I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18147

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

Tymon Słoczyński
Nski S. Derya Uysal
Jeffrey M. Wooldridge

SEPTEMBER 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18147

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

Tymon Słoczyński
Brandeis University

Nski S. Derya Uysal
LMU Munich

Jeffrey M. Wooldridge
Michigan State University

SEPTEMBER 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

How should researchers adjust for covariates? We show that if the propensity score is estimated using a specific covariate balancing approach, inverse probability weighting (IPW), augmented inverse probability weighting (AIPW), and inverse probability weighted regression adjustment (IPWRA) estimators are numerically equivalent for the average treatment effect (ATE), and likewise for the average treatment effect on the treated (ATT). The resulting weights are inherently normalized, making normalized and unnormalized IPW and AIPW identical. We discuss implications for instrumental variables and difference-in-differences estimators and illustrate with two applications how these numerical equivalences simplify analysis and interpretation.

JEL Classification: C20, C21, C23, C26

Keywords: covariate balancing, difference-in-differences, double robustness, instrumental variables, inverse probability tilting, treatment effects, weighting

Corresponding author:
Tymon Słoczyński
Department of Economics & International Business School
Brandeis University
415 South Street, MS 021
Waltham, MA 02453
USA

E-mail: tslocz@brandeis.edu

* This paper was presented at Brandeis University, Ege University, the 2025 Annual Conference of the International Association for Applied Econometrics, the 2025 Workshop in Econometrics at TU Dortmund University, the 2024 LMU Munich - Todai Econometrics Workshop, the 2024 Econometric Society European Meetings, the 2024 BSE Summer Forum Workshop on Microeconometrics and Policy Evaluation, the 2023 BU/BC Greenline Econometrics Workshop, and the 2023 Annual Meeting of the Southern Economic Association. We thank the audiences at those seminars and conferences, as well as Bernie Black, Brant Callaway, Bruno Ferman, Bryan Graham, Pedro Sant'Anna, Rahul Singh, Max Tabord-Meehan, Joachim Winter, Yinchu Zhu, and José Zubizarreta, for helpful conversations and comments. We also thank Lennart Ohly for excellent research assistance.

1 Introduction

Covariate adjustment is central to causal inference, yet the choice of method remains contested. Much recent research has highlighted the shortcomings of a number of well-established estimation methods in reproducing suitable averages of heterogeneous treatment effects. A key lesson from this literature is that additive linear models may often fail to properly adjust for covariates when those covariates are relevant for identification. This concern arises not only under unconfoundedness (e.g., Słoczyński, 2022; Goldsmith-Pinkham, Hull, and Kolesár, 2024; Chen, 2025), but also in instrumental variables (e.g., Słoczyński, 2024; Blandhol, Bonney, Mogstad, and Torgovitsky, 2025) and difference-in-differences settings (e.g., Caetano and Callaway, 2024).

When considering alternatives to standard methods, researchers face a wide range of options, including regression adjustment, matching, weighting, and doubly robust estimators, as well as related approaches based on machine learning. Such variety is not necessarily desirable: the existence of a common standard facilitates comparability across studies, while additional researcher degrees of freedom invite specification searching (Simmons, Nelson, and Simonsohn, 2011; Vivalt, 2019; Ferman, Pinto, and Possebom, 2020).

In this paper, we build on the fact that many of these alternative methods require first-step estimation of the propensity score. We assume that a researcher would be willing to commit to estimating the propensity score using a particular method of moments approach with desirable properties, namely the inverse probability tilting (IPT) estimator of Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2012, 2016). (To be clear, a researcher using IPT still needs to choose a model for the propensity score, perhaps logit or probit, but would then estimate this model using the method of moments instead of maximum likelihood.) Our main contribution is to demonstrate that commitment to using IPT substantially reduces the choice set (i.e., the number of alternative estimators) available to researchers. Specifically, we show that using the IPT moment conditions to estimate the propensity score leads to numerical equivalence between members of several classes of estimators of average treatment effects: inverse probability weighting (IPW), augmented inverse probability weighting (AIPW), and inverse probability weighted regression adjustment (IPWRA), with the latter two classes using linear models for potential outcome means. Our equivalence results are very general, as they are valid for any propensity score model having the index form (e.g., logit or probit). In addition, they apply to both normalized and unnormalized versions of IPW and AIPW, as well as to estimators of both the

average treatment effect (ATE) and the average treatment effect on the treated (ATT).

Our equivalence results are meaningful insofar as IPT is appealing in its own right. Indeed, a major reason for this appeal is that the moment conditions underlying IPT impose a desirable property known as "exact balancing." Specifically, IPT estimates the parameters of the propensity score model such that, after reweighting, the mean covariate values are identical across key groups: between the (weighted) treatment group, the (weighted) comparison group, and the (unweighted) full sample when estimating the ATE, and between the (unweighted) treatment group and the (weighted) comparison group when estimating the ATT. In this sense, IPT is a prototypical "covariate balancing" estimator, ensuring that causal comparisons are made only between groups with identical (mean) characteristics. As shown by Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2012, 2016), IPT estimators of the ATE and ATT also enjoy local efficiency and double robustness properties. That is, if both the propensity score model chosen by the researcher (e.g., logit or probit) and the linear specification of potential outcome means are correct, the estimators are semiparametrically efficient; if only one is correctly specified, the estimators remain consistent. As we review below, the IPT estimator of the ATT is also identical to the subsequent proposals of Hainmueller (2012) and Imai and Ratkovic (2014).[1]

We also translate our equivalence results to instrumental variables and difference-in-differences settings. In both contexts, weighting and doubly robust estimators have played a prominent role in the recent literature, and our results again simplify the set of alternative estimators available to empirical researchers. In particular, our results imply that the doubly robust estimator proposed by Sant'Anna and Zhao (2020), which uses IPT-estimated propensity scores, is numerically equivalent to the simple IPW estimator of Abadie (2005) with IPT weights. It follows that in implementations of Sant'Anna and Zhao (2020) it is redundant to estimate the model for the untreated potential change in outcomes.

We illustrate our findings with two empirical applications. In the first application, we revisit the study of the causal effects of cash transfers on longevity in Aizer, Eli, Ferrie, and Lleras-Muney (2016). In line with an earlier replication in Słoczyński (2022), we conclude that there is insufficient evidence to reject the null hypothesis of zero average effects. Our numerical equivalences simplify analysis and interpretation, and none of the IPT estimates is significantly different from zero despite usually having smaller standard errors than the corresponding estimates based on maximum likelihood. In the second applica-

[1] Specifically, IPT coincides with Hainmueller's (2012) entropy balancing estimator when the propensity score is estimated with the logit model.

tion, we replicate Sant'Anna and Zhao's (2020) analysis of the NSW–CPS data, which was previously analyzed by LaLonde (1986) and many others. We show that many of the estimates reported by Sant'Anna and Zhao (2020) would have been identical to their preferred estimates had they used IPT to estimate the propensity score in all cases.

We also supplement this paper with a companion Stata package, `teffects2`, available at the Statistical Software Components (SSC) Archive. Our package implements IPW, AIPW, and IPWRA estimators of the ATE and ATT under unconfoundedness, with several approaches to estimate the weights, including IPT. The package can also be used to estimate the ATT in difference-in-differences settings after a suitable transformation of the outcome variable. This provides a novel implementation of the doubly robust difference-in-differences (DRDID) estimator proposed by Sant'Anna and Zhao (2020).

Literature Review

This paper builds on a body of research on estimating average treatment effects under the assumption of unconfoundedness. We focus on three classes of estimators: inverse probability weighting (IPW), as in Hirano, Imbens, and Ridder (2003), augmented inverse probability weighting (AIPW), as in Robins, Rotnitzky, and Zhao (1994), and inverse probability weighted regression adjustment (IPWRA), as in Wooldridge (2007) and Słoczyński and Wooldridge (2018). These classes of estimators, as well as several others, were surveyed by Imbens and Wooldridge (2009), Abadie and Cattaneo (2018), and Uysal (2024).

Each of the classes of estimators we consider requires a first-step estimation of the propensity score. This is typically done using maximum likelihood estimation (MLE) of a standard binary response model for treatment assignment (e.g., logit or probit). However, this estimation approach does not guarantee that any desirable balancing properties are satisfied in finite samples. In contrast, various "covariate balancing" estimators of the propensity score are explicitly constructed with these properties in mind; they are also tailored to the specific parameter of interest (e.g., ATE or ATT) to improve the statistical properties of the corresponding treatment effect estimator.

Following the early work on inverse probability tilting (IPT) by Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2012, 2016), many papers have proposed alternative covariate balancing procedures for estimating average treatment effects. Hainmueller (2012) suggested estimating the inverse probability weights directly—subject to balancing and normalizing constraints—rather than estimating the propensity score first and then inverting it to obtain the weights. Known as "entropy balancing," this procedure

was designed to estimate the ATT and was later shown to be identical to IPT when the latter uses the logit model (Zhao and Percival, 2017; Tan, 2020). Imai and Ratkovic (2014) proposed using different moment conditions than those in Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2012) when estimating the ATE; however, the resulting estimator lacks some desirable theoretical properties of IPT, such as double robustness. On the other hand, Imai and Ratkovic's (2014) moment conditions for estimating the ATT are the same as in IPT, which implies that the resulting estimator is also equivalent to IPT (as well as to entropy balancing when the logit model is used). Zubizarreta (2015) relaxed the exact balancing requirements of earlier methods and proposed estimating weights that minimize variance subject to approximate balancing constraints. Zhao (2019) unified and generalized much of the earlier work by introducing a covariate balancing framework based on optimizing loss functions tailored to a given estimand. Sant'Anna, Song, and Xu (2022) proposed estimating the propensity score by maximizing balance across the entire covariate distribution rather than in selected functions of the covariates.

Most of the early work on covariate balancing focused on addressing missing data problems and estimating average treatment effects under unconfoundedness. However, recent research has also applied similar ideas to estimating various parameters of interest in difference-in-differences (Sant'Anna and Zhao, 2020; Callaway and Sant'Anna, 2021) and instrumental variables settings (Heiler, 2022; Sant'Anna, Song, and Xu, 2022; Singh and Sun, 2024; Słoczyński, Uysal, and Wooldridge, 2025).

This paper is also related to the important work of Robins, Sued, Lei-Gomez, and Rotnitzky (2007), Kline (2011), Chattopadhyay and Zubizarreta (2023), and Bruns-Smith, Dukes, Feller, and Ogburn (2025) demonstrating numerical equivalences between regression adjustment and weighting estimators of average treatment effects, under the constraint that both the potential outcome means and the weights are linear in covariates. While the weights may generally be approximated as a linear function of a high-dimensional dictionary, as in Chernozhukov, Newey, and Singh (2022) and Bruns-Smith, Dukes, Feller, and Ogburn (2025), the corresponding parametric restriction is unlikely to be plausible in low-dimensional settings, especially since it is equivalent to assuming an inverse linear model for the propensity score. In such settings, some of the estimated weights are likely to be negative, which invalidates the sample boundedness property of the resulting estimator (cf. Robins, Sued, Lei-Gomez, and Rotnitzky, 2007).

In this paper, we extend and generalize this earlier work by demonstrating that the numerical equivalences between the IPW, AIPW, and IPWRA estimators are driven by the

IPT moment conditions rather than parametric restrictions on the propensity score or the weights. Unlike our paper, the results in Robins, Sued, Lei-Gomez, and Rotnitzky (2007), Kline (2011), Chattopadhyay and Zubizarreta (2023), and Bruns-Smith, Dukes, Feller, and Ogburn (2025) are specific to the inverse linear model for the propensity score. (Most of these results are also limited to IPW.) Under this strong parametric restriction, which our paper does not make, IPW, AIPW, and IPWRA with IPT moment conditions are also numerically equivalent to (linear) regression adjustment.

Plan of the Paper

We organize the paper as follows. In Section 2, we review the estimation problems solved by the IPT weights for the ATE and 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 IPT-based estimators of the ATT. We emphasize that since we are establishing numerical equivalences, we do not need to, and do not, state the assumptions under which the estimators are consistent. These have been covered elsewhere and are well known. In Section 4, we discuss the consequences that the algebraic equivalence results have for estimating local average treatment effects with instrumental variables and heterogeneous treatment effects in difference-in-differences settings. In Section 5, we discuss our empirical applications. In Section 6, we conclude. Our proofs are provided in the Appendix. In the Supplemental Appendix, we briefly discuss implementation of IPT in R and Stata.

2 Covariate Balancing

In a general missing data setting, Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2012, 2016) introduced inverse probability tilting (IPT) as a method for estimating the propensity score, along with other parameters of interest. In this section, we review the estimation problems solved by this method in the binary treatment case.

Let W denote the binary treatment indicator, and define the propensity score as $\mathrm{P}(W = 1|\mathbf{X} = \mathbf{x})$. Assume an index model, $p(\mathbf{x}\gamma)$, where $\mathbf{x}$ is $1 \times K$, γ is $K \times 1$, and $x_1 \equiv 1$ ensures an intercept. Although $p(\mathbf{x}\gamma)$ is typically taken to be logit, our results apply more generally to any index model, including probit, complementary log-log, linear, and inverse linear models. Let $Y(0)$ and $Y(1)$ denote the potential outcomes. Recall that the ATE and 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 in this paper pertain to algebraic equivalences, and therefore, we do not discuss the identification of population parameters.

To fix ideas, consider the case where $p(\mathbf{x}\gamma)$ is logit, i.e., $p(\mathbf{x}\gamma)=\exp(\mathbf{x}\gamma)/\left[1+\exp(\mathbf{x}\gamma)\right]$. In practice, the logit model is often conflated with its estimation via maximum likelihood, in which case, for a sample of size N, the maximum likelihood estimator $\hat{\gamma}_{mle}$ solves the first-order condition:

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

IPT replaces this condition with a different set of moment equations for estimating γ. While our discussion below is not limited to the logit case, the point is that even when the logit model is used, estimation need not rely on maximum likelihood; it can proceed via the method of moments instead. When W is a treatment indicator, the IPT moment conditions proposed by Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2012) for estimating $\mathrm{E}[Y(1)]$ are

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

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

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

and these equations define the IPT estimator of γ, $\hat{\gamma}_{1,ipt}$, regardless of the specific model chosen for $\mathrm{P}(W=1|\mathbf{X}=\mathbf{x})$. Note that we have put a "1" subscript on $\hat{\gamma}_{1,ipt}$ because, in the treatment effects setting, there is another set of moment conditions for estimating $\mathrm{E}[Y(0)]$ that leads to a different IPT estimator of γ. Again, by iterated expectations,

$$\mathrm{E}\left[\frac{1-W}{1-p(\mathbf{X}\gamma)}\mathbf{X}'\right] \;=\; \mathrm{E}(\mathbf{X}'), \tag{2.4}$$

and this leads to the sample analog:

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

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

$$N^{-1} \sum_{i=1}^{N} \frac{W_i}{p(\mathbf{X}_i \hat{\gamma}_{1,ipt})} \;=\; 1 \tag{2.6}$$

and

$$N^{-1} \sum_{i=1}^{N} \frac{1 - W_i}{1 - p(\mathbf{X}_i \hat{\gamma}_{0,ipt})} \;=\; 1. \tag{2.7}$$

These two equations are key, as the summands in (2.6) are the weights for estimating $\mathrm{E}[Y(1)]$ in IPW estimation, and those in (2.7) are the weights used in estimating $\mathrm{E}[Y(0)]$, as we review in Section 3. Equations (2.6) and (2.7) show that the IPT weights are automatically normalized for estimating the ATE. That is, the sample mean of the weights is not stochastic but instead equal to one by construction. These equations also indicate that the IPT estimator of the ATE will require estimating the propensity score twice, with one set of predicted probabilities used to estimate $\mathrm{E}[Y(1)]$ and another to estimate $\mathrm{E}[Y(0)]$.

For estimating the ATT, the moment equations used by Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2016) 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

$$\begin{aligned} N_1^{-1} \sum_{i=1}^{N} W_i \mathbf{X}_i' \;&=\; \left(\frac{N_1}{N} \right)^{-1} N^{-1} \sum_{i=1}^{N} \frac{p(\mathbf{X}_i \hat{\gamma}_{0,ipt})\,(1-W_i)}{1 - p(\mathbf{X}_i \hat{\gamma}_{0,ipt})} \cdot \mathbf{X}_i' \\ &=\; N_1^{-1} \sum_{i=1}^{N} \frac{p(\mathbf{X}_i \hat{\gamma}_{0,ipt})\,(1-W_i)}{1 - p(\mathbf{X}_i \hat{\gamma}_{0,ipt})} \cdot \mathbf{X}_i' \end{aligned}$$

or

$$\bar{\mathbf{X}}_1' \;=\; N_1^{-1} \sum_{i=1}^{N} \frac{p(\mathbf{X}_i \hat{\gamma}_{0,ipt})\,(1-W_i)}{1 - p(\mathbf{X}_i \hat{\gamma}_{0,ipt})} \cdot \mathbf{X}_i', \tag{2.8}$$

where $\bar{\mathbf{X}}_1 = N_1^{-1} \sum_{i=1}^{N} W_i \mathbf{X}_i$. Because $1 \in \mathbf{X}_i$, (2.8) implies

$$N_1 \;=\; \sum_{i=1}^{N} \frac{p(\mathbf{X}_i \hat{\gamma}_{0,ipt})\,(1-W_i)}{1 - p(\mathbf{X}_i \hat{\gamma}_{0,ipt})},$$

which implies that the weights used in the IPW estimation of the ATT sum to the number of treated units. In other words, like in the case of the ATE, the IPT weights for estimating

the ATT are automatically normalized.

It may seem surprising that we use $\hat{\gamma}_{0,ipt}$ to denote the IPT estimator of γ in the context of estimating the ATT, given that we used the same notation in equations (2.5) and (2.7) above. This is fully warranted, however, because this estimator is in fact the same as the IPT estimator defined by equation (2.5). Indeed, as shown by Tan (2020), the moment conditions in (2.5), which balance the weighted covariates of the comparison group with those of the overall sample, are algebraically equivalent to the conditions in (2.8), which instead balance them with the treated group, but using a different set of weights. To see this equivalence, we can rewrite the sample moment conditions in (2.5) as

$$N^{-1}\sum_{i=1}^{N}\left(\frac{1-W_i}{1-p(\mathbf{X}_i\hat{\gamma}_{0,ipt})}-1\right)\cdot\mathbf{X}_i' \;=\; \mathbf{0},$$

which, after simple algebra, can be expressed as

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

and this, in turn, is easily seen as equivalent to equation (2.8).

It is also useful to briefly compare the IPT moment conditions for estimating the ATE with a subsequent proposal by Imai and Ratkovic (2014), known as the "covariate balancing propensity score (CBPS)," which uses different moment conditions to obtain a single estimator of γ. Again, if the propensity score is correctly specified then, by iterated expectations,

$$\mathrm{E}(\mathbf{X}') \;=\; \mathrm{E}\left[\frac{W}{p(\mathbf{X}\gamma)}\mathbf{X}'\right] \;=\; \mathrm{E}\left[\frac{1-W}{1-p(\mathbf{X}\gamma)}\mathbf{X}'\right]. \tag{2.9}$$

Rather than using the implications of (2.9) separately, which is what IPT does, Imai and Ratkovic (2014) use the second equality to obtain the following sample moment conditions:

$$N^{-1}\sum_{i=1}^{N}\frac{W_i}{p(\mathbf{X}_i\hat{\gamma}_{cbps})}\mathbf{X}_i' \;=\; N^{-1}\sum_{i=1}^{N}\frac{1-W_i}{1-p(\mathbf{X}_i\hat{\gamma}_{cbps})}\mathbf{X}_i'. \tag{2.10}$$

After simple algebra, the moment conditions can be expressed as

$$\sum_{i=1}^{N}\left(\frac{W_i-p(\mathbf{X}_i\hat{\gamma}_{cbps})}{p(\mathbf{X}_i\hat{\gamma}_{cbps})\left[1-p(\mathbf{X}_i\hat{\gamma}_{cbps})\right]}\right)\mathbf{X}_i' \;=\; \mathbf{0}. \tag{2.11}$$

Comparing (2.11) with (2.1), we can see that the CBPS approach—when applied to the logit model—weights the MLE moment conditions by the estimated inverse conditional

variance, $\text{Var}(W_i|\mathbf{X}_i)$.

Because the first element of $\mathbf{X}_i$ is unity, (2.10) also implies that

$$\sum_{i=1}^{N} \frac{W_i}{p(\mathbf{X}_i\hat{\gamma}_{cbps})} = \sum_{i=1}^{N} \frac{1-W_i}{1-p(\mathbf{X}_i\hat{\gamma}_{cbps})}. \tag{2.12}$$

Equation (2.12) shows that the weights appearing in the IPW estimates of $\text{E}[Y(1)]$ and $\text{E}[Y(0)]$ sum to the same value, but that common value is not necessarily the sample size, N. In other words, when these are used as weights in IPW, the CBPS weights are not automatically normalized.

Finally, when estimating the ATT, Imai and Ratkovic (2014) suggest using the moment conditions in equation (2.8), following the approach of Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2016). This implies that the IPT and CBPS estimators of the ATT are the same. When using the logit model, as shown by Tan (2020), both approaches are also numerically identical to the entropy balancing estimator of Hainmueller (2012).[2]

3 Equivalence of Estimators

In this section, we establish numerical equivalences among three different classes of 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 alternatives to OLS estimation of an additive linear model when unconfoundedness is assumed to hold: IPW, AIPW, and IPWRA. As we establish algebraic equivalence, we do not impose assumptions beyond those necessary for the existence of estimates for a given sample. This simply means that the estimated propensity scores are strictly between zero and one for all i.

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

$$\hat{\tau}_{ate,ipt} = \hat{\mu}_{1,ipt} - \hat{\mu}_{0,ipt} = N^{-1}\sum_{i=1}^{N} \frac{W_iY_i}{p(\mathbf{X}_i\hat{\gamma}_{1,ipt})} - N^{-1}\sum_{i=1}^{N} \frac{(1-W_i)Y_i}{1-p(\mathbf{X}_i\hat{\gamma}_{0,ipt})}, \tag{3.13}$$

where the subscript "ipt" indicates the use of IPT weights. See, e.g., Wooldridge (2010, Section 21.3) for a variant of this estimator with MLE-based weights and Egel, Graham,

[2]The three estimators of the ATT will no longer coincide if, in the case of CBPS, the IPT moment conditions are combined with the first-order conditions of the maximum likelihood estimator (the so-called "overidentified CBPS"). Likewise, the entropy balancing estimator will differ from IPT and CBPS if its implementation constrains higher moments of the covariates to be balanced, too.

and Pinto (2008) and Graham, Pinto, and Egel (2012) for IPT. We know from (2.6) and (2.7) that the weights in both weighted averages are automatically 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,

$$\hat{\mu}_{1,aipt} \;=\; N^{-1}\sum_{i=1}^{N} \frac{W_i\left(Y_i - \mathbf{X}_i\hat{\beta}_1\right)}{p(\mathbf{X}_i\hat{\gamma}_{1,ipt})} \;+\; N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\hat{\beta}_1, \tag{3.14}$$

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 (3.14) 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.[3]

For μ_0, the AIPT estimator is

$$\hat{\mu}_{0,aipt} \;=\; N^{-1}\sum_{i=1}^{N} \frac{(1-W_i)\left(Y_i - \mathbf{X}_i\hat{\beta}_0\right)}{1 - p(\mathbf{X}_i\hat{\gamma}_{0,ipt})} \;+\; N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\hat{\beta}_0, \tag{3.15}$$

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,

$$\min_{\mathbf{b}_1} N^{-1}\sum_{i=1}^{N}\frac{W_i}{\hat{p}_i}\left(Y_i - \mathbf{X}_i\mathbf{b}_1\right)^2, \tag{3.16}$$

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 (3.16), μ_1 is estimated by averaging the fitted values across all observations, as in the case of linear regression adjustment:

$$\hat{\mu}_{1,iptra} \;=\; N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\tilde{\beta}_1 \;=\; \bar{\mathbf{X}}\tilde{\beta}_1. \tag{3.17}$$

The IPTRA estimator of μ_0, $\hat{\mu}_{0,iptra}$, uses the untreated units with weights $\left(1 - p(\mathbf{X}_i\hat{\gamma}_{0,ipt})\right)^{-1}$, 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 inverse probability weights are obtained using MLE, CBPS, or some other

[3]Normalization is less important in AIPW than in IPW. Knaus (2024) shows that "unnormalized" AIPW, unlike unnormalized IPW, can still be expressed as a weighted average of observed outcomes with weights that sum to one. This normalization is automatic under standard implementations of outcome regressions.

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 weights: IPW, normalized IPW (NIPW), AIPW, normalized AIPW (NAIPW), and IPWRA. (IPWRA is always normalized.) Strikingly, when IPT weights are used instead, all of these estimates are identical.

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*

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

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 linearly, then three prominent 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 the equivalence of several prominent estimators of the ATT when the IPT weights from equation (2.8) 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 IPT weights, is

$$\hat{\mu}_{0|1,ipt} \;=\; N_1^{-1}\sum_{i=1}^{N}\frac{p(\mathbf{X}_i\hat{\gamma}_{0,ipt})\,(1-W_i)\,Y_i}{1-p(\mathbf{X}_i\hat{\gamma}_{0,ipt})}. \tag{3.18}$$

As noted earlier, the weights in (3.18) sum to the number of treated units; thus, they are automatically normalized. The same weights also appear in the AIPW estimator. Therefore, the normalized and unnormalized IPW estimators coincide, as do the normalized and unnormalized AIPW estimators. Specifically, the AIPW estimator of $\mu_{0|1}$ is

$$\begin{aligned}\hat{\mu}_{0|1,aipt} &= N_1^{-1}\sum_{i=1}^{N}\frac{\hat{p}_i\,(1-W_i)}{1-\hat{p}_i}\left(Y_i - \mathbf{X}_i\hat{\beta}_0\right) + N_1^{-1}\sum_{i=1}^{N}W_i\mathbf{X}_i\hat{\beta}_0 \\ &= N_1^{-1}\sum_{i=1}^{N}\frac{\hat{p}_i\,(1-W_i)\,Y_i}{1-\hat{p}_i} - N_1^{-1}\sum_{i=1}^{N}\frac{\hat{p}_i\,(1-W_i)\,\mathbf{X}_i\hat{\beta}_0}{1-\hat{p}_i} + \bar{\mathbf{X}}_1\hat{\beta}_0, \end{aligned} \tag{3.19}$$

where $\hat{p}_i = p(\mathbf{X}_i\hat{\gamma}_{0,ipt})$ are now the IPT 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 $\mu_{0|1}$, using the weights from (2.8), is

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

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

$$\min_{\mathbf{b}_0} \; N^{-1}\sum_{i=1}^{N} \frac{\hat{p}_i\,(1-W_i)}{1-\hat{p}_i}\,(Y_i - \mathbf{X}_i\mathbf{b}_0)^2\,, \tag{3.20}$$

where $\hat{p}_i = p(\mathbf{X}_i\hat{\gamma}_{0,ipt})$. We have the following equivalence result.

Proposition 3.2. *Let $\hat{\gamma}_{0,ipt}$ be the estimators solving* (2.8) *with $\hat{p}_i = p(\mathbf{X}_i\hat{\gamma}_{0,ipt}) < 1$ for all i. Then, the IPW, AIPW, and IPWRA estimates of $\mu_{0|1}$, using the IPT weights and linear conditional means in the latter two cases, are identical. Therefore, the three estimates of τ_{att} are identical.*

Similar to Proposition 3.1, the implication of Proposition 3.2 is that if one uses the IPT weights in estimating $\mu_{0|1}$ as well as a linear model for $\mathrm{E}[Y(0)|\mathbf{X}]$, then the IPW, AIPW, and IPWRA estimators of the ATT are numerically identical; moreover, the IPW and AIPW versions are automatically normalized.

4 Implications for Instrumental Variables and Difference-in-Differences Settings

In this section, we briefly discuss the implications of the results in Section 3 for estimating local average treatment effects with instrumental variables and heterogeneous treatment effects in difference-in-differences settings.

4.1 Instrumental Variables

The results in Section 3 have implications for estimators of the local average treatment effect (LATE) and the local average treatment effect on the treated (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 treatment variable. We assume it to be binary, although this can be easily relaxed. We also have a binary instrumental variable, Z.

It follows from Frölich (2007) that 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 weights, estimators of the LATE based on IPW, AIPW, and IPWRA are all identical. The inverse probability weights, in this case, for both the numerator and the denominator, are based on 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 generally unspecified, here it may be impractical to use the linear model in the denominator when W is binary. See Słoczyński, Uysal, and Wooldridge (2022) for using other doubly robust estimators to exploit the binary nature of W and maybe special features of Y. In such cases, however, the numerical equivalence results no longer hold.

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 IPT weights, as in Proposition 3.2, then it follows immediately that the estimators of τ_{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 ATT are identical when using these weights.

4.2 Difference-in-Differences

Some popular estimators in difference-in-differences (DID) settings are based on applying standard treatment effect estimators after suitably transforming the outcome variable. For example, Abadie (2005), with two time periods, proposes applying IPW to the differences

$Y_{i2} - Y_{i1}$, where Y_{it} is the outcome for unit i in period t. Abadie (2005) uses maximum likelihood estimation of the propensity score to construct the inverse probability weights. Sant'Anna and Zhao (2020) instead develop a doubly robust estimator of the ATT. The estimator uses a structure similar to equation (3.19), although Sant'Anna and Zhao (2020) also normalize the weights and replace the OLS estimates of the conditional mean of the untreated potential change in outcomes with the WLS estimates similar to equation (3.20); they also use the IPT weights from equation (2.8). Strikingly, the results in Section 3 imply that all these additional modifications have no impact on the final estimate of the ATT when the IPT weights are used; in other words, the simple IPW estimator in Abadie (2005) is numerically identical to Sant'Anna and Zhao (2020) when both use the IPT moment conditions to estimate the propensity score.

Similar conclusions hold with many periods and staggered interventions. Callaway and Sant'Anna (2021) extend Sant'Anna and Zhao (2020) to estimate ATTs by treatment cohort (i.e., the first period of treatment), g, and calendar time, t. To estimate these ATTs, τ_{gt}, Callaway and Sant'Anna (2021) apply different versions of AIPW and IPWRA to differences $Y_{it} - Y_{i,g-1}$, where $Y_{i,g-1}$ is the outcome in the period just before the first treatment period for treatment cohort g. Callaway and Sant'Anna (2021) emphasize that the comparison group can either consist of the never treated (NT) cohort or the NT cohort plus other cohorts that are first treated in period $t+1$ or later ("not yet treated"). One of the estimators recommended by Callaway and Sant'Anna (2021) is AIPW with the IPT weights from equation (2.8). It follows immediately that applying IPW, AIPW, or IPWRA to $Y_{it} - Y_{i,g-1}$ with treatment indicator D_{ig} (indicating treatment cohort) and controls $\mathbf{X}_i$ delivers identical estimates of the τ_{gt} when IPT weights are used. This is true even when $t < g-1$, which provides event study graphs for studying the existence of pre-trends.

An alternative transformation in the staggered intervention settings uses data on all pre-treatment outcomes by removing the average of the outcomes over all pre-treatment periods: $\dot{Y}_{itg} \equiv Y_{it} - (g-1)^{-1}\sum_{s=1}^{g-1} Y_{is}$. As shown in Lee and Wooldridge (2024), under standard no anticipation and conditional parallel trends assumptions, one can apply various treatment effect estimators to the cross-sectional data $\left\{\left(\dot{Y}_{itg}, D_{ig}, \mathbf{X}_i\right) : i = 1,...,N\right\}$ to consistently estimate τ_{gt}. Again, the results in Section 3 immediately imply that the IPW, AIPW, and IPWRA estimators with the IPT weights from equation (2.8) are all identical when applied to these data once one chooses a suitable comparison group.

5 Empirical Applications

In this section, we illustrate our findings with two empirical applications, beginning with a replication of a prominent study of the causal effects of cash transfers on longevity (Aizer, Eli, Ferrie, and Lleras-Muney, 2016) and concluding with a reanalysis of the empirical application in Sant'Anna and Zhao (2020).

5.1 The Effects of Cash Transfers on Longevity

Aizer, Eli, Ferrie, and Lleras-Muney (2016) study the long-run impacts of the Mothers' Pension (MP) program, which was the first government-sponsored welfare program in the prewar U.S. The outcome studied by the authors is the log age at death of children of the program participants. A key strength of the original study is in its careful construction of the comparison group, which consists only of mothers who were initially deemed eligible for participation but were later rejected. Still, Słoczyński (2022) argues that some of the conclusions of this paper are not robust to treatment effect heterogeneity.

In our application, we use the same data as Aizer, Eli, Ferrie, and Lleras-Muney (2016) and Słoczyński (2022). We consider three covariate specifications and two sources of information on dates of death: program records and death certificates. In our first specification, we control for cohort and state fixed effects. In our second specification, in line with Aizer, Eli, Ferrie, and Lleras-Muney (2016), we replace state fixed effects with a battery of individual, county, and state characteristics. In our final specification, we reintroduce state fixed effects without dropping any other covariates.[4]

Table 1 reports a number of estimates of the effects of cash transfers on longevity. As in Aizer, Eli, Ferrie, and Lleras-Muney (2016), the OLS estimates from an additive model, which controls for program participation and covariates but not interactions between the two, strongly suggest that cash transfers positively influenced the longevity of the children of their beneficiaries. However, in line with the replication in Słoczyński (2022), the majority of the estimates of the ATE and ATT are smaller or much smaller than the OLS estimates and not statistically significant. At the same time, there is a clear difference between the two panels of Table 1 that report weighting and doubly robust estimators based on MLE and IPT weights. In the case of MLE, there is a wide variation in estimates, which range from 0.0014 to 0.0597 for the ATE and from –0.0014 to 0.0645 for the ATT. Conditional on

[4]The final specification in Aizer, Eli, Ferrie, and Lleras-Muney (2016) uses county rather than state fixed effects but is otherwise identical. In our application, using county fixed effects is not feasible because, in several counties, every eligible applicant was treated, resulting in a failure of overlap.

Table 1: Replication of Aizer, Eli, Ferrie, and Lleras-Muney (2016)

	(1)		(2)		(3)		(4)	
OLS	0.0157***		0.0158***		0.0163***		0.0146**	
	(0.0058)		(0.0059)		(0.0059)		(0.0059)	
	ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT
RA	0.0105*	0.0096	0.0100	0.0092	0.0100	0.0090	0.0080	0.0069
	(0.0062)	(0.0063)	(0.0068)	(0.0071)	(0.0071)	(0.0075)	(0.0071)	(0.0075)
	MLE weights							
	ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT
IPW	0.0597***	0.0645***	0.0438	0.0391	0.0113	0.0002	0.0093	–0.0014
	(0.0164)	(0.0181)	(0.0886)	(0.1002)	(0.1216)	(0.1381)	(0.1222)	(0.1387)
NIPW	0.0107*	0.0099	0.0064	0.0047	0.0025	0.0001	0.0014	–0.0006
	(0.0061)	(0.0063)	(0.0072)	(0.0077)	(0.0072)	(0.0077)	(0.0072)	(0.0076)
AIPW	0.0102*	0.0093	0.0056	0.0040	0.0042	0.0023	0.0024	0.0007
	(0.0062)	(0.0064)	(0.0069)	(0.0073)	(0.0072)	(0.0076)	(0.0072)	(0.0077)
NAIPW	0.0102*	0.0093	0.0056	0.0039	0.0042	0.0023	0.0024	0.0007
	(0.0062)	(0.0064)	(0.0069)	(0.0073)	(0.0072)	(0.0076)	(0.0072)	(0.0076)
IPWRA	0.0101	0.0092	0.0072	0.0058	0.0050	0.0028	0.0037	0.0019
	(0.0062)	(0.0064)	(0.0066)	(0.0069)	(0.0069)	(0.0073)	(0.0069)	(0.0073)
	IPT weights							
	ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT
IPW	0.0101	0.0092	0.0076	0.0063	0.0057	0.0040	0.0047	0.0032
	(0.0062)	(0.0064)	(0.0066)	(0.0069)	(0.0067)	(0.0071)	(0.0067)	(0.0071)
NIPW	0.0101	0.0092	0.0076	0.0063	0.0057	0.0040	0.0047	0.0032
	(0.0062)	(0.0064)	(0.0066)	(0.0069)	(0.0067)	(0.0071)	(0.0067)	(0.0071)
AIPW	0.0101	0.0092	0.0076	0.0063	0.0057	0.0040	0.0047	0.0032
	(0.0062)	(0.0064)	(0.0066)	(0.0069)	(0.0067)	(0.0071)	(0.0067)	(0.0071)
NAIPW	0.0101	0.0092	0.0076	0.0063	0.0057	0.0040	0.0047	0.0032
	(0.0062)	(0.0064)	(0.0066)	(0.0069)	(0.0067)	(0.0071)	(0.0067)	(0.0071)
IPWRA	0.0101	0.0092	0.0076	0.0063	0.0057	0.0040	0.0047	0.0032
	(0.0062)	(0.0064)	(0.0066)	(0.0069)	(0.0067)	(0.0071)	(0.0067)	(0.0071)
State fixed effects		✓				✓		✓
Cohort fixed effects		✓		✓		✓		✓
Individual controls				✓		✓		✓
State characteristics				✓		✓		✓
County characteristics				✓		✓		✓
Observations	7,860		7,859		7,859		7,857	

Notes: The source of data is Aizer, Eli, Ferrie, and Lleras-Muney (2016). The outcome of interest is the log age at death, as reported in program records (specifications 1–3) or on the death certificate (specification 4). "OLS" corresponds to the OLS estimation of an additive linear model. "RA" corresponds to the regression adjustment estimator, which is based on the OLS estimation of a fully interacted linear model. The remaining estimators use a logit model for the propensity score and are defined in Section 3. Standard errors are in parentheses. *Statistically significant at the 10% level; **at the 5% level; ***at the 1% level.

choosing a specific covariate specification, the choice of an estimator can have a profound impact on the researcher's conclusion. On the other hand, in the case of IPT, this choice is entirely inconsequential, as there are no differences across estimators conditional on a particular specification choice. This illustrates Propositions 3.1 and 3.2, which demonstrate the underlying numerical equivalences. Moreover, in the case of IPT, the standard errors are usually slightly smaller than in the case of the corresponding estimates based on MLE.

5.2 The Effects of a Training Program on Earnings

A large number of papers, originating with LaLonde (1986), combine experimental data from the evaluation of the National Supported Work (NSW) program with a nonexperimental comparison group from the Current Population Survey (CPS) or the Panel Study of Income Dynamics (PSID). The premise of this literature is that a successful nonexperimental estimation method should closely replicate the experimental estimate of the effect of the program when combining the original treatment group with an artificial comparison group (see, e.g., LaLonde, 1986; Dehejia and Wahba, 1999) or the "effect" of zero when combining the latter with the original control group (see, e.g., Smith and Todd, 2005).

In our application, we closely follow a recent reanalysis of these data in Sant'Anna and Zhao (2020), who restrict their attention to samples combining the CPS comparison group and variants of the original control group, previously analyzed by LaLonde (1986), Dehejia and Wahba (1999) (the "DW" sample), and Smith and Todd (2005) (the "early RA" sample). As is standard in this literature, the outcome of interest is real earnings in 1978. Because Sant'Anna and Zhao (2020) focus on various difference-in-differences estimators, they often use the transformed outcome, $Y_{i2} - Y_{i1}$; here, this is equal to the difference between real earnings in 1978 and real earnings in 1975. The baseline covariates include age, years of education, real earnings in 1974, and indicator variables for less than 12 years of education, being married, being Black, and being Hispanic. Other covariate specifications also include additional higher-order and interaction terms.

Table 2 replicates every estimate and standard error in Sant'Anna and Zhao's Table 3, while also reporting a number of additional results.[5] The bottom line is that weighting and doubly robust estimators perform quite well in replicating the true effect of zero; except for

[5] "TWFE" corresponds to $\hat{\tau}^{fe}$ in Sant'Anna and Zhao (2020). This is the OLS estimate from a panel data specification with real earnings in 1975 and 1978, a "treatment" indicator, and unit and year fixed effects. "RA" corresponds to $\hat{\tau}^{reg}$ in Sant'Anna and Zhao (2020). IPW with MLE weights corresponds to $\hat{\tau}^{ipw,p}$. NIPW with MLE weights corresponds to $\hat{\tau}^{ipw,p}_{std}$. NAIPW with MLE weights corresponds to $\hat{\tau}^{dr,p}$. Finally, all the estimates in the "IPT weights" panel are identical to Sant'Anna and Zhao's preferred estimator, $\hat{\tau}^{dr,p}_{imp}$.

Table 2: Replication of Sant'Anna and Zhao (2020)

	LaLonde sample			DW sample			Early RA sample		
TWFE	868**	868**	868**	2,092***	2,092***	2,092***	1,136	1,136	1,136
	(353)	(359)	(352)	(459)	(472)	(458)	(730)	(752)	(728)
RA	–1,301***	–830**	–1,041***	–230	402	27	–831	–264	–498
	(350)	(360)	(358)	(408)	(426)	(428)	(583)	(596)	(591)
	MLE weights								
IPW	–1,108***	–732	–685	188	–34	97	–516	–495	–337
	(409)	(534)	(523)	(459)	(845)	(793)	(611)	(781)	(740)
NIPW	–1,022**	–564	–558	155	481	502	–515	–223	–165
	(398)	(487)	(485)	(452)	(672)	(653)	(607)	(718)	(700)
AIPW	–859**	–613	–575	247	409	584	–434	–244	–124
	(399)	(513)	(504)	(449)	(779)	(727)	(605)	(753)	(718)
NAIPW	–871**	–626	–597	253	408	514	–434	–246	–148
	(396)	(496)	(491)	(451)	(691)	(663)	(605)	(724)	(701)
IPWRA	–908**	–590	–599	247	531	533	–443	–173	–143
	(394)	(467)	(470)	(452)	(581)	(577)	(607)	(682)	(677)
	IPT weights								
IPW	–901**	–591	–599	253	520	524	–441	–176	–144
	(394)	(467)	(470)	(452)	(588)	(582)	(607)	(683)	(677)
NIPW	–901**	–591	–599	253	520	524	–441	–176	–144
	(394)	(467)	(470)	(452)	(588)	(582)	(607)	(683)	(677)
AIPW	–901**	–591	–599	253	520	524	–441	–176	–144
	(394)	(467)	(470)	(452)	(588)	(582)	(607)	(683)	(677)
NAIPW	–901**	–591	–599	253	520	524	–441	–176	–144
	(394)	(467)	(470)	(452)	(588)	(582)	(607)	(683)	(677)
IPWRA	–901**	–591	–599	253	520	524	–441	–176	–144
	(394)	(467)	(470)	(452)	(588)	(582)	(607)	(683)	(677)
Linear	✓			✓			✓		
DW		✓			✓			✓	
ADW			✓			✓			✓
Observations	16,417	16,417	16,417	16,252	16,252	16,252	16,134	16,134	16,134

Notes: The source of data is LaLonde (1986). The outcome of interest is real earnings in 1978. "TWFE" corresponds to the OLS estimation of a panel data specification with outcomes measured in 1975 and 1978, a "treatment" indicator, as well as unit and year fixed effects. "RA" corresponds to the regression adjustment estimator, which is based on the OLS estimation of a fully interacted linear model with transformed outcomes. The remaining estimators use transformed outcomes and a logit model for the propensity score, and are defined in Section 3. "Linear" corresponds to a specification in which all covariates are included linearly. "DW" corresponds to a specification that includes all the covariates in the linear specification as well as an indicator for zero earnings in 1974, age squared, age cubed divided by 1000, years of schooling squared, and an interaction between years of schooling and real earnings in 1974. "ADW" corresponds to a specification that includes all the covariates in the DW specification as well as interactions between married and real earnings in 1974 and between married and zero earnings in 1974. Standard errors are in parentheses.

*Statistically significant at the 10% level; **at the 5% level; ***at the 1% level.

the simplest covariate specification applied to the LaLonde sample, none of these estimates are significantly different from the true effect. In addition, Sant'Anna and Zhao (2020) argue that their preferred estimator based on IPT weights tends to have smaller standard errors than estimators based on MLE weights. While this is true, Table 2 also illustrates our previous point that Sant'Anna and Zhao's (2020) estimator might be unnecessarily complex; when using the IPT weights, even the simplest "unnormalized" IPW estimator is numerically equivalent to it. Our standard errors, obtained together with the point estimates using our companion Stata package, `teffects2`, are also identical to those reported by Sant'Anna and Zhao (2020).

6 Conclusion

Applied researchers face many ways to adjust for covariates, but in this paper, we show that several popular estimators are in fact identical under a simple condition. Specifically, our results assume that the propensity score is estimated using inverse probability tilting, a method of moments approach developed by Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2012, 2016). Estimators based on or equivalent to this approach have already become popular in difference-in-differences settings (cf. Sant'Anna and Zhao, 2020) and outside economics (cf. Hainmueller, 2012; Imai and Ratkovic, 2014). Our results, simplifying the set of alternative estimators available to researchers, offer a novel rationale for adopting this approach in various contexts, such as under unconfoundedness and in instrumental variables and difference-in-differences settings.

Appendix

Proof of Proposition 3.1 Consider estimating μ_1; the argument for μ_0 follows in the same way. First, with $\hat{p}_i \equiv p(\mathbf{X}_i\hat{\gamma}_{1,ipt})$, equation (2.6) 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 (3.13). Now, consider the AIPT estimator in (3.14). 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 (3.17). Given that $1 \in \mathbf{X}_i$, the first-order condition for the WLS estimator of β_1, following from (3.16), 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 $\hat{p}_i \equiv p(\mathbf{X}_i\hat{\gamma}_{0,ipt})$. Using simple algebra, the AIPW estimator of $\mu_{0|1}$ using IPT weights, given in (3.19), can

be written as

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

where the final equality follows from equation (2.8), the IPT moment conditions for estimating the ATT.

For the IPWRA estimator using the IPT weights, note that the first-order condition for $\tilde{\beta}_0$, following from (3.20), 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 (2.8),

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

This completes the proof.

References

ABADIE, A. (2005): "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 72(1), 1–19.

ABADIE, A., AND M. D. CATTANEO (2018): "Econometric Methods for Program Evaluation," *Annual Review of Economics*, 10, 465–503.

AIZER, A., S. ELI, J. FERRIE, AND A. LLERAS-MUNEY (2016): "The Long-Run Impact of Cash Transfers to Poor Families," *American Economic Review*, 106(4), 935–971.

BLANDHOL, C., J. BONNEY, M. MOGSTAD, AND A. TORGOVITSKY (2025): "When Is TSLS Actually LATE?," NBER Working Paper No. 29709.

BRUNS-SMITH, D., O. DUKES, A. FELLER, AND E. L. OGBURN (2025): "Augmented Balancing Weights as Linear Regression," *Journal of the Royal Statistical Society, Series B*, forthcoming.

CAETANO, C., AND B. CALLAWAY (2024): "Difference-in-Differences When Parallel Trends Holds Conditional on Covariates," arXiv preprint arXiv:2406.15288.

CALLAWAY, B., AND P. H. C. SANT'ANNA (2021): "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225(2), 200–230.

CHATTOPADHYAY, A., AND J. R. ZUBIZARRETA (2023): "On the Implied Weights of Linear Regression for Causal Inference," *Biometrika*, 110(3), 615–629.

CHEN, J. (2025): "Potential Weights and Implicit Causal Designs in Linear Regression," arXiv preprint arXiv:2407.21119.

CHERNOZHUKOV, V., W. K. NEWEY, AND R. SINGH (2022): "Automatic Debiased Machine Learning of Causal and Structural Effects," *Econometrica*, 90(3), 967–1027.

DEHEJIA, R. H., AND S. WAHBA (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94(448), 1053–1062.

EGEL, D., B. S. GRAHAM, AND C. C. D. X. PINTO (2008): "Inverse Probability Tilting and Missing Data Problems," NBER Working Paper No. 13981.

FERMAN, B., C. PINTO, AND V. POSSEBOM (2020): "Cherry Picking with Synthetic Controls," *Journal of Policy Analysis and Management*, 39(2), 510–532.

FRÖLICH, M. (2007): "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 139(1), 35–75.

GOLDSMITH-PINKHAM, P., P. HULL, AND M. KOLESÁR (2024): "Contamination Bias in Linear Regressions," *American Economic Review*, 114(12), 4015–4051.

GRAHAM, B. S., C. C. D. X. PINTO, AND D. EGEL (2012): "Inverse Probability Tilting for Moment Condition Models with Missing Data," *Review of Economic Studies*, 79(3), 1053–1079.

——— (2016): "Efficient Estimation of Data Combination Models by the Method of Auxiliary-to-Study Tilting (AST)," *Journal of Business & Economic Statistics*, 34(2), 288–301.

HAINMUELLER, J. (2012): "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political Analysis*, 20(1), 25–46.

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

HIRANO, K., G. W. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71(4), 1161–1189.

IMAI, K., AND M. RATKOVIC (2014): "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society, Series B*, 76(1), 243–263.

IMBENS, G. W., AND J. M. WOOLDRIDGE (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86.

KLINE, P. (2011): "Oaxaca-Blinder as a Reweighting Estimator," *American Economic Review: Papers & Proceedings*, 101(3), 532–537.

KNAUS, M. C. (2024): "Treatment Effect Estimators as Weighted Outcomes," arXiv preprint arXiv:2411.11559.

LALONDE, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76(4), 604–620.

LEE, S. J., AND J. M. WOOLDRIDGE (2024): "A Simple Transformation Approach to Difference-in-Differences Estimation for Panel Data," working paper.

ROBINS, J. M., A. ROTNITZKY, AND L. P. ZHAO (1994): "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89(427), 846–866.

ROBINS, J. M., M. SUED, Q. LEI-GOMEZ, AND A. ROTNITZKY (2007): "Comment: Performance of Double-Robust Estimators When "Inverse Probability" Weights Are Highly Variable," *Statistical Science*, 22(4), 544–559.

SANT'ANNA, P. H. C., X. SONG, AND Q. XU (2022): "Covariate Distribution Balance via Propensity Scores," *Journal of Applied Econometrics*, 37(6), 1093–1120.

SANT'ANNA, P. H. C., AND J. ZHAO (2020): "Doubly Robust Difference-in-Differences Estimators," *Journal of Econometrics*, 219(1), 101–122.

SIMMONS, J. P., L. D. NELSON, AND U. SIMONSOHN (2011): "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science*, 22(11), 1359–1366.

SINGH, R., AND L. SUN (2024): "Double Robustness for Complier Parameters and a Semi-Parametric Test for Complier Characteristics," *Econometrics Journal*, 27(1), 1–20.

SŁOCZYŃSKI, T. (2022): "Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights," *Review of Economics and Statistics*, 104(3), 501–509.

——— (2024): "When Should We (Not) Interpret Linear IV Estimands as LATE?," arXiv preprint arXiv:2011.06695.

SŁOCZYŃSKI, T., S. D. UYSAL, AND J. M. WOOLDRIDGE (2022): "Doubly Robust Estimation of Local Average Treatment Effects Using Inverse Probability Weighted Regression Adjustment," arXiv preprint arXiv:2208.01300.

——— (2025): "Abadie's Kappa and Weighting Estimators of the Local Average Treatment Effect," *Journal of Business & Economic Statistics*, 43(1), 164–177.

SŁOCZYŃSKI, T., AND J. M. WOOLDRIDGE (2018): "A General Double Robustness Result for Estimating Average Treatment Effects," *Econometric Theory*, 34(1), 112–133.

SMITH, J. A., AND P. E. TODD (2005): "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?," *Journal of Econometrics*, 125, 305–353.

TAN, Z. (2020): "Regularized Calibrated Estimation of Propensity Scores with Model Misspecification and High-Dimensional Data," *Biometrika*, 107(1), 137–158.

UYSAL, S. D. (2024): "Estimation of Causal Effects with a Binary Treatment Variable: A Unified M-Estimation Framework," *Journal of Econometric Methods*, 13(1), 145–204.

VIVALT, E. (2019): "Specification Searching and Significance Inflation Across Time, Methods and Disciplines," *Oxford Bulletin of Economics and Statistics*, 81(4), 797–816.

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

——— (2010): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge–London, 2nd edn.

ZHAO, Q. (2019): "Covariate Balancing Propensity Score by Tailored Loss Functions," *Annals of Statistics*, 47(2), 965–993.

ZHAO, Q., AND D. PERCIVAL (2017): "Entropy Balancing Is Doubly Robust," *Journal of Causal Inference*, 5(1), 20160010.

ZUBIZARRETA, J. R. (2015): "Stable Weights That Balance Covariates for Estimation with Incomplete Outcome Data," *Journal of the American Statistical Association*, 110(511), 910–922.

SUPPLEMENTAL APPENDIX FOR "COVARIATE BALANCING AND THE EQUIVALENCE OF WEIGHTING AND DOUBLY ROBUST ESTIMATORS OF AVERAGE TREATMENT EFFECTS"

TYMON SŁOCZYŃSKI* S. DERYA UYSAL†
JEFFREY M. WOOLDRIDGE‡

This appendix explains how to obtain IPT propensity score estimates in Stata and R, as well as how to install and use our companion Stata package, `teffects2`. This package implements IPW, AIPW, and IPWRA estimators of the ATE and ATT under unconfoundedness, with several approaches to estimate the weights, including IPT. The package can also be used to estimate the ATT in difference-in-differences settings after a suitable transformation of the outcome variable. Throughout this appendix, as well as in `teffects2`, we restrict our attention to the logit model. In what follows, among other things, we will show how to estimate this model using the method of moments approach of Egel, Graham, and Pinto (2008) and Graham, Pinto, and Egel (2012, 2016) instead of maximum likelihood.

Implementation in Stata

This code illustrates IPT estimation by reproducing the estimate in column 1 of Table 2, which corresponds to the first entry in column 2 of Table 3 in Sant'Anna and Zhao (2020). The parameter of interest is the ATT, and the estimation procedure is based on the sample moment conditions in (2.5). The code can easily be modified for use in other applications.

First, we show how to reproduce this estimate using `teffects2`. To download this package, type

```
ssc install teffects2, all
```

in the Command window. Then, run the following code:

*Department of Economics, Brandeis University, tslocz@brandeis.edu
†Department of Economics, LMU Munich, derya.uysal@econ.lmu.de
‡Department of Economics, Michigan State University, wooldri1@msu.edu

```
* Load the data
use lalonde, clear

* Restrict attention to the NSW control and CPS comparison units
keep if (dataset == 0 | dataset == 4) & treated == 0

* Recode the NSW control units as "treated" (cf. Smith and Todd, 2005)
replace treated = 1 if dataset == 0

* Specify outcome, treatment, and control variables
local Y diff
local W treated
local X age educ re74 nodegree married black hispanic

* Estimate the ATT using teffects2
teffects2 ipw (`Y') (`W' `X', ipt), atet
teffects2 aipw (`Y' `X') (`W' `X', ipt), atet
teffects2 ipwra (`Y' `X') (`W' `X', ipt), atet
```

As implied by Proposition 3.2, the estimates (and standard errors) obtained with `teffects2 ipw`, `teffects2 aipw`, and `teffects2 ipwra` are identical, except for negligible differences due to floating-point precision. The output also matches the IPT estimate in column 1 of Table 2 as well as the first entry in column 2 of Table 3 in Sant'Anna and Zhao (2020). For example, with `teffects2 ipwra`, we obtain:

```
. teffects2 ipwra (`Y' `X') (`W' `X', ipt), atet

Treatment effect estimation                     Number of obs     =     16,417
Estimator      : IPW regression adjustment
Outcome model  : linear
Treatment model: logit IPT
------------------------------------------------------------------------------
             |               Robust
        diff | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         ATT |  -901.2702   393.6127    -2.29   0.022    -1672.737   -129.8036
      POmean |   2964.636   254.5088    11.65   0.000     2465.808    3463.464
------------------------------------------------------------------------------
```

In line with Stata's official `teffects` command, which only allows maximum likelihood estimation of the propensity score, `POmean` corresponds to an estimate of the mean untreated outcome (when estimating the ATE) or the mean untreated outcome among the treated (when estimating the ATT, as in the example above).

Second, we show how to obtain the underlying propensity score estimates, $p(\mathbf{X}_i\hat{\gamma}_{0,ipt})$, and how to reproduce the estimate of the ATT from scratch, i.e., without using `teffects2`.

```
* Download the data
use https://tslocz.github.io/lalonde.dta, clear

* Restrict attention to the NSW control and CPS comparison units
keep if (dataset == 0 | dataset == 4) & treated == 0

* Recode the NSW control units as "treated" (cf. Smith and Todd, 2005)
replace treated = 1 if dataset == 0

* Standardize nonbinary covariates
egen age_std = std(age)
egen educ_std = std(educ)
egen re74_std = std(re74)

* Specify outcome, treatment, and control variables
local Y diff
local W treated
local X age_std educ_std re74_std nodegree married black hispanic

* Set up the method of moments procedure
local eq0 (eq0: ((1 - `W') * (1 + exp({that0: `X' _cons})) - 1))
local inst0 instruments(eq0: `X')

* Obtain the IPT propensity score estimates
gmm `eq0', `inst0'
predict double xb0, xb equation(that0)
generate double p_hat0 = logistic(xb0)

* Estimate the ATT
generate double term = (p_hat0 * (1 - `W') * `Y') / (1 - p_hat0)
```

```
summarize term
scalar m1_hat = r(mean)
summarize ‘Y’ if ‘W’ == 1
scalar m2_hat = r(mean)
summarize ‘W’
scalar m3_hat = r(mean)
scalar att = m2_hat - m1_hat / m3_hat
```

The final estimate, implementing the IPW estimator of the ATT with the IPT weights, matches the `teffects2` estimate above, as well as the appropriate estimates in Table 2 and Sant'Anna and Zhao (2020):

```
. display att
-901.27028
```

Although this is not necessary to estimate the ATT, a researcher interested in the ATE also needs to obtain $p(\mathbf{X}_i\hat{\gamma}_{1,ipt})$ using the sample moment conditions in (2.3). To compute these estimates, the code above should be modified as follows:

```
* Set up the method of moments procedure
local eq1 (eq1: (‘W’ * (1 + exp({that1: ‘X’ _cons})) / exp({that1:}) - 1))
local inst1 instruments(eq1: ‘X’)

* Obtain the IPT propensity score estimates
gmm ‘eq1’, ‘inst1’
predict double xb1, xb equation(that1)
generate double p_hat1 = logistic(xb1)
```

In this example, `gmm` fails to converge with default settings, but does converge under some alternative optimization routines. The fact that obtaining $p(\mathbf{X}_i\hat{\gamma}_{1,ipt})$ is more difficult than obtaining $p(\mathbf{X}_i\hat{\gamma}_{0,ipt})$ should be treated as informative rather than problematic, as it reflects the underlying identification challenge—in the LaLonde (1986) data, it is simply very difficult to reweight the experimental subjects to resemble the CPS participants on average.

Implementation in R

As above, we show how to use IPT to reproduce the ATT estimate in column 1 of Table 2, which corresponds to the first entry in column 2 of Table 3 in Sant'Anna and Zhao (2020).

```
# Install and load the add-on package geex
install.packages("geex")
library(geex)

# Download the data
lalonde <- read.csv("https://tslocz.github.io/lalonde.csv")

# Restrict attention to the NSW control and CPS comparison units
nswcps <- subset(lalonde, (dataset %in% c(0, 4)) & treated == 0)

# Recode the NSW control units as "treated" (cf. Smith and Todd, 2005)
nswcps$W <- as.integer(nswcps$dataset == 0)

# Standardize nonbinary covariates
nswcps$age_std  <- as.numeric(scale(nswcps$age))
nswcps$educ_std <- as.numeric(scale(nswcps$educ))
nswcps$re74_std <- as.numeric(scale(nswcps$re74))

# Specify outcome, treatment, and control variables
Y <- nswcps$diff
W <- nswcps$W
X <- model.matrix(
  ~ age_std + educ_std + re74_std + nodegree + married + black + hispanic,
  data = nswcps
)

# Set up supporting objects
df <- data.frame(W = W, X, check.names = FALSE)
X_cols <- colnames(X)

# Specify starting values
p <- length(X_cols)
gamma_start <- numeric(p)

# Set up the method of moments procedure
score_eq0 <- function(data) {
```

```
  W_i <- data$W
  X_i <- as.vector(as.matrix(data[X_cols]))
  function(theta) {
    eta_i <- sum(X_i * theta)
    p_i   <- plogis(eta_i)
    ((1 - W_i) / (1 - p_i) - 1) * X_i
  }
}

# Obtain the IPT propensity score estimates
mest_eq0 <- m_estimate(
  estFUN = score_eq0,
  data   = df,
  root_control = setup_root_control(start = gamma_start)
)
gamma0  <- as.numeric(coef(mest_eq0))
p_hat0  <- as.vector(plogis(X %*% gamma0))

# Estimate the ATT
term    <- (p_hat0 * (1 - W) * Y) / (1 - p_hat0)
m1_hat  <- mean(term)
m2_hat  <- mean(Y[W == 1])
m3_hat  <- mean(W)
att     <- m2_hat - m1_hat / m3_hat
```

The resulting estimate, implementing the IPW estimator of the ATT with the IPT weights, matches both Stata estimates above, as well as the appropriate estimates in Table 2 and Sant'Anna and Zhao (2020):

```
> att
[1] -901.2702
```

To obtain $p(\mathbf{X}_i\hat{\gamma}_{1,ipt})$, the code above should be modified as follows:

```
# Set up the method of moments procedure
score_eq1 <- function(data) {
  W_i <- data$W
  X_i <- as.vector(as.matrix(data[X_cols]))
```

```
  function(theta) {
    eta_i <- sum(X_i * theta)
    p_i   <- plogis(eta_i)
    (W_i / p_i - 1) * X_i
  }
}

# Obtain the IPT propensity score estimates
mest_eq1 <- m_estimate(
  estFUN = score_eq1,
  data   = df,
  root_control = setup_root_control(start = gamma_start)
)
gamma1  <- as.numeric(coef(mest_eq1))
p_hat1  <- as.vector(plogis(X %*% gamma1))
```

As in Stata, obtaining $p(\mathbf{X}_i\hat{\gamma}_{1,ipt})$ is challenging; however, convergence is achievable with alternative solver choices and better starting values, such as the MLE coefficients.

Appendix References

EGEL, D., B. S. GRAHAM, AND C. C. D. X. PINTO (2008): "Inverse Probability Tilting and Missing Data Problems," NBER Working Paper No. 13981.

GRAHAM, B. S., C. C. D. X. PINTO, AND D. EGEL (2012): "Inverse Probability Tilting for Moment Condition Models with Missing Data," *Review of Economic Studies*, 79(3), 1053–1079.

——— (2016): "Efficient Estimation of Data Combination Models by the Method of Auxiliary-to-Study Tilting (AST)," *Journal of Business & Economic Statistics*, 34(2), 288–301.

LALONDE, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76(4), 604–620.

SANT'ANNA, P. H. C., AND J. ZHAO (2020): "Doubly Robust Difference-in-Differences Estimators," *Journal of Econometrics*, 219(1), 101–122.