Estimation of Causal Effects using Propensity Score

Derya Uysal

November 18, 2025

Research Presentation
based on "Covariate Balancing and the Equivalence of Weighting and Doubly Robust Estimators of Average Treatment Effects"

with Tymon Słoczyński & Jeffrey M. Wooldridge

University of Innsbruck

The goal and the problem

- There is a huge interest in using observational data to estimate the effects of treatments on outcomes.
    - **Do training programs increase participants' wages?**
    - Does remote work increase employee productivity?
    - Does financial literacy training reduce household debt problems?
    - ...
- Treatment assignment in observational studies is not random!
- Treated subjects often differ systematically from those of untreated subjects
    - Randomized experiments
    - Statistical Methods based on non-experimental data

  $\Rightarrow$ **There is a wide range of statistical methods!**

> In this paper, we provide new insights which are helpful in reducing the number of alternative methods in a meaningful way.

The goal and the problem

- There is a huge interest in using observational data to estimate the effects of treatments on outcomes.
  - **Do training programs increase participants' wages?**
  - Does remote work increase employee productivity?
  - Does financial literacy training reduce household debt problems?
  - ...
- Treatment assignment in observational studies is not random!
- Treated subjects often differ systematically from those of untreated subjects
  - Randomized experiments
  - Statistical Methods based on non-experimental data

  $\Rightarrow$ **There is a wide range of statistical methods!**

> In this paper, we provide new insights which are helpful in reducing the number of alternative methods in a meaningful way.

The goal and the problem

- There is a huge interest in using observational data to estimate the effects of treatments on outcomes.
  - **Do training programs increase participants' wages?**
  - Does remote work increase employee productivity?
  - Does financial literacy training reduce household debt problems?
  - ...
- Treatment assignment in observational studies is not random!
- Treated subjects often differ systematically from those of untreated subjects
  - Randomized experiments
  - Statistical Methods based on non-experimental data

  ⇒ **There is a wide range of statistical methods!**

> In this paper, we provide new insights which are helpful in reducing the number of alternative methods in a meaningful way.

The goal and the problem

- There is a huge interest in using observational data to estimate the effects of treatments on outcomes.
  - **Do training programs increase participants' wages?**
  - Does remote work increase employee productivity?
  - Does financial literacy training reduce household debt problems?
  - ...
- Treatment assignment in observational studies is not random!
- Treated subjects often differ systematically from those of untreated subjects
  - Randomized experiments
  - Statistical Methods based on non-experimental data

  $\Rightarrow$ **There is a wide range of statistical methods!**

> In this paper, we provide new insights which are helpful in reducing the number of alternative methods in a meaningful way.

The goal and the problem

- There is a huge interest in using observational data to estimate the effects of treatments on outcomes.
  - **Do training programs increase participants' wages?**
  - Does remote work increase employee productivity?
  - Does financial literacy training reduce household debt problems?
  - ...
- Treatment assignment in observational studies is not random!
- Treated subjects often differ systematically from those of untreated subjects
  - Randomized experiments
  - Statistical Methods based on non-experimental data

  ⇒ **There is a wide range of statistical methods!**

> In this paper, we provide new insights which are helpful in reducing the number of alternative methods in a meaningful way.

The goal and the problem

- There is a huge interest in using observational data to estimate the effects of treatments on outcomes.
  - **Do training programs increase participants' wages?**
  - Does remote work increase employee productivity?
  - Does financial literacy training reduce household debt problems?
  - ...
- Treatment assignment in observational studies is not random!
- Treated subjects often differ systematically from those of untreated subjects
  - Randomized experiments
  - Statistical Methods based on non-experimental data

$\Rightarrow$ **There is a wide range of statistical methods!**

> In this paper, we provide new insights which are helpful in reducing the number of alternative methods in a meaningful way.

The goal and the problem

- There is a huge interest in using observational data to estimate the effects of treatments on outcomes.
    - **Do training programs increase participants' wages?**
    - Does remote work increase employee productivity?
    - Does financial literacy training reduce household debt problems?
    - ...
- Treatment assignment in observational studies is not random!
- Treated subjects often differ systematically from those of untreated subjects
    - Randomized experiments
    - Statistical Methods based on non-experimental data

  $\Rightarrow$ **There is a wide range of statistical methods!**

> In this paper, we provide new insights which are helpful in reducing the number of alternative methods in a meaningful way.

The goal and the problem

- There is a huge interest in using observational data to estimate the effects of treatments on outcomes.
  - **Do training programs increase participants' wages?**
  - Does remote work increase employee productivity?
  - Does financial literacy training reduce household debt problems?
  - ...
- Treatment assignment in observational studies is not random!
- Treated subjects often differ systematically from those of untreated subjects
  - Randomized experiments
  - Statistical Methods based on non-experimental data

  $\Rightarrow$ **There is a wide range of statistical methods!**

In this paper, we provide new insights which are helpful in reducing the number of alternative methods in a meaningful way.

Potential Outcome Framework

Binary Treatment: (participation in a job training program or not)

$$W = \begin{cases} 1, & \text{participation in a job training program} \\ 0, & \text{otherwise} \end{cases}$$

Potential Outcome:

- $Y(1)$: wages in case of participation
- $Y(0)$: wages in case of non participation

Observed Outcome: (observed wages)

$$Y = \begin{cases} Y(1), & \text{if } W = 1 \\ Y(0), & \text{if } W = 0 \end{cases}$$

or, in a more compact notation: $Y = (1 - W)Y(0) + WY(1)$.

X denotes the vector of observable characteristics like age, gender, race, education, etc.

Potential Outcome Framework

Binary Treatment: (participation in a job training program or not)

$$W = \begin{cases} 1, & \text{participation in a job training program} \\ 0, & \text{otherwise} \end{cases}$$

Potential Outcome:

- $Y(1)$: wages in case of participation
- $Y(0)$: wages in case of non participation

Observed Outcome: (observed wages)

$$Y = \begin{cases} Y(1), & \text{if } W = 1 \\ Y(0), & \text{if } W = 0 \end{cases}$$

or, in a more compact notation: $Y = (1 - W)Y(0) + WY(1)$.

X denotes the vector of observable characteristics like age, gender, race, education, etc.

Potential Outcome Framework

Binary Treatment: (participation in a job training program or not)

$$W = \begin{cases} 1, & \text{participation in a job training program} \\ 0, & \text{otherwise} \end{cases}$$

Potential Outcome:

- $Y(1)$: wages in case of participation
- $Y(0)$: wages in case of non participation

Observed Outcome: (observed wages)

$$Y = \begin{cases} Y(1), & \text{if } W = 1 \\ Y(0), & \text{if } W = 0 \end{cases}$$

or, in a more compact notation: $Y = (1 - W)Y(0) + WY(1)$.

X denotes the vector of observable characteristics like age, gender, race, education, etc.

Treatment Effect Parameters

Average Treatment Effect (ATE)

$$\tau_{ate} = \mathbb{E}[Y(1) - Y(0)]. \tag{1}$$

Average Treatment Effect on Treated (ATT)

$$\tau_{att} = \mathbb{E}[Y(1) - Y(0) | W = 1]. \tag{2}$$

Treatment Effect Parameters

Average Treatment Effect (ATE)

$$\tau_{ate} = \mathbb{E}[Y(1) - Y(0)]. \tag{1}$$

Average Treatment Effect on Treated (ATT)

$$\tau_{att} = \mathbb{E}[Y(1) - Y(0)|W = 1]. \tag{2}$$

Motivation

- Under standard identification assumptions, several competing estimators exist for the same parameters of interest.

Motivation

- Under standard identification assumptions, several competing estimators exist for the same parameters of interest.
- Many of them involve weighting by the inverse of the propensity score.

Motivation

- Under standard identification assumptions, several competing estimators exist for the same parameters of interest.
- Many of them involve weighting by the inverse of the propensity score.
- What is propensity score?

Motivation

- Under standard identification assumptions, several competing estimators exist for the same parameters of interest.
- Many of them involve weighting by the inverse of the propensity score.
- What is propensity score?
  - The propensity score (PS) is a **widely used concept** in causal inference.

Motivation

- Under standard identification assumptions, several competing estimators exist for the same parameters of interest.
- Many of them involve weighting by the inverse of the propensity score.
- What is propensity score?
  - The propensity score (PS) is a **widely used concept** in causal inference.
  - It measures the **probability of the treatment** given the confounders, i.e. probability of participating in the training program conditional individual characteristics like age, education, etc..

Motivation

- Under standard identification assumptions, several competing estimators exist for the same parameters of interest.
- Many of them involve weighting by the inverse of the propensity score.
- What is propensity score?
  - The propensity score (PS) is a **widely used concept** in causal inference.
  - It measures the **probability of the treatment** given the confounders, i.e. probability of participating in the training program conditional individual characteristics like age, education, etc..
  - Formally:

$$p(x) = \Pr(W_i = 1 \mid X_i = x)$$

Motivation

- Under standard identification assumptions, several competing estimators exist for the same parameters of interest.
- Many of them involve weighting by the inverse of the propensity score.
- What is propensity score?
  - The propensity score (PS) is a **widely used concept** in causal inference.
  - It measures the **probability of the treatment** given the confounders, i.e. probability of participating in the training program conditional individual characteristics like age, education, etc..
  - Formally:

    $$p(x) = \Pr(W_i = 1 \mid X_i = x)$$

  - Propensity-score weights create a synthetic sample $\Rightarrow$ mimics a randomized experiment

Motivation

- Under standard identification assumptions, several competing estimators exist for the same parameters of interest.
- Many of them involve weighting by the inverse of the propensity score.
- What is propensity score?
    - The propensity score (PS) is a **widely used concept** in causal inference.
    - It measures the **probability of the treatment** given the confounders, i.e. probability of participating in the training program conditional individual characteristics like age, education, etc..
    - Formally:
    $$p(x) = \Pr(W_i = 1 \mid X_i = x)$$
    - Propensity-score weights create a synthetic sample ⇒ mimics a randomized experiment
    - But, PS is usually not known and has to be estimated

Contribution in a nutshell

We consider three classes of estimation approaches of average treatment effects involving weighting by the propensity score:

- inverse probability weighting (IPW), as in Hirano, Imbens, and Ridder (2003),
- augmented inverse probability weighting (AIPW), as in Robins, Rotnitzky, and Zhao (1994),
- inverse probability weighted regression adjustment (IPWRA), as in Wooldridge (2007), Uysal (2016), Słoczyński and Wooldridge (2018).

Relying these approaches one can construct, in fact, **five different estimators** for each treatment effect parameter.

Contribution in a nutshell

- Equivalences for the ATE and ATT: We show that if ones uses a particular method of moments approach to estimate the unknown propensity score model, then the five competing estimators deliver the same estimates for the ATE and ATT.
- Equivalences for other settings: We also show that these equivalences have interesting implications, for two popular settings and treatment effect parameters:
  - Instrumental Variable (IV) setting and the LATE (Local Average Treatment Effect)
  - Difference-in Differences (DID) setting and the ATT

Contribution in a nutshell

- Equivalences for the ATE and ATT: We show that if ones uses a particular method of moments approach to estimate the unknown propensity score model, then the five competing estimators deliver the same estimates for the ATE and ATT.
- Equivalences for other settings: We also show that these equivalences have interesting implications, for two popular settings and treatment effect parameters:
  - Instrumental Variable (IV) setting and the LATE (Local Average Treatment Effect)
  - Difference-in Differences (DID) setting and the ATT

The equivalence results are interesting in their own right, but they also narrow the menu of options available to applied researchers significantly!

Today's Roadmap

Estimation

Equivalence of estimators

Implications for other settings

Estimation

Propensity score (PS)

- Remember the propensity score is the probability of receiving the treatment conditional on the confounders, e.g. probability of getting the training conditional on the observable characteristics.
- Propensity scores serve two main purposes:

    1. Modeling the participation decision conditional on confounders: The propensity score represents the probability of receiving a treatment given observed covariates.
    2. Balancing Property: Conditional on the propensity score, the distribution of observed covariates is similar between treated and untreated subjects.

- Model the propensity score as a function of the covariates:
    - Assume an index model, $p(\mathbf{x}\gamma)$, where $\mathbf{x}$ is $1 \times K$, γ is $K \times 1$, and $x_1 = 1$, so that the linear index always includes an intercept.
    - In most applications, $p(\mathbf{x}\gamma) = \exp(\mathbf{x}\gamma) / [1 + \exp(\mathbf{x}\gamma)]$, i.e. logistic probability function

Maximum Likelihood Approach

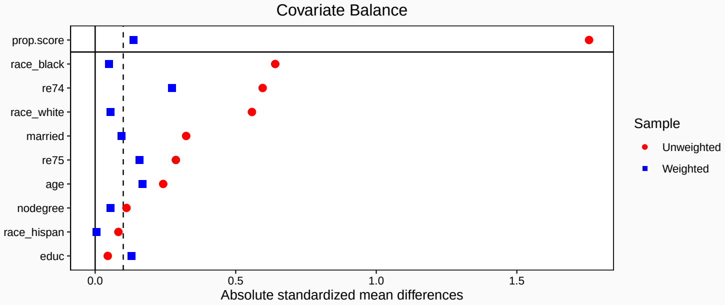
- Propensity score (PS) estimation has historically relied on maximum likelihood estimation (MLE) of a standard binary response model, such as logit or probit.
  - Estimate γ by ML $\Rightarrow \hat{\gamma}_{mle} \Rightarrow$ $p(\boldsymbol{X}_i \hat{\gamma}_{mle})$

Maximum Likelihood Approach

- Propensity score (PS) estimation has historically relied on maximum likelihood estimation (MLE) of a standard binary response model, such as logit or probit.
  - Estimate γ by ML $\Rightarrow \hat{\gamma}_{mle} \Rightarrow$ $p(\boldsymbol{X}_i \hat{\gamma}_{mle})$
- While MLE is natural choice when the goal is to estimate the parameters in the propensity score model, it does not necessarily satisfy balancing property (in finite samples).

Example: Lalonde data

- **Treatment (W):** Participation in job training program
- **Outcome (Y):** Post-program earnings in 1978 (`re78`)
- **Covariates (X):** `age`, `educ`, `race`, `married`, `nodegree`, `re74`, `re75`



Standardized mean differences (SMDs) before (unweighted) and after PS weighting. PS is estimated by maximum likelihood. Dashed line at $|\text{SMD}| = 0.1$ is a common balance threshold. The "propensity score" row shows balance of the estimated treatment probability.

Inverse Probability Tilting Approach

- To improve estimation the propensity score, thus the estimation of the ATE, different methods are proposed.
- One approach involves using balancing moment conditions to estimate the propensity score, i.e. to estimate γ, and use these estimated scores to construct the weights (Graham, Pinto and Egel, 2012)

Inverse Probability Tilting Approach

- To improve estimation the propensity score, thus the estimation of the ATE, different methods are proposed.
- One approach involves using balancing moment conditions to estimate the propensity score, i.e. to estimate γ, and use these estimated scores to construct the weights (Graham, Pinto and Egel, 2012)
- The inverse probability tilting (IPT) moment conditions proposed by Graham, Pinto and Egel (2012) are as follows:

Inverse Probability Tilting Approach

- To improve estimation the propensity score, thus the estimation of the ATE, different methods are proposed.
- One approach involves using balancing moment conditions to estimate the propensity score, i.e. to estimate γ, and use these estimated scores to construct the weights (Graham, Pinto and Egel, 2012)
- The inverse probability tilting (IPT) moment conditions proposed by Graham, Pinto and Egel (2012) are as follows:

For the treated:

$$\mathrm{E}\left[\frac{W}{p(\mathbf{x}\gamma)}\boldsymbol{X}'\right] = \mathrm{E}\left[\boldsymbol{X}'\right] \tag{3}$$

and their sample analog is:

$$N^{-1}\sum_{i=1}^{N}\frac{W_i\boldsymbol{X}_i}{p(\boldsymbol{X}_i\hat{\gamma}_{1,ipt})} = \bar{\boldsymbol{X}} \tag{4}$$

These equations define the IPT estimator of γ, $\hat{\gamma}_{1,ipt}$.

Inverse Probability Tilting Approach

- To improve estimation the propensity score, thus the estimation of the ATE, different methods are proposed.
- One approach involves using balancing moment conditions to estimate the propensity score, i.e. to estimate γ, and use these estimated scores to construct the weights (Graham, Pinto and Egel, 2012)
- The inverse probability tilting (IPT) moment conditions proposed by Graham, Pinto and Egel (2012) are as follows:

For the treated:

$$\mathrm{E}\left[\frac{W}{p(\mathbf{x}\gamma)}\boldsymbol{X}'\right] = \mathrm{E}\left[\boldsymbol{X}'\right] \tag{3}$$

and their sample analog is:

$$N^{-1}\sum_{i=1}^{N}\frac{W_i\boldsymbol{X}_i}{p(\boldsymbol{X}_i\hat{\gamma}_{1,ipt})} = \bar{\boldsymbol{X}} \tag{4}$$

These equations define the IPT estimator of γ, $\hat{\gamma}_{1,ipt}$.

For the controls:

$$\mathrm{E}\left[\frac{1-W}{1-p(\mathbf{x}\gamma)}\boldsymbol{X}'\right] = \mathrm{E}\left[\boldsymbol{X}'\right], \tag{5}$$

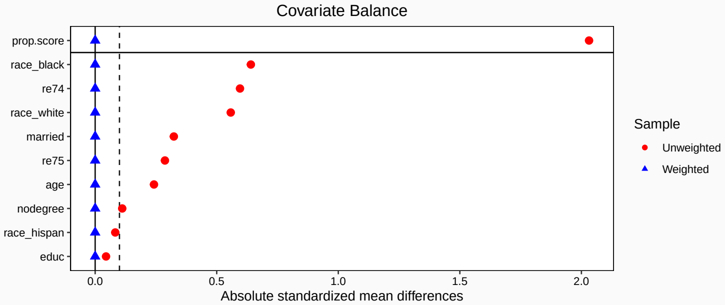
with the corresponding sample analog:

$$N^{-1}\sum_{i=1}^{N}\frac{(1-W_i)\boldsymbol{X}_i}{1-p(\boldsymbol{X}_i\hat{\gamma}_{0,ipt})} = \bar{\boldsymbol{X}}. \tag{6}$$

These equations define the IPT estimator $\hat{\gamma}_{0,ipt}$ of γ.

Example: Lalonde data

- **Treatment (W):** Participation in job training program
- **Outcome (Y):** Post-program earnings in 1978 (`re78`)
- **Covariates (X):** `age`, `educ`, `race`, `married`, `nodegree`, `re74`, `re75`



Standardized mean differences (SMDs) before (unweighted) and after PS weighting. PS is estimated by IPT (Graham, Pinto and Egel, 2012). Dashed line at $|\text{SMD}| = 0.1$ is a common balance threshold. The "propensity score" row shows balance of the estimated treatment probability.

Estimators of the ATE with MLE based weights

Models: $\Pr[W=1|X=x] = p(\mathbf{x}\boldsymbol{\gamma})$, $\mathrm{E}[Y(1)|X=x] = x\beta_1$, $\mathrm{E}[Y(0)|X=x] = x\beta_0$

Method	ATE estimator $\hat{\tau}_{ate}$	Outcome model estimation
IPW	$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{W_iY_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})} - \frac{(1-W_i)Y_i}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})}\right)$	none
NIPW	$\sum_{i=1}^{N}\frac{\frac{W_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})}}{\sum_{i=1}^{N}\frac{W_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})}}Y_i - \sum_{i=1}^{N}\frac{\frac{1-W_i}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})}}{\sum_{i=1}^{N}\frac{1-W_i}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})}}Y_i$	none
AIPW	$N^{-1}\sum_{i=1}^{N}\frac{W_i(Y_i-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_1)}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})} + N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_1$ $-N^{-1}\sum_{i=1}^{N}\frac{(1-W_i)(Y_i-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_0)}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})} + N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_0.$	$\hat{\boldsymbol{\beta}}_1 = \arg\min_{\boldsymbol{b}}\frac{1}{N}\sum_{i=1}^{N}W_i\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_1\right)^2,$ $\hat{\boldsymbol{\beta}}_0 = \arg\min_{\boldsymbol{b}}\frac{1}{N}\sum_{i=1}^{N}(1-W_i)\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_0\right)^2$
NAIPW	similar normalization as in NIPW	same as AIPW
IPWRA	$N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_1 - N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_0.$	$\tilde{\boldsymbol{\beta}}_1 = \arg\min_{\boldsymbol{b}}\frac{1}{N}\sum_{i=1}^{N}\frac{W_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})}\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_1\right)^2,$ $\tilde{\boldsymbol{\beta}}_0 = \arg\min_{\boldsymbol{b}}\frac{1}{N}\sum_{i=1}^{N}\frac{1-W_i}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{mle})}\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_0\right)^2$

Estimators of the ATE with IPT based weights

Models: $\Pr[W=1|X=x]=p(\boldsymbol{x\gamma})$, $\mathrm{E}[Y(1)|X=x]=x\beta_1$, $\mathrm{E}[Y(0)|X=x]=x\beta_0$

Method	ATE estimator $\hat{\tau}_{ate}$	Outcome model estimation
IPW	$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{W_iY_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{1,ipt})}-\frac{(1-W_i)Y_i}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{0,ipt})}\right)$	none
NIPW	ex-ante normalized. i.e., $\sum_{i=1}^{N}\frac{W_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{1,ipt})}=\sum_{i=1}^{N}\frac{(1-W_i)}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{0,ipt})}=N$	none
AIPW	$N^{-1}\sum_{i=1}^{N}\frac{W_i(Y_i-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_1)}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{1,ipt})}+N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_1$ $-N^{-1}\sum_{i=1}^{N}\frac{(1-W_i)(Y_i-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_0)}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{0,ipt})}+N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_0.$	$\hat{\boldsymbol{\beta}}_1=\arg\min_{\boldsymbol{b}}\frac{1}{N}\sum_{i=1}^{N}W_i\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_1\right)^2,$ $\hat{\boldsymbol{\beta}}_0=\arg\min_{\boldsymbol{b}}\frac{1}{N}\sum_{i=1}^{N}(1-W_i)\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_0\right)^2$
NAIPW	ex-ante normalized	same as AIPW
IPWRA	$N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_1-N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_0.$	$\tilde{\boldsymbol{\beta}}_1=\arg\min_{\boldsymbol{b}}\frac{1}{N}\sum_{i=1}^{N}\frac{W_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{1,ipt})}\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_1\right)^2,$ $\tilde{\boldsymbol{\beta}}_0=\arg\min_{\boldsymbol{b}}\frac{1}{N}\sum_{i=1}^{N}\frac{1-W_i}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{0,ipt})}\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_0\right)^2$

Estimators of the ATE with IPT based weights

Models: $\Pr[W=1|X=x] = p(\mathbf{x}\boldsymbol{\gamma})$, $\mathrm{E}[Y(1)|X=x] = x\beta_1$, $\mathrm{E}[Y(0)|X=x] = x\beta_0$

Method	ATE estimator $\hat{\tau}_{ate}$	Outcome model estimation
IPW	$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{W_i Y_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{1,ipt})} - \frac{(1-W_i)Y_i}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{0,ipt})}\right)$	none
NIPW	ex-ante normalized. i.e., $\sum_{i=1}^{N}\frac{W_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{1,ipt})} = \sum_{i=1}^{N}\frac{(1-W_i)}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{0,ipt})} = N$	none
AIPW	$N^{-1}\sum_{i=1}^{N}\frac{W_i(Y_i-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_1)}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{1,ipt})} + N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_1$ $-N^{-1}\sum_{i=1}^{N}\frac{(1-W_i)(Y_i-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_0)}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{0,ipt})} + N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_0.$	$\hat{\boldsymbol{\beta}}_1 = \arg\min_{\boldsymbol{b}} \frac{1}{N}\sum_{i=1}^{N} W_i\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_1\right)^2,$ $\hat{\boldsymbol{\beta}}_0 = \arg\min_{\boldsymbol{b}} \frac{1}{N}\sum_{i=1}^{N}(1-W_i)\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_0\right)^2$
NAIPW	ex-ante normalized	same as AIPW
IPWRA	$N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_1 - N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_0.$	$\tilde{\boldsymbol{\beta}}_1 = \arg\min_{\boldsymbol{b}} \frac{1}{N}\sum_{i=1}^{N}\frac{W_i}{p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{1,ipt})}\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_1\right)^2,$ $\tilde{\boldsymbol{\beta}}_0 = \arg\min_{\boldsymbol{b}} \frac{1}{N}\sum_{i=1}^{N}\frac{1-W_i}{1-p(\boldsymbol{X}_i\hat{\boldsymbol{\gamma}}_{0,ipt})}\left(Y_i-\boldsymbol{X}_i\boldsymbol{b}_0\right)^2$

Equivalence of estimators

Equivalence of estimators #1

Proposition 1

If one uses IPT weights to estimate τ_{ate} where conditional means $\mathrm{E}\left[\, Y(0) \middle|\, \boldsymbol{X} \right]$ and $\mathrm{E}\left[\, Y(1) \middle|\, \boldsymbol{X} \right]$ are modeled linearly, then all five commonly used ATE estimators are identical. By implication, the IPW and AIPW estimators are automatically normalized.

Equivalence of estimators #1

Proposition 1

If one uses IPT weights to estimate τ_{ate} where conditional means $\mathrm{E}\left[\, Y(0) |\, \boldsymbol{X} \right]$ and $\mathrm{E}\left[\, Y(1) |\, \boldsymbol{X} \right]$ are modeled linearly, then all five commonly used ATE estimators are identical. By implication, the IPW and AIPW estimators are automatically normalized.

More formally;

Let $\hat{\gamma}_{1,ipt}$ be the IPT estimates of γ for the treated group, with $\hat{p}_i = p(\boldsymbol{X}_i \hat{\gamma}_{1,ipt}) > 0$ for all i and $\hat{\gamma}_{0,ipt}$ be the IPT estimates of γ for the treated group, with $\hat{p}_i = p(\boldsymbol{X}_i \hat{\gamma}_{0,ipt}) > 0$ for all i. Then IPW, NIPW, AIPW, and IPWRA estimates of τ_{ate} are identical.

Estimators of the ATT with IPT based weights

Recall that

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

and the first term can be consistently estimated by the sample mean of Y_i over the treated units, $\bar{Y}_1$. Only $\mu_{0|1}$ needs to be estimated.

Method	ATT estimator $\hat{\mu}_{0\|1}$	Outcome model estimation
IPW	$N_1^{-1} \sum_{i=1}^{N} \frac{p(\boldsymbol{X}_i \hat{\boldsymbol{\gamma}}_{0,ipt})\,(1-Wi)\,Y_i}{1 - p(\boldsymbol{X}_i \hat{\boldsymbol{\gamma}}_{0,ipt})}.$	none
NIPW	ex-ante normalized. i.e., $\sum_{i=1}^{N} \frac{p(\boldsymbol{X}_i \hat{\boldsymbol{\gamma}}_{0,ipt})\,(1-Wi)\,Y_i}{1 - p(\boldsymbol{X}_i \hat{\boldsymbol{\gamma}}_{0,ipt})} = N_1$	none
AIPW	$N_1^{-1} \sum_{i=1}^{N} \frac{p(\boldsymbol{X}_i \hat{\boldsymbol{\gamma}}_{0,ipt})\,(1-Wi)}{1 - p(\boldsymbol{X}_i \hat{\boldsymbol{\gamma}}_{0,ipt})} \left(Y_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}_0\right) +$ $N_1^{-1} \sum_{i=1}^{N} W_i \boldsymbol{X}_i \hat{\boldsymbol{\beta}}_0$	$\hat{\boldsymbol{\beta}}_0 = \arg\min_{\boldsymbol{b}} \frac{1}{N} \sum_{i=1}^{N} (1-W_i)\left(Y_i - \boldsymbol{X}_i \boldsymbol{b}_0\right)^2$
NAIPW	ex-ante normalized	same as AIPW
IPWRA	$N_1^{-1} \sum_{i=1}^{N} W_i \boldsymbol{X}_i \tilde{\boldsymbol{\beta}}_0$	$\tilde{\boldsymbol{\beta}}_0 = \arg\min_{\boldsymbol{b}_0} N^{-1} \sum_{i=1}^{N} \frac{p(\boldsymbol{X}_i \hat{\boldsymbol{\gamma}}_{0,ipt})\,(1-W_i)}{1 - p(\boldsymbol{X}_i \hat{\boldsymbol{\gamma}}_{0,ipt})} \left(Y_i - \boldsymbol{X}_i \boldsymbol{b}_0\right)^2$

Equivalence of estimators #2

Proposition 2

If one uses IPT weights to estimate $\mathrm{E}\left[\, Y(0) \,|\, W = 1\right]$ and $\mathrm{E}\left[\, Y(0) \,|\, \boldsymbol{X}\right]$ is modeled linearly, then all five commonly used ATT estimators yield identical results. By implication, the IPW and AIPW estimators are automatically normalized.

Equivalence of estimators #2

Proposition 2

If one uses IPT weights to estimate $\mathrm{E}\left[\, Y(0) \middle|\, W = 1\right]$ and $\mathrm{E}\left[\, Y(0) \middle|\, \boldsymbol{X}\right]$ is modeled linearly, then all five commonly used ATT estimators yield identical results. By implication, the IPW and AIPW estimators are automatically normalized.

Formally;

Let $\hat{\gamma}_{0,ipt}$ be the estimators solving moment equations for the control observations with $\hat{p}_i = p(\boldsymbol{X}_i \hat{\gamma}_{0,ipt}) < 1$ for all i. Then the IPW, NIPW, AIPW, NAIPW, and IPWRA estimates of $\mu_{0|1}$ using the IPT weights, and linear conditional means in the latter two cases, are identical. Therefore, the five estimates of τ_{att} are identical.

Implications for other settings

Implications for IV

As before, W is a binary treatment. Now we also have a binary instrumental variable, Z.

Local Average Treatment Effect (LATE)

$$\tau_{late} = \mathbb{E}[Y_1 - Y_0 | W_1 > W_0] \tag{7}$$

- It follows from Frölich (2007) that many estimators of the LATE are ratios of estimators:

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

- It follows from Proposition 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, NIPW, AIPW, NAIPW 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:

$$\Pr(Z_i = 1 \mid X_i = x) = q(X\delta)$$

Implications for difference-in-differences (DID)

- Some popular estimators in DID settings are based on applying standard treatment effect estimators - after suitably transforming the outcome variable.
- Abadie (2005) proposed an IPW estimator of the ATT.
- Sant'Anna and Zhao (2020) proposed AIPW/RA type doubly robust ("DR DID") estimators of this parameter.

$\Rightarrow$ When both use IPT base propensity score estimates, Sant'Anna and Zhao (2020) is identical to Abadie (2005).

- Similar conclusions hold in settings with multiple periods and staggered interventions.

Implications for difference-in-differences (DID)

- Some popular estimators in DID settings are based on applying standard treatment effect estimators - after suitably transforming the outcome variable.
- Abadie (2005) proposed an IPW estimator of the ATT.
- Sant'Anna and Zhao (2020) proposed AIPW/RA type doubly robust ("DR DID") estimators of this parameter.

⇒ When both use IPT base propensity score estimates, Sant'Anna and Zhao (2020) is identical to Abadie (2005).

- Similar conclusions hold in settings with multiple periods and staggered interventions.

Bottom line: IPT based covariate balancing approach unifies competing DID estimators as well as LATE estimators.

Conclusion

- We provide several numerical equivalence results for estimators of average treatment effects
- When the propensity score is estimated via inverse probability tilting (IPT) (Egel et al., 2008; Graham et al., 2012, 2016):
  - IPW, NIPW, AIPW, NAIPW and IPWRA estimators yield identical results across diverse settings
- Our findings offer a unifying rationale for IPT:
  Robust, efficient, and simplifies the choice among estimators
- We introduce new Stata and R packages `teffects2`,
  - both of which allow estimation of ATE and ATT using IPT-based weights.

Scan the QR code to access our paper

Empirical Illustrations

Aizer, Eli, Ferrie, and Lleras-Muney (2016)

- the long-run impacts of the Mothers' Pension (MP) program on longevity
- three covariate specifications and two sources of information on dates of death: program records and death certificates
- We estimate the ATE and ATT by IPW, AIPW, IPWRA and the normalized versions of first two
- with MLE weights: range from 0.0014 to 0.0597 for the ATE and from –0.0014 to 0.0645 for the ATT
- with IPT weights: the choice of an estimator is inconsequential
- in the case of IPT, the standard errors are usually slightly smaller than in the case of the corresponding estimates based on MLE.

Empirical Illustrations

Sant'Anna and Zhao (2020)

- reanalysis of well known Lalonde (1986) data
- treatment: participation in the National Supported Work (NSW) program
- outcome: the difference between real earnings in 1978 and real earnings in 1975
- baseline covariates: age, years of education, real earnings in 1974, and indicator variables for less than12 years of education, being married, being Black, and being Hispanic.
- all the estimates with IPT weights are identical to Sant'Anna and Zhao's preferred estimator
- have smaller standard errors than estimators based on MLE weights