

# Difference-in-Differences

MIXTAPE SESSION

---



# Roadmap

Basic DiD

Simple case, no covariates

IPW

DRDiD

# Introducing DiD

- DiD in its modern form dates back to Orley Ashenfelter and David Card at Princeton in the late 70s and mid 80s (Card says they invent the term in mid 1980s)
- Unclear when key identifying assumptions like parallel trends are worked out (as originally there is no potential outcomes in play)
- Attractive elements included natural experimentation, panel data, and did not require randomization
- US context may also have been such that it had a lot of upside
- But let's start at the beginning with story and example

# Miasma Debates

- Health policy debates in the 19th century is where we DiD pop up, mainly as a matter of logical deduction
- Dominant disease theory in 19th century was *miasma* – disease caused by smelly vapor
- Helped fuel a lot of sanitation reforms, but inadequate for explaining many diseases
- “All models are wrong but some are useful” – George Box)
- Keep in mind – microorganisms would not be identified until much later, partly caused by poor resolution in microscopes (Freedman 2007)

# Miasma I: Ignaz Semmelweis

- 1840s, Vienna maternity wards had high postpartum infections in one wing compared to other wings
- One division had doctors and trainee doctors, but another had midwives and trainee midwives
- Ignaz Semmelweis notes the difference in 1841 when hospitals moved to “anatomical” training involving cadavers (Jakeila lecture 4, DiD)
- New training happens to one but not the other and Semmelweis thinks the mortality is caused by working with cadavers
- Proposes washing hands with chlorine in 1847 in the midwives’ wing and uses a DiD design of pre and post

## Miasma II: John Snow and cholera

- More well known involves John Snow, an important inventor of a device that delivered chloroform to patients for surgery (major advancement in anesthesia)
- He was also an amateur epidemiologist in 19th century, and these days it's more common to hear him be named the father of DiD
- Believed cholera was spread through the Thames water supply which contradicted dominant theory about "dirty air" transmission
- Grand experiment: Lambeth moves its pipe between 1849 and 1854; Southwark and Vauxhall delay
- How can he use this event to test his hypothesis? Three ways: simple comparisons, interrupted time series of the difference in differences (DiD)

# Simple cross-sectional design

*Table:* Lambeth and Southwark and Vauxhall, 1854

<b>Company</b>	<b>Cholera mortality</b>
Lambeth	$Y = L + D$
Southwark and Vauxhall	$Y = SV$

$$\widehat{\delta}_{cs} = D + (L - SV)$$

This is selection bias with respect to  $Y^0$ .

# Interrupted time series design

Table: Lambeth, 1849 and 1854

<b>Company</b>	<b>Time</b>	<b>Cholera mortality</b>
Lambeth	1849	$Y = L$
	1854	$Y = L + (T + D)$

$$\hat{\delta}_{its} = D + T$$

This is selection bias with respect to time,  $T \neq 0$

# Difference-in-differences

Table: Lambeth and Southwark and Vauxhall, 1849 and 1854

<b>Companies</b>	<b>Time</b>	<b>Outcome</b>	$D_1$	$D_2$
Lambeth	Before	$Y = L$	$T_L + D$	$D$
	After	$Y = L + T_L + D$		
Southwark and Vauxhall	Before	$Y = SV$	$T_{SV}$	$D$
	After	$Y = SV + T_{SV}$		

$$\hat{\delta}_{did} = D + (T_L - T_{SV})$$

This is selection bias with respect to  $Y^0$  trends

## DiD equation, 2x2 and parallel trends

- I tend to group DiD based on whether the design will be based on two things:
  1. **Parallel trends:**  $T_L = T_{SV}$  for identification
  2. **DiD equation:** (After minus before for treated) - (After minus before for comparison group)
- Goodman-Bacon (2018; 2021) calls the DiD equation the “2x2” and I will too sometimes, but they mean the same thing
- Both tend to show up repeatedly in each of the methods we examine in this workshop
- It’s how I differentiate DiD from others like the broadly defined panel estimators

## DiD equation with sample averages

$$\widehat{\delta}_{kU}^{2x2} = \left( \bar{y}_k^{post(k)} - \bar{y}_k^{pre(k)} \right) - \left( \bar{y}_U^{post(k)} - \bar{y}_U^{pre(k)} \right)$$

## DiD equation with population expectations

$$\hat{\delta}_{kU}^{2x2} = \left( E[Y_k|Post] - E[Y_k|Pre] \right) - \left( E[Y_U|Post] - E[Y_U|Pre] \right)$$

# Potential outcomes and the switching equation

$$\widehat{\delta}_{kU}^{2x2} = \underbrace{\left( E[Y_k^1|Post] - E[Y_k^0|Pre] \right) - \left( E[Y_U^0|Post] - E[Y_U^0|Pre] \right)}_{\text{Switching equation}} + \underbrace{E[Y_k^0|Post] - E[Y_k^0|Post]}_{\text{Adding zero}}$$

## Parallel trends bias

$$\begin{aligned}\hat{\delta}_{kU}^{2x2} &= \underbrace{E[Y_k^1|Post] - E[Y_k^0|Post]}_{\text{ATT}} \\ &\quad + \underbrace{\left[ E[Y_k^0|Post] - E[Y_k^0|Pre] \right] - \left[ E[Y_U^0|Post] - E[Y_U^0|Pre] \right]}_{\text{Non-parallel trends bias in 2x2 case}}\end{aligned}$$

Note that this is what we saw in the table earlier –  $T_L \neq T_{SV}$  and it's not measurable because  $T_L$  is confounded by the ATT itself

## Similar role $Y^0$ in selection bias

Cross-sectional comparisons and selection bias with respect to  $Y^0$

$$\begin{aligned} E[Y|D=1] - E[Y|D=0] &= ATE + E[Y^0|D=1] - E[Y^0|D=0] \\ &= +(1-\pi)(ATT - ATU) \end{aligned}$$

DiD comparisons and selection bias with respect to  $Y^0$

$$\begin{aligned} \widehat{\delta}_{kU}^{2x2} &= \underbrace{E[Y_k^1|Post] - E[Y_k^0|Post]}_{ATT} \\ &\quad + \underbrace{\left[ E[Y_k^0|Post] - E[Y_k^0|Pre] \right] - \left[ E[Y_U^0|Post] - E[Y_U^0|Pre] \right]}_{\text{Non-parallel trends bias in 2x2 case}} \end{aligned}$$

# Independence

## Independence assumption

Treatment is assigned to a population independent of that population's potential outcomes

$$(Y^0, Y^1) \perp\!\!\!\perp D$$

This is random or quasi-random assignment and ensures mean potential outcomes for the treatment group and control group are the same. Also ensures other variables are distributed the same for a large sample.

$$E[Y^0|D = 1] = E[Y^0|D = 0]$$

$$E[Y^1|D = 1] = E[Y^1|D = 0]$$

## Non-randomization identification

- But we won't use randomization for identification with DiD
- This is one of the appeals of DiD – we aren't using physical randomization
- But note, independence works and gives confidence "because we know how the science works" – Don Rubin
- There is no such "science" with parallel trends, so we tend to focus on a variety of ad hoc tests like event studies for confidence
- But otherwise, it's actually quite similar, and shows that identification does not always depend on randomization

# OLS Specification

- Simple difference in means can be computed with OLS, too, but not as easy to explicitly see the parallel trends so I'll show you graphically
- Properly specified OLS model will also identify the ATT when there is only two groups and no covariates
- Often preferred because
  - OLS estimates the ATT under parallel trends
  - Easy to calculate the standard errors
  - Easy to include multiple periods
- But some issues emerge with differential timing, time varying covariates and continuous treatments

# Minimum wages

- Card and Krueger (1994) have a famous study estimating causal effect (ATT) of minimum wages on employment
- Exploited a policy change in New Jersey between February and November in mid-1990s where minimum wage was increased
- Neighbor PA did not (remember no spillover assumption)
- Found no effect on minimum wage; if anything it was the opposite effect
- Caused some feuding with Buchanan saying people who believed it enough to adopt it were “camp following whores” (Wall Street Journal op-ed during Clinton admin)

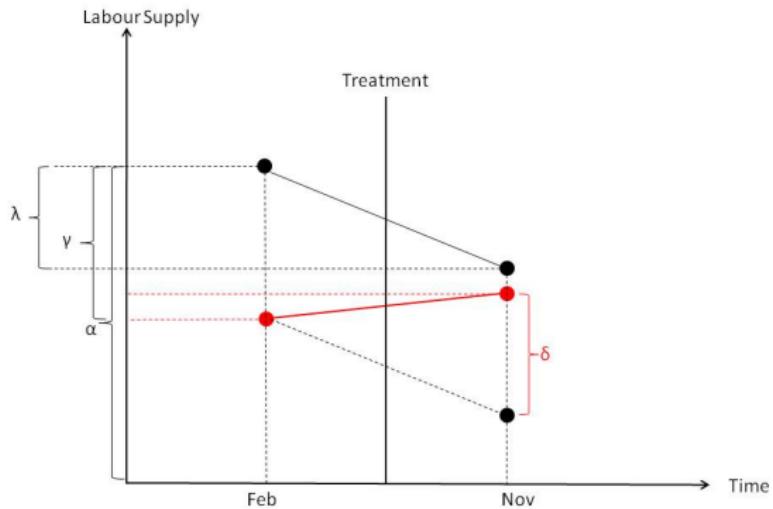
# OLS specification of the DiD equation

- The equivalent regression includes time and group fixed effects:

$$Y_{its} = \alpha + \gamma NJ_s + \lambda d_t + \delta (NJ \times d)_{st} + \varepsilon_{its}$$

- NJ is a dummy equal to 1 if the observation is from NJ
- d is a dummy equal to 1 if the observation is from November (the post period)
- This equation takes the following values
  - PA Pre:  $\alpha$
  - PA Post:  $\alpha + \lambda$
  - NJ Pre:  $\alpha + \gamma$
  - NJ Post:  $\alpha + \gamma + \lambda + \delta$
- DiD equation:  $(NJ \text{ Post} - NJ \text{ Pre}) - (PA \text{ Post} - PA \text{ Pre}) = \delta$

$$Y_{ist} = \alpha + \gamma N J_s + \lambda d_t + \delta (N J \times d)_{st} + \varepsilon_{ist}$$



## OLS with two way fixed effects

Under parallel trends, OLS estimates the ATT. Researchers often will use OLS with time-varying covariates, but this is not advised as it is only unbiased under more restrictive assumptions which we discuss next (though see new working paper by Cattaneo, et al. 2022 on controlling for time-varying covariates).

## Inverse probability weighting DiD

- Abadie (2005) proposed a DiD estimator that could incorporate covariates and get an unbiased estimate of the ATT
- Researcher needs treatment and comparison group observed before and after treatment
- If treatment group units are selected based on their (observed) covariates, then baseline covariates are also needed

## Time varying versus time invariant covariates

- In a DiD, we may need to control for X because treatment is only conditional on X
- But in TWFE, all time invariant covariates are absorbed by the unit fixed effects – only time varying covariates will survive TWFE
- But time varying covariates place restrictions, as we will see, on the DGP and run the threat of conditioning on outcomes if they were changed by the treatment
- Abadie proposes using only the covariates at baseline to form weights in the simple DiD formula

## Three step method

1. Compute each unit's "after minus before" which is the DD part
2. Then estimate a propensity score which you'll use to weight each unit
3. Finally, compare weighted changes in "after minus before" for treatment versus comparison groups

You can have heterogeneous treatment effects, but not differential timing

# Terms

- $t$  is year of treatment which doesn't vary across units (so no differential timing)
- $Y^1$  and  $Y^0$  are potential outcomes (counterfactual versus actual)
- $D$  is 1 or 0 based on group and time
- $b$  is the “baseline” which is similar to CS using  $g$  as the one year pre-treatment
- $X$  are “baseline” covariates **only** – they do not vary over time, which means propensity scores are estimated off the  $b$  period **only**

# Assumptions

Kind of common for this propensity score literature to only have two assumptions. But usually the first conditional independence. Now it is parallel trends because this is DD

1. Conditional parallel trends

$$E[Y_t^0 - Y_b^0 | D = 1, X_b] - E[Y_t^0 - Y_t^0 | D = 0, X_b]$$

(Notice the  $b$  subscript. What is that you think?)

2. Common support

$$\Pr(D = 1) > 0; \Pr(D = 1 | X) < 1$$

Let's see a picture of common support that I drew. Apologies it's horrible

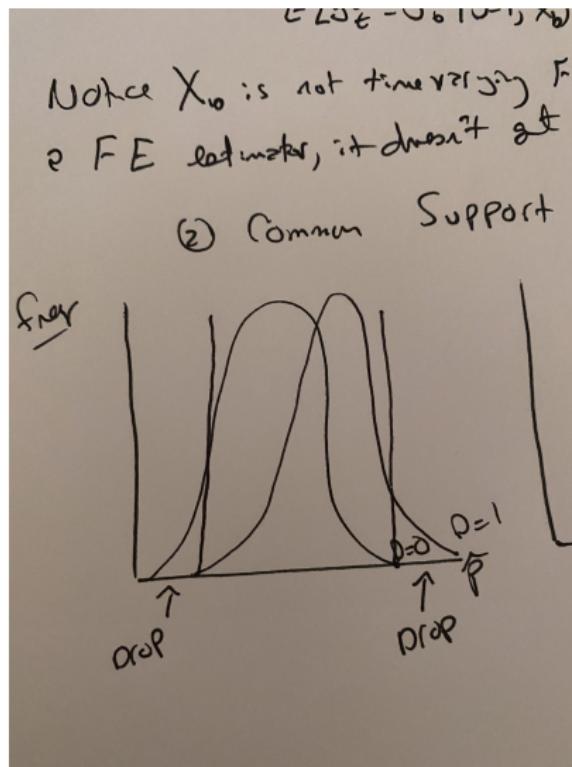
## Common support

As we are identifying the ATT, we only need common support with respect to treated units

Your identify assumptions are always with respect to the missing covariates in other words and for the ATT, you are missing  $Y^0$  for the treatment group

If we were estimating ATU, we'd be missing  $Y^1$  for controls and need common support ( $Y$  in treatment for all ranges of control), and for ATE we'd need both

# Visualizing propensity score to get common support



# Definition and estimation

Defining the ATT parameter of interest

$$ATT = E[Y_t^1 - Y_t^0 | D_t = 1] \quad (1)$$

Abadie's estimator

$$E \left[ \frac{Y_t - Y_b}{Pr(D_t = 1)} \times \frac{D_t - Pr(D = 1|X_b)}{1 - Pr(D = 1|X_b)} \right] \quad (2)$$

# Propensity scores

- Paper is titled “Semi-parametric DiD” because Abadie imposes structure on the polynomials used to construct the propensity score
- You can use OLS linear probability models or series logit estimation

## Estimating propensity scores

It's common to hear people say that we don't know the propensity score; we can only estimate it. Same here – we approximate it with regressions

$$\widehat{Pr}(X_b) = \widehat{\gamma}_0 + \widehat{\gamma}_1 X + \widehat{\gamma}_2 X^2 + \dots \varepsilon \quad (3)$$

$$\widehat{Pr}(X_b) = F(\widehat{\gamma}_0 + \widehat{\gamma}_1 X + \widehat{\gamma}_2 X^2 + \dots) \quad (4)$$

## Commentary on paper's influence

Abadie has a great year in 2003: synthetic control in AER, semiparametric DiD in Restud, semiparametric IV (his JMP) in JoE

Semiparametric DiD has over 2,000 cites

But I'm not sure it was widely adopted by the *applied* community. It seems like the *applied community* starts paying attention to the econometric contributions much later, and Abadie (2003) is gaining a renaissance at the moment because of the next paper

Nonetheless, there is code in Stata called -absdid- but an unsatisfying part of the semiparametric piece is that the results change with respect to not just the estimation method (OLS or logit), but also with respect to the order in which the covariates appear

# Doubly Robust Difference-in-differences

- DR models control for covariates twice – once using the propensity score, once using outcomes adjusted by regression – and are unbiased so long as:
  - The regression specification for the outcome is correctly specified
  - The propensity score specification is correctly specified
- Sant'Anna and Zhao (2020) incorporated DR into DiD by combining inverse probability weighting and outcome regression into a single DiD model
- It's in the engine of Callaway and Sant'Anna (2020) that we discuss later so it merits close study
- One of my favorite lesser known of the new DiD papers

# Patterns in econometrician reasoning

1. Define the target parameter first (as opposed to writing down model first)
2. Identification (e.g., parallel trends)
3. Estimation
4. Aggregation
5. Inference

## Defining the target parameter

Major part of the new econometrics is to always start with the target parameter and build to it using estimation and identification that “works”

$$\delta = E[Y_{it}^1 - Y_{it}^0 | D_i = 1]$$

## Identification assumptions I: Data

Assumption 1: Assume panel data or repeated cross-sectional data

Handling repeated cross-sectional data is possible but assumes modularity which is a kind of stability assumption, but I'll use panel representation.

Cross-sections will be potentially violated with changing sample compositions (see my Mixtape on the Napster example).

## Identification assumptions II: Modification to parallel trends

Assumption 2: Conditional parallel trends

Counterfactual trends for the treatment group are the same as the control group for all values of  $X$

$$E[Y_1^0 - Y_0^0 | X, D = 1] = E[Y_1^0 - Y_0^0 | X, D = 0]$$

## Identification assumptions III: Common support

### Assumption 3: Common support

For some  $e > 0$ , the probability of being in the treatment group is greater than  $e$  and the probability of being in the treatment group conditional on  $X$  is  $\leq 1 - e$ .

Intuition of assumption 3: Called overlap or common support. Means there is at least a small fraction of the population that is treated and that for every value of the covariates  $X$  there is at least a small chance that the unit is not treated. It's called common support when it's a propensity score but it's just about the distribution of treatment and control across values of  $X$ . Very common when dealing with covariate comparisons as otherwise you're extrapolating (curse of dimensionality)

## Estimating DD with Assumptions 1-3

- Assumptions 1-3 gives us a couple of options of estimating the DiD
- We can either use the outcome regression (OR) approach of Heckman, et al 1997
- Or we can use the inverse probability weighting (IPW) approach of Abadie (2005)

## Outcome regression

This is the Heckman, et al. (1997) approach where the outcome evolution is modeled with a regression

$$\widehat{\delta}^{OR} = \overline{Y}_{1,1} - \left[ \overline{Y}_{1,0} + \frac{1}{n^T} \sum_{i|D_i=1} (\widehat{\mu}_{0,1}(X_i) - \widehat{\mu}_{0,0}(X_i)) \right]$$

where  $\overline{Y}$  is the sample average of  $Y$  among units in the treatment group at time  $t$  and  $\widehat{\mu}(X)$  is an estimator of the true, but unknown,  $m_{d,t}(X)$  which is by definition equal to  $E[Y_t|D = d, X = x]$ .

# Outcome regression

$$\hat{\delta}^{OR} = \bar{Y}_{1,1} - \left[ \bar{Y}_{1,0} + \frac{1}{n^T} \sum_{i|D_i=1} (\hat{\mu}_{0,1}(X_i) - \hat{\mu}_{0,0}(X_i)) \right]$$

1. Regress changes  $\Delta Y$  on  $X$  among untreated groups using baseline covariates only
2. Get fitted values of the regression using all  $X$  from  $D = 1$  only.  
Average those
3. Calculate change in this fitted  $Y$  among treated with the average fitted values

## Inverse probability weighting

This is the Abadie (2005) approach where we use weighting

$$\hat{\delta}^{ipw} = \frac{1}{E_N[D]} E \left[ \frac{D - \hat{p}(X)}{1 - \hat{p}(X)} (Y_1 - Y_0) \right]$$

where  $\hat{p}(X)$  is an estimator for the true propensity score. Reduces the dimensionality of  $X$  into a single scalar.

## These models cannot be ranked

- Outcome regression needs  $\hat{\mu}(X)$  to be correctly specified, whereas
- Inverse probability weighting needs  $\hat{p}(X)$  to be correctly specified
- It's hard to "rank" these two in practice with regards to model misspecification because each is inconsistent when their own models are misspecified
- Well why don't we just use TWFE? I've never heard anyone complain about including covariates in TWFE and I've been doing it my entire adult life, so we're good right?
- Depends on if you want to assume three more things.

## TWFE

Here's the TWFE specification:

$$Y_{it} = \alpha_1 + \alpha_2 T_t + \alpha_3 D_i + \delta(T_i \times D_t) + \varepsilon_{it}$$

Just add in covariates then right?

$$Y_{it} = \alpha_1 + \alpha_2 T_t + \alpha_3 D_i + \delta(T_i \times D_t) + \theta \cdot X_{it} + \varepsilon_{it}$$

Sure! If you're willing to impose three *more* assumptions

# Decomposing TWFE with covariates

TWFE places restrictions on the DGP. Previous TWFE regression under assumptions 1-3 implies the following:

$$E[Y_1^1 | D = 1, X] = \alpha_1 + \alpha_2 + \alpha_3 + \delta + \theta X$$

Conditional parallel trends implies

$$E[Y_1^0 - Y_0^0 | D = 1, X] = E[Y_1^0 - Y_0^0 | D = 0, X]$$

$$E[Y_1^0 | D = 1, X] - E[Y_0^0 | D = 1, X] = E[Y_1^0 | D = 0, X] - E[Y_0^0 | D = 0, X]$$

$$E[Y_1^0 | D = 1, X] = E[Y_0^0 | D = 1, X] + E[Y_1^0 | D = 0, X] - E[Y_0^0 | D = 0, X]$$

$$E[Y_1^0 | D = 1, X] = E[Y_0 | D = 1, X] + E[Y_1 | D = 0, X] - E[Y_0 | D = 0, X]$$

## Switching equation substitution

Last line from the switching equation. This gives us:

$$E[Y_1^0 | D = 1, X] = \alpha_1 + \alpha_2 + \alpha_3 + \theta X$$

Now compare this with our earlier  $Y^1$  expression

$$E[Y_1^1 | D = 1, X] = \alpha_1 + \alpha_2 + \alpha_3 + \delta + \theta X$$

We can define our target parameter, the ATT, now in terms of the fixed effects representation

## Collecting terms

TWFE representation of our conditional expectations of the potential outcomes

$$E[Y_1^1|D = 1, X] = \alpha_1 + \alpha_2 + \alpha_3 + \delta + \theta_1 X$$

$$E[Y_1^0|D = 1, X] = \alpha_1 + \alpha_2 + \alpha_3 + \theta_2 X$$

Substitute these into our target parameter

$$\begin{aligned} ATT &= E[Y_1^1|D = 1, X] - E[Y_1^0|D = 1, X] \\ &= (\alpha_1 + \alpha_2 + \alpha_3 + \delta + \theta_1 X) - (\alpha_1 + \alpha_2 + \alpha_3 + \theta_2 X) \\ &= \delta + (\theta_1 X - \theta_2 X) \end{aligned}$$

What if  $\theta_1 X \neq \theta_2 X$ ?

## Assumption 4: Homogeneous treatment effects in $X$

TWFE requires homogenous treatment effects in  $X$  (i.e., the treatment effect is the same for all  $X$ )

If  $X$  is sex, then effects are the same for males and females.

If  $X$  is continuous, like income, then the effect is the same whether someone makes \$1 or \$1 million.

## X-specific trends

TWFE also places restrictions on covariate trends for the two groups too. Take conditional expectations of our TWFE equation.

$$E[Y_1|D = 1] = \alpha_1 + \alpha_2 + \alpha_3 + \delta + \theta X_{11}$$

$$E[Y_0|D = 1] = \alpha_1 + \alpha_3 + \theta X_{10}$$

$$E[Y_1|D = 0] = \alpha_1 + \alpha_2 + \theta X_{01}$$

$$E[Y_0|D = 0] = \alpha_1 + \theta X_{00}$$

## X-specific trends

Now take the DiD formula:

$$\delta^{DD} = \left( (\alpha_1 + \alpha_2 + \alpha_3 + \delta + \theta X_{11}) - (\alpha_1 + \alpha_3 + \theta X_{10}) \right) - \left( (\alpha_1 + \alpha_2 + \theta X_{01}) - (\alpha_1 + \theta X_{00}) \right)$$

Eliminating terms, we get:

$$\delta^{DD} = \delta + (\theta X_{11} - \theta X_{10}) - (\theta X_{01} - \theta X_{00})$$

Second line requires that trends in X for treatment group equal trends in X for control group.

## Assumption 5 and 6

We need “no  $X$ -specific trends” for the treatment group (assumption 5) and comparison group (assumption 6)

**Intuition:** No  $X$ -specific trends means the evolution of potential outcome  $Y^0$  is the same regardless of  $X$ . This would mean you cannot allow rich people to be on a different trend than poor people, for instance.

Without these six, in general TWFE will not identify ATT.

## Why not both?

- Let's review the problem. What if you claim you need  $X$  for conditional parallel trends?
- You have three options:
  1. Outcome regression (Heckman, et al. 1997) – needs Assumptions 1-3
  2. Inverse probability weighting (Abadie 2005) – needs Assumptions 1-3
  3. TWFE (everybody everywhere all the time) – needs Assumptions 1-6
- Problem is 1 and 2 need the models to be correctly specified
- Doubly robust combines them to give us insurance; we now get two chances to be wrong, as opposed to just one
- I'm going to only stick to the panel data expressions bc all repeated cross-section does add in some terms (and I've not written up semiparametric bounds yet)

## Notation

$p(x)$  : propensity score model

$$\Delta Y = Y_1 - Y_0 = Y_{post} - Y_{pre}$$

$\mu_{d,\Delta} = \mu_{d,1}(X) - \mu_{d,0}(X)$ , where  $\mu(X)$  is a model for

$$m_{d,t} = E[Y_t | D = d, X = x]$$

So that means  $\mu_{0,\Delta}$  is just the control group's change in average  $Y$  for each  $X = x$

## Population DR DiD model for panel data

$$\delta^{dr} = E \left[ \left( \frac{D}{E[D]} - \frac{\frac{p(X)(1-D)}{(1-p(X))}}{E \left[ \frac{p(X)(1-D)}{(1-p(X))} \right]} \right) (\Delta Y - \mu_{0,\Delta}(X)) \right]$$

Notice how the model controls for  $X$ : you're weighting the adjusted outcomes using the propensity score

The reason you control for  $X$  twice is because you don't know which model is right. DR DiD frees you from making a choice without making you pay too much for it

# Efficiency

- Authors exploit all the restrictions implied by the assumptions to construct semiparametric bounds
- This is where the influence function comes in, which those who have studied the DID code closely may have noticed
- One of the main results of the paper is that the DR DiD estimator is also DR for inference
- Let's skip to Monte Carlos

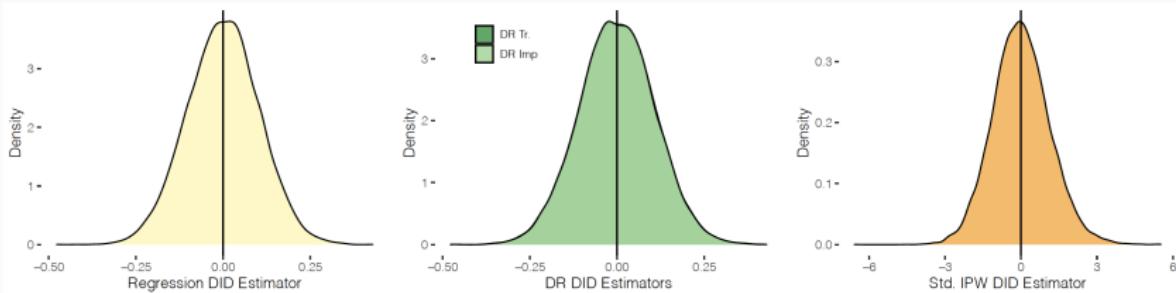
## Monte Carlo details

- Compare DR with TWFE, OR and IPW
- Sample size is 1,000
- 10,000 Monte Carlo experiments
- Propensity score estimated with logit; OR estimated using linear specification

*Table:* Monte Carlo Simulations, DGP1, Both OR and Propensity score correct

	<b>Bias</b>	<b>RMSE</b>	<b>SE</b>	<b>Coverage</b>	<b>CI length</b>
TWFE	-20.9518	21.1227	2.5271	0.000	9.9061
OR	-0.0012	0.1005	0.1010	0.9500	0.3960
IPW	0.0257	2.7743	2.6636	0.9518	10.4412
DR	-0.0014	0.1059	0.1052	0.9473	0.4124

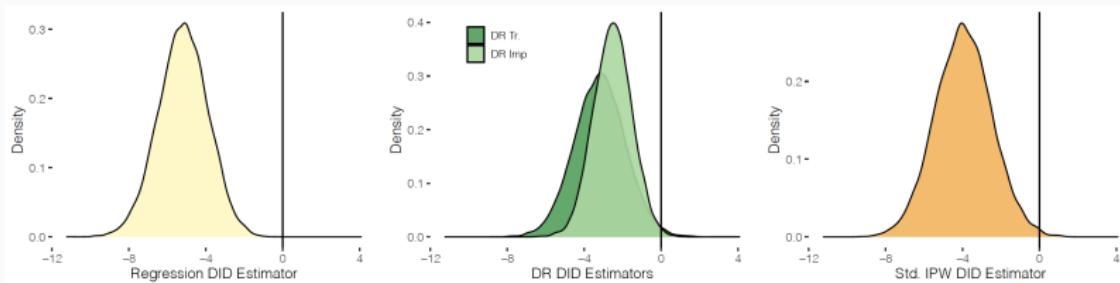
**Figure 1:** Monte Carlo for DID estimators, DGP1: Both pscore and OR are correctly specified



*Table:* Monte Carlo Simulations, DGP4, Neither OR and Propensity score correct

	<b>Bias</b>	<b>RMSE</b>	<b>SE</b>	<b>Coverage</b>	<b>CI length</b>
TWFE	-16.3846	16.5383	3.6268	0.000	14.2169
OR	-5.2045	5.3641	1.2890	0.0145	5.0531
IPW	-1.0846	2.6557	2.3746	0.9487	9.3084
DR	-3.1878	3.4544	1.2946	0.3076	5.0749

**Figure 4:** Monte Carlo for DID estimators, DGP4: Both OR and PS are misspecified



# Code

There is code in R and Stata

- Stata: **drdid**
- R: **drdid**

Remember – it's for 2x2 with covariates (i.e., one treatment group)

## Concluding remarks

- These two papers mark a different approach than is often the case for applied researchers who simply estimate regression models and hope they recover “reasonably weighted” causal effects
- These new DiD start with target parameter and identification then build estimation
- TWFE, as it turns out, is not mostly harmless