Covariate Balancing and the Equivalence of Weighting and Doubly Robust Estimators of Average Treatment Effects*

Tymon Słoczyński
Brandeis University

S. Derya Uysal
Ludwig Maximilian University of Munich

Jeffrey M. Wooldridge
Michigan State University

This version: October 27, 2023

Abstract: We show that when the propensity score is estimated using a suitable covariate balancing procedure, the commonly used inverse probability weighting (IPW) estimator, augmented inverse probability weighting (AIPW) with linear conditional mean, and inverse probability weighted regression adjustment (IPWRA) with linear conditional mean are all numerically the same for estimating the average treatment effect (ATE) or the average treatment effect on the treated (ATT). Further, suitably chosen covariate balancing weights are automatically normalized, which means that normalized and unnormalized versions of IPW and AIPW are identical. For estimating the ATE, the weights that achieve the algebraic equivalence of IPW, AIPW, and IPWRA are based on propensity scores estimated using the inverse probability tilting (IPT) method of Graham, Pinto and Egel (2012). For the ATT, the weights are obtained using the covariate balancing propensity score (CBPS) method developed in Imai and Ratkovic (2014). These equivalences also make covariate balancing methods attractive when the treatment is confounded and one is interested in the local average treatment effect.

*For helpful comments, we thank Pedro Sant'Anna.

1 Introduction

Treatment effect estimation methods that directly use an estimated propensity score (PS) have historically relied on maximum likelihood estimation (MLE) of a standard binary response model, such as logit or probit. While MLE is natural when the goal is to estimate the parameters in the propensity score model, it is not necessarily preferred when the PS estimators are used to obtain an object of interest such as the average treatment effect (ATE) or the average treatment effect on the treated (ATT). In fact, when one allows that the PS model might be misspecified, one can obtain estimates of average treatment effects that are more resilient using PS estimates other than the MLE. Graham et al. (2012) introduce the method of "inverse probability tilting" (IPT), where the parameters in the PS model are obtained from moment conditions that "balance" the means of the control variables, $\mathbf{X}$. Because two potential outcome means are used to obtain the ATE, different PS parameters are estimated using the treated and untreated units. Graham et al. (2012) show that the IPT estimator has a satisfying asymptotic local efficiency property. For our purposes, the asymptotic properties of the IPT estimator are not directly relevant, as we derive algebraic equivalences when IPT weights are used to obtain popular ATE estimators. These are the usual inverse probability weighting (IPW) estimators that originated with Horvitz and Thompson (1952), augmented inverse probability weighting (AIPW) with linear mean functions (Robins, Rotnitzky and Zhao, 1994), and inverse probability weighted regression adjustment (IPWRA) with linear mean functions (Hirano and Imbens, 2001; Wooldridge, 2007; Słoczyński and Wooldridge, 2018).

At this point, it is useful to head off a point of possible confusion. In studying higher-order properties of IPT and competitors, Graham et al. (2012) present a framework that includes IPT and the typical AIPW and IPWRA estimators. Importantly, the latter two are based on the PS estimated by MLE, and so the three estimators are generally different (and AIPW is not normalized). Here we establish that if the same IPT weights are used in all estimators, there is no algebraic difference between them.

An alternative to IPT for estimating the ATE was proposed by Imai and Ratkovic (2014). Rather than estimating separate PS parameters for the treated and untreated units, Imai and Ratkovic (2014) obtain a single set of parameters—just like the MLE approach. The estimates are chosen so that weighted averages of the covariates across treated and untreated groups are the same. Imai and Ratkovic (2014) refer to the resulting estimates of the PS as the "covariate balancing propensity score" (CBPS) estimates. It turns out that the CBPS estimates do not lead to normalized weights when used in IPW, and the IPW, AIPW, and IPWRA procedures are not generally identical. Nevertheless, Imai and Ratkovic (2014) also propose CBPS weights for estimating the ATT. The estimated PS parameters are chosen so that a weighted average of the covariates over the untreated units equals the average of the covariates over the treated units. This property leads to normalized weights for estimating the ATT. As we show in Section 3, it also leads to equivalence of IPW, AIPW, and IPWRA when all three use CBPS weights and the latter two use a linear regression function for the outcome.

The rest of the paper is organized as follows. In Section 2, we review the estimation problems solved by the IPT weights for the ATE and the CBPS weights for the ATT, and discuss some simple implications. In Section 3, we derive the equivalences among various IPT-based estimators of the ATE and then the CBPS-based estimators of the ATT. In Section 4, we briefly discuss the implications that the algebraic equivalence results have for estimating the local average treatment effect (LATE) and the local average treatment effect on the treated (LATT). We emphasize that, because in all cases we are establishing numerical equivalences, we need not, and do not, state assumptions under which the estimators are consistent. These have been covered elsewhere and are well known.

2 Balancing Propensity Scores

In a general missing data setting, Graham et al. (2012) introduced exponential tilting as a way of estimating the propensity score along with other parameters of interest in the context of missing data. Imai and Ratkovic (2014) introduced the covariate balancing propensity score in the context of treatment effect estimation. In this section we review the estimation problems solved by these methods in the binary treatment case.

Let W be the binary treatment indicator and define the propensity score $\mathrm{P}(W = 1|\mathbf{X} = \mathbf{x})$. Assume an index model, $p(\mathbf{x}\gamma)$, where $\mathbf{x}$ is $1\times K$, γ is $K\times 1$, and $x_1 \equiv 1$ so that the linear index always includes an intercept. In most applications, $p(\cdot)$ is the logistic function, and this was assumed in Imai and Ratkovic (2014). The potential outcomes are $Y(0)$ and $Y(1)$. Recall that the ATE and the ATT are defined as

$$\tau_{ate} \;=\; \mathrm{E}\left[Y(1) - Y(0)\right]$$

and

$$\tau_{att} \;=\; \mathrm{E}\left[Y(1) - Y(0)|W = 1\right].$$

However, we emphasize that the results of this paper have to do with algebraic equivalences, and so we do not discuss identification of population parameters.

When $p(\mathbf{x}\gamma) = \exp(\mathbf{x}\gamma)/[1+\exp(\mathbf{x}\gamma)]$, given a sample of size N, it is well known that the first-order condition solved by the maximum likelihood estimator, $\hat{\gamma}_{mle}$, is

$$\sum_{i=1}^{N} \mathbf{X}_i'\left[W_i - p(\mathbf{X}_i\hat{\gamma}_{mle})\right] \;=\; \mathbf{0}. \tag{2.1}$$

If we define residuals as $W_i - p(\mathbf{X}_i\hat{\gamma}_{mle})$, conditions (2.1) imply that the residuals always average to zero and are uncorrelated with each $X_{j,i}$ in the sample—properties shared by the OLS estimator of a linear model. (Typically, we would think of having a random sample from the population; but, again, the results in this paper are algebraic, and the mechanism generating the data plays no role.)

The inverse probability tilting (IPT) moment conditions proposed by Graham et al. (2012) are different. When W is a selection indicator meaning that the elements of $\mathbf{X}$ are observed, Graham et al. (2012) propose using the moment conditions

$$\mathrm{E}\left[\frac{W}{p(\mathbf{X}\gamma)}\mathbf{X}'\right] = \mathrm{E}(\mathbf{X}'), \tag{2.2}$$

which follow immediately by iterated expectations when $p(\mathbf{X}\gamma) = \mathrm{P}(W=1|\mathbf{X}) = \mathrm{E}(W|\mathbf{X})$. (If we were considering identification, we would need to assume, at a minimum, that $p(\mathbf{X}\gamma) > 0$ with probability one.) The sample analog of (2.2) is

$$N^{-1}\sum_{i=1}^{N}\frac{W_i\mathbf{X}_i}{p(\mathbf{X}_i\hat{\gamma}_{1,ate,ipt})} = \bar{\mathbf{X}}, \tag{2.3}$$

and these equations define the IPT estimator of γ, $\hat{\gamma}_{1,ate,ipt}$. Note that we have put a "1" subscript on $\hat{\gamma}_{1,ate,ipt}$ because, in the treatment effects setting, there is another set of moment conditions that leads to a different IPT estimator of γ. Again by iterated expectations,

$$\mathrm{E}\left[\frac{1-W}{1-p(\mathbf{X}\gamma)}\mathbf{X}'\right] = \mathrm{E}(\mathbf{X}'),$$

and this leads to the sample analog:

$$N^{-1}\sum_{i=1}^{N}\frac{(1-W_i)\,\mathbf{X}_i}{1-p(\mathbf{X}_i\hat{\gamma}_{0,ate,ipt})} = \bar{\mathbf{X}}.$$

In general, $\hat{\gamma}_{0,ate,ipt} \neq \hat{\gamma}_{1,ate,ipt}$. However, because $1 \in \mathbf{X}_i$, it follows immediately that

$$\sum_{i=1}^{N}\frac{W_i}{p(\mathbf{X}_i\hat{\gamma}_{1,ate,ipt})} = N \tag{2.4}$$

and

$$\sum_{i=1}^{N}\frac{1-W_i}{1-p(\mathbf{X}_i\hat{\gamma}_{0,ate,ipt})} = N. \tag{2.5}$$

These two equations are key, as the summands in (2.4) are the weights for estimating $\mathrm{E}[Y(1)]$ in IPW estimation and those in (2.5) are the weights used in estimating $\mathrm{E}[Y(0)]$, as we review in Section 3. Equations (2.4) and (2.5) show that the IPT weights are automatically normalized (or self-balancing, or stabilized) for estimating the ATE. That is, the sample mean of the weights is not stochastic but instead equal to one by construction.

Imai and Ratkovic (2014) use different moment conditions to obtain a single estimator of γ when the focus is on the ATE. Again, if the propensity score is correctly specified then, by iterated expectations,

$$\mathrm{E}(\mathbf{X}') \;=\; \mathrm{E}\left[\frac{W}{p(\mathbf{X}\gamma)}\mathbf{X}'\right] \;=\; \mathrm{E}\left[\frac{1-W}{1-p(\mathbf{X}\gamma)}\mathbf{X}'\right]. \tag{2.6}$$

Rather than using the implications of (2.6) separately, which is what IPT does, Imai and Ratkovic (2014) use the second equality to obtain the following sample moment conditions:

$$N^{-1}\sum_{i=1}^{N}\frac{W_i}{p(\mathbf{X}_i\hat{\gamma}_{ate,cbps})}\mathbf{X}_i' \;=\; N^{-1}\sum_{i=1}^{N}\frac{1-W_i}{1-p(\mathbf{X}_i\hat{\gamma}_{ate,cbps})}\mathbf{X}_i'. \tag{2.7}$$

After simple algebra, the moment conditions can be expressed as

$$\sum_{i=1}^{N}\left(\frac{W_i - p(\mathbf{X}_i\hat{\gamma}_{ate,cbps})}{p(\mathbf{X}_i\hat{\gamma}_{ate,cbps})\left[1-p(\mathbf{X}_i\hat{\gamma}_{ate,cbps})\right]}\right)\mathbf{X}_i' \;=\; \mathbf{0}. \tag{2.8}$$

Comparing (2.8) with (2.1), we can see that the CBPS approach effectively weights the MLE moment conditions by the estimated inverse conditional variance, $\mathrm{Var}(W_i|\mathbf{X}_i)$.

Because the first element of $\mathbf{X}_i$ is unity, (2.7) implies that

$$\sum_{i=1}^{N}\frac{W_i}{p(\mathbf{X}_i\hat{\gamma}_{ate,cbps})} \;=\; \sum_{i=1}^{N}\frac{1-W_i}{1-p(\mathbf{X}_i\hat{\gamma}_{ate,cbps})}. \tag{2.9}$$

Equation (2.9) shows that the weights appearing in the IPW estimates of $\mathrm{E}[Y(1)]$ and $\mathrm{E}[Y(0)]$ sum to the same value, but that common value is not necessarily the sample size, N. Therefore, when these are used as weights in IPW, the CBPS weights are not automatically normalized.

For estimating the ATT, the moment equations used by Imai and Ratkovic (2014) are

$$\mathrm{E}(\mathbf{X}'|W=1) \;=\; \frac{1}{\rho}\cdot\mathrm{E}(W\cdot\mathbf{X}') \;=\; \frac{1}{\rho}\cdot\mathrm{E}\left[\frac{p(\mathbf{X}\gamma)\,(1-W)}{1-p(\mathbf{X}\gamma)}\cdot\mathbf{X}'\right],$$

where $\rho = \mathrm{P}(W=1)$. Using $\hat{\rho} = N_1/N$, where N_1 is the number of treated units, the K sample moment conditions are

$$\begin{aligned} N_1^{-1}\sum_{i=1}^{N}W_i\mathbf{X}_i' \;&=\; \left(\frac{N_1}{N}\right)^{-1}N^{-1}\sum_{i=1}^{N}\frac{p(\mathbf{X}_i\hat{\gamma}_{att,cbps})\,(1-W_i)}{1-p(\mathbf{X}_i\hat{\gamma}_{att,cbps})}\cdot\mathbf{X}_i' \\ &=\; N_1^{-1}\sum_{i=1}^{N}\frac{p(\mathbf{X}_i\hat{\gamma}_{att,cbps})\,(1-W_i)}{1-p(\mathbf{X}_i\hat{\gamma}_{att,cbps})}\cdot\mathbf{X}_i' \end{aligned}$$

or

$$\bar{\mathbf{X}}_1' \;=\; N_1^{-1}\sum_{i=1}^{N}\frac{p(\mathbf{X}_i\hat{\gamma}_{att,cbps})\,(1-W_i)}{1-p(\mathbf{X}_i\hat{\gamma}_{att,cbps})}\cdot\mathbf{X}_i', \tag{2.10}$$

where $\bar{\mathbf{X}}_1 = N_1^{-1} \sum_{i=1}^{N} W_i \mathbf{X}_i$. Because $1 \in \mathbf{X}_i$, (2.10) implies

$$N_1 \;=\; \sum_{i=1}^{N} \frac{p(\mathbf{X}_i \hat{\gamma}_{att,cbps})\,(1-W_i)}{1-p(\mathbf{X}_i \hat{\gamma}_{att,cbps})},$$

which implies that the weights used in IPW estimation of the ATT sum to the number of treated units. In other words, the CBPS weights for estimating the ATT are automatically normalized.

3 Equivalence of Estimators

We now establish numerical equivalences of three different estimators that incorporate inverse probability weighting, starting with estimators of the ATE.

3.1 Estimators of the ATE

The three estimators we consider are among the most popular when unconfoundedness is assumed to hold: IPW, AIPW, and IPWRA. As we are establishing algebraic equivalence, we do not impose assumptions beyond those needed for existence of the estimates for a given sample. This simply means the estimated propensity scores are strictly between zero and one for all i.

The IPW estimator of τ_{ate} using the IPT weights is

$$\hat{\tau}_{ate,ipt} \;=\; \hat{\mu}_{1,ipt} \;-\; \hat{\mu}_{0,ipt} \;=\; N^{-1} \sum_{i=1}^{N} \frac{W_i Y_i}{p(\mathbf{X}_i \hat{\gamma}_{1,ipt})} \;-\; N^{-1} \sum_{i=1}^{N} \frac{(1-W_i)\,Y_i}{1-p(\mathbf{X}_i \hat{\gamma}_{0,ipt})}, \tag{3.11}$$

where "ate" has been dropped from the estimators of γ for notational simplicity and we use "ipt" to indicate that the IPT weights are being used. See, for example, Wooldridge (2010, Section 21.3) for a variant of this estimator with MLE-based weights and Graham et al. (2012) for IPT. We know by (2.4) and (2.5) that the weights in both of the weighted averages sum to unity and so the weights are already normalized.

The AIPW estimator with IPT weights, which we refer to as AIPT, is also the difference in estimates of $\mu_1 \equiv \mathrm{E}[Y(1)]$ and $\mu_0 \equiv \mathrm{E}[Y(0)]$; that is, $\hat{\tau}_{ate,aipt} = \hat{\mu}_{1,aipt} - \hat{\mu}_{0,aipt}$. For μ_1,

$$\hat{\mu}_{1,aipt} \;=\; N^{-1} \sum_{i=1}^{N} \frac{W_i \left(Y_i - \mathbf{X}_i \hat{\beta}_1\right)}{p(\mathbf{X}_i \hat{\gamma}_{1,ipt})} \;+\; N^{-1} \sum_{i=1}^{N} \mathbf{X}_i \hat{\beta}_1, \tag{3.12}$$

where remember that $1 \in \mathbf{X}_i$. Although it is not important for the equivalence result, the estimates $\hat{\beta}_1$ typically come from an OLS regression of Y_i on $\mathbf{X}_i$ using $W_i = 1$ (treated units). The first term in (3.12) is a weighted average of the resulting residuals over the treated units. The weights are exactly those appearing in $\hat{\mu}_{1,ipt}$ and are therefore normalized.

For μ_0, the AIPT estimator is

$$\hat{\mu}_{0,aipt} \;=\; N^{-1}\sum_{i=1}^{N}\frac{(1-W_i)\left(Y_i-\mathbf{X}_i\hat{\beta}_0\right)}{1-p(\mathbf{X}_i\hat{\gamma}_{0,ipt})} \;+\; N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\hat{\beta}_0, \tag{3.13}$$

where $\hat{\beta}_0$ are probably the OLS estimates from a regression of Y_i on $\mathbf{X}_i$ using $W_i = 0$.

The third estimator we consider is the IPWRA estimator with IPT weights, which we refer to as IPTRA. For μ_1, we first solve a weighted least squares (WLS) problem,

$$\min_{\mathbf{b}_1} N^{-1}\sum_{i=1}^{N}\frac{W_i}{\hat{p}_i}\left(Y_i-\mathbf{X}_i\mathbf{b}_1\right)^2, \tag{3.14}$$

where $\hat{p}_i = p(\mathbf{X}_i\hat{\gamma}_{1,ipt})$ are the IPT propensity score estimates. Given the WLS estimates $\tilde{\beta}_1$ from (3.14), μ_1 is estimated by averaging the fitted values across all observations, as in the case of linear regression adjustment:

$$\hat{\mu}_{1,iptra} \;=\; N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\tilde{\beta}_1 \;=\; \bar{\mathbf{X}}\tilde{\beta}_1. \tag{3.15}$$

The IPTRA estimator of μ_0, $\hat{\mu}_{0,iptra}$, uses the untreated units with weights $(1-\hat{p}_i)^{-1}$, and produces $\tilde{\beta}_0$. The final IPTRA estimator of the ATE is given by $\hat{\tau}_{ate,iptra} = \hat{\mu}_{1,iptra} - \hat{\mu}_{0,iptra} = \bar{\mathbf{X}}\tilde{\beta}_1 - \bar{\mathbf{X}}\tilde{\beta}_0$.

When the PS weights are obtained using MLE, or some other method of moments procedure, $\hat{\tau}_{ate,ipw}$, $\hat{\tau}_{ate,aipw}$, and $\hat{\tau}_{ate,ipwra}$ are generally different. In fact, $\hat{\tau}_{ate,ipw}$ and $\hat{\tau}_{ate,aipw}$ do not generally use normalized weights, and so one could have five different estimates using the same estimated PS weights: IPW, normalized IPW (NIPW), AIPW, normalized AIPW (NAIPW), and IPWRA. (IPWRA is always normalized.) Strikingly, when IPT weights are used instead, all of the estimates are the same.

Proposition 3.1. *Let $\hat{\gamma}_{1,ipt}$ be the estimates from the IPT estimation in equation* (2.3)*, with $\hat{p}_i = p(\mathbf{X}_i\hat{\gamma}_{1,ipt}) > 0$ for all i. Then $\hat{\mu}_{1,ipt}$, $\hat{\mu}_{1,aipt}$, and $\hat{\mu}_{1,iptra}$ are numerically identical. The same is true of $\hat{\mu}_{0,ipt}$, $\hat{\mu}_{0,aipt}$, and $\hat{\mu}_{0,iptra}$, which means that*

$$\hat{\tau}_{ate,ipt} \;=\; \hat{\tau}_{ate,aipt} \;=\; \hat{\tau}_{ate,iptra}.$$

Proof. See Appendix A. □

The implication of Proposition 3.1 is that, if one uses the IPT weights in estimating both μ_0 and μ_1, where conditional means $\mathrm{E}[Y(0)|\mathbf{X}]$ and $\mathrm{E}[Y(1)|\mathbf{X}]$ are modeled linear, then three commonly used estimators of the ATE are numerically identical; moreover, the IPW and AIPW versions are automatically normalized.

3.2 Estimators of the ATT

We now establish equivalence of several common estimators of the ATT when the CBPS weights are used. Recall that

$$\tau_{att} \ = \ \mathrm{E}[Y(1)|W=1] - \mathrm{E}[Y(0)|W=1] \ \equiv \ \mu_{1|1} - \mu_{0|1},$$

and the first term is always consistently estimated using the sample mean of Y_i over the treated units, $\bar{Y}_1$. The IPW estimator for the second term, using the CBPS weights (which are automatically normalized), is

$$\hat{\mu}_{0|1,ipwcbps} \ = \ N_1^{-1} \sum_{i=1}^{N} \frac{p(\mathbf{X}_i \hat{\gamma}_{cbps})\,(1-W_i)\,Y_i}{1-p(\mathbf{X}_i \hat{\gamma}_{cbps})}. \tag{3.16}$$

As mentioned in Section 2, the weights in (3.16) sum to unity. These are the same weights that appear in the AIPW estimator, and so the unnormalized and normalized AIPW estimators are also the same. Specifically, the AIPW estimator of $\mu_{0|1}$ is

$$\begin{aligned}
\hat{\mu}_{0|1,aipwcbps} \ &= \ N_1^{-1} \sum_{i=1}^{N} \frac{\hat{p}_i\,(1-W_i)}{1-\hat{p}_i} \left(Y_i - \mathbf{X}_i \hat{\beta}_0\right) \ + \ N_1^{-1} \sum_{i=1}^{N} W_i \mathbf{X}_i \hat{\beta}_0 \\
&= \ N_1^{-1} \sum_{i=1}^{N} \frac{\hat{p}_i\,(1-W_i)\,Y_i}{1-\hat{p}_i} \ - \ N_1^{-1} \sum_{i=1}^{N} \frac{\hat{p}_i\,(1-W_i)\,\mathbf{X}_i \hat{\beta}_0}{1-\hat{p}_i} \ + \ \bar{\mathbf{X}}_1 \hat{\beta}_0,
\end{aligned} \tag{3.17}$$

where $\hat{p}_i = p(\mathbf{X}_i \hat{\gamma}_{cbps})$ are now the CBPS propensity score estimates and $\hat{\beta}_0$ is typically the OLS estimator from regressing Y_i on $\mathbf{X}_i$ using $W_i = 0$.

Finally, the IPWRA estimator of $\mu_{0|1}$ is

$$\hat{\mu}_{0|1,ipwracbps} \ = \ \bar{\mathbf{X}}_1 \tilde{\beta}_0,$$

where $\tilde{\beta}_0$ now solves the WLS problem:

$$\min_{\mathbf{b}_0} \ N^{-1} \sum_{i=1}^{N} \frac{\hat{p}_i\,(1-W_i)}{1-\hat{p}_i} \left(Y_i - \mathbf{X}_i \mathbf{b}_0\right)^2, \tag{3.18}$$

where $\hat{p}_i = p(\mathbf{X}_i \hat{\gamma}_{cbps})$. We have the following equivalence result.

Proposition 3.2. *Let $\hat{\gamma}_{cbps}$ be the estimators solving* (2.10) *with $\hat{p}_i = p(\mathbf{X}_i \hat{\gamma}_{cbps}) < 1$ for all i. Then the IPW, AIPW, and IPWRA estimates of $\mu_{0|1}$ using the CBPS weights, and linear conditional means in the latter two cases, are identical. Therefore, the three estimates of τ_{att} are identical.*

Proof. See Appendix A. □

4 Implications for Estimators of the LATE and LATT

The previous results for the ATE and ATT have implications for estimators of the LATE and LATT when using control variables $\mathbf{X}$; a recent treatment is Słoczyński, Uysal and Wooldridge (2022), which we follow here. As before, W is a binary treatment. Now we also have a binary instrumental variable, Z.

As discussed in Słoczyński et al. (2022), many estimators of the LATE are ratios of estimators of the ATE,

$$\hat{\tau}_{late} \;=\; \frac{\hat{\tau}_{ate,Y|Z}}{\hat{\tau}_{ate,W|Z}},$$

where $\hat{\tau}_{ate,Y|Z}$ is an estimator of the ATE where Y is the outcome, Z plays the role of the treatment, and the covariates $\mathbf{X}$ are used to account for confounders of Z. Again, we are only concerned with equivalences and not statistical properties. The denominator, $\hat{\tau}_{ate,W|Z}$, is an estimated ATE where W is the outcome and Z again is the treatment indicator, with covariates $\mathbf{X}$. It follows from Proposition 3.1 that when linear conditional means are used for both Y and W, and IPT is used for the PS weights, estimators of the LATE based on IPW, AIPW, and IPWRA are all identical. The propensity score weights in this case, for both the numerator and the denominator, are for the instrument propensity score:

$$\mathrm{P}(Z=1|\mathbf{X}) \;=\; q(\mathbf{X}\delta).$$

It should be noted, however, that unlike in the case of the ATE, where the nature of Y is unspecified, here it may be impractical to use the linear model in the denominator given the binary nature of W. See Słoczyński et al. (2022) for using other doubly robust estimators to exploit the binary nature of W and maybe special features of Y.

We do not have a general equivalence of the different estimators using CBPS weights, and the normalized and unnormalized estimators of the ATE are generally different. However, in the case of the LATE, the weights used in the numerator and denominator are the same—namely, $\frac{Z_i}{q(\mathbf{X}_i\hat{\delta}_{ate,cbps})}$ and $\frac{1-Z_i}{1-q(\mathbf{X}_i\hat{\delta}_{ate,cbps})}$. By (2.9), with the obvious change in notation, the weights for $Z_i=1$ and $Z_i=0$ sum to the same value (but not unity in general). We can factor this common value out of the expressions for the ATEs in both the numerator and denominator, and then cancel. It follows that the IPW and NIPW estimators of the LATE are identical, as previously observed by Heiler (2022), and the same is true of AIPW and NAIPW. (Recall that IPWRA is implicitly normalized.)

Estimators of the LATT that incorporate control variables $\mathbf{X}$ can be written as the ratio of estimators of the ATT, where the instrument plays the role of the treatment variable:

$$\hat{\tau}_{latt} \;=\; \frac{\hat{\tau}_{att,Y|Z}}{\hat{\tau}_{att,W|Z}},$$

where $\hat{\tau}_{att,Y|Z}$ and $\hat{\tau}_{att,W|Z}$ are both estimators of the ATT with "treatment" variable Z and outcome variables Y and W, respectively. If these estimators use the appropriate CBPS weights, as in Proposition 3.2, then it follows immediately that the versions of $\hat{\tau}_{latt}$ based on IPW, AIPW with linear regression functions, and IPWRA with linear regression functions are all numerically the same. Also, recall that the normalized and unnormalized estimators of the ATTs are identical when using these weights.

Appendix A Proofs

Proof of Proposition 3.1. Consider estimating μ_1; the argument for μ_0 follows in the same way. First, with $\hat{p}_i \equiv p(\mathbf{X}_i\hat{\gamma}_{1,ate,ipt})$, equation (2.4) implies that

$$N^{-1}\sum_{i=1}^{N} W_i/\hat{p}_i \;=\; 1.$$

The IPT estimator of μ_1 is the first term in (3.11). Now, consider the AIPT estimator in (3.12). Simple algebra shows it can be expressed as

$$\begin{aligned}
\hat{\mu}_{1,aipt} &= N^{-1}\sum_{i=1}^{N}\frac{W_iY_i}{\hat{p}_i} - N^{-1}\sum_{i=1}^{N}\frac{W_i\mathbf{X}_i\hat{\beta}_1}{\hat{p}_i} + N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\hat{\beta}_1 \\
&= \hat{\mu}_{1,ipt} - N^{-1}\sum_{i=1}^{N}\frac{W_i\mathbf{X}_i\hat{\beta}_1}{\hat{p}_i} + \bar{\mathbf{X}}\hat{\beta}_1 \\
&= \hat{\mu}_{1,ipt} + \left[\bar{\mathbf{X}} - N^{-1}\sum_{i=1}^{N}\frac{W_i\mathbf{X}_i}{\hat{p}_i}\right]\hat{\beta}_1 \\
&= \hat{\mu}_{1,ipt},
\end{aligned}$$

where the last equality uses (2.3) (with an intercept explicitly included).

Now, consider the IPTRA estimator in (3.15). Given that $1 \in \mathbf{X}_i$, the first-order condition for the WLS estimator of β_1, following from (3.14), is easily seen to imply that

$$N^{-1}\sum_{i=1}^{N}\frac{W_iY_i}{\hat{p}_i} \;=\; \left(N^{-1}\sum_{i=1}^{N}\frac{W_i\mathbf{X}_i}{\hat{p}_i}\right)\tilde{\beta}_1.$$

The term on the left is, again, $\hat{\mu}_{1,ipt}$. For the term on the right, use the IPT moment conditions in (2.3), as before:

$$\hat{\mu}_{1,ipt} \;=\; N^{-1}\sum_{i=1}^{N}\frac{W_iY_i}{\hat{p}_i} \;=\; \left(N^{-1}\sum_{i=1}^{N}\frac{W_i\mathbf{X}_i}{\hat{p}_i}\right)\tilde{\beta}_1 \;=\; \bar{\mathbf{X}}\tilde{\beta}_1 \;=\; \hat{\mu}_{1,iptra}.$$

Repeating the same argument for μ_0 completes the proof.

Proof of Proposition 3.2. Recall that $\mu_{0|1} \equiv \mathrm{E}[Y(0)|W=1]$. Redefine $\hat{p}_i$ as $\hat{p}_i \equiv p(\mathbf{X}_i\hat{\gamma}_{att,cbps})$. Using simple algebra, the AIPW estimator of $\mu_{0|1}$ using CBPS weights, given in (3.17), can be written as

$$\begin{aligned}
\hat{\mu}_{0|1,aipwcbps} &= N_1^{-1}\sum_{i=1}^{N}\frac{\hat{p}_i(1-W_i)Y_i}{1-\hat{p}_i} - N_1^{-1}\sum_{i=1}^{N}\frac{\hat{p}_i(1-W_i)\mathbf{X}_i\hat{\beta}_0}{1-\hat{p}_i} + \bar{\mathbf{X}}_1\hat{\beta}_0 \\
&= \hat{\mu}_{0|1,ipwcbps} + \left[\bar{\mathbf{X}}_1 - N_1^{-1}\sum_{i=1}^{N}\frac{\hat{p}_i(1-W_i)\mathbf{X}_i}{1-\hat{p}_i}\right]\hat{\beta}_0 \\
&= \hat{\mu}_{0|1,ipwcbps},
\end{aligned}$$

where the final equality follows from (2.10), the CBPS moment conditions for estimating the ATT.

For IPWRA using CBPS weights, note that the first-order condition for $\tilde{\beta}_0$, following from (3.18), is

$$\sum_{i=1}^{N}\frac{\hat{p}_i(1-W_i)}{1-\hat{p}_i}\mathbf{X}_i'\left(Y_i - \mathbf{X}_i\tilde{\beta}_0\right) = \mathbf{0}.$$

Focusing on the first element $1 \in \mathbf{X}_i$ and dividing by N_1, we can write

$$N_1^{-1}\sum_{i=1}^{N}\frac{\hat{p}_i(1-W_i)Y_i}{1-\hat{p}_i} = N_1^{-1}\sum_{i=1}^{N}\frac{\hat{p}_i(1-W_i)\mathbf{X}_i\tilde{\beta}_0}{1-\hat{p}_i}$$

or, again using (2.10),

$$\hat{\mu}_{0|1,ipwcbps} = \bar{\mathbf{X}}_1\tilde{\beta}_0 = \hat{\mu}_{0|1,ipwracbps}.$$

This completes the proof.

References

Graham, Bryan S., Cristine Campos de Xavier Pinto, and Daniel Egel, "Inverse Probability Tilting for Moment Condition Models with Missing Data," *Review of Economic Studies*, 2012, *79* (3), 1053–1079.

Heiler, Phillip, "Efficient Covariate Balancing for the Local Average Treatment Effect," *Journal of Business & Economic Statistics*, 2022, *40* (4), 1569–1582.

Hirano, Keisuke and Guido W. Imbens, "Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catheterization," *Health Services & Outcomes Research Methodology*, 2001, *2* (3–4), 259–278.

Horvitz, Daniel G. and Donovan J. Thompson, "A Generalization of Sampling without Replacement from a Finite Universe," *Journal of the American Statistical Association*, 1952, *47* (260), 663–685.

Imai, Kosuke and Marc Ratkovic, "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society, Series B*, 2014, *76* (1), 243–263.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao, "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 1994, *89* (427), 846–866.

Słoczyński, Tymon and Jeffrey M. Wooldridge, "A General Double Robustness Result for Estimating Average Treatment Effects," *Econometric Theory*, 2018, *34* (1), 112–133.

_ , S. Derya Uysal, and Jeffrey M. Wooldridge, "Doubly Robust Estimation of Local Average Treatment Effects Using Inverse Probability Weighted Regression Adjustment," arXiv:2208.01300v1, 2022.

Wooldridge, Jeffrey M., "Inverse Probability Weighted Estimation for General Missing Data Problems," *Journal of Econometrics*, 2007, *141* (2), 1281–1301.

_ , *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge–London: MIT Press, 2010.