# STAGGERINGLY PROBLEMATIC: POTENTIAL ISSUES WITH STAGGERED DID AND THEIR SOLUTIONS

JOHN M. BARRIOS | WASH U - OLIN

Please Reference as:
Barrios, John. 2021. "Staggeringly Problematic: A Primer on Staggered DiD for Accounting Researchers"

# OVERVIEW

I. Quick DiD refresher
   a) Canonical Model
   b) Staggered -Time Variation in Treatment

II. Issues with Staggered DiD
   a) Simulation of main issues

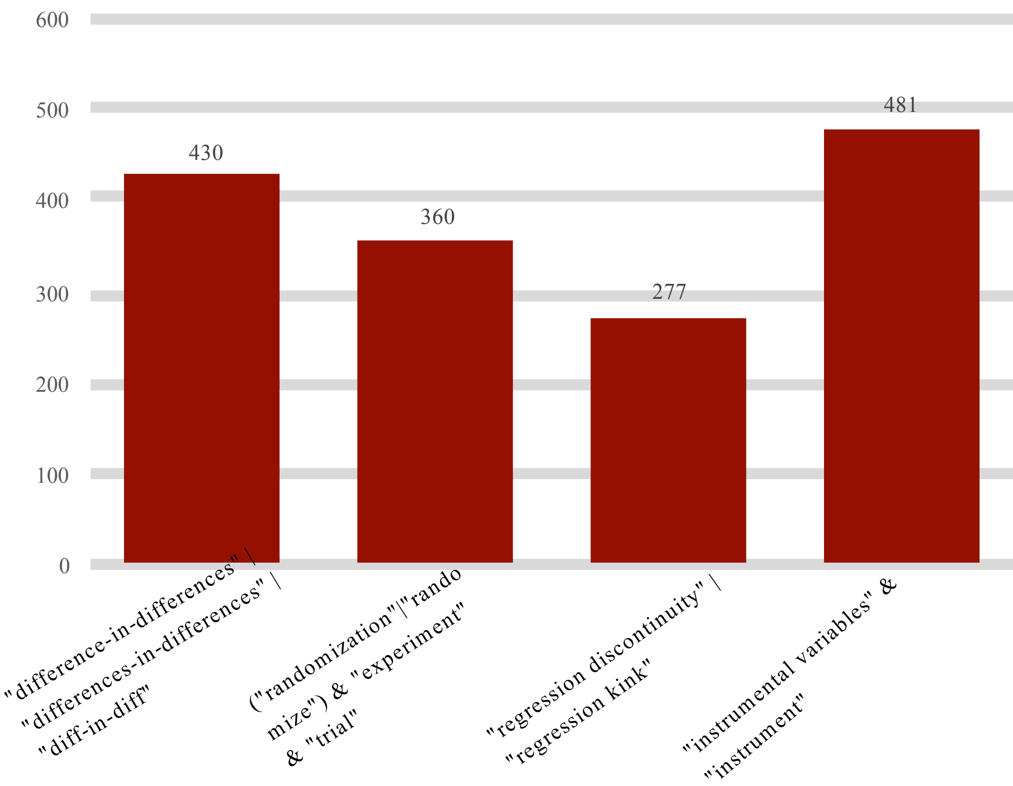
III. Overview of potential solutions - Applications
   a) Bacon Goodman – timing diagnostics and weights
   b) Stacking Event Studies

## THINK OF THIS AS A PRIMER ON STAGGERED DiD

**Keywords in NBER Papers Since 2012**

2014/2015 AER/QJE/JPE/ReStud/JHE/JDE published 93 DD papers:

**49% had timing variation**




BARRIOS (2021)

# DIFF – IN – DIFF

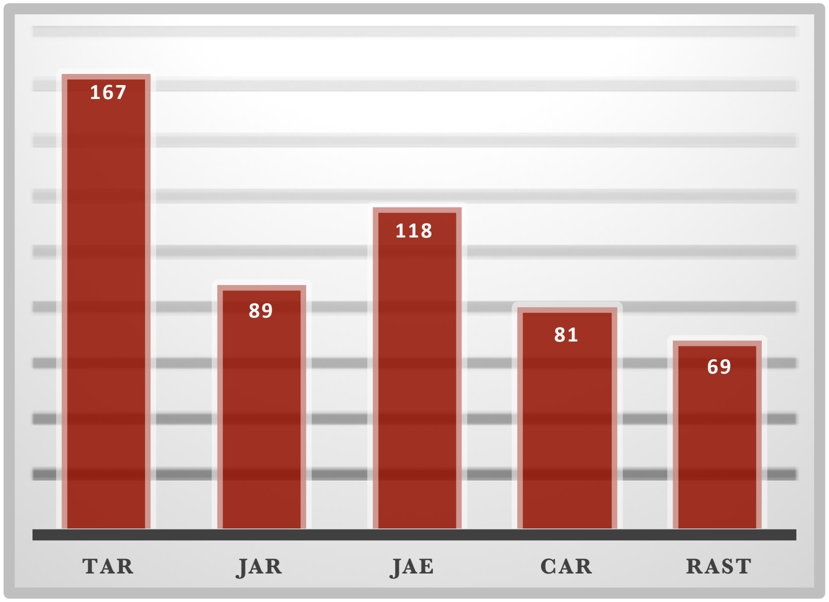
One of the most popular identification strategies in applied work today

- More than 430 working papers in the past 10 year using DiD
  - Economics and Finance

**Keywords in Published Papers Since 2011**

2011/2020 TAR/JAE/JAR/CAR/RAST published 500+ DD papers:

**30% had timing variation mentioned**




"difference-in-differences" | "differences-in-differences" | "diff-in-diff"

# DIFF – IN – DIFF

One of the most popular identification strategies in applied work today

- More than 430 working papers in the past 10 year using DiD
  - Economics and Finance
- More than 500 published papers
  - Top 5 Accounting journals 2011-2020

Angrist and Krueger (1999):

$$\{E[Y_i \mid c = \text{Miami}, t = 1981] - E[Y_i \mid c = \text{Comparison}, t = 1981]\}$$
$$-\{E[Y_i \mid c = \text{Miami}, t = 1979] - E[Y_i \mid c = \text{Comparison}, t = 1979]\} = \delta. \quad (21)$$

Athey and Imbens (2006):

i.e., $\varepsilon_i \perp (G_i, T_i)$, and is normalized to have mean zero. The standard DID estimand is

$$(2) \qquad \tau^{\mathrm{DID}} = [\mathbb{E}[Y_i|G_i = 1, T_i = 1] - \mathbb{E}[Y_i|G_i = 1, T_i = 0]] - [\mathbb{E}[Y_i|G_i = 0, T_i = 1] - \mathbb{E}[Y_i|G_i = 0, T_i = 0]].$$

Cameron and Trivedi (2007):

Then the OLS estimator reduces to

$$\widehat{\phi} = \Delta\bar{y}^{\mathrm{tr}} - \Delta\bar{y}^{\mathrm{nt}}. \quad (22.43)$$

This estimator is called the **differences-in-differences (DID) estimator**, since one estimates the time difference for the treated and untreated groups and then takes the difference in the time differences.

Angrist and Pischke (2009):

The population difference-in-differences,

$$\{E[Y_{ist}|s = NJ, t = Nov] - E[Y_{ist}|s = NJ, t = Feb]\}$$
$$-\{E[Y_{ist}|s = PA, t = Nov] - E[Y_{ist}|s = PA, t = Feb]\} = \delta,$$

BARRIOS (2021)

# DIFF – IN – DIFF

Canonical 2x2 DiD

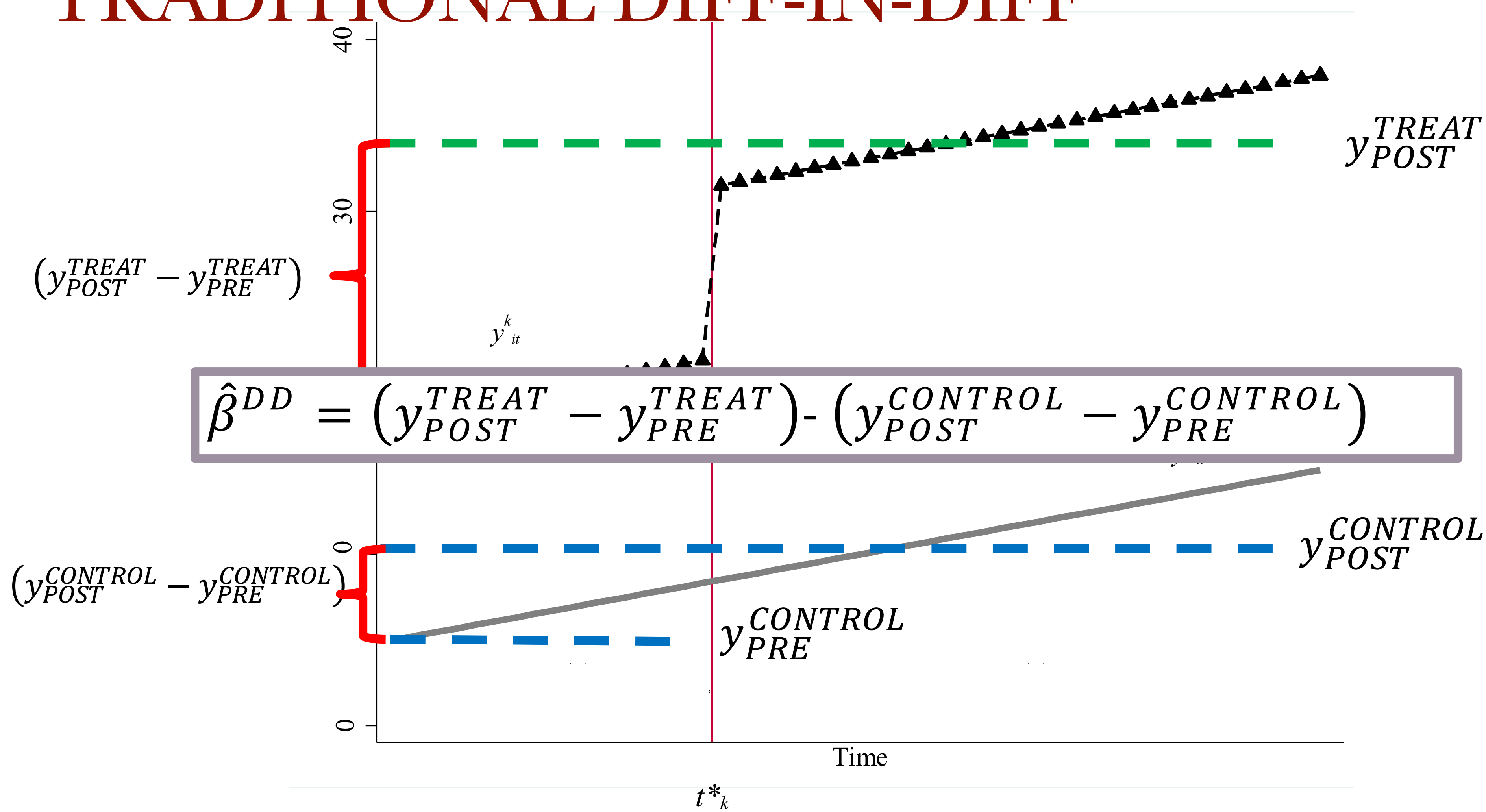
Examine one group ***treated*** at time ***t*** against ***control*** group that was not treated

- Two Groups
- Single Treatment
- Compares the difference in their changes

$$\beta^{DD} = (y^{TREAT}_{POST} - y^{TREAT}_{PRE}) - (y^{CONTROL}_{POST} - y^{CONTROL}_{PRE})$$

Attempts to mimic random assignment with treatment and "comparison" sample

# TRADITIONAL DIFF-IN-DIFF

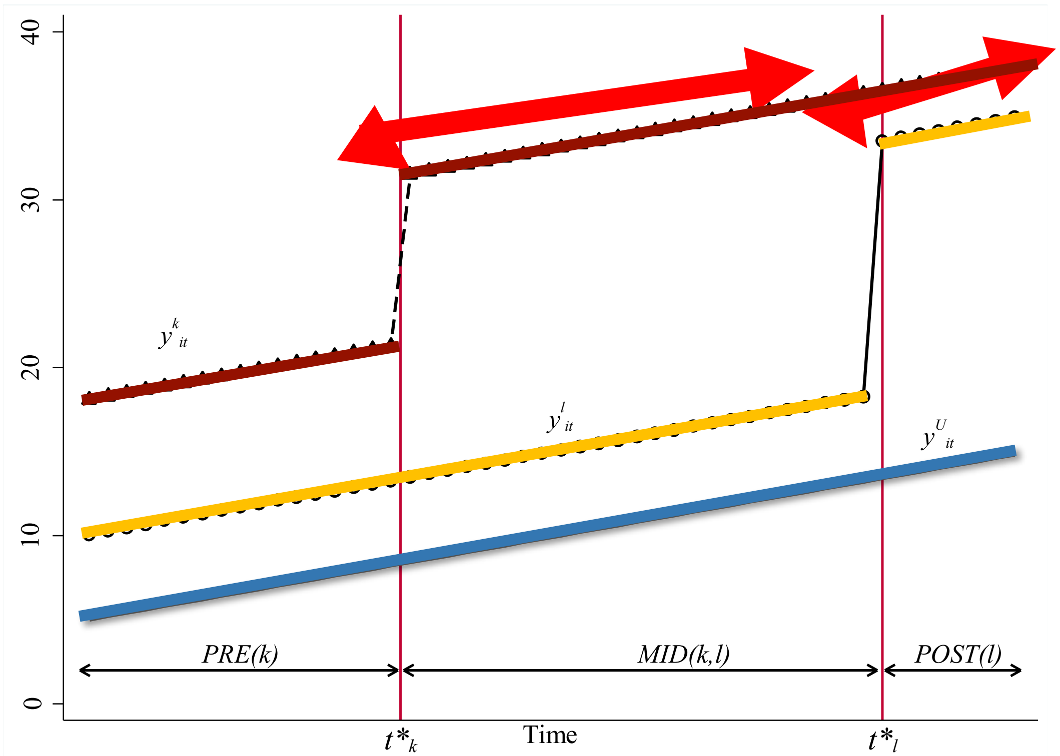



$$\hat{\beta}^{DD} = (y^{TREAT}_{POST} - y^{TREAT}_{PRE}) - (y^{CONTROL}_{POST} - y^{CONTROL}_{PRE})$$

# FROM TRADITIONAL 2X2 TO STAGGERED DIFF-IN-DIFF

Most DiD applications exploit variation in the timing of treatments across groups and units

$D_{it}$ turns on at different times $t^*$




BARRIOS (2021)

# VARIATION IN TIMING

These are cases where the treatment turns on at a different time for different units

- IPOS
- Big-4 Audits
- Law Changes
- Natural disasters
- Mass layoffs, …

# FROM TRADITIONAL 2X2 TO STAGGERED DIFF-IN-DIFF

Most DiD applications exploit variation in the timing of treatments across groups and units

Yet to test, most paper run a two-way fixed effects (TWFE) estimator

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

The coefficient that comes from this specification is ***not as easily interpretable as the 2x2 coefficient*** discussed before

Numerous papers have now begun to examine the interpretation of this coefficient

# RECENT RESEARCH ON DiD WITH TIMING


Abraham, Sarah, and Liyang Sun. 2018. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects."  Working Paper.

Athey, Susan, and Guido W. Imbens. 2018. "Design-based Analysis in Difference-in-Differences Settings with Staggered Adoption."  Working Paper.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2003. "Some Evidence on Race, Welfare Reform, and Household Income."  The American Economic Review 93 (2):293-298. doi: 10.2307/3132242.

Borusyak, Kirill, and Xavier Jaravel. 2017. "Revisiting Event Study Designs."  Harvard University Working Paper.

Callaway, Brantly, and Pedro Sant'Anna. 2018. "Difference-in-Differences With Multiple Time Periods and an Application on the Minimum Wage and Employment."  Working Paper.

de Chaisemartin, C., and X. D'HaultfŒuille. 2018a. "Fuzzy Differences-in-Differences."  The Review of Economic Studies 85 (2):999-1028. doi: 10.1093/restud/rdx049.

de Chaisemartin, C.;, and X. D'HaultfŒuille. 2018b. "Two-way fixed effects estimators with heterogeneous treatment effects."  Working Paper.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro. 2018. "Pre-event Trends in the Panel Event-study Design."  National Bureau of Economic Research Working Paper Series No. 24565. doi: 10.3386/w24565.

Gibbons, Charles, E., Juan Carlos Suárez Serrato, and Michael Urbancic, B. 2018. Broken or Fixed Effects? In Journal of Econometric Methods.

Goodman-Bacon, Andrew. 2018. "Difference-in-Differences with Variation in Treatment Timing."  *National Bureau of Economic Research Working Paper Series* No. 25018. doi: 10.3386/w25018.

Imai, Kosuke, In Song Kim, and Erik Wang. 2018. "Matching Methods for Causal Inference with Time-Series Cross-Section Data."  Working Paper.

Krolikowski, Pawel. 2017. "Choosing a Control Group for Displaced Workers."  ILR Review:0019793917743707. doi: 10.1177/0019793917743707.

Sloczynski, Tymon. 2017. "A General Weighted Average Representation of the Ordinary and Two-Stage Least Squares Estimands."  Working Paper.

Strezhnev, Anton. 2018. "Semiparametric Weighting Estimators for Multi-Period Difference-in-Differences Designs."  *Working Paper*.

Wooldridge, Jeffrey M. 2005. "Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models."  The Review of Economics and Statistics 87 (2):385-390.

# Two-Way Fixed Effects Estimator

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

Unit fixed effects

Time fixed effects

Treatment dummy

**What is $\hat{\beta}^{DD}$?**

This has been unclear, but we do have good intuition for subsamples



## EXAMPLE OF THE TWFE ESTIMATOR IN THE CASE OF THREE GROUPS:

1. **Never Treated**
2. **Treated Early**
3. **Treated Later**

# FROM TRADITIONAL 2X2 TO STAGGERED DIFF-IN-DIFF

Most DiD applications exploit variation in the timing of treatments across groups and units

Yet most paper run a two-way fixed effects (TWFE) estimator when there are more than two units and periods

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

The coefficient that comes from this not easily interpretable parameter as the 2x2 coefficient discussed before

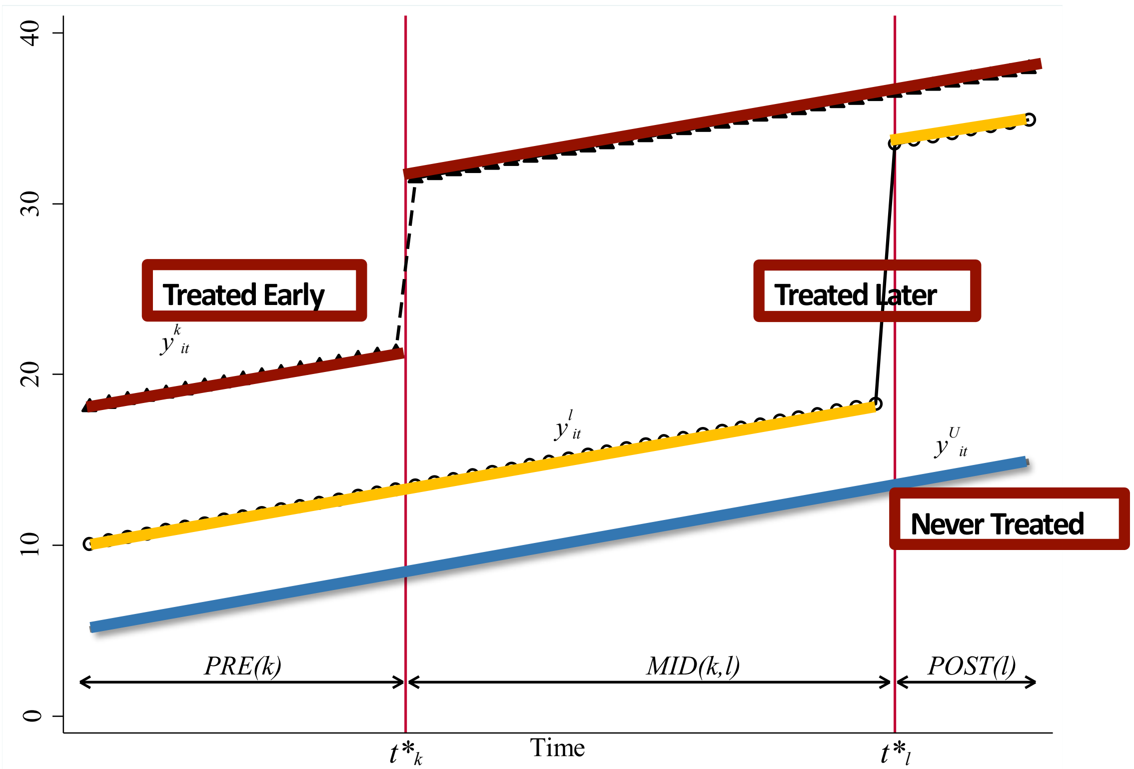
Consensus that the coefficient is a weighted average of many different 2x2 treatment effects

These weights are often negative and non-intuitive requiring accounting researchers to further examine what the $\hat{\beta}^{DD}$ is reflecting

Follow Goodman Bacon 2019 to explain the interpretation of $\hat{\beta}^{DD}$ as

**a weighted average of all possible 2x2 DiD estimators in the dataset**

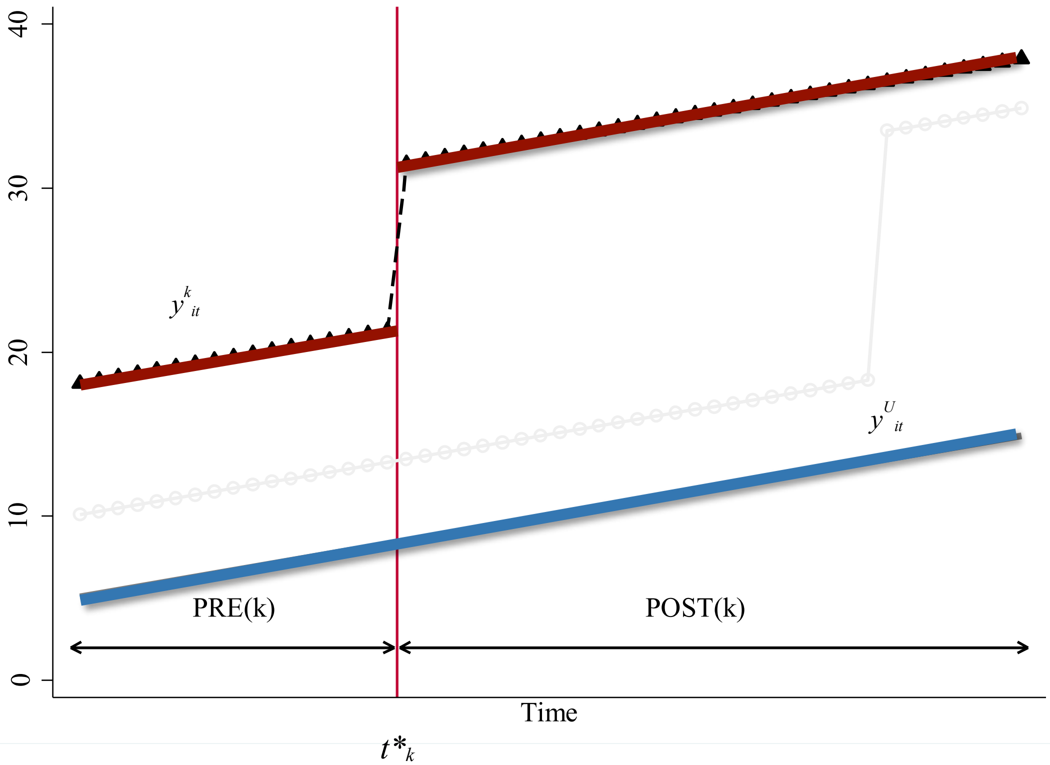
2x2 DDs: subsamples with two groups (treat/control) and two periods (pre/post)




BARRIOS (2021)

# EXAMPLE OF THE TWFE ESTIMATOR IN THE CASE OF THREE GROUPS:

1. Never Treated
2. Treated Early
3. Treated Later



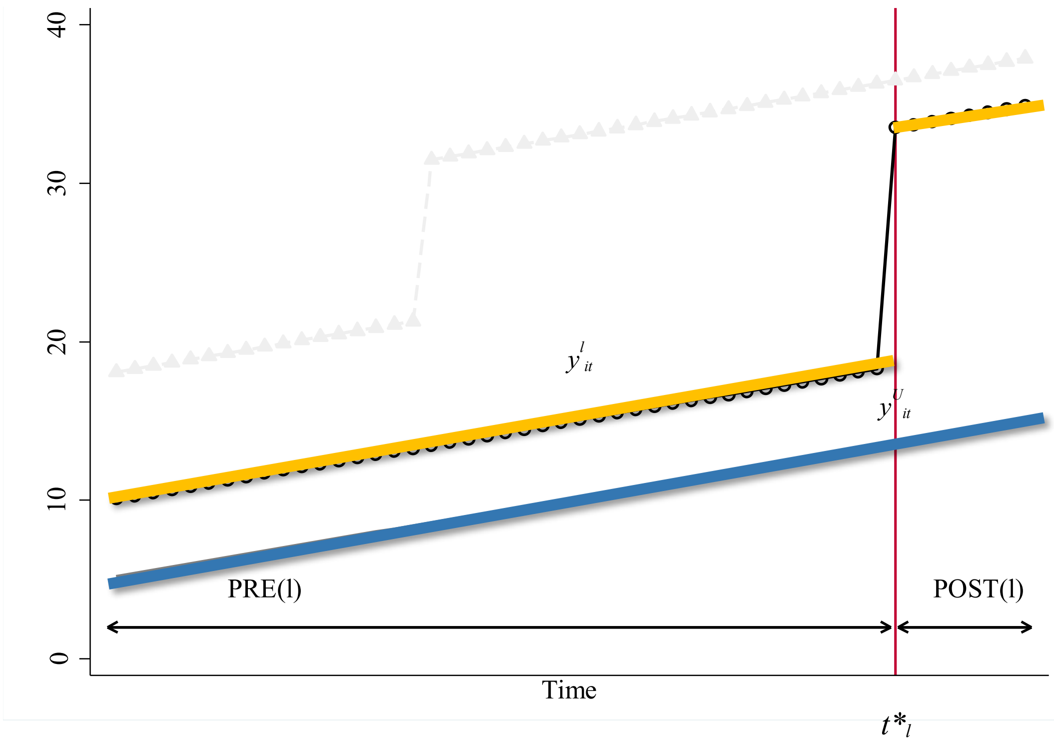

BARRIOS (2021)

# $1^{ST}$ - 2X2 COMPARISON

## TREATED EARLY VS NEVER TREATED

$$\widehat{\beta}^{2x2}_{kU}$$

$$\widehat{\beta}^{2x2}_{kU} = \left(y_k^{POST(k)} - y_k^{PRE(k)}\right) - \left(y_U^{POST(k)} - y_U^{PRE(k)}\right)$$



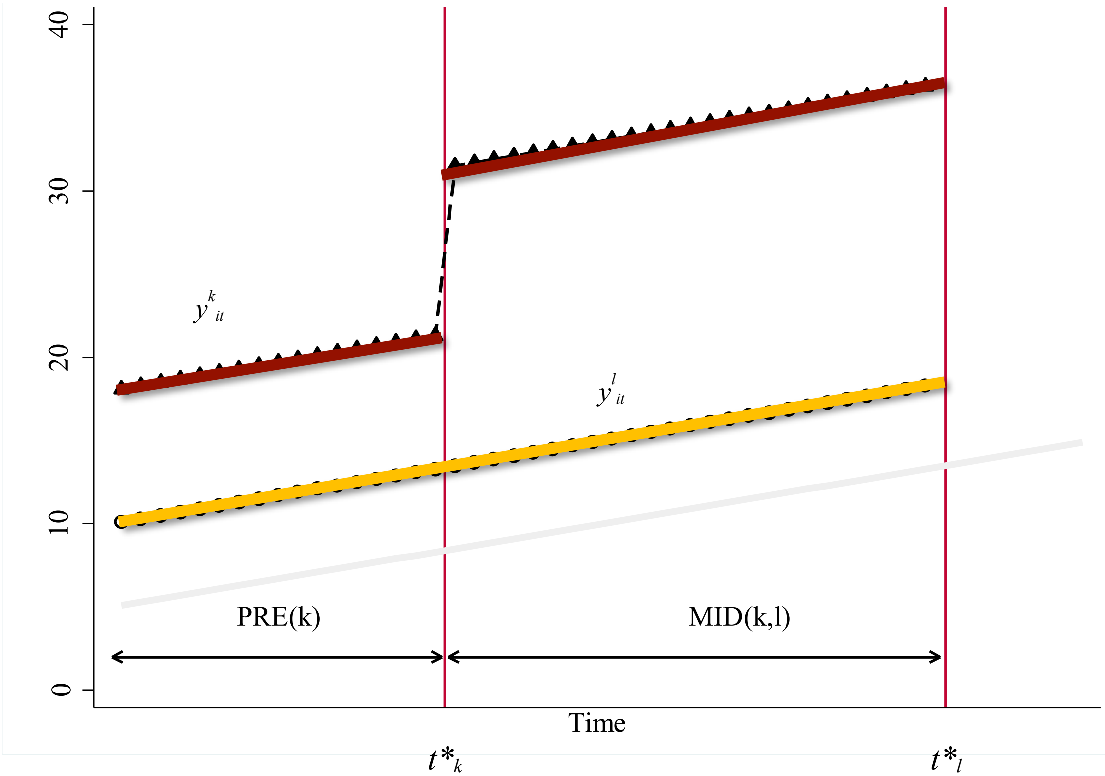

BARRIOS (2021)

# $2^{ND}$ - 2X2 COMPARISON

# TREATED LATE VS NEVER TREATED

$$\widehat{\beta}_{\ell U}^{2x2}$$

$$\widehat{\beta}_{\ell U}^{2x2} = \left(y_{\ell}^{POST(\ell)} - y_{\ell}^{PRE(\ell)}\right) - \left(y_{U}^{POST(\ell)} - y_{U}^{PRE(\ell)}\right)$$



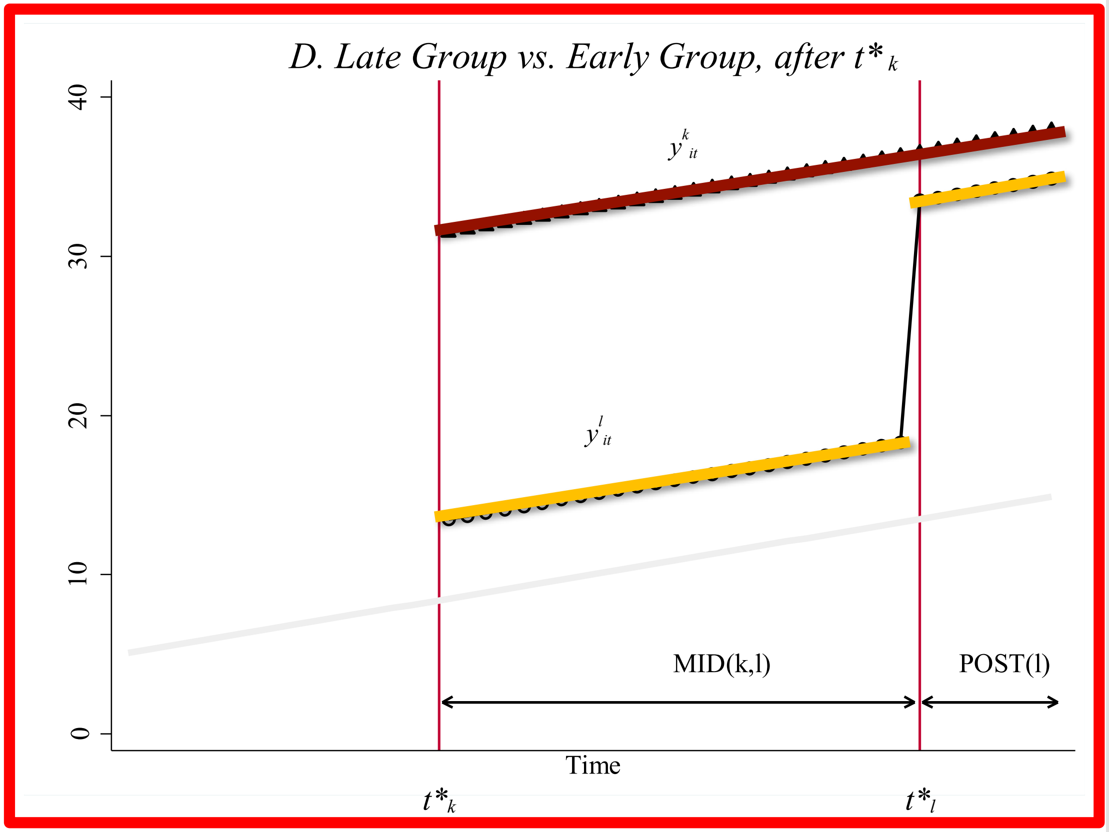

BARRIOS (2021)

# 3RD - 2X2 COMPARISON

## TREATED EARLY VS TREATED LATE PRE t*

$$\widehat{\beta}^{2x2,k}_{k\ell}$$

$$\widehat{\beta}^{2x2,k}_{k\ell} = \left(y_k^{MID(k,\ell)} - y_k^{PRE(k)}\right) - \left(y_\ell^{MID(k,\ell)} - y_\ell^{PRE(k)}\right)$$




BARRIOS (2021)

# $4^{TH}$ - 2X2 COMPARISON

## TREATED LATE VS TREATED EARLY AFTER $t^*$

$$\widehat{\beta}_{k\ell}^{2x2,\ell}$$

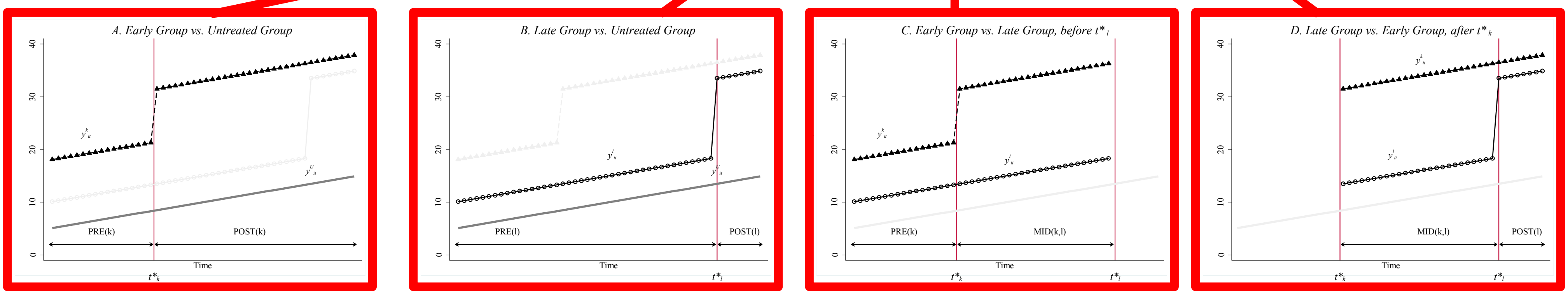
$$\widehat{\beta}_{k\ell}^{2x2,\ell} = \left(y_{\ell}^{Post(l)} - y_{\ell}^{MID(k,\ell)}\right) - \left(y_{k}^{Post(\ell)} - y_{k}^{MID(k,\ell)}\right)$$

# DIFFERENCE-IN-DIFFERENCES DECOMPOSITION (3 GROUP CASE)

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

For three groups:

$$\hat{\beta}^{DD} = s_{kU}\boxed{\hat{\beta}^{2x2}_{kU}} + s_{\ell U}\boxed{\hat{\beta}^{2x2}_{\ell U}} + s^{k}_{k\ell}\boxed{\hat{\beta}^{2x2,k}_{k\ell}} + s^{k}_{k\ell}\boxed{\hat{\beta}^{2x2,\ell}_{k\ell}}$$



# WHAT DO WE LEARN FROM THE DECOMPOSITION
# COMPARISON OF THE 2X2S

1. Helps clarify what comparisons are being made when we run the regression:

    "switchers vs untreated"?

    "early switchers vs late switchers"?

    "late vs early" (this is less obvious)?

   It's all of these

2. "What is the control group?"

   Every group acts as a control (sometimes)

3. Help clarifies theory
   We understand the estimand (ATET)
   Identification assumption (common trends) for each 2x2
    **late vs. early comparisons are biased if treatment effects vary over time**

# DIFFERENCE-IN-DIFFERENCES DECOMPOSITION (3 GROUP CASE)

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

For three groups:

$$\hat{\beta}^{DD} = \boxed{s_{kU}}\hat{\beta}^{2x2}_{kU} + \boxed{s_{\ell U}}\hat{\beta}^{2x2}_{\ell U} + \boxed{s^{k}_{k\ell}}\hat{\beta}^{2x2,k}_{k\ell} + \boxed{s^{\ell}_{k\ell}}\hat{\beta}^{2x2,\ell}_{k\ell}$$

*A. Early Group vs. Untreated Group*

**Size:**

$$(n_k + n_U)^2$$
$$\times$$

**Variance:**

$$n_{kU}(1 - n_{kU})D_k(1 - D_k)$$

*B. Late Group vs. Untreated Group*

**Size:**

$$(n_\ell + n_U)^2$$
$$\times$$

**Variance:**

$$n_{\ell U}(1 - n_{\ell U})D_\ell(1 - D_\ell)$$

*C. Early Group vs. Late Group, before $t^*_\ell$*

**Size:**

$$((n_k + n_\ell)(1 - D_\ell))^2$$
$$\times$$

**Variance:**

$$n_{k\ell}(1 - n_{k\ell})\frac{D_k - D_\ell}{1 - D_\ell}\frac{1 - D_k}{1 - D_\ell}$$

*D. Late Group vs. Early Group, after $t^*_k$*

**Size:**

$$((n_k + n_\ell)D_k)^2$$
$$\times$$

**Variance:**

$$n_{k\ell}(1 - n_{k\ell})\frac{D_\ell}{D_k}\frac{D_k - D_\ell}{D_k}$$

**Weights:** $\dfrac{(\text{subsample share})^2(\text{subsample variance of FE–adjusted } D)}{\text{total variance of FE–adjusted } D}$

# WHAT DO WE LEARN FROM THE DECOMPOSITION WEIGHTS

1. Relative importance of each kind of comparison.
   More important if bigger sample size or treated closer to middle of the panel (i.e. bigger variance).

"How much" comes from timing vs comparisons to untreated.

2. Importance of specific 2x2 DDs.
   Sometimes a few of the 2x2 DD terms dominate.

3. Clarifies theory
   The estimand and ID assumption are "variance weighted";
   Can compare estimand to "parameters of interest" and conduct a proper balance test

# POTENTIAL ISSUES WITH TWFE ESTIMATION AND STAGGERED DID

**1st** Two-way fixed effects estimates of DiD that rely staggered treatments only recover the average treatment effect when treatment effects are homogeneous.

When treatment effects are ***heterogeneous*** across firms, OLS *over-weights units with more variance in treatment status* in order to achieve a more precise estimate of the treatment effect

Thus, if you want the *average treatment effect on the treated -* some sort of *re-weighting is required*

**2nd** DiD estimates are biased when treatment effects change over time within unit.

Occurs because already treated units serve as controls in some of the 2X2 DiDs underlying the weighted average

Thus, when treatment effects are not constant over time **using already treated units as controls biases estimates of the treatment effect**

Basically, you are *differencing a change in the treatment effect* by using the already treated units

Now I use a few simulations to show how these issues may arise

# POTENTIAL ISSUES WITH TWFE ESTIMATION AND STAGGERED DID

**1st** Two-way fixed effects estimates of DiD that rely staggered treatments only recover the average treatment effect when treatment effects are homogeneous.

- When treatment effects are ***heterogeneous*** across firms, OLS *over-weights units with more variance in treatment status* in order to achieve a more precise estimate of the treatment effect
- Thus, if you want the *average treatment effect on the treated* - some sort of *re-weighting is required*

**2nd** DiD estimates are biased when treatment effects change over time within unit.

- Occurs because already treated units serve as controls in some of the 2X2 DiDs underlying the weighted average
- Thus, when treatment effects are not constant over time **using already treated units as controls biases estimates of the treatment effect**
  - Basically, you are *differencing a change in the treatment effect* by using the already treated units

Now I use a few simulations to show how these issues may arise

# POTENTIAL ISSUES WITH TWFE ESTIMATION AND STAGGERED DID

**1st** Two-way fixed effects estimates of DiD that rely staggered treatments only recover the average treatment effect when treatment effects are homogeneous.

- When treatment effects are ***heterogeneous*** across firms, OLS *over-weights units with more variance in treatment status* in order to achieve a more precise estimate of the treatment effect
- Thus, if you want the *average treatment effect on the treated -* some sort of *re-weighting is required*

**2nd** DiD estimates are biased when treatment effects change over time within unit.

- Occurs because already treated units serve as controls in some of the 2X2 DiDs underlying the weighted average
- Thus, when treatment effects are not constant over time **using already treated units as controls biases estimates of the treatment effect**
  - Basically, you are *differencing a change in the treatment effect* by using the already treated units

Now I use a few simulations to show how these issues may arise

# SIMULATION SET-UP

Data generating process

Linear process where: $y_{it} = \alpha_i + \alpha_t + \beta_{it} + \epsilon_{it}$

$\alpha_i$ and $\alpha_t$ follow a $N(0,1)$ and $\epsilon_{it}$ follows a N (0, .25)

$\beta_{it}$ can be assigned either as:

**1 Period Treatment:**

50 states and 1000 firms randomly drawn from the 50 states

Half the states receive Treatment in 1995: Treat indicator is 1 starting in 1995 and after for the treated state

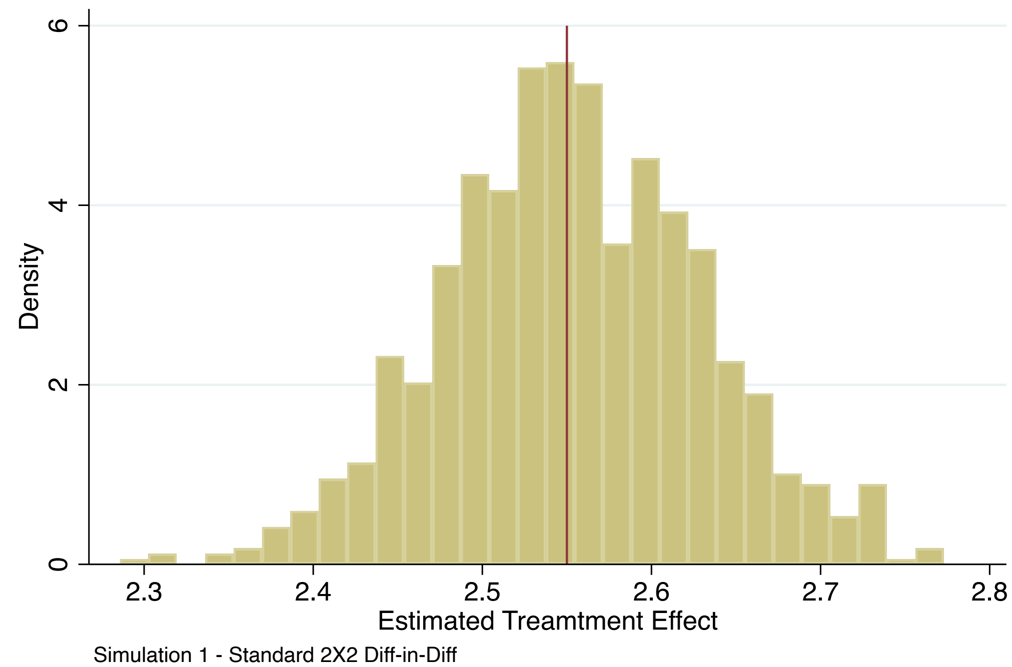
Every firm in a treated state, has a firm-specific treatment effect $\mu_i$ is from a N (0.3, 0.04)

Treatment effects are trend breaks: the accumulated treatment effect: $\mu_i \times (\text{year} - 1995 + 1)$ for years after 1995

Then I estimate a two way fixed effect regression to get an estimated $\beta$

Simulate this data 1,000 and plot the distribution of estimates $\hat{\beta}$ and the true effect (red line)

# 1 PERIOD 2X2 DIFF-IN-DIFF SIMULATION



# SIMULATION SET-UP

Data generating process

Linear process where: $y_{it} = \alpha_i + \alpha_t + \beta_{it} + \epsilon_{it}$

$\alpha_i$ and $\alpha_t$ follow a $N(0,1)$ and $\epsilon_{it}$ follows a N (0,.25)

$\beta_{it}$ can be assigned either as:

**Staggered Treatment:**

40 states are randomly assigned into four treatment cohorts of size 250 depending on year of treatment

Years of assignment are 1986, 1992, 1998, and 2004

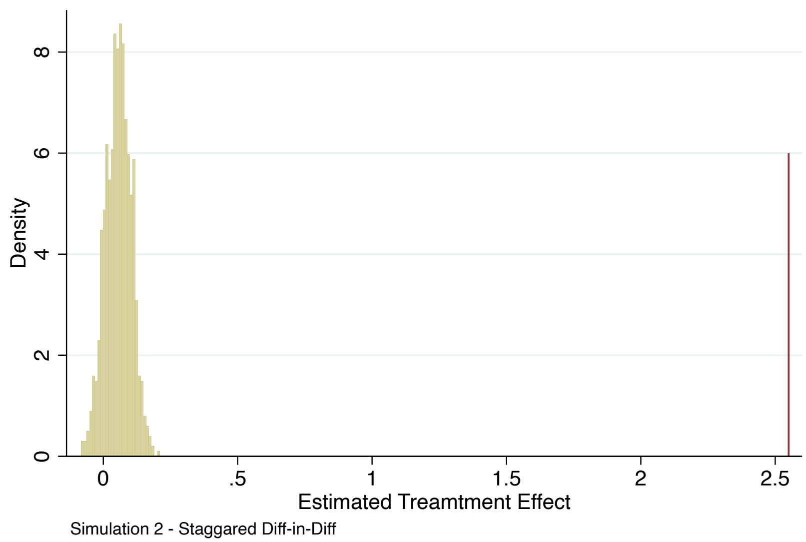
The treatment process is identical to the one before except:

$\beta_{it} = \mu_i \times (\text{year} - \tau_c + 1)$

where $\tau_c$ is the treatment year cohort for the firms (1986, 1992 etc).

Then I estimate a two way fixed effect regression to get an estimated $\hat{\beta}$

Simulate this data 1,000 and plot the distribution of estimates $\hat{\beta}$ and the true effect (red line)



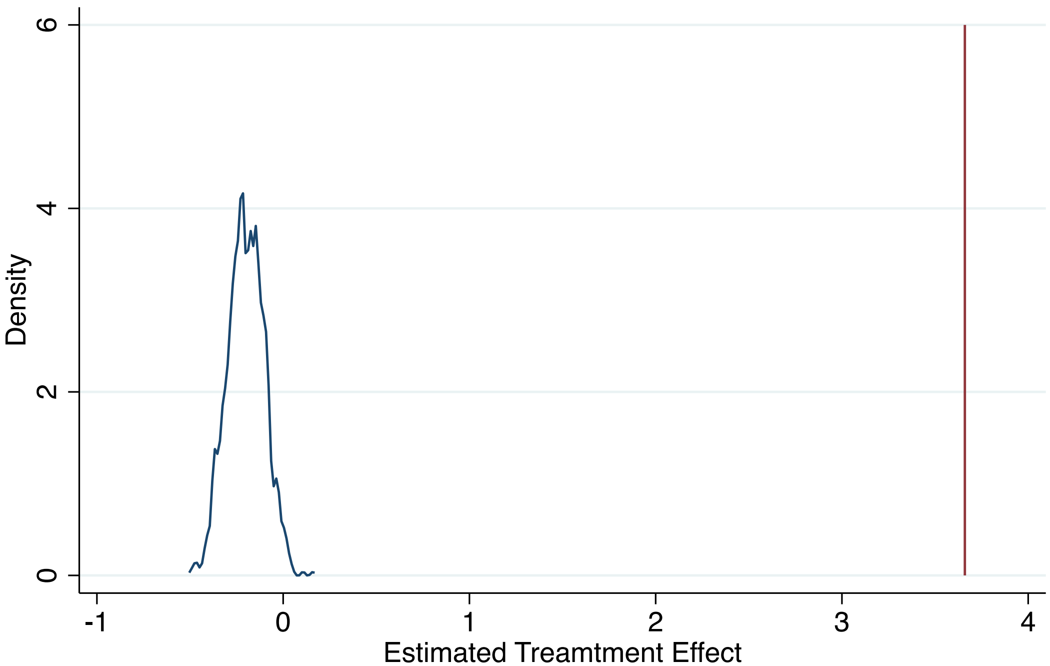

# STAGGERED TREATMENT DIFF-IN-DIFF SIMULATION

Issue: we use prior treated firms as controls

- When the treatment effect is "dynamic", i.e. takes more than one period to be incorporated
- Thus you ***subtract*** the treatment effects from prior treated units from the estimate of future control firms

This biases your estimates towards zero

- when all the treatment effects are the same



Simulation 3 - Staggared Diff-in-Diff TE larger for Early States

# SEVERELY BIASING THE $\hat{\beta}$ SIMULATION

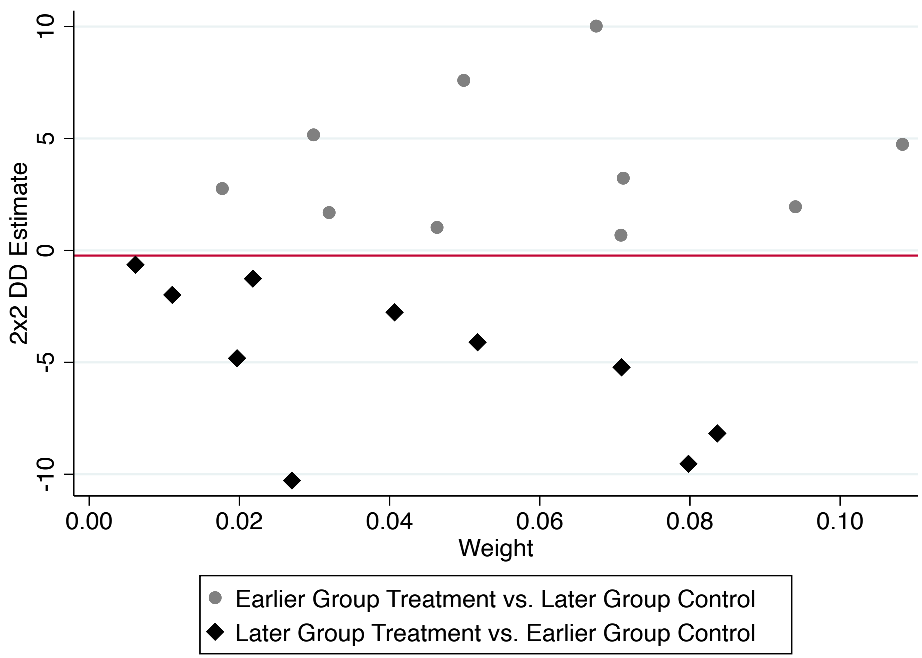
If treatment effects for early treated firms are larger than treatment effects for later treated firms

Flips sign of $\beta$

Firms are randomly assigned to one of 50 states

The 50 states are randomly assigned into one of 5 Groups.

- The average treatment effect multiple decreases over time
- (G1 - .8, G2 - .4, G3 - .3, G4 - .2, G5 - .1)




# SEVERELY BIASING THE $\hat{\beta}$ SIMULATION

If treatment effects for early treated units are larger than the treatment effects on later treated units

Flips sign of $\beta$

Can see by using the Bacon Decomposition to plot the different 2x2 betas against their weights

The ***early vs late*** comparisons are **positive**

While the ***late vs early*** comparisons are **negative**

**Callaway & Sant'Anna**

Develop an **inverse propensity weighted long-difference in cohort-specific average treatment effects** between treated and untreated units for a given treatment cohort

**Abraham & Sun**

Extension of the standard event-study TWFE model where we **saturate the event time indicators (i.e. t = -2, -1, ...) with indicators for the treatment initiation year group and sum to overall aggregate event time indicators by cohort size**

**Cengiz et al. – Stacking Procedure**

Akin to matching treated to control firms that do not experience treatment within the pre- and post-window as defined by the researcher

Stack each comparison

Run the same TWFE estimates as in standard DiD

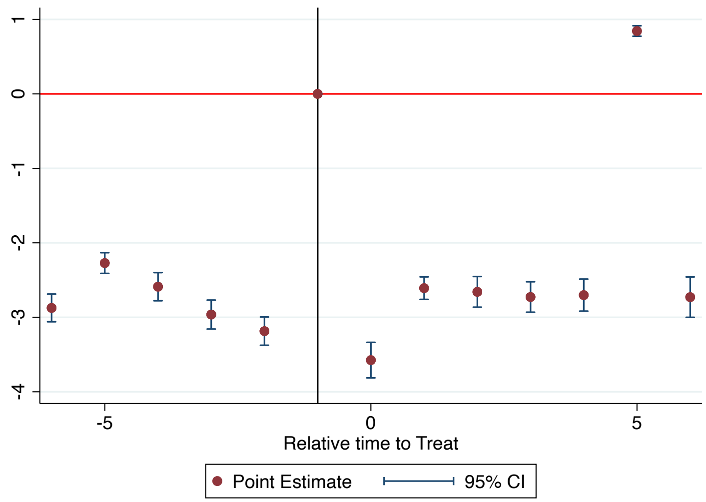
- But include **interactions for the cohort-specific dataset with all of the fixed effects, controls, and clusters**.

# PROPOSED SOLUTIONS

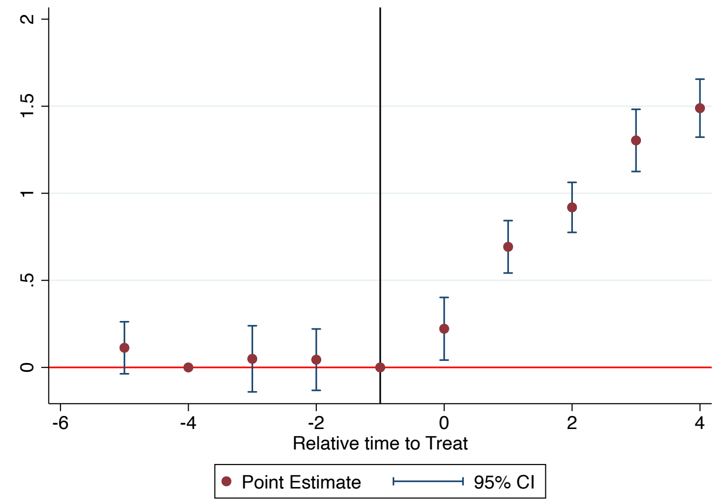
BARRIOS (2021

# HOW THEY HELP

## UNADJUSTED EVENT STUDY




## ADJUSTED – STACKING DESIGN




BARRIOS (2021

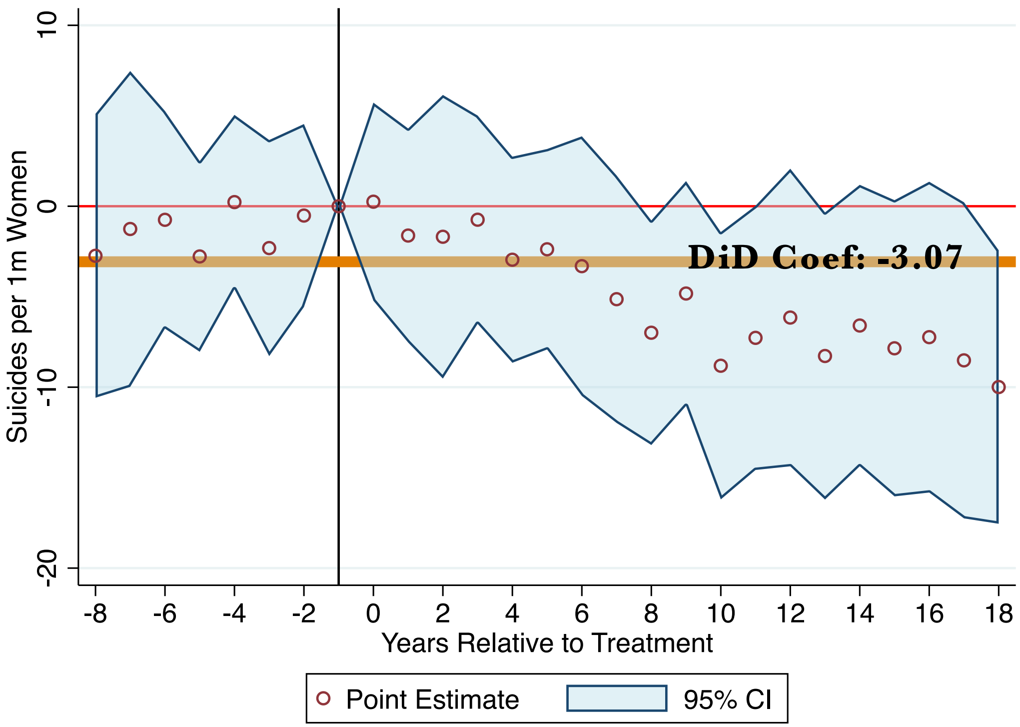
# REAL WORLD EXAMPLES

# 1) THE EFFECT OF UNILATERAL DIVORCE ON SUICIDE

# 2) ACCOUNTING – IN THE PAPER

BARRIOS (2021)

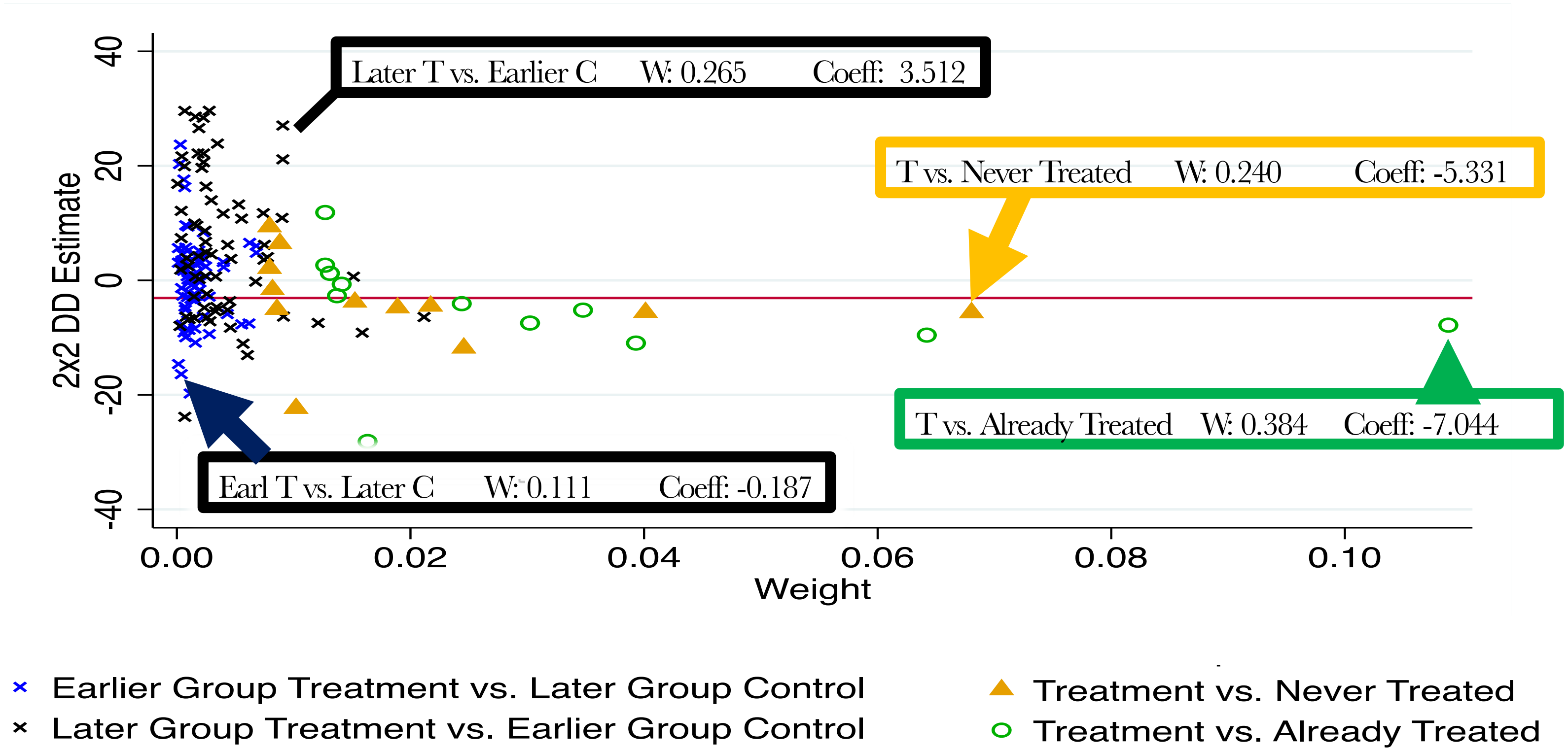
| No-Fault Divorce Year | Number of States | Share of States | Treatment Share |
|---|---|---|---|
| **Non-Reform States** | 5 | 0.10 | . |
| **Pre-64 Reform States** | 8 | 0.16 | . |
| **1969** | 2 | 0.04 | 0.85 |
| **1970** | 2 | 0.04 | 0.82 |
| **1971** | 7 | 0.14 | 0.79 |
| **1972** | 3 | 0.06 | 0.76 |
| **1973** | 10 | 0.20 | 0.73 |
| **1974** | 3 | 0.06 | 0.70 |
| **1975** | 2 | 0.04 | 0.67 |
| **1976** | 1 | 0.02 | 0.64 |
| **1977** | 3 | 0.06 | 0.61 |
| **1980** | 1 | 0.02 | 0.52 |
| **1984** | 1 | 0.02 | 0.39 |
| **1985** | 1 | 0.02 | 0.36 |




BARRIOS (2021

# PANEL EVENT STUDY

# GRAPHING THE DECOMPOSITION




BARRIOS (2021)

# HELPFUL TIPS WHEN IMPLEMENTING STAGGERED DESIGNS

It is important to check whether common trends are satisfied in any panel data set being used for DiD analysis. ***Event–Study!!***

- Crucial to use event study design in these panel tests
- Can ***trust a flat pre period!***

When treatment turns on at different times:

- Group the event study coefficients by cohort or fit a trend break instead of a constant DiD estimator
- Another way is to re-center and stack all the possible two-by-two DD comparisons with clean pre and post period for control units

Look at the weights and 2X2 DiD from the decomposition:

- Can describe in your test how many terms account for "most" of the weight.
- Check if the comparisons to "already treated" groups are very different than the other comparisons – can help give measure of bias
- You can take out the bias from time-varying effects by hand - Average the 2x2 DDs ***except*** or those that use already treated units as controls weighting by the decomposition weights.
  - The estimate will not be biased by time-varying treatment effects (*although it will generally be skewed toward short-run effects)*

BARRIOS (2021)

**Bacondecomp** - implements a decomposition of the difference-in-differences (DD) estimator with variation in treatment timing

**Flexpaneldid** - causal analysis of treatments with varying start dates and treatment durations within panel data

**Eventdd** - panel event study corresponding to a difference-in-difference style model. It generates plots of the leads and lags of the treatment

# HELPFUL STATA PACKAGES

Packages used in the presentation and paper to examine staggered DiD issues.

Also can make your life easier on your papers!

# REFERENCES


Goodman-Bacon, Andrew. 2018. "Difference-in-Differences with Variation in Treatment Timing." *National Bureau of Economic Research Working Paper Series* No. 25018. doi: 10.3386/w25018.

Stevenson, Betsey, and Justin Wolfers. 2006. "Bargaining in the Shadow of the Law: Divorce Laws and Family Distress." *The Quarterly Journal of Economics* 121 (1):267-288.

Baker, Andrew. 2019. "Difference-in-Differences Methodology." https://andrewcbaker.netlify.com/2019/09/25/difference-in-differences-methodology/.

Textbooks and Survey Articles that describe 2x2 DD:

Angrist, Joshua D., and Alan B. Krueger. 1999. "Chapter 23 - Empirical Strategies in Labor Economics." In Handbook of Labor Economics, edited by Orley C. Ashenfelter and David Card, 1277-1366. Elsevier.

Angrist, Joshua David, and Jörn-Steffen Pischke. 2009. Mostly harmless econometrics : an empiricist's companion. Princeton: Princeton University Press.

Angrist, Joshua David, and Jörn-Steffen Pischke. 2015. Mastering 'metrics : the path from cause to effect. Princeton ; Oxford: Princeton University Press.

Cameron, Adrian Colin, and P. K. Trivedi. 2005. Microeconometrics : methods and applications. Cambridge ; New York: Cambridge University Press.

Heckman, James J., Robert J. Lalonde, and Jeffrey A. Smith. 1999. "Chapter 31 - The Economics and Econometrics of Active Labor Market Programs." In Handbook of Labor Economics, edited by Orley C. Ashenfelter and David Card, 1865-2097. Elsevier.

Meyer, Bruce D. 1995. "Natural and Quasi-Experiments in Economics." Journal of Business & Economic Statistics 13 (2):151-161. doi: 10.2307/1392369.

Wooldridge, Jeffrey M. 2010. Econometric analysis of cross section and panel data. 2nd ed. Cambridge, Mass.: MIT Press.