Causal Inference using Difference-in-Differences

Lecture 5: How covariates can make your DiD more plausible

Pedro H. C. Sant'Anna
Emory University

January 2025

Summary of previous lectures

Canonical DiD setup

- So far, we have considered the canonical DiD setup.
  - 2 time periods: $t = 1$ (before treatment) and $t = 2$ (after treatment).
  - 2 groups: $G = 2$ (treated at period 2) and $G = \infty$ (untreated by period 2).
- Main parameter of interest: Average Treatment Effect among Treated units

$$ATT \equiv \underbrace{\mathbb{E}\left[Y_{t=2}(2)\,|G=2\right]}_{\text{estimable from the data}} - \underbrace{\mathbb{E}\left[Y_{t=2}(\infty)\,|G=2\right]}_{\text{counterfactual component}}$$

Canonical DiD setup

Identification of the ATT is achieved via two main assumptions:

Assumption (No-Anticipation)

For all units i, $Y_{i,t}(g) = Y_{i,t}(\infty)$ *for all groups in their pre-treatment periods, i.e., for all* $t < g$.

Assumption (Parallel Trends Assumption)

$$\mathbb{E}\left[Y_{i,t=2}(\infty)|G_i = 2\right] - \mathbb{E}\left[Y_{i,t=1}(\infty)|G_i = 2\right] = \mathbb{E}\left[Y_{i,t=2}(\infty)|G_i = \infty\right] - \mathbb{E}\left[Y_{i,t=1}(\infty)|G_i = \infty\right]$$

PS: We are taking SUTVA for granted from now onwards (NOT without loss of generality, though)

"Brute force" DiD estimator

- Canonical DiD Estimator:

$$\widehat{\theta}_n^{DiD} = \left(\bar{Y}_{g=2,t=2} - \bar{Y}_{g=2,t=1}\right) - \left(\bar{Y}_{g=\infty,t=2} - \bar{Y}_{g=\infty,t=1}\right),$$

where $\bar{Y}_{g=d,t=j}$ is the sample mean of the outcome Y for units in group d in time period j,

$$\bar{Y}_{g=d,t=j} = \frac{1}{N_{g=d,t=j}} \sum_{i=1}^{N \cdot T} Y_i 1\{G_i = d\} 1\{T_i = j\},$$

with

$$N_{g=d,t=j} = \sum_{i=1}^{N \cdot T} 1\{G_i = d\} 1\{T_i = j\},$$

G_i and T_i are group and time dummy, respectively, and Y_i is the "poolled" outcome data.

"TWFE" DiD estimator

- In practice, most of us would rely on the following TWFE regression specification to estimate the *ATT*:

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{\equiv ATT} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + \varepsilon_{i,t},$$

where $\mathbb{E}[\varepsilon_{i,t} | G_i, T_i] = 0$ *almost surely.*

What if the Parallel Trends Assumption is not plausible?

Conditional Parallel Trends

Unconditional parallel trends assumption

- So far, covariates have played no role in our analysis
- But what if units with different observed characteristics were to evolve differently in the absence of treatment?
  - *Effect of Minimum wage on employment:* is it sensible to assume that, in the absence of treatment, employment in states in the NE of the US would have evolved similarly as in states in the South of the US?
  - *Effect of training on earnings:* is it reasonable to assume that earnings among young workers would have evolved similarly to older workers in the absence of treatment?
- In general, the PTA may be implausible if pre-treatment characteristics that are thought to be associated with the dynamics of the outcome variable are "unbalanced" between the treated and the untreated group. (Abadie, 2005).

How can we “relax” the PTA and

allow for “covariate-specific” trends?

Conditional Parallel Trends Assumption

- In order to “relax” the PTA, we can assume that it holds **only after** conditioning on a vector of observed pre-treatment covariates

Assumption (Conditional Parallel Trends Assumption)

$$\mathbb{E}\left[Y_{t=2}(\infty)|G=2,X\right]-\mathbb{E}\left[Y_{t=1}(\infty)|G=2,X\right]=\mathbb{E}\left[Y_{t=2}(\infty)|G=\infty,X\right]-\mathbb{E}\left[Y_{t=1}(\infty)|G=\infty,X\right]\quad a.s.$$

- The conditional PT assumption states that, in the absence of treatment, conditional on X, the evolution of the outcome among the treated units is, on average, the same as the evolution of the outcome among the untreated units.

- **It allows for covariate-specific trends!**

Strong overlap

- When covariates are available, we will introduce an additional assumption stating that every unit has a strictly positive probability of being in the untreated group.

Assumption (Strong Overlap Assumption)

The conditional probability of belonging to the treatment group, given observed characteristics X, is uniformly bounded away from 1.

That is, for some $\epsilon > 0$, $\mathbb{P}[G = 2|X] < 1 - \epsilon$ *almost surely.*

- The covariates X here are the same as those used to justify the conditional PT assumption!
- For identification purposes, we can take $\epsilon = 0$. For (standard) inference, though, we would have problems without relying on "extrapolation"; see, e.g., Khan and Tamer (2010).

How do the conditional PTA and

and strong overlap help us, DiDistas?

Identification of ATT under conditional parallel trends and overlap

1) First, recall the conditional PT assumption:

$$\mathbb{E}\left[Y_{t=2}(\infty)|G=2,X\right]-\mathbb{E}\left[Y_{t=1}(\infty)|G=2,X\right]=\mathbb{E}\left[Y_{t=2}(\infty)|G=\infty,X\right]-\mathbb{E}\left[Y_{t=1}(\infty)|G=\infty,X\right].$$

2) By simple manipulation, we can write it as

$$\mathbb{E}\left[Y_{t=2}\left(\infty\right)|G=2,X\right] \quad = \quad \mathbb{E}\left[Y_{t=1}\left(\infty\right)|G=2,X\right]+\left(\mathbb{E}\left[Y_{t=2}\left(\infty\right)|G=\infty,X\right]-\mathbb{E}\left[Y_{t=1}\left(\infty\right)|G=\infty,X\right]\right)$$

3) Now, exploiting No-Anticipation, SUTVA, and strong overlap:

$$\mathbb{E}\left[Y_{t=2}\left(\infty\right)|G=2,X\right] \quad = \quad \underbrace{\mathbb{E}\left[Y_{t=1}\left(2\right)|G=2,X\right]}_{by\ No-Anticipation}+\left(\mathbb{E}\left[Y_{t=2}\left(\infty\right)|G_{=}\infty,X\right]-\mathbb{E}\left[Y_{t=1}\left(\infty\right)|G=\infty,X\right]\right)$$

$$\mathbb{E}\left[Y_{t=2}\left(\infty\right)|G=2,X\right] \quad = \quad \underbrace{\mathbb{E}\left[Y_{t=1}|G=2,X\right]+\left(\mathbb{E}\left[Y_{t=2}|G=\infty,X\right]-\mathbb{E}\left[Y_{t=1}|G=\infty,X\right]\right)}_{by\ SUTVA+overlap}$$

Conditional Parallel Trends and the conditional ATT

- Let's define the **Conditional ATT**: $ATT(X) \equiv \mathbb{E}\left[Y_{t=2}(2) - Y_{t=2}(\infty) | G = 2, X\right]$.
- Now, combining the results of previous slides, we have that, under SUTVA + No-Anticipation + Conditional PT assumptions, it follows that:

$$
\begin{aligned}
\text{ATT}(\text{X}) &= \mathbb{E}\left[Y_{t=2}|G=2,X\right] - \left(\mathbb{E}\left[Y_{t=1}|G=2,X\right] + \left(\mathbb{E}\left[Y_{t=2}|G=\infty,X\right] - \mathbb{E}\left[Y_{t=1}|G=\infty,X\right]\right)\right) \\
&= \left(\mathbb{E}\left[Y_{t=2}|G=2,X\right] - \mathbb{E}\left[Y_{t=1}|G=2,X\right]\right) - \left(\mathbb{E}\left[Y_{t=2}|G=\infty,X\right] - \mathbb{E}\left[Y_{t=1}|G=\infty,X\right]\right)
\end{aligned}
$$

- We can identify the conditional ATT function - **a very rich object!**
- This also implies that the unconditional ATT is identified - all we have to do is to integrate X among treated units:

$$ATT = \mathbb{E}\left[ATT(X)|G=2\right]$$

Conditional Parallel Trends and the unconditional ATT

In terms of estimable pieces, we get that

$$\begin{aligned}
\textbf{ATT} &= \mathbb{E}\left[\left(\mathbb{E}\left[Y_{t=2} | G=2, X\right] - \mathbb{E}\left[Y_{t=1} | G=2, X\right]\right) - \left(\mathbb{E}\left[Y_{t=2} | G=\infty, X\right] - \mathbb{E}\left[Y_{t=1} | G=\infty, X\right]\right) \middle| G=2\right] \\
&= \left(\mathbb{E}\left[Y_{t=2} | G=2\right] - \mathbb{E}\left[Y_{t=1} | G=2\right]\right) - \mathbb{E}\left[\left(\mathbb{E}\left[Y_{t=2} | G=\infty, X\right] - \mathbb{E}\left[Y_{t=1} | G=\infty, X\right]\right) \middle| G=2\right]
\end{aligned}$$

where the second equality follows from the Law of Iterated Expectations and covariates and group indicators being stationary (which hold by construction on a balanced panel; we will come back to this in a bit).

Can we use a simple regression here?

Usage of simple TWFE linear regressions with covariates

Usage of simple TWFE linear regressions with covariates

The temptation

TWFE DiD estimator

- Under unconditional PTA, we have shown that we can use the TWFE regression to recover the *ATT*:

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{\equiv ATT} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + \varepsilon_{i,t},$$

  where $\mathbb{E}[\varepsilon_{i,t}|G_i, T_i] = 0$ *almost surely.*

- It is **very tempting** to “extrapolate” and use the “more general” TWFE regression specification:

$$Y_{i,t} = \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 1\{G_i = 2\} + \tilde{\lambda}_0 1\{T_i = 2\} + \underbrace{\tilde{\beta}_0^{twfe}}_{????} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + X_i'\tilde{\alpha}_{0,2} + \tilde{\varepsilon}_{i,t},$$

  where $\mathbb{E}[\tilde{\varepsilon}_{i,t}|G_i, T_i, X_i] = 0$ *almost surely.*

Is $\tilde{\beta}_0^{twfe}$ "similar" to the ATT?

Usage of simple TWFE linear regressions with covariates

Simulation exercise

Monte Carlo simulation exercise

- This is a great point to illustrate the power of simulations to assess if "intuitive" extensions are sensible.
- Here, knowing the "truth" help us to hold our methods accountable.
- In this particular exercise, we will use a Data generating process similar to those of Kang and Schafer (2007)
- Samples sizes $n = 1,000$
- For each design, we consider $10,000$ Monte Carlo experiments
- Available data are $\{Y_{t=2}, Y_{t=1}, D, X\}_{i=1}^{n}$, where $D_i = 1\{G_i = 2\}$.

Monte Carlo simulation exercise

- Covariates are generated as $X_j \sim N(0,1)$, $j = 1, 2, 3, 4$.
- Let $X = (X_1, X_2, X_3, X_4)$, and

$$f_{reg}(X) = 210 + 27.4 \cdot X_1 + 13.7 \cdot (X_2 + X_3 + X_4)$$
$$f_{ps}(X) = 0.75 \cdot (-X_1 + 0.5 \cdot X_2 - 0.25 \cdot X_3 - 0.1 \cdot X_4)$$

- Also, let

$$v(X, D) \overset{d}{\sim} N(D \cdot f_{reg}(X), 1)$$
$$\varepsilon_{t=1} \overset{d}{\sim} N(0,1)$$
$$\varepsilon_{t=2}(2) \overset{d}{\sim} N(0,1)$$
$$\varepsilon_{t=2}(\infty) \overset{d}{\sim} N(0,1)$$
$$U \overset{d}{\sim} U(0,1)$$

How we generate potential outcomes and group indicators

$$
\begin{aligned}
Y_{i,t=1}(\infty) &= f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{i,t=1} \\
Y_{i,t=2}(\infty) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{i,t=2}(\infty) \\
Y_{i,t=2}(2) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{i,t=2}(2) \\
p(X_i) &= \frac{\exp(f_{ps}(X_i))}{1 + \exp(f_{ps}(X_i))} \\
D_i &= 1\{p(X_i) \geq U\}
\end{aligned}
$$

In this setup, $ATT(X) = 0$ $a.s.$

Simulation results

We estimate $\tilde{\beta}_0^{twfe}$ from the following specification:

$$Y_{i,t} = \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 1\{G_i = 2\} + \tilde{\lambda}_0 1\{T_i = 2\} + \tilde{\beta}_0^{twfe} \left(1\{G_i = 2\} \cdot 1\{T_i = 2\}\right) + X_i' \tilde{\alpha}_{0,2} + \tilde{\varepsilon}_{i,t},$$

- Average of the $\widehat{\tilde{\beta}}_0^{twfe}$ in the simulations: -16.36 (**very biased!**)
- Coverage probability of 95% Confidence Interval: 0 (**does not control size!**)

Simulation results

Figure 1: Monte Carlo for TWFE-based estimators

Why there is so much bias here?

Usage of simple TWFE linear regressions with covariates

The problems of the simple TWFE specification with covariates

Simple TWFE DiD regression estimator with covariates

- The TWFE specification is given by

$$Y_{i,t} = \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 1\{G_i = 2\} + \tilde{\lambda}_0 1\{T_i = 2\} + \tilde{\beta}_0^{twfe}\left(1\{G_i = 2\} \cdot 1\{T_i = 2\}\right) + X_i'\tilde{\alpha}_{0,2} + \tilde{\varepsilon}_{i,t},$$

where $\mathbb{E}[\varepsilon_{i,t}|G_i, T_i, X_i] = 0$ *almost surely.*

- Now, let's play with its terms:

$$\begin{aligned}
\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 1, X_i] &= \tilde{\alpha}_{0,1} + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\lambda}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 1, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + \tilde{\lambda}_0 + \tilde{\beta}_0^{twfe} + X_i'\tilde{\alpha}_{0,2}
\end{aligned}$$

Simple TWFE regression estimator with covariates

- Set of moment restrictions:

$$\begin{aligned}
\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 1, X_i] &= \tilde{\alpha}_{0,1} + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\lambda}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 1, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + \tilde{\lambda}_0 + \tilde{\beta}_0^{twfe} + X_i'\tilde{\alpha}_{0,2}
\end{aligned}$$

- Let's analyze the implications of these moment restrictions, one by one.

EMORY UNIVERSITY

Simple TWFE regression estimator with covariates

- Set of moment restrictions:

$$\begin{aligned}
\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 1, X_i] &= \tilde{\alpha}_{0,1} + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\lambda}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 1, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + \tilde{\lambda}_0 + \tilde{\beta}_0^{twfe} + X_i'\tilde{\alpha}_{0,2}
\end{aligned}$$

- First, notice that

$$\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 1, X_i] = \tilde{\lambda}_0$$

Evolution of the outcome among untreated units does not depend on X!

Simple TWFE regression estimator with covariates

- Set of moment restrictions:

$$\begin{aligned}
\mathbb{E}[Y_{i,t}|G_i=\infty, T_i=1, X_i] &= \tilde{\alpha}_{0,1} + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i=\infty, T_i=2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\lambda}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i=2, T_i=1, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i=2, T_i=2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + \tilde{\lambda}_0 + \tilde{\beta}_0^{twfe} + X_i'\tilde{\alpha}_{0,2}
\end{aligned}$$

- Second, notice that

$$\mathbb{E}[Y_{i,t}|G_i=2, T_i=2, X_i] - \mathbb{E}[Y_{i,t}|G_i=2, T_i=1, X_i] = \tilde{\lambda}_0 + \tilde{\beta}_0^{twfe}$$

Evolution of the outcome among treated units does not depend on X!

Simple TWFE regression estimator with covariates

- Set of moment restrictions:

$$
\begin{aligned}
\mathbb{E}[Y_{i,t}|G_i=\infty, T_i=1, X_i] &= \tilde{\alpha}_{0,1} + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i=\infty, T_i=2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\lambda}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i=2, T_i=1, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + X_i'\tilde{\alpha}_{0,2} \\
\mathbb{E}[Y_{i,t}|G_i=2, T_i=2, X_i] &= \tilde{\alpha}_{0,1} + \tilde{\gamma}_0 + \tilde{\lambda}_0 + \tilde{\beta}_0^{twfe} + X_i'\tilde{\alpha}_{0,2}
\end{aligned}
$$

- Lastly, notice that, under conditional PT, No-Anticipation, and SUTVA,

$$
\begin{aligned}
ATT(X) &= \left(\mathbb{E}[Y_{i,t}|G_i=2, T_i=2, X_i] - \mathbb{E}[Y_{i,t}|G_i=2, T_i=1, X_i]\right) \\
&\quad - \left(\mathbb{E}[Y_{i,t}|G_i=\infty, T_i=2, X_i] - \mathbb{E}[Y_{i,t}|G_i=\infty, T_i=1, X_i]\right) \\
&= \tilde{\beta}_0^{twfe}
\end{aligned}
$$

Average Treatment effects are homogeneous between covariate subpopulations!

TWFE with covariates

Key to success:

Separate identification from estimation/inference!

How can we do it?

Alternative Estimands

Semi and nonparametric DiD procedures

Once you separate identification from estimation procedures, we realize that DiD with covariates has many faces!

Alternative Estimands

Regression adjustment

Regression adjustment procedure

- The "first face" of DiD procedure is already familiar.
- Originally proposed by Heckman, Ichimura and Todd (1997); Heckman, Ichimura, Smith and Todd (1998)
- Idea is to work directly from the identifying assumptions.
- We have already seen that, under conditional PT, No-anticipation, and SUTVA,

$$
\begin{aligned}
ATT &= \mathbb{E}\left[\left(\underbrace{\mathbb{E}\left[Y_{t=2}|G=2,X\right]}_{=m_{t=2}^{G=2}(X)} - \underbrace{\mathbb{E}\left[Y_{t=1}|G=2,X\right]}_{=m_{t=1}^{G=2}(X)}\right) - \left(\underbrace{\mathbb{E}\left[Y_{t=2}|G=\infty,X\right]}_{=m_{t=2}^{G=\infty}(X)} - \underbrace{\mathbb{E}\left[Y_{t=1}|G=\infty,X\right]}_{=m_{t=1}^{G=\infty}(X)}\right)\middle| G=2\right] \\
&= \mathbb{E}\left[\left(m_{t=2}^{G=2}(X) - m_{t=1}^{G=2}(X)\right) - \left(m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X)\right)\middle| G=2\right]
\end{aligned}
$$

Regression adjustment procedure

- Our life is a bit easier once we have that

$$ATT = \mathbb{E}\left[\left(m_{t=2}^{G=2}(X) - m_{t=1}^{G=2}(X)\right) - \left(m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X)\right) \middle| G = 2\right].$$

- Now, it is a matter of estimating the unknown **regression functions** $m_t^G(X)$ with your favorite estimation method - it can be parametric, semiparametric, or nonparametric!

Regression adjustment procedure

- Our life is a bit easier once we have that

$$ATT = \mathbb{E}\left[\left(m_{t=2}^{G=2}(X) - m_{t=1}^{G=2}(X)\right) - \left(m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X)\right)\Big| G = 2\right].$$

- For example, let $\mu_{t=s}^{G=g}(X) = X'\beta_{0,t=s}^{G=g}$ be a working model for $m_{t=s}^{G=g}(X)$.

- We can then estimate the betas in each subsample using OLS, compute the fitted values using all covariates values among treated units, and then average the combination of these fitted values:

$$\widehat{ATT}_n^{reg} = \mathbb{E}_n\left[\left(\widehat{\mu}_{t=2}^{G=2}(X) - \widehat{\mu}_{t=1}^{G=2}(X)\right) - \left(\widehat{\mu}_{t=2}^{G=\infty}(X) - \widehat{\mu}_{t=1}^{G=\infty}(X)\right)\Big| G = 2\right].$$

Regression adjustment with panel data

- Our life can be even easier if we have access to panel data:

Assumption (Panel Data Sampling Scheme)

The data $\{Y_{i,t=1}, Y_{i,t=2}, G_i, X_i\}_{i=i}^{n}$ *is a random sample of the population of interest.*

- Observing $Y_{t=1}$ and $Y_{t=2}$ for the same units allows us to simplify the formulas a lot!

$$
\begin{aligned}
ATT &= \mathbb{E}\left[\left(\mathbb{E}\left[Y_{t=2}|G=2,X\right] - \mathbb{E}\left[Y_{t=1}|G=2,X\right]\right) - \left(\mathbb{E}\left[Y_{t=2}|G=\infty,X\right] - \mathbb{E}\left[Y_{t=1}|G=\infty,X\right]\right)|G=2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y_{t=2}-Y_{t=1}|G=2,X\right] - \mathbb{E}\left[Y_{t=2}-Y_{t=1}|G=\infty,X\right]|G=2\right] \\
&= \mathbb{E}\left[Y_{t=2}-Y_{t=1}|G=2\right] - \mathbb{E}\left[\mathbb{E}\left[Y_{t=2}-Y_{t=1}|G=\infty,X\right]|G=2\right] \\
&= \mathbb{E}\left[Y_{t=2}-Y_{t=1}|G=2\right] - \mathbb{E}\left[m_{\Delta}^{G=\infty}(X)\,|G=2\right]
\end{aligned}
$$

- Only have to model one conditional expectation:

$$m_{\Delta}^{G=\infty}(X) \equiv \mathbb{E}\left[Y_{t=2}-Y_{t=1}|G=\infty,X\right]$$

Regression adjustment with stationary repeated cross-section data

- Sometimes, we only have access to (**stationary**) repeated cross-section data:

Assumption (Repeated Cross-Section Data Sampling Scheme)

The pooled repeated cross-section data $\{Y_i, G_i, T_i, X_i\}_{i=1}^{n}$ *consist of iid draws from the mixture distribution*

$$\begin{aligned} P\left(Y \leq y, X \leq x, G = g, T = t\right) &= 1\{t = 2\} \cdot \lambda \cdot P\left(Y_{t=2} \leq y, X \leq x, G = g | T = 2\right) \\ &\quad + 1\{t = 1\} \cdot (1 - \lambda)\, P\left(Y_{t=1} \leq y, X \leq x, G = g | T = 1\right), \end{aligned}$$

where $(y, x, g, t) \in \mathbb{R} \times \mathbb{R}^k \times \{2, \infty\} \times \{1, 2\}$, $\lambda = \mathbb{P}\left(T = 2\right) \in (0, 1)$.

Furthermore, $(G, X)\,|T = 1 \overset{d}{\sim} (G, X)\,|T = 2$, *i.e., there are no compositional changes over time.*

- Question: Would it be possible to allow compositional changes? What would change? How would you proceed?

Regression adjustment with stationary repeated cross-section data

- In this case, the formula can also be simplified (but not as much as in the case of panel data):

$$\begin{aligned} ATT &= \mathbb{E}\left[\left(m_{t=2}^{G=2}(X) - m_{t=1}^{G=2}(X)\right) - \left(m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X)\right) \middle| G=2\right] \\ &= \left(\mathbb{E}\left[Y|G=2, T=2\right] - \mathbb{E}\left[Y|G=2, T=1\right]\right) - \mathbb{E}\left[\left(m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X)\right) \middle| G=2\right] \end{aligned}$$

- We have to model conditional expectations only for untreated units:

$$m_{t=s}^{G=\infty}(X) = \mathbb{E}\left[Y|G=\infty, T=s, X\right], \quad s=1,2$$

Regression-adjusted DiD estimators

rely on researchers ability to model

the outcome evolution.

Alternative Estimands

Inverse Probability Weighting procedure

Inverse probability weighting procedures

- The "second face" of semi/nonparametric DiD procedures avoids directly modeling the outcome evolution.
- Instead, it models the propensity score, i.e., prob of belonging to the group $G = 2$: $p(X) \equiv \mathbb{P}(G = 2|X) = \mathbb{P}(D = 1|X)$, where $D = 1\{G = 2\}$.
- Originally proposed by Abadie (2005):

$$ATT^{ipw,p} = \frac{\mathbb{E}\left[\left(D - \frac{(1-D)p(X)}{1-p(X)}\right)(Y_{t=2} - Y_{t=1})\right]}{\mathbb{E}[D]},$$

$$ATT^{ipw,rc} = \frac{\mathbb{E}\left[\left(D - \frac{(1-D)p(X)}{1-p(X)}\right)\frac{1\{T=2\} - \lambda}{\lambda}Y\right]}{\mathbb{E}[D]},$$

where $\lambda = \mathbb{E}[1\{T = 2\}]$.

Inverse probability weighting procedures

$$ATT^{ipw,p} = \frac{\mathbb{E}\left[\left(D - \frac{(1-D)p\left(X\right)}{1-p(X)}\right)\left(Y_{t=2} - Y_{t=1}\right)\right]}{\mathbb{E}\left[D\right]},$$

$$ATT^{ipw,rc} = \frac{\mathbb{E}\left[\left(D - \frac{(1-D)p\left(X\right)}{1-p(X)}\right)\frac{T-\lambda}{\lambda}Y\right]}{\mathbb{E}\left[D\right]},$$

where $\lambda = \mathbb{E}\left[T\right]$.

- These formulas suggest a simple two-step estimation procedure, too!
  1. Choose your favorite method to estimate the unknown propensity score $p(X)$.
  2. Plug in the estimated fitted propensity score values into the *ATT* equation, and replace the population expectations with their sample analogue.

Inverse probability weighting procedures

- For example, let $\pi(X) = \Lambda(X) \equiv \frac{exp(X'\gamma_0)}{1+exp(X'\gamma_0)}$ be a working model for the propensity score

- We can estimate γ_0 using the logit maximum likelihood estimator.

- Let $\widehat{\pi}(X) = \frac{exp(X'\widehat{\gamma}_0)}{1+exp(X'\widehat{\gamma}_n)}$

- Abadie's proposed *ATT* estimator with panel data is

$$\widehat{ATT}_n^{ipw,p} = \frac{\mathbb{E}_n\left[\left(D - \frac{(1-D)\widehat{\pi}(X)}{1-\widehat{\pi}(X)}\right)(Y_{t=2} - Y_{t=1})\right]}{\mathbb{E}_n[D]}.$$

Hájek-based Inverse probability weighting procedures

- One potential drawback of Abadie's IPW DiD estimator is that their weights are not "normalized", i.e., they do not sum up to one.
- More formally, Abadie's IPW DiD estimator is of the Horvitz and Thompson (1952) type.
- We know from the survey literature that Hájek (1971)-type estimators can be more stable, as they use "normalized" weights.
- Building on this insight, Sant'Anna and Zhao (2020) built on Abadie (2005) and considered the Hájek (1971)-type IPW DiD estimands.

Hájek-based Inverse probability weighting with panel

Sant'Anna and Zhao (2020) considered the following estimand when Panel data are available:

$$
\begin{aligned}
ATT^{ipw,p}_{std} &= \mathbb{E}\left[\left(w^{p}_{G=2}(D) - w^{p}_{G=\infty}(D,X;p)\right)\left(Y_{t=2} - Y_{t=1}\right)\right] \\
&= \mathbb{E}\left[\left(\frac{D}{\mathbb{E}[D]} - \frac{\dfrac{p(X)(1-D)}{1-p(X)}}{\mathbb{E}\left[\dfrac{p(X)(1-D)}{1-p(X)}\right]}\right)\left(Y_{t=2} - Y_{t=1}\right)\right],
\end{aligned}
$$

where

$$
w^{p}_{G=2}(D) = \frac{D}{\mathbb{E}[D]}, \qquad \textit{and} \qquad w^{p}_{G=\infty}(D,X;g) = \frac{g(X)(1-D)}{1-g(X)} \Bigg/ \mathbb{E}\left[\frac{g(X)(1-D)}{1-g(X)}\right]
$$

Hájek-based Inverse probability weighting with repeated cross-section

Sant'Anna and Zhao (2020) considered the following estimand when stationary RCS data are available:

$$ATT_{std}^{ipw,rc} = \mathbb{E}\left[\left(w_{G=2}^{rc}\left(D,T\right) - w_{G=\infty}^{rc}\left(D,T,X;p\right)\right) \cdot Y\right]$$

where

$$\begin{aligned} w_{G=2}^{rc}\left(D,T\right) &= w_{G=2,t=2}^{rc}\left(D,T\right) - w_{G=2,t=1}^{rc}\left(D,T\right), \\ w_{G=\infty}^{rc}\left(D,T,X;g\right) &= w_{G=\infty,t=2}^{rc}\left(D,T,X;g\right) - w_{G=\infty,t=1}^{rc}\left(D,T,X;g\right), \end{aligned}$$

and, for $s = 1, 2$,

$$w_{G=2,t=s}^{rc}\left(D,T\right) = \frac{D \cdot 1\left\{T=s\right\}}{\mathbb{E}\left[D \cdot 1\left\{T=s\right\}\right]},$$

$$w_{G=\infty,t=s}^{rc}\left(D,T,X;g\right) = \frac{g(X)\left(1-D\right) \cdot 1\left\{T=s\right\}}{1-g(X)} \Bigg/ \mathbb{E}\left[\frac{g(X)\left(1-D\right) \cdot 1\left\{T=s\right\}}{1-g(X)}\right].$$

IPW-adjusted DiD estimators

rely on researchers ability to model

the propensity score.


OUTCOME REGRESSION
PROPENSITY SCORE
JAKE-CLARK.TUMBLR
imgflip.com

Alternative Estimands

Doubly robust DiD estimators

Doubly robust DiD procedures

- **Combine both outcome regression and IPW** approaches to form more robust estimators.
- Originally proposed by Sant'Anna and Zhao (2020)
- Estimators are **Doubly Robust consistent**: they are consistent for the ATT if either (but not necessarily both)
  - Regression working models for outcome dynamics are correctly specified
  - Propensity score working model is correctly specified

Doubly robust DiD procedure with panel

Sant'Anna and Zhao (2020) considered the following doubly robust estimand when panel data are available:

$$
\begin{aligned}
ATT^{dr,p} &= \mathbb{E}\left[\left(w^p_{G=2}(D) - w^p_{G=\infty}(D,X;p)\right)\left((Y_{t=2}-Y_{t=1}) - \left(m^{G=\infty}_{t=2}(X) - m^{G=\infty}_{t=1}(X)\right)\right)\right] \\
&= \mathbb{E}\left[\left(\frac{D}{\mathbb{E}[D]} - \frac{\frac{p(X)(1-D)}{1-p(X)}}{\mathbb{E}\left[\frac{p(X)(1-D)}{1-p(X)}\right]}\right)\left((Y_{t=2}-Y_{t=1}) - \left(m^{G=\infty}_{t=2}(X) - m^{G=\infty}_{t=1}(X)\right)\right)\right],
\end{aligned}
$$

where

$$
w^p_{G=2}(D) = \frac{D}{\mathbb{E}[D]}, \quad \text{and} \quad w^p_{G=\infty}(D,X;g) = \frac{g(X)(1-D)}{1-g(X)} \Big/ \mathbb{E}\left[\frac{g(X)(1-D)}{1-g(X)}\right]
$$

Doubly robust DiD procedure with panel

- Sant'Anna and Zhao (2020) also shown that $ATT^{dr,p}$ is semiparametrically (locally) efficient.

- If all working models are correctly specified, the DR DiD estimator for the $ATT^{dr,p}$ is "the most precise estimator" (minimum asymptotic variance) among all (regular) estimators that does not rely on additional functional form restrictions.

- Sant'Anna and Zhao (2020) also discuss how to get further improved DR DiD estimators by "carefully" choosing first-step estimators for the regression adjustment and propensity score working models.

- For the sake of time, we will not go into detail on these.

Doubly robust DiD procedure with repeated cross-section

Sant'Anna and Zhao (2020) considered two different doubly robust estimands when RCS data are available.

$$ATT_1^{dr,rc} = \mathbb{E}\left[\left(w_{G=2}^{rc}\left(D,T\right) - w_{G=\infty}^{rc}\left(D,T,X;p\right)\right)\cdot\left(Y - \left(m_{G=\infty,t=2}^{rc}\left(X\right) - m_{G=\infty,t=1}^{rc}\left(X\right)\right)\right)\right]$$

where

$$\begin{aligned} w_{G=2}^{rc}\left(D,T\right) &= w_{G=2,t=2}^{rc}\left(D,T\right) - w_{G=2,t=1}^{rc}\left(D,T\right), \\ w_{G=\infty}^{rc}\left(D,T,X;g\right) &= w_{G=\infty,t=2}^{rc}\left(D,T,X;g\right) - w_{G=\infty,t=1}^{rc}\left(D,T,X;g\right), \end{aligned}$$

and, for $s = 1, 2$, $g = 2, \infty$, we have that $m_{G=g,t=s}^{rc}\left(x\right) \equiv \mathbb{E}[Y|G=g, T=s, X=x]$,

$$w_{G=2,t=s}^{rc}\left(D,T\right) = \frac{D\cdot 1\left\{T=s\right\}}{\mathbb{E}\left[D\cdot 1\left\{T=s\right\}\right]},$$

$$w_{G=\infty,t=s}^{rc}\left(D,T,X;g\right) = \frac{g(X)\left(1-D\right)\cdot 1\left\{T=s\right\}}{1-g(X)} \Bigg/ \mathbb{E}\left[\frac{g(X)\left(1-D\right)\cdot 1\left\{T=s\right\}}{1-g(X)}\right].$$

Doubly robust DiD procedure with repeated cross-section

Sant'Anna and Zhao (2020) second DR DiD estimand also relies on outcome regression models for the treated unit:

$$ATT_2^{dr,rc} = ATT_1^{dr,rc}$$

$$+ \left(\mathbb{E}\left[m^{rc}_{G=2,t=2}(X) - m^{rc}_{G=\infty,t=2}(X)\middle| D=1\right] - \mathbb{E}\left[m^{rc}_{G=2,t=2}(X) - m^{rc}_{G=\infty,t=2}(X)\middle| D=1, T=2\right]\right)$$

$$- \left(\mathbb{E}\left[m^{rc}_{G=2,t=1}(X) - m^{rc}_{G=\infty,t=1}(X)\middle| D=1\right] - \mathbb{E}\left[m^{rc}_{G=2,t=1}(X) - m^{rc}_{G=\infty,t=1}(X)\middle| D=1, T=1\right]\right),$$

Doubly robust DiD procedure with repeated cross-section

- Both DR DiD estimators for RCS data are consistent for the ATT under the same conditions:

- Even if the regression model for the outcome evolution for the treated group is misspecified, $ATT_2^{dr,rc}$ is consistent for the ATT (provided that either the pscore or the regression models for outcome evolution among untreated units are correctly specified).

- However, in general, $ATT_2^{dr,rc}$ is more efficient than $ATT_1^{dr,rc}$.

- In fact, Sant'Anna and Zhao (2020) shown that $ATT_2^{dr,rc}$ is (locally) semiparametrically efficient.

Let’s see how these work in a simulation exercise

Monte Carlo simulations

Simulations

- Data generating processes are similar to those considered in the TFWE example
- We compare DR DiD estimators with IPW (standardized and non-standardized), outcome regression, and TWFE estimators
- Samples sizes $n = 1,000$
- For each design, we consider $10,000$ Monte Carlo experiments
- Available data are $\{Y_{t=2}, Y_{t=1}, D, Z\}_{i=1}^{n}$, where $D_i = 1\{G_i = 2\}$.
- We estimate the pscore assuming a logit specification, and the outcome regression models assuming a linear specification

DGPs

- Since we want to check the effect of model misspecifications, we will generate covariates slightly different than before.
- Let $Z_j = \left(\tilde{Z} - \mathbb{E}\left[\tilde{Z}\right]\right) / \sqrt{Var\left(\tilde{Z}\right)}, j = 1, 2, 3, 4$, where

$$\begin{aligned}
\tilde{Z}_1 &= \exp\left(\frac{X_1}{2}\right) \\
\tilde{Z}_2 &= \frac{X_2}{1 + \exp\left(X_1\right)} + 10 \\
\tilde{Z}_3 &= \left(\frac{X_1 X_3}{25} + 0.6\right)^3 \\
\tilde{Z}_4 &= \left(X_2 + X_4 + 20\right)^2
\end{aligned}$$

and $X_j \sim N\left(0, 1\right), j = 1, 2, 3, 4$.

DGPs

- For a generic $W = (W_1, W_2, W_3, W_4)$, let

$$
\begin{aligned}
f_{reg}(W) &= 210 + 27.4 \cdot W_1 + 13.7 \cdot (W_2 + W_3 + W_4) \\
f_{ps}(W) &= 0.75 \cdot (-W_1 + 0.5 \cdot W_2 - 0.25 \cdot W_3 - 0.1 \cdot W_4)
\end{aligned}
$$

- Also, let

$$
\begin{aligned}
&v(X, D) \overset{d}{\sim} N(D \cdot f_{reg}(X), 1) \\
&v(Z, D) \overset{d}{\sim} N(D \cdot f_{reg}(Z), 1) \\
&\varepsilon_{t=1} \overset{d}{\sim} N(0, 1) \\
&\varepsilon_{t=2}(2) \overset{d}{\sim} N(0, 1) \\
&\varepsilon_{t=2}(\infty) \overset{d}{\sim} N(0, 1) \\
&U \overset{d}{\sim} U(0, 1)
\end{aligned}
$$

DGPs

- We now consider four different DGPs
- *DGP1*:

$$
\begin{aligned}
Y_{i,t=1}(\infty) &= f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{i,t=1} \\
Y_{i,t=2}(\infty) &= 2 \cdot f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{i,t=2}(\infty) \\
Y_{i,t=2}(2) &= 2 \cdot f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{i,t=2}(\infty) \\
p(Z_i) &= \frac{\exp(f_{ps}(Z_i))}{1 + \exp(f_{ps}(Z_i))} \\
D_i &= 1\{p(Z_i) \geq U\}
\end{aligned}
$$

- Both the pscore and the OR models are correctly specified

DGPs

- *DGP2:*

$$\begin{aligned}
Y_{i,t=1}(\infty) &= f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{i,t=1} \\
Y_{i,t=2}(\infty) &= 2 \cdot f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{i,t=2}(\infty) \\
Y_{i,t=2}(2) &= 2 \cdot f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{i,t=2}(\infty) \\
p(X_i) &= \frac{\exp(f_{ps}(X_i))}{1 + \exp(f_{ps}(X_i))} \\
D_i &= 1\{p(X_i) \geq U\}
\end{aligned}$$

- Only the OR model is correctly specified

DGPs

- *DGP3:*

$$
\begin{aligned}
Y_{i,t=1}(\infty) &= f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{i,t=1} \\
Y_{i,t=2}(\infty) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{i,t=2}(\infty) \\
Y_{i,t=2}(2) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{i,t=2}(\infty) \\
p(Z_i) &= \frac{\exp(f_{ps}(Z_i))}{1 + \exp(f_{ps}(Z_i))} \\
D_i &= 1\{p(Z_i) \geq U\}
\end{aligned}
$$

- Only the pscore model is correctly specified

DGPs

- *DGP4*:

$$
\begin{aligned}
Y_{i,t=1}(\infty) &= f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{i,t=1} \\
Y_{i,t=2}(\infty) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{i,t=2}(\infty) \\
Y_{i,t=2}(2) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{i,t=2}(\infty) \\
p(X_i) &= \frac{\exp(f_{ps}(X_i))}{1 + \exp(f_{ps}(X_i))} \\
D_i &= 1\{p(X_i) \geq U\}
\end{aligned}
$$

- Both the pscore and the OR models are misspecified

Table 1: Monte Carlo Simulations, DGP1: Both pscore and OR are correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-20.9518	21.1227	2.5271	0.0000	9.9061
$\widehat{\tau}^{reg}$	-0.0012	0.1005	0.1010	0.9500	0.3960
$\widehat{\tau}^{ipw,p}$	0.0257	2.7743	2.6636	0.9518	10.4412
$\widehat{\tau}^{ipw,p}_{std}$	0.0075	1.1320	1.0992	0.9476	4.3090
$\widehat{\tau}^{dr,p}$	-0.0014	0.1059	0.1052	0.9473	0.4124
$\widehat{\tau}^{dr,p}_{imp}$	-0.0013	0.1057	0.1043	0.9451	0.4088

Figure 2: Monte Carlo for DID estimators, DGP1: Both pscore and OR are correctly specified

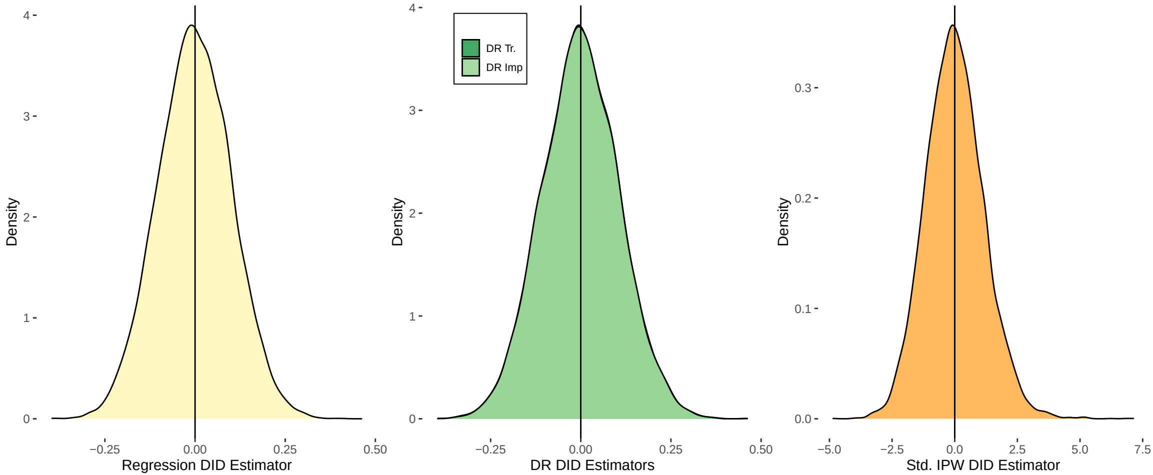
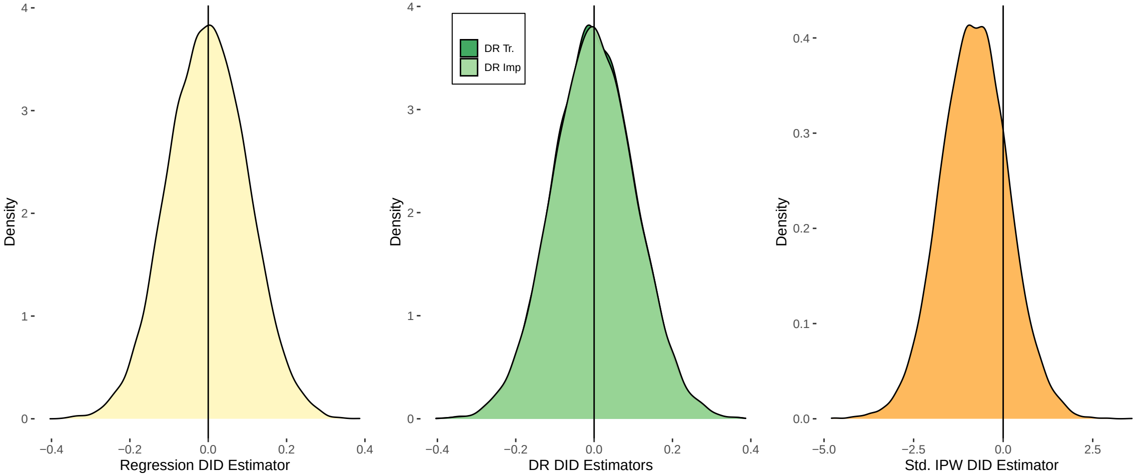


Table 2: Monte Carlo Simulations, DGP2: Only OR is correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-19.2859	19.4683	2.5754	0.0000	10.0955
$\widehat{\tau}^{reg}$	-0.0008	0.0997	0.1004	0.9492	0.3937
$\widehat{\tau}^{ipw,p}$	2.0100	3.2982	2.5049	0.8376	9.8193
$\widehat{\tau}^{ipw,p}_{std}$	-0.7942	1.2253	0.9241	0.8564	3.6226
$\widehat{\tau}^{dr,p}$	-0.0008	0.1036	0.1031	0.9469	0.4043
$\widehat{\tau}^{dr,p}_{imp}$	-0.0007	0.1042	0.1030	0.9445	0.4039

Figure 3: Monte Carlo for DID estimators, DGP2: Only OR is correctly specified



EMORY UNIVERSITY

Table 3: Monte Carlo Simulations, DGP3: Only PS is correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-13.1703	13.3638	3.5611	0.0035	13.9596
$\widehat{\tau}^{reg}$	-1.3843	1.8684	1.2286	0.8001	4.8159
$\widehat{\tau}^{ipw,p}$	0.0114	3.1982	3.0043	0.9468	11.7769
$\widehat{\tau}^{ipw,p}_{std}$	-0.0299	1.4270	1.3990	0.9447	5.4840
$\widehat{\tau}^{dr,p}$	-0.0513	1.2142	1.1768	0.9416	4.6132
$\widehat{\tau}^{dr,p}_{imp}$	-0.0709	1.0151	0.9842	0.9423	3.8581

Figure 4: Monte Carlo for DID estimators, DGP3: Only PS is correctly specified

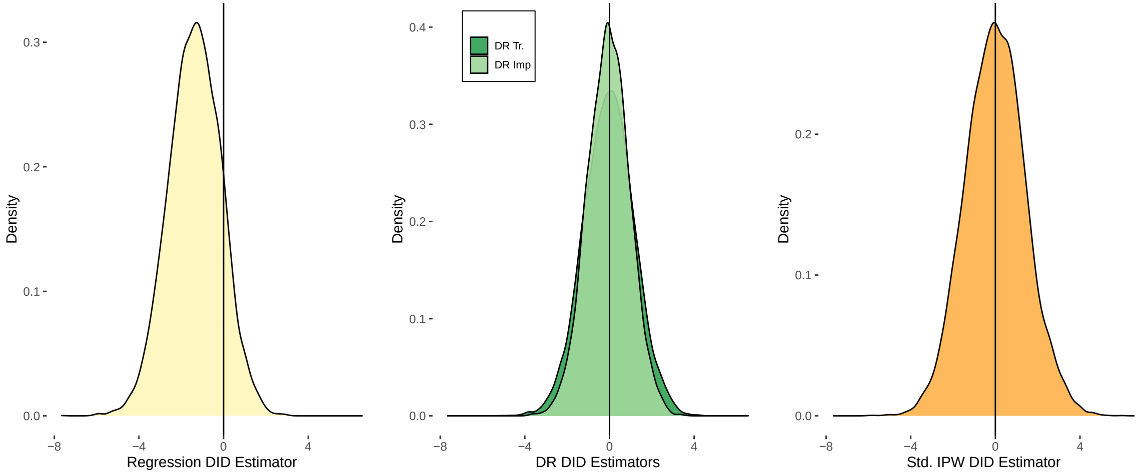
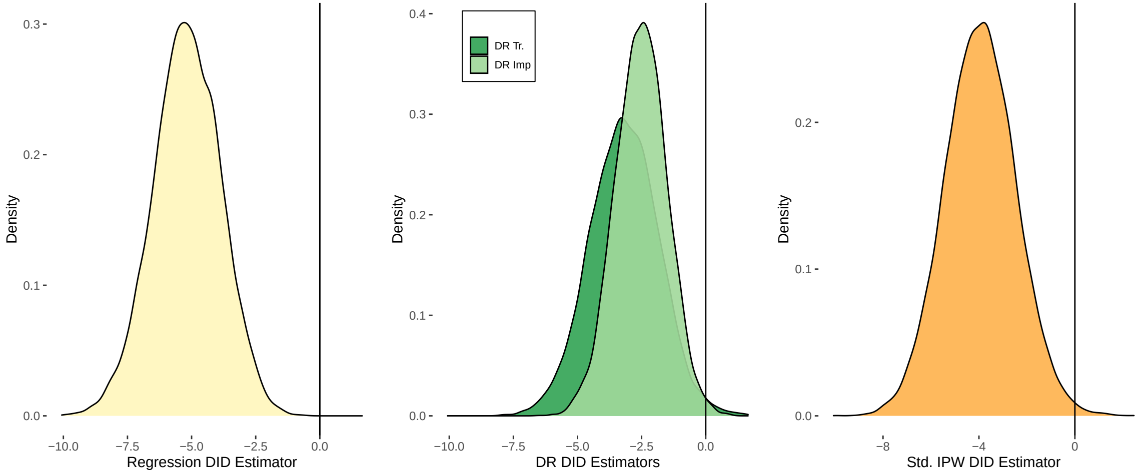


Table 4: Monte Carlo Simulations, DGP4: Both OR and PS are misspecified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-16.3846	16.5383	3.6268	0.0000	14.2169
$\widehat{\tau}^{reg}$	-5.2045	5.3641	1.2890	0.0145	5.0531
$\widehat{\tau}^{ipw,p}$	-1.0846	2.6557	2.3746	0.9487	9.3084
$\widehat{\tau}^{ipw,p}_{std}$	-3.9538	4.2154	1.4585	0.2282	5.7172
$\widehat{\tau}^{dr,p}$	-3.1878	3.4544	1.2946	0.3076	5.0749
$\widehat{\tau}^{dr,p}_{imp}$	-2.5291	2.7202	0.9837	0.2737	3.8561

Figure 5: Monte Carlo for DID estimators, DGP4: Both OR and PS are misspecified



Monte Carlo simulations for repeated cross-section data

- Same DGPs as before, but now, we observe a sample from $T = 2$ or $T = 1$ with probability 0.5.

Table 5: Monte Carlo Simulations, DGP1: Both the pscore and the OR are correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-20.7916	21.0985	3.5705	0.0002	13.9962
$\widehat{\tau}^{reg}$	0.0263	7.5878	7.5702	0.9510	29.6751
$\widehat{\tau}^{ipw,rc}$	-0.6619	55.9708	55.5516	0.9493	217.7621
$\widehat{\tau}^{ipw,rc}_{std}$	-0.0502	9.6477	9.5815	0.9487	37.5596
$\widehat{\tau}^{dr,rc}_{1}$	0.0129	3.0414	3.0340	0.9504	11.8934
$\widehat{\tau}^{dr,rc}_{2}$	0.0041	0.2159	0.2102	0.9441	0.8239
$\widehat{\tau}^{dr,rc}_{1,imp}$	0.0136	3.0413	3.0337	0.9507	11.8921
$\widehat{\tau}^{dr,rc}_{2,imp}$	0.0047	0.2163	0.2049	0.9371	0.8032

Figure 6: Monte Carlo for DID estimators, DGP1: Both the pscore and the OR are correctly specified

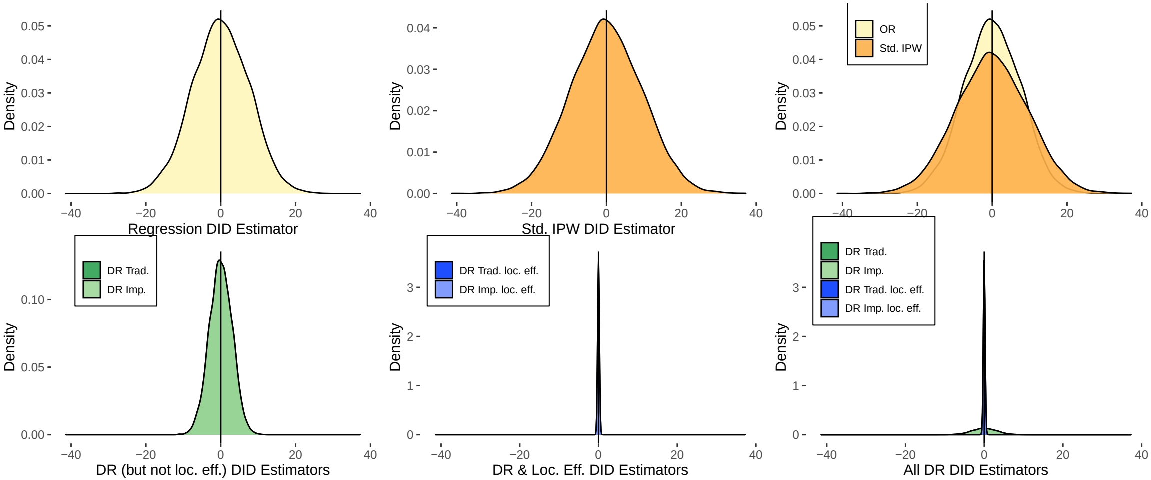
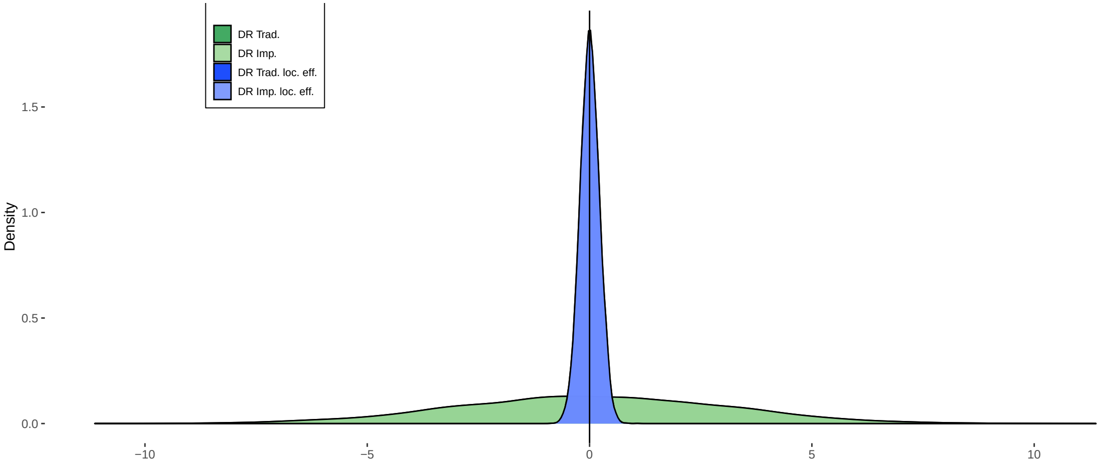


Figure 7: Monte Carlo for DID estimators, DGP1: Both the pscore and the OR are correctly specified



EMORY UNIVERSITY

Figure 8: Monte Carlo for DID estimators, DGP1: Both the pscore and the OR are correctly specified

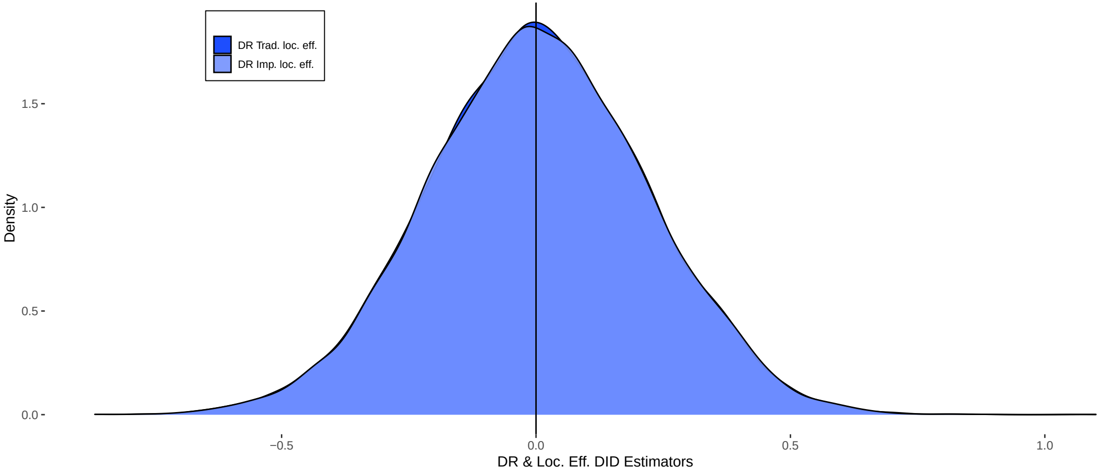
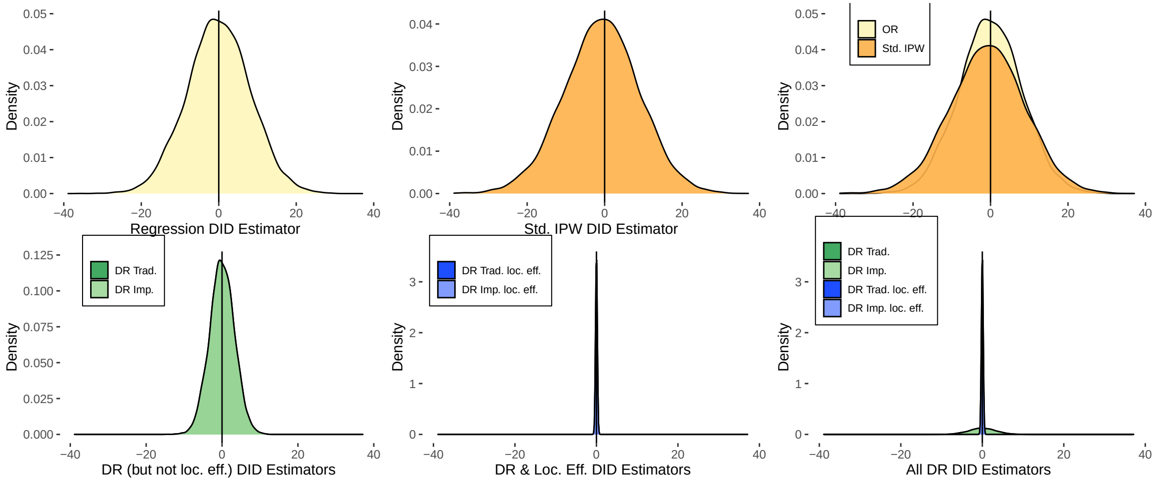


Table 6: Monte Carlo Simulations, DGP2: Only the OR is correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-19.1783	19.5289	3.6345	0.0005	14.2472
$\widehat{\tau}^{reg}$	-0.0244	8.1906	8.1493	0.9481	31.9454
$\widehat{\tau}^{ipw,rc}$	1.8203	55.0496	54.9614	0.9491	215.4486
$\widehat{\tau}^{ipw,rc}_{std}$	-0.8119	9.8141	9.7018	0.9459	38.0310
$\widehat{\tau}^{dr,rc}_{1}$	-0.0102	3.2814	3.2651	0.9486	12.7991
$\widehat{\tau}^{dr,rc}_{2}$	-0.0002	0.2108	0.2054	0.9454	0.8051
$\widehat{\tau}^{dr,rc}_{1,imp}$	-0.0095	3.2818	3.2650	0.9488	12.7989
$\widehat{\tau}^{dr,rc}_{2,imp}$	0.0002	0.2127	0.2030	0.9403	0.7958

Figure 9: Monte Carlo for DID estimators, DGP2: Only the OR is correctly specified



EMORY UNIVERSITY

Figure 10: Monte Carlo for DID estimators, DGP2: Only the OR is correctly specified

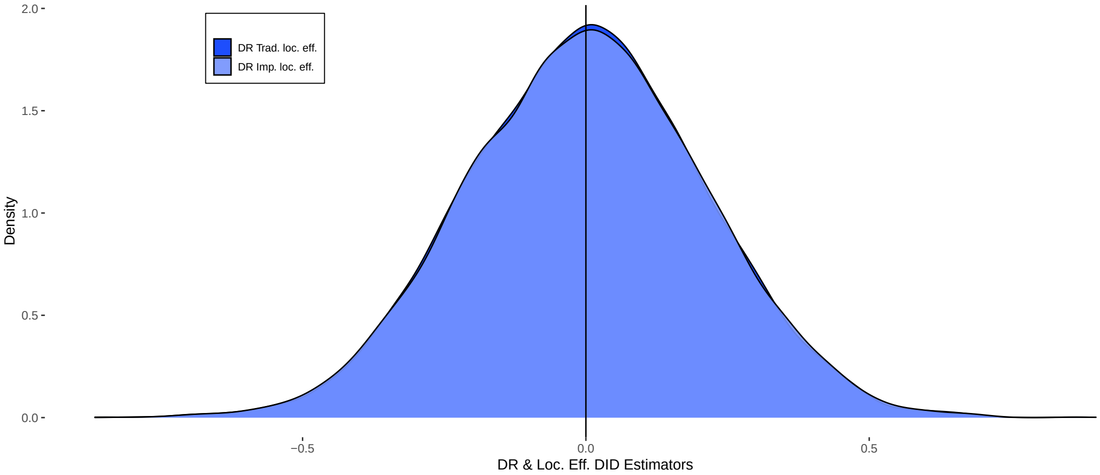


Table 7: Monte Carlo Simulations, DGP3: Only the PS is correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-13.1310	14.0577	5.0424	0.2598	19.7664
$\widehat{\tau}^{reg}$	-1.3763	8.1367	8.0046	0.9421	31.3782
$\widehat{\tau}^{ipw,rc}$	-0.9734	57.2618	56.9005	0.9465	223.0501
$\widehat{\tau}^{ipw,rc}_{std}$	0.0508	9.4283	9.3068	0.9431	36.4826
$\widehat{\tau}^{dr,rc}_{1}$	-0.0855	5.6917	5.6276	0.9453	22.0602
$\widehat{\tau}^{dr,rc}_{2}$	-0.0289	4.7419	4.6585	0.9416	18.2613
$\widehat{\tau}^{dr,rc}_{1,imp}$	-0.1191	4.8371	4.7970	0.9450	18.8042
$\widehat{\tau}^{dr,rc}_{2,imp}$	-0.0762	4.0623	3.9669	0.9436	15.5503

Figure 11: Monte Carlo for DID estimators, DGP3: Only the PS is correctly specified

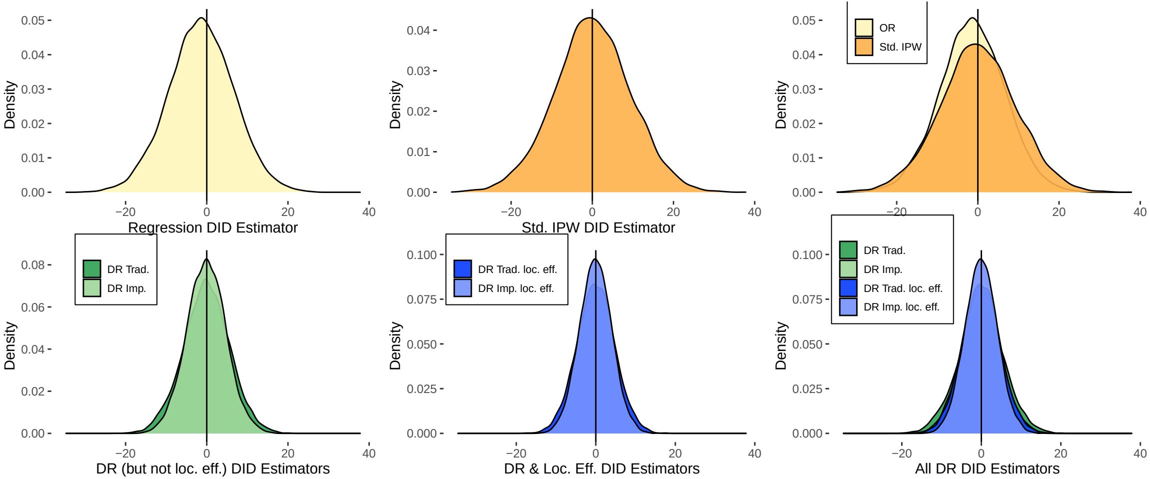


Figure 12: Monte Carlo for DID estimators, DGP3: Only the PS is correctly specified

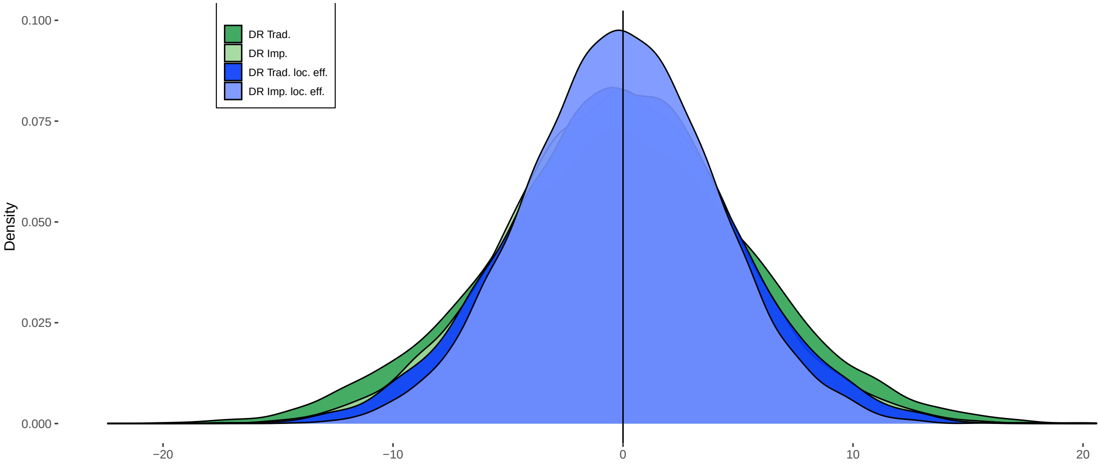


Figure 13: Monte Carlo for DID estimators, DGP3: Only the PS is correctly specified

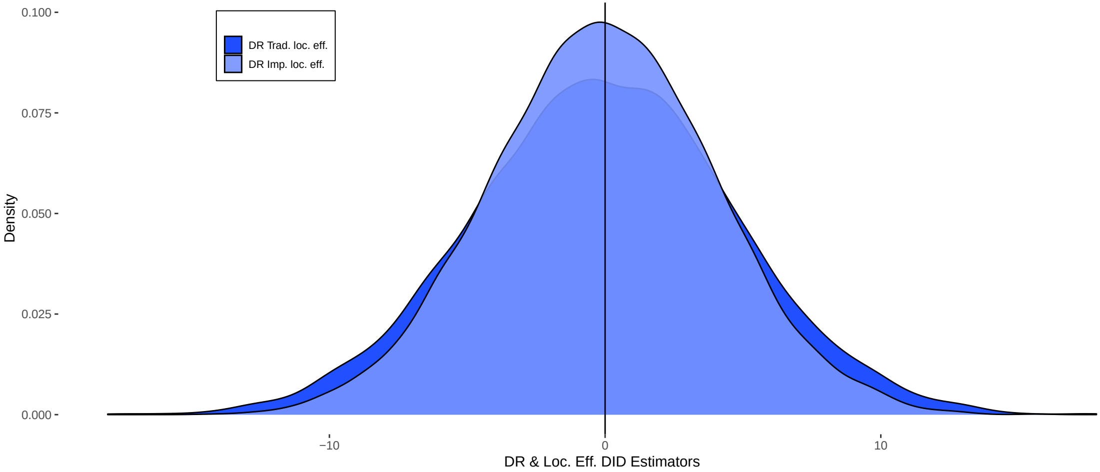


Table 8: Monte Carlo Simulations, DGP4: Both the OR and the PS are misspecified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-16.3305	17.1263	5.1307	0.1138	20.1123
$\widehat{\tau}^{reg}$	-5.3378	9.9773	8.5196	0.9075	33.3969
$\widehat{\tau}^{ipw,rc}$	-1.3912	55.1777	55.6717	0.9518	218.2330
$\widehat{\tau}^{ipw,rc}_{std}$	-4.1487	10.5195	9.6864	0.9304	37.9707
$\widehat{\tau}^{dr,rc}_{1}$	-3.3422	7.0709	6.1963	0.9157	24.2897
$\widehat{\tau}^{dr,rc}_{2}$	-3.2751	6.0158	4.8876	0.8863	19.1593
$\widehat{\tau}^{dr,rc}_{1,imp}$	-2.6888	5.5642	4.8416	0.9134	18.9790
$\widehat{\tau}^{dr,rc}_{2,imp}$	-2.6138	4.8453	3.9673	0.8923	15.5519

Figure 14: Monte Carlo for DID estimators, DGP4: Both the OR and the PS are misspecified

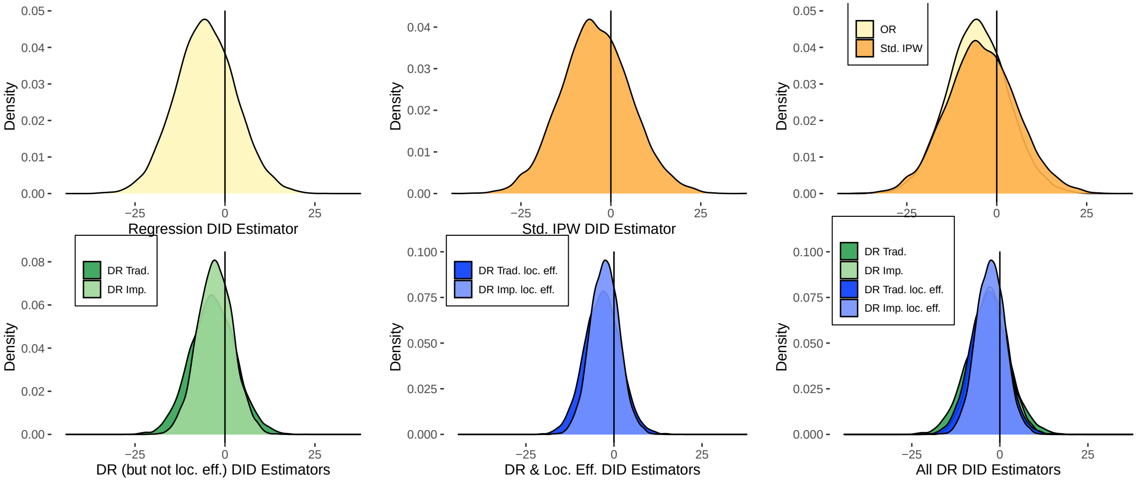


Figure 15: Monte Carlo for DID estimators, DGP4: Both the OR and the PS are misspecified

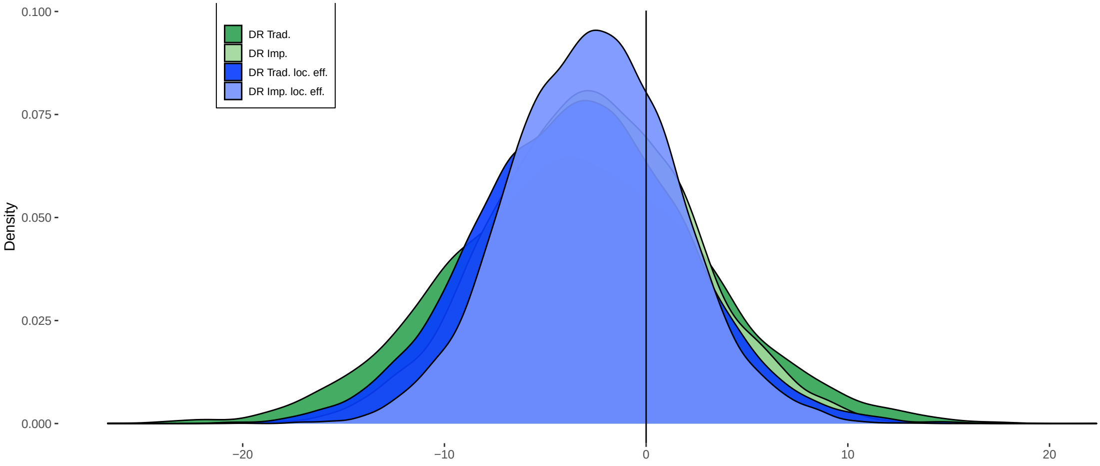
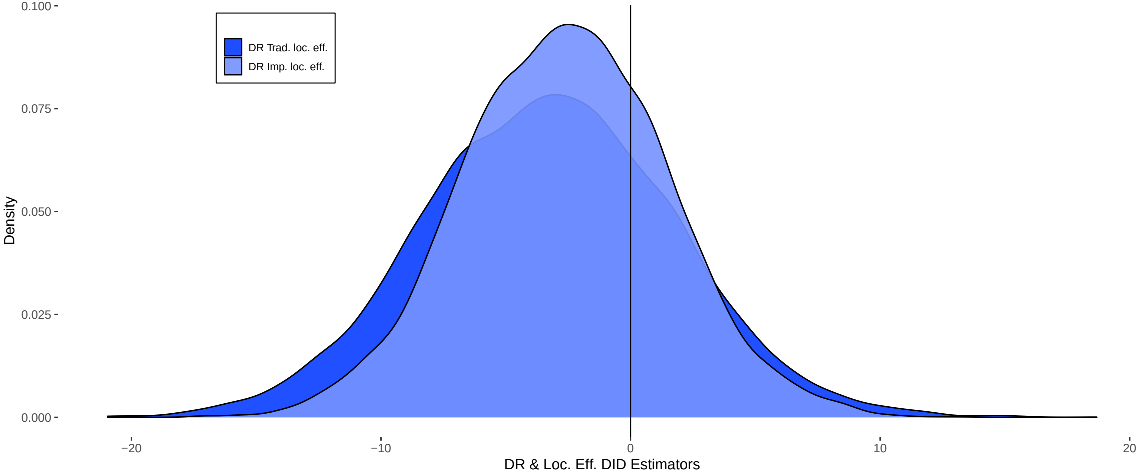


Figure 16: Monte Carlo for DID estimators, DGP4: Both the OR and the PS are misspecified



What are the main take-away messages?

Take-way messages

DiD procedures with covariates

- We can include covariates into DiD to allow for covariate-specific trends
- Covariates should not be post-treatment variables
- There are several "correct" ways of implementing conditional DiD:
  - Regression adjustments
  - Inverse probability weighting
  - Doubly Robust (augmented inverse probability weighting)
- TWFE, though, can be severely biased.
- DR DiD is my preferred method:
  - More robust against model misspecifications
  - Can be semiparametrically efficient (confidence intervals are tighter)

Empirical application

Empirical application

- Let's switch to R/Stata so we can see how to do all these things!

References

Abadie, Alberto, "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 2005, *72* (1), 1–19.

Hájek, J., "Discussion of 'An essay on the logical foundations of survey sampling, Part I', by D. Basu," in V. P. Godambe and D. A. Sprott, eds., *Foundations of Statistical Inference*, Toronto: Holt, Rinehart, and Winston, 1971.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd, "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 1998, *66* (5), 1017–1098.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd, "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, October 1997, *64* (4), 605–654.

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

Kang, Joseph D. Y. and Joseph L. Schafer, "Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data.," *Statistical Science*, 2007, *22* (4), 569–573.

Khan, Shakeeb and Elie Tamer, "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 2010, *78* (6), 2021–2042.

Sant'Anna, Pedro H. C. and Jun Zhao, "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, November 2020, *219* (1), 101–122.