AAVAR	1.96	5.09	5.78	3.19	2.26	2.26	2.32	2.19
	1600	BIAS	0.00	0.02	0.03	0.02	0.00	0.00	0.00	0.00
		MCVAR	0.51	1.29	1.45	0.87	0.59	0.59	0.59	0.59
		MCMSE	0.51	1.32	1.52	0.91	0.59	0.59	0.59	0.59
		AAVAR	0.49	1.19	1.37	0.82	0.57	0.57	0.58	0.57
3/1	100	BIAS	0.00	0.05	0.38	0.41	0.00	0.00	0.00	0.00
		MCVAR	15.00	68.74	122.66	54.83	20.21	17.61	17.28	18.16
		MCMSE	15.00	68.96	136.84	71.79	20.21	17.61	17.28	18.17
		AAVAR	12.27	54.68	42.86	16.09	16.31	11.79	13.64	9.43
	400	BIAS	0.00	0.02	0.08	0.10	0.00	0.00	0.00	0.00
		MCVAR	3.20	12.65	34.57	11.99	4.34	4.23	4.16	4.22
		MCMSE	3.20	12.68	35.22	13.05	4.34	4.23	4.16	4.22
		AAVAR	3.06	11.04	26.40	6.92	3.80	3.72	4.06	3.16
	1600	BIAS	0.00	0.01	0.02	0.03	0.00	0.00	0.00	0.00
		MCVAR	0.71	3.08	8.97	2.76	1.01	1.01	1.01	1.02
		MCMSE	0.71	3.08	9.01	2.84	1.01	1.01	1.01	1.02
		AAVAR	0.76	2.85	7.78	2.20	0.99	0.99	1.01	0.94

Note: MC Results for the DGP2 with heterogeneous treatment and bounded Xs. The regression model is correctly specified, but the propensity score model is wrongly specified. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Ratio refers to the treated-control ratio. BIAS: average bias over Monte Carlo simulations. MCVAR: Monte Carlo Variance × 100. MCMSE: Monte Carlo Mean Squared Error × 100. AAVAR: Average of the estimated variances based on the asymptotic variance of the M-estimators × 100.

Table B5: MC Results for correctly specified propensity score and misspecified regression models, homogeneous Treatment, Bounded X

Ratio	N		REG	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2
1/3	100	BIAS	-0.02	0.10	0.33	0.38	0.00	-0.01	0.00	0.00
		MCVAR	14.44	293.95	128.09	60.07	18.20	16.59	16.37	17.35
		MCMSE	14.47	294.98	138.78	74.19	18.21	16.59	16.37	17.35
		AAVAR	11.78	266.70	49.93	19.03	13.97	11.08	12.89	8.80
	400	BIAS	-0.01	0.02	0.06	0.08	0.00	0.00	0.00	0.00
		MCVAR	3.10	42.85	31.07	12.23	3.89	3.85	3.84	3.90
		MCMSE	3.11	42.89	31.43	12.93	3.89	3.85	3.84	3.90
		AAVAR	2.93	39.26	25.10	7.74	3.39	3.36	3.62	2.96
	1600	BIAS	-0.01	0.01	0.02	0.02	0.00	0.00	0.00	0.00
		MCVAR	0.76	10.32	7.85	2.88	0.99	0.99	0.99	0.99
		MCMSE	0.76	10.34	7.88	2.93	0.99	0.99	0.99	0.99
		AAVAR	0.73	9.38	6.95	2.37	0.88	0.88	0.90	0.85
1/1	100	BIAS	-0.01	0.02	0.08	0.13	0.00	0.00	0.00	0.00
		MCVAR	7.21	34.17	30.65	16.21	8.52	8.38	8.23	8.28
		MCMSE	7.21	34.19	31.22	18.03	8.52	8.38	8.23	8.28
		AAVAR	6.68	31.73	24.04	10.43	7.09	6.98	7.62	6.32
	400	BIAS	-0.01	0.01	0.02	0.03	0.00	0.00	0.00	0.00
		MCVAR	1.69	5.72	5.45	3.24	1.92	1.91	1.91	1.92
		MCMSE	1.69	5.73	5.47	3.32	1.92	1.92	1.91	1.92
		AAVAR	1.68	5.67	5.36	3.01	1.85	1.85	1.89	1.81
	1600	BIAS	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		MCVAR	0.42	1.38	1.30	0.80	0.49	0.49	0.49	0.49
		MCMSE	0.43	1.38	1.30	0.80	0.49	0.49	0.49	0.49
		AAVAR	0.42	1.30	1.25	0.76	0.47	0.47	0.47	0.46
3/1	100	BIAS	-0.02	0.04	0.32	0.37	0.00	0.00	0.00	0.00
		MCVAR	14.79	71.73	126.91	59.77	18.28	16.94	16.82	17.82
		MCMSE	14.80	71.90	137.28	73.63	18.28	16.94	16.82	17.82
		AAVAR	11.80	58.09	50.16	18.99	13.20	11.05	12.84	8.75
	400	BIAS	-0.01	0.01	0.06	0.08	0.00	0.00	0.00	0.00
		MCVAR	3.10	14.13	31.95	12.20	3.90	3.84	3.81	3.87
		MCMSE	3.11	14.14	32.31	12.86	3.90	3.84	3.81	3.87
		AAVAR	2.95	11.98	25.11	7.85	3.44	3.40	3.67	2.99
	1600	BIAS	-0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00
		MCVAR	0.77	2.99	6.91	2.49	0.95	0.95	0.95	0.95
		MCMSE	0.78	2.99	6.91	2.51	0.95	0.95	0.95	0.95
		AAVAR	0.73	3.01	7.02	2.37	0.89	0.89	0.91	0.86

Note: MC Results for the DGP3 with homogeneous treatment and bounded Xs. The regression model is misspecified, but the propensity score model is correctly specified. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Ratio refers to the treated-control ratio. BIAS: average bias over Monte Carlo simulations. MCVAR: Monte Carlo Variance × 100. MCMSE: Monte Carlo Mean Squared Error × 100. AAVAR: Average of the estimated variances based on the asymptotic variance of the M-estimators × 100.

Table B6: MC Results for correctly specified propensity score and misspecified regression models, heterogeneous Treatment, Bounded X

Ratio	**N**		**REG**	**IPW1**	**IPW2**	**IPW3**	**DR1a**	**DR1b**	**DR1c**	**DR2**
1/3	100	BIAS	-0.03	0.07	0.26	0.31	0.00	0.00	0.00	0.00
		MCVAR	23.00	241.70	103.03	57.30	28.39	25.92	25.78	27.31
		MCMSE	23.05	242.19	109.73	66.70	28.39	25.92	25.78	27.31
		AAVAR	18.14	234.97	44.98	23.45	22.06	16.98	19.63	13.68
	400	BIAS	-0.02	0.03	0.06	0.07	0.00	0.00	0.00	0.00
		MCVAR	4.53	34.14	24.64	11.94	5.46	5.38	5.36	5.42
		MCMSE	4.56	34.21	24.98	12.46	5.46	5.38	5.36	5.42
		AAVAR	4.53	31.66	20.42	8.47	5.10	5.07	5.48	4.49
	1600	BIAS	-0.02	0.01	0.02	0.02	0.00	0.00	0.00	0.00
		MCVAR	1.14	7.40	5.58	2.54	1.29	1.28	1.29	1.30
		MCMSE	1.17	7.42	5.62	2.57	1.29	1.28	1.29	1.30
		AAVAR	1.13	7.41	5.49	2.44	1.29	1.29	1.32	1.26
1/1	100	BIAS	-0.01	0.02	0.07	0.12	0.00	0.00	0.00	0.00
		MCVAR	10.03	31.44	30.79	18.40	11.35	11.20	11.05	11.12
		MCMSE	10.04	31.46	31.27	19.90	11.35	11.20	11.05	11.12
		AAVAR	9.53	28.53	24.54	12.81	9.85	9.74	10.48	8.99
	400	BIAS	-0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00
		MCVAR	2.53	5.48	5.81	3.89	2.73	2.72	2.71	2.72
		MCMSE	2.54	5.48	5.82	3.93	2.73	2.72	2.71	2.72
		AAVAR	2.39	5.49	5.69	3.61	2.55	2.55	2.59	2.51
	1600	BIAS	-0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00
		MCVAR	0.60	1.33	1.34	0.92	0.64	0.64	0.64	0.64
		MCMSE	0.61	1.34	1.35	0.92	0.64	0.64	0.64	0.64
		AAVAR	0.60	1.28	1.34	0.91	0.64	0.64	0.65	0.64
3/1	100	BIAS	-0.02	0.05	0.33	0.38	0.00	0.00	0.00	0.00
		MCVAR	17.38	72.12	130.76	63.44	20.91	19.53	19.35	20.23
		MCMSE	17.39	72.35	141.64	77.52	20.91	19.53	19.35	20.23
		AAVAR	13.96	57.95	52.19	20.58	15.36	13.21	15.06	10.78
	400	BIAS	-0.01	0.00	0.05	0.08	0.00	0.00	0.00	0.00
		MCVAR	3.67	15.05	33.64	12.94	4.57	4.48	4.42	4.47
		MCMSE	3.68	15.05	33.91	13.56	4.57	4.48	4.42	4.47
		AAVAR	3.47	13.01	26.59	8.39	3.95	3.91	4.19	3.49
	1600	BIAS	-0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00
		MCVAR	0.83	3.37	7.90	2.87	1.05	1.04	1.04	1.05
		MCMSE	0.84	3.37	7.90	2.89	1.05	1.04	1.04	1.05
		AAVAR	0.86	3.19	7.21	2.51	1.01	1.01	1.02	0.98
		AAVAR	0.73	3.01	7.02	2.37	0.89	0.89	0.91	0.86

Note: MC Results for the DGP3 with heterogeneous treatment and bounded Xs. The regression model is misspecified, but the propensity score model is correctly specified. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Ratio refers to the treated-control ratio. BIAS: average bias over Monte Carlo simulations. MCVAR: Monte Carlo Variance × 100. MCMSE: Monte Carlo Mean Squared Error × 100. AAVAR: Average of the estimated variances based on the asymptotic variance of the M-estimators × 100.

Table B7: MC Results for misspecified propensity score regression models, homogeneous Treatment, Bounded X

Ratio	N		REG	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2
1/3	100	BIAS	0.09	0.22	0.53	0.57	0.09	0.09	0.10	0.10
		MCVAR	15.14	602.52	144.62	63.96	20.87	18.03	17.79	19.04
		MCMSE	15.89	607.55	172.20	96.17	21.76	18.93	18.71	20.01
		AAVAR	12.44	1076.04	50.54	18.56	18.11	11.88	14.11	9.20
	400	BIAS	0.09	0.14	0.20	0.22	0.10	0.10	0.10	0.10
		MCVAR	3.20	60.25	41.69	14.28	4.44	4.33	4.28	4.38
		MCMSE	4.05	62.14	45.66	19.32	5.38	5.28	5.26	5.37
		AAVAR	3.09	52.85	30.62	7.82	3.90	3.82	4.20	3.21
	1600	BIAS	0.10	0.12	0.13	0.13	0.09	0.09	0.09	0.09
		MCVAR	0.77	13.42	10.17	3.18	1.12	1.11	1.10	1.11
		MCMSE	1.68	14.77	11.88	4.92	2.00	1.99	2.00	2.01
		AAVAR	0.77	13.01	9.48	2.63	1.02	1.02	1.05	0.97
1/1	100	BIAS	0.09	0.14	0.24	0.30	0.10	0.10	0.10	0.10
		MCVAR	7.77	53.18	43.59	20.50	9.96	9.63	9.36	9.44
		MCMSE	8.66	55.15	49.25	29.76	10.90	10.58	10.36	10.44
		AAVAR	7.18	47.94	30.50	11.37	7.95	7.70	8.69	6.72
	400	BIAS	0.10	0.12	0.14	0.15	0.10	0.10	0.10	0.10
		MCVAR	1.85	8.84	8.47	4.27	2.27	2.27	2.26	2.27
		MCMSE	2.80	10.34	10.47	6.49	3.25	3.24	3.24	3.25
		AAVAR	1.80	8.31	7.86	3.65	2.12	2.11	2.19	2.04
	1600	BIAS	0.09	0.12	0.13	0.12	0.09	0.09	0.09	0.09
		MCVAR	0.46	1.92	1.89	1.02	0.58	0.58	0.59	0.59
		MCMSE	1.33	3.37	3.49	2.41	1.47	1.47	1.48	1.48
		AAVAR	0.45	1.89	1.85	0.94	0.53	0.53	0.54	0.53
3/1	100	BIAS	0.09	0.15	0.51	0.56	0.10	0.10	0.10	0.10
		MCVAR	15.46	83.32	145.58	64.52	20.75	18.40	18.20	19.38
		MCMSE	16.30	85.67	171.82	95.88	21.73	19.38	19.22	20.44
		AAVAR	12.40	62.83	51.35	18.42	15.46	11.76	13.97	9.02
	400	BIAS	0.09	0.12	0.19	0.22	0.10	0.10	0.10	0.10
		MCVAR	3.29	17.55	42.90	14.74	4.71	4.55	4.47	4.57
		MCMSE	4.18	18.96	46.61	19.69	5.66	5.52	5.49	5.59
		AAVAR	3.10	14.77	30.77	7.89	3.98	3.88	4.27	3.23
	1600	BIAS	0.09	0.10	0.12	0.12	0.09	0.09	0.09	0.09
		MCVAR	0.80	3.65	9.30	2.84	1.14	1.13	1.14	1.14
		MCMSE	1.66	4.72	10.67	4.32	2.00	2.00	2.00	2.00
		AAVAR	0.78	3.77	9.60	2.65	1.05	1.05	1.09	0.99

Note: MC Results for the DGP4 with homogeneous treatment and bounded Xs. Both regression and propensity score models are misspecified. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Ratio refers to the treated-control ratio. BIAS: average bias over Monte Carlo simulations. MCVAR: Monte Carlo Variance × 100. MCMSE: Monte Carlo Mean Squared Error × 100. AAVAR: Average of the estimated variances based on the asymptotic variance of the M-estimators × 100.

Table B8: MC Results for misspecified propensity score and regression models, heterogeneous Treatment, Bounded X

Ratio	N		REG	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2
1/3	100	BIAS	0.16	0.27	0.52	0.56	0.18	0.18	0.18	0.18
		MCVAR	23.87	344.78	117.94	62.88	32.64	28.31	28.10	29.96
		MCMSE	26.52	351.85	144.49	94.20	35.77	31.43	31.32	33.33
		AAVAR	19.08	380.83	45.97	23.76	28.01	18.18	21.44	14.31
	400	BIAS	0.17	0.20	0.26	0.28	0.17	0.17	0.17	0.17
		MCVAR	4.82	45.89	32.25	14.29	6.63	6.44	6.35	6.46
		MCMSE	7.59	50.05	38.77	21.85	9.52	9.35	9.33	9.46
		AAVAR	4.75	41.28	24.80	9.04	5.84	5.74	6.34	4.84
	1600	BIAS	0.16	0.19	0.20	0.20	0.16	0.16	0.16	0.16
		MCVAR	1.18	9.96	7.59	3.06	1.54	1.54	1.56	1.57
		MCMSE	3.81	13.47	11.54	6.91	4.20	4.20	4.23	4.25
		AAVAR	1.19	9.89	7.31	2.81	1.52	1.52	1.58	1.45
1/1	100	BIAS	0.14	0.19	0.28	0.34	0.15	0.15	0.15	0.15
		MCVAR	10.94	44.48	43.14	23.12	13.41	13.00	12.74	12.86
		MCMSE	13.01	48.14	50.94	34.44	15.64	15.25	15.07	15.20
		AAVAR	10.24	38.01	30.34	14.00	11.09	10.76	11.94	9.60
	400	BIAS	0.14	0.16	0.18	0.18	0.14	0.14	0.14	0.14
		MCVAR	2.77	7.56	8.11	4.80	3.20	3.19	3.18	3.19
		MCMSE	4.75	10.12	11.23	8.19	5.12	5.12	5.14	5.14
		AAVAR	2.57	7.54	7.98	4.31	2.92	2.91	3.00	2.83
	1600	BIAS	0.14	0.17	0.17	0.16	0.14	0.14	0.14	0.14
		MCVAR	0.65	1.89	1.97	1.16	0.74	0.74	0.73	0.74
		MCMSE	2.65	4.63	4.98	3.88	2.75	2.75	2.75	2.75
		AAVAR	0.65	1.76	1.90	1.12	0.74	0.74	0.74	0.73
3/1	100	BIAS	0.11	0.18	0.54	0.58	0.12	0.12	0.12	0.12
		MCVAR	18.12	97.78	151.40	67.89	24.02	21.13	20.83	21.95
		MCMSE	19.39	100.86	180.64	102.09	25.38	22.54	22.32	23.50
		AAVAR	14.60	79.31	53.03	19.91	19.08	13.99	16.25	11.17
	400	BIAS	0.12	0.13	0.21	0.23	0.11	0.11	0.12	0.12
		MCVAR	3.86	17.81	43.10	14.93	5.16	5.03	4.92	5.01
		MCMSE	5.24	19.60	47.36	20.39	6.47	6.35	6.26	6.36
		AAVAR	3.65	15.40	32.80	8.57	4.47	4.39	4.79	3.76
	1600	BIAS	0.12	0.13	0.14	0.15	0.12	0.12	0.12	0.12
		MCVAR	0.87	4.20	11.07	3.41	1.20	1.19	1.20	1.20
		MCMSE	2.25	5.85	13.09	5.70	2.58	2.58	2.60	2.60
		AAVAR	0.90	3.95	9.71	2.75	1.17	1.16	1.19	1.11

Note: MC Results for the DGP4 with heterogeneous treatment and bounded Xs. Both regression and propensity score models are misspecified. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Ratio refers to the treated-control ratio. BIAS: average bias over Monte Carlo simulations. MCVAR: Monte Carlo Variance × 100. MCMSE: Monte Carlo Mean Squared Error × 100. AAVAR: Average of the estimated variances based on the asymptotic variance of the M-estimators × 100.

Table B9: MC Results: DGP1, Unbounded X, Control-Treated Ratio 1:1

			Homogeneous Treatment								Heterogeneous Treatment							
			REG	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2	REG	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2
Entire	100	BIAS	0.000	0.005	0.058	0.080	0.000	-0.001	-0.001	-0.001	0.000	0.034	0.123	0.168	0.000	-0.001	-0.001	-0.001
Sample		VAR	0.059	1.477	0.787	0.476	0.091	0.078	0.073	0.074	0.119	1.723	0.826	0.470	0.151	0.138	0.133	0.134
		MSE	0.059	1.477	0.790	0.483	0.091	0.078	0.073	0.074	0.119	1.724	0.841	0.498	0.151	0.138	0.133	0.134
	400	BIAS	0.000	-0.002	0.015	0.036	0.000	0.000	0.000	0.000	0.000	0.002	0.031	0.072	0.000	0.000	0.000	0.000
		VAR	0.015	0.362	0.256	0.156	0.022	0.021	0.019	0.019	0.029	0.466	0.288	0.146	0.037	0.035	0.034	0.034
		MSE	0.015	0.362	0.256	0.157	0.022	0.021	0.019	0.019	0.029	0.466	0.289	0.151	0.037	0.035	0.034	0.034
	1600	BIAS	0.000	-0.005	0.002	0.008	0.000	0.000	0.000	0.000	0.000	-0.006	0.003	0.022	0.000	0.000	0.000	0.000
		VAR	0.003	0.067	0.063	0.049	0.005	0.005	0.004	0.004	0.006	0.085	0.071	0.045	0.009	0.009	0.008	0.008
		MSE	0.003	0.067	0.063	0.049	0.005	0.005	0.004	0.004	0.006	0.086	0.071	0.046	0.009	0.009	0.008	0.008
Trim 1	100	BIAS	-0.001	-0.002	0.021	0.017	-0.001	-0.001	-0.001	-0.001	-0.004	0.018	0.045	0.036	-0.004	-0.004	-0.004	-0.004
		VAR	0.070	0.511	0.484	0.494	0.074	0.074	0.075	0.075	0.179	0.502	0.525	0.520	0.183	0.183	0.183	0.184
		MSE	0.070	0.511	0.484	0.494	0.074	0.074	0.075	0.075	0.179	0.503	0.527	0.521	0.183	0.183	0.183	0.184
	400	BIAS	0.000	-0.004	0.006	0.006	0.000	0.000	0.000	0.000	0.000	0.001	0.014	0.016	0.001	0.001	0.001	0.001
		VAR	0.015	0.119	0.110	0.107	0.018	0.018	0.018	0.018	0.044	0.140	0.129	0.119	0.046	0.046	0.046	0.046
		MSE	0.015	0.119	0.110	0.107	0.018	0.018	0.018	0.018	0.044	0.140	0.130	0.119	0.046	0.046	0.046	0.046
	1600	BIAS	0.000	-0.007	-0.002	0.000	0.000	0.000	0.000	0.000	0.000	-0.008	-0.002	0.002	0.000	0.000	0.000	0.000
		VAR	0.003	0.038	0.035	0.032	0.004	0.004	0.004	0.004	0.010	0.047	0.041	0.035	0.011	0.011	0.011	0.011
		MSE	0.003	0.038	0.035	0.032	0.004	0.004	0.004	0.004	0.010	0.047	0.041	0.035	0.011	0.011	0.011	0.011
Trim 2	100	BIAS	-0.001	-0.011	-0.005	0.008	-0.002	-0.002	-0.002	-0.002	-0.001	-0.015	-0.002	0.022	-0.002	-0.002	-0.002	-0.002
		VAR	0.065	0.380	0.348	0.324	0.069	0.069	0.069	0.069	0.137	0.428	0.385	0.347	0.142	0.142	0.141	0.141
		MSE	0.065	0.380	0.348	0.324	0.069	0.069	0.069	0.069	0.137	0.428	0.385	0.348	0.142	0.142	0.141	0.141
	400	BIAS	0.000	-0.003	0.000	0.003	0.001	0.001	0.001	0.001	0.001	-0.001	0.004	0.009	0.002	0.002	0.002	0.002
		VAR	0.015	0.087	0.080	0.078	0.017	0.017	0.017	0.017	0.033	0.100	0.090	0.083	0.034	0.034	0.034	0.034
		MSE	0.015	0.087	0.080	0.078	0.017	0.017	0.017	0.017	0.033	0.100	0.090	0.083	0.034	0.034	0.034	0.034
	1600	BIAS	0.000	-0.003	0.001	0.002	0.000	0.000	0.000	0.000	0.000	-0.003	0.001	0.003	0.000	0.000	0.000	0.000
		VAR	0.003	0.022	0.020	0.020	0.003	0.003	0.003	0.003	0.007	0.025	0.023	0.021	0.008	0.008	0.008	0.008
		MSE	0.003	0.022	0.020	0.020	0.003	0.003	0.003	0.003	0.007	0.025	0.023	0.021	0.008	0.008	0.008	0.008

Note: MC Results for the DGP1 with homogeneous and heterogeneous treatment where control-treated ratio is 1:1 where Xs are drawn from a normal distribution. Both regression and propensity score models are correct. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Entire sample means that no trimming rule is applied. Trim1 and Trim2 are the first and second trimming rules described in the text.

Table B10: MC Results: DGP1, Unbounded X, Control-Treated Ratio 1:3

			Homogeneous Treatment								Heterogeneous Treatment							
			REG	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2	REG	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2
Entire Sample	100	BIAS	0.002	0.072	0.118	0.135	0.001	0.001	0.002	0.000	0.002	0.152	0.322	0.347	0.001	0.001	0.002	0.000
		VAR	0.107	4.157	1.751	1.441	0.173	0.139	0.133	0.142	0.164	3.908	1.374	0.895	0.232	0.197	0.191	0.200
		MSE	0.107	4.163	1.765	1.459	0.173	0.139	0.133	0.142	0.164	3.931	1.479	1.016	0.232	0.197	0.191	0.200
	400	BIAS	-0.003	0.026	0.045	0.058	-0.003	-0.003	-0.001	-0.001	-0.003	0.034	0.106	0.148	-0.003	-0.003	-0.001	-0.001
		VAR	0.024	1.137	0.641	0.463	0.049	0.038	0.034	0.035	0.039	1.309	0.573	0.278	0.063	0.053	0.049	0.050
		MSE	0.024	1.137	0.643	0.467	0.049	0.038	0.034	0.035	0.039	1.310	0.584	0.300	0.063	0.053	0.049	0.050
	1600	BIAS	0.001	0.000	0.006	0.018	0.001	0.000	0.000	0.000	0.001	-0.003	0.019	0.058	0.001	0.000	0.000	0.000
		VAR	0.005	0.251	0.207	0.148	0.011	0.010	0.009	0.009	0.009	0.300	0.206	0.088	0.015	0.014	0.013	0.013
		MSE	0.005	0.251	0.207	0.148	0.011	0.010	0.009	0.009	0.009	0.300	0.206	0.092	0.015	0.014	0.013	0.013
Trim 1	100	BIAS	0.001	0.042	0.038	0.032	0.001	0.001	0.001	0.000	0.423	0.510	0.525	0.505	0.422	0.422	0.422	0.422
		VAR	0.110	1.073	1.072	1.146	0.115	0.116	0.117	0.119	0.264	0.758	0.796	0.806	0.269	0.269	0.271	0.273
		MSE	0.110	1.074	1.073	1.147	0.115	0.116	0.117	0.119	0.444	1.018	1.073	1.062	0.448	0.448	0.450	0.451
	400	BIAS	-0.002	0.010	0.009	0.007	-0.003	-0.003	-0.003	-0.003	0.238	0.269	0.277	0.270	0.238	0.238	0.238	0.238
		VAR	0.023	0.326	0.307	0.310	0.029	0.028	0.029	0.029	0.066	0.265	0.235	0.209	0.072	0.072	0.073	0.073
		MSE	0.023	0.326	0.307	0.310	0.029	0.028	0.029	0.029	0.124	0.337	0.312	0.283	0.129	0.129	0.130	0.129
	1600	BIAS	0.000	0.012	0.011	0.012	0.001	0.001	0.001	0.001	0.128	0.147	0.150	0.150	0.129	0.129	0.129	0.129
		VAR	0.005	0.115	0.108	0.106	0.007	0.007	0.007	0.007	0.016	0.098	0.084	0.066	0.018	0.018	0.018	0.018
		MSE	0.005	0.115	0.108	0.106	0.007	0.007	0.007	0.007	0.032	0.120	0.107	0.089	0.035	0.035	0.035	0.035
Trim 2	100	BIAS	0.001	0.001	0.016	0.024	0.001	0.001	0.001	0.001	0.789	0.783	0.812	0.830	0.788	0.788	0.788	0.788
		VAR	0.094	0.692	0.686	0.669	0.099	0.098	0.098	0.099	0.231	0.618	0.633	0.575	0.235	0.235	0.235	0.235
		MSE	0.094	0.692	0.686	0.669	0.099	0.098	0.098	0.099	0.854	1.231	1.293	1.265	0.857	0.857	0.857	0.857
	400	BIAS	-0.002	-0.017	-0.011	-0.008	-0.002	-0.002	-0.002	-0.002	0.786	0.773	0.783	0.788	0.786	0.786	0.786	0.786
		VAR	0.021	0.162	0.157	0.154	0.023	0.023	0.023	0.023	0.055	0.145	0.146	0.134	0.057	0.057	0.057	0.057
		MSE	0.021	0.163	0.158	0.154	0.023	0.023	0.023	0.023	0.673	0.743	0.759	0.755	0.675	0.675	0.675	0.675
	1600	BIAS	0.000	0.000	0.001	0.002	0.000	0.000	0.000	0.000	0.789	0.789	0.792	0.793	0.789	0.789	0.789	0.789
		VAR	0.005	0.039	0.038	0.037	0.005	0.005	0.005	0.005	0.014	0.036	0.037	0.033	0.014	0.014	0.014	0.014
		MSE	0.005	0.039	0.038	0.037	0.005	0.005	0.005	0.005	0.637	0.659	0.664	0.662	0.638	0.638	0.638	0.638

Note: MC Results for the DGP1 with homogeneous and heterogeneous treatment where control-treated ratio is 1:3 where Xs are drawn from a normal distribution. Both regression and propensity score models are correct. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Entire sample means that no trimming rule is applied. Trim1 and Trim2 are the first and second trimming rules described in the text.

Table B11: MC Results: DGP1, Unbounded X, Control-Treated Ratio 3:1

			Homogeneous Treatment								Heterogeneous Treatment							
			REG	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2	REG	IPW1	IPW2	IPW3	DR1a	DR1b	DR1c	DR2
Entire Sample	100	BIAS	0.000	-0.039	0.103	0.114	-0.002	-0.001	0.000	0.000	0.000	-0.028	0.124	0.149	-0.002	-0.001	0.000	0.000
		VAR	0.108	4.086	1.753	1.434	0.171	0.138	0.134	0.143	0.167	4.330	1.880	1.548	0.230	0.197	0.193	0.202
		MSE	0.108	4.088	1.764	1.447	0.171	0.138	0.134	0.143	0.167	4.331	1.895	1.571	0.230	0.197	0.193	0.202
	400	BIAS	-0.001	-0.041	0.014	0.034	-0.001	-0.001	-0.001	-0.001	-0.001	-0.041	0.015	0.047	-0.001	-0.001	-0.001	-0.001
		VAR	0.024	0.968	0.643	0.497	0.040	0.035	0.033	0.034	0.039	1.016	0.667	0.519	0.055	0.050	0.048	0.049
		MSE	0.024	0.970	0.643	0.498	0.040	0.035	0.033	0.034	0.039	1.017	0.667	0.521	0.055	0.050	0.048	0.049
	1600	BIAS	0.000	-0.011	0.012	0.029	-0.003	-0.002	0.000	0.000	0.000	-0.009	0.014	0.034	-0.003	-0.002	0.000	0.000
		VAR	0.005	0.240	0.221	0.154	0.010	0.009	0.008	0.008	0.008	0.250	0.225	0.157	0.013	0.012	0.011	0.011
		MSE	0.005	0.240	0.221	0.155	0.010	0.009	0.008	0.008	0.008	0.250	0.226	0.159	0.013	0.012	0.011	0.011
Trim 1	100	BIAS	-0.001	-0.035	0.049	0.038	0.000	0.000	0.000	0.000	-0.425	-0.445	-0.364	-0.373	-0.424	-0.424	-0.425	-0.424
		VAR	0.111	1.155	1.066	1.136	0.117	0.117	0.119	0.120	0.269	1.168	1.244	1.304	0.274	0.274	0.276	0.278
		MSE	0.111	1.157	1.068	1.137	0.117	0.117	0.119	0.120	0.450	1.367	1.376	1.444	0.454	0.455	0.457	0.458
	400	BIAS	-0.003	-0.010	0.014	0.011	-0.004	-0.004	-0.004	-0.004	-0.246	-0.252	-0.226	-0.228	-0.247	-0.247	-0.247	-0.247
		VAR	0.023	0.327	0.307	0.312	0.027	0.027	0.027	0.027	0.065	0.360	0.371	0.371	0.070	0.070	0.070	0.070
		MSE	0.023	0.327	0.307	0.312	0.027	0.027	0.027	0.027	0.126	0.424	0.422	0.423	0.131	0.131	0.131	0.131
	1600	BIAS	0.000	0.000	0.012	0.013	0.001	0.001	0.000	0.000	-0.130	-0.131	-0.118	-0.116	-0.130	-0.130	-0.130	-0.130
		VAR	0.005	0.106	0.103	0.099	0.007	0.007	0.007	0.007	0.017	0.121	0.121	0.117	0.019	0.019	0.019	0.019
		MSE	0.005	0.106	0.103	0.099	0.007	0.007	0.007	0.007	0.034	0.138	0.135	0.130	0.036	0.036	0.036	0.036
Trim 2	100	BIAS	-0.001	0.010	0.011	0.020	0.000	0.000	0.000	0.000	-0.786	-0.774	-0.769	-0.749	-0.785	-0.785	-0.785	-0.785
		VAR	0.093	0.765	0.664	0.648	0.097	0.097	0.097	0.097	0.236	0.880	0.787	0.765	0.240	0.240	0.240	0.241
		MSE	0.093	0.765	0.664	0.649	0.097	0.097	0.097	0.097	0.854	1.480	1.379	1.326	0.856	0.856	0.857	0.857
	400	BIAS	0.000	-0.002	0.003	0.007	0.000	0.000	0.000	0.000	-0.782	-0.786	-0.778	-0.770	-0.782	-0.782	-0.782	-0.782
		VAR	0.021	0.173	0.156	0.153	0.022	0.022	0.022	0.022	0.055	0.206	0.189	0.183	0.057	0.057	0.057	0.057
		MSE	0.021	0.173	0.156	0.153	0.022	0.022	0.022	0.022	0.667	0.824	0.795	0.777	0.669	0.669	0.668	0.668
	1600	BIAS	0.000	-0.006	-0.008	-0.007	0.000	0.000	0.000	0.000	-0.787	-0.795	-0.796	-0.794	-0.786	-0.786	-0.786	-0.786
		VAR	0.004	0.042	0.037	0.036	0.005	0.005	0.005	0.005	0.013	0.048	0.045	0.043	0.013	0.013	0.013	0.013
		MSE	0.004	0.042	0.037	0.037	0.005	0.005	0.005	0.005	0.632	0.681	0.680	0.674	0.633	0.633	0.633	0.633

Note: MC Results for the DGP1 with homogeneous and heterogeneous treatment where control-treated ratio is 3:1 where Xs are drawn from a normal distribution. Both regression and propensity score models are correct. Meanwhile, 16000, 4000 and 1000 Monte Carlo replications are used for the sample sizes 100, 400 and 1600, respectively. Entire sample means that no trimming rule is applied. Trim1 and Trim2 are the first and second trimming rules described in the text.