# `drdid` and `csdid`:

## Doubly robust DID with multiple time periods

F. Rios-Avila [1] P.H.C. Sant'Anna[2] B. Callaway[3] A. Naqvi[4]

[1]Levy Economics Institute

[2]Microsoft and Vanderbilt University

[3]University of Georgia

[4]International Institute for Applied Systems Analysis

2021 Stata: Economics Virtual Symposium

# Special Thanks


- This project would not have reached this point without the help of many.

- Special thanks to:
  - Austin Nichols (Abt Associates)
  - Enrique Pinzon (Stata Corp)
  - Miklos Koren (Central European University)

# Table of Contents

# Introduction

- DID is one of the most popular methods of applied researchers aiming to analyze Causal Effects.
  - The canonical DID (2x2) compares the changes in the outcome of treated units with changes observed among non-treated/control units.
  - Under the Parallel Trends Assumption (PTA), differences in those changes (DiD) identify the Average Treatment Effects of the treated units (ATT).
- Empirical research typically deviates from the Canonical design because:
  - Researchers have access to many periods. $(T > 2)$
  - Treatment may occur at different points in time. $(G > 1)$
  - Groups may be different in terms of observed characteristics. $(X's)$

# Introduction

- Simple Solution: Identify the ATT ($\theta$) by
  - Adding fixed effects for individuals or cohorts.
  - Adding time fixed effects.
  - Adding controls

$$y = \alpha_i + \delta_t + \beta X + \theta^{twfe} * (Eff\_Tr) + u$$

- Recent research (de Chaisemartin and D'Haultoeuille, 2020; Goodman-Bacon, 2021; Borusyak and Jaravel, 2017) has shown that this simple generalization (A.K.A. TWFE) may not be adequate to identify an ATT when effects are heterogeneous.

- While others (Abadie, 2005; Heckman et al 1997; Imbens and Wooldrige, 2009; Sant'Anna and Zhao, 2020) have proposed many alternatives to better account for differences in controls.

# Introduction

With this framework, today I'll focus on the following topics:

- The identification of the ATT using canonical DID designs, the role of covariates (based on Sant'Anna and Zhao (2020) - SZ ), and its implementation via `drdid`.
- The problems related with the simple TWFE generalization of the DID estimators. (based on Goodman-Bacon (2021))
- The estimation of the DID effects with Multiple Periods (based on Callaway and Sant'Anna (2021)-CS), and its implementation via `csdid`.

# Table of Contents

# Canonical 2x2 DID

Setup

- Assume that we have access to a panel data where all units are followed over two periods: $t = 0, 1$.
- All units fall within two groups: Treated or untreated/control, status which cannot be changed ($D_i$).

Under full heterogeneity, potential and observed outcomes for a unit $i$ at time $t$ can be written as:

$$y_{i,t}(W) = \mu_i + \gamma_i t + \theta_i W \times t$$

$$y_{i,t} = D_i y_{i,t}(1) + (1 - D_i) y_{i,t}(0)$$

$$y_{i,0} = y_{i,0}(1) = y_{i,0}(0)$$

# Canonical 2x2 DID

In this framework, the treatment effect for unit $i$ at $t = 1$ is the difference between both Potential Outcomes:

$$\theta_i = y_{i,1}(1) - y_{i,1}(0)$$

And the ATT is:

$$\begin{aligned} ATT = E(\theta_i | D_i = 1) &= E_1(\theta_i) \\ &= E_1(y_{i,1}) - E_1(y_{i,1}(0)) \\ &= E_1(y_{i,1}) - (E_1(y_{i,0}) + E_1(\gamma_i)) \end{aligned}$$

If we can estimate $E_1(\gamma_i)$, we can estimate the ATT.

# Canonical 2x2 DID

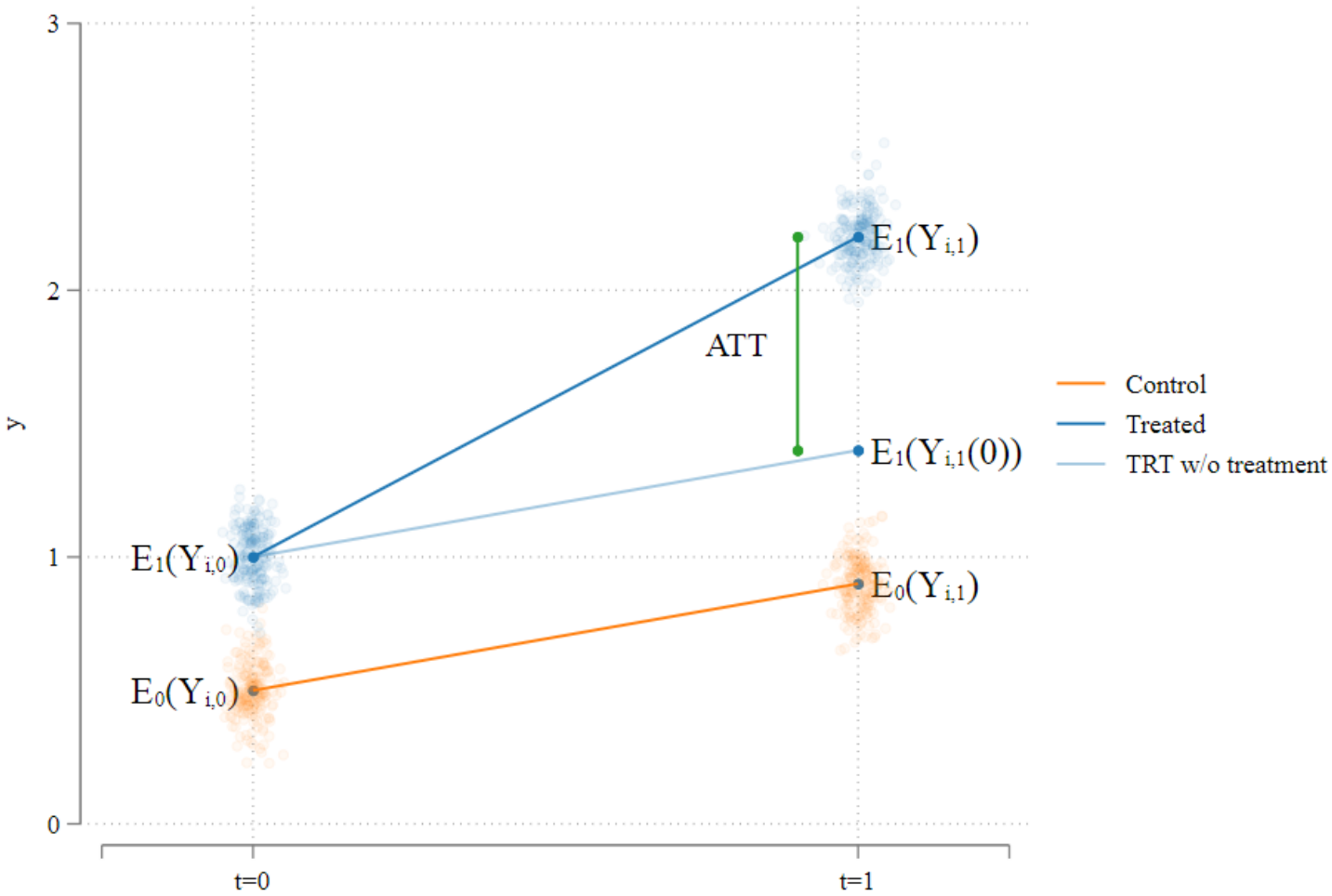
Thus, we rely on the Parallel trends Assumption (PTA)

$$E_1(\gamma_i) = E_0(\gamma_i)$$
$$E_1(y_{i,1}(0) - y_{i,0}) = E_0(y_{i,1} - y_{i,0})$$

This assumes that the growth experienced by the control group should be the same as the growth of the treated group in absence of the treatment.

$$ATT = [E_1(y_{i,1} - y_{i,0})] - [E_0(y_{i,1} - y_{i,0})]$$

# Canonical 2x2 DID



# 2x2 DID with Covariates

Simple 2x2 DID identifies the ATT under Unconditional PTA.
This may be appropriate if:

- The treated and control groups have similar characteristics; or if
- The change in outcome $\gamma_i$ does not depend on characteristics.

SZ assumes that the PTA holds only when considering groups with the same characteristics (Conditional PTA):

$$E_1(\gamma_i|X) = E_0(\gamma_i|X)$$
$$E_1(y_{i,1}(0) - y_{i,0}|X) = E_0(y_{i,1} - y_{i,0}|X) = \gamma(x)$$

Thus, the ATT estimation should adjust for differences in characteristics.

$$ATT = [E_1(y_{i,1} - y_{i,0})] - [E_1((E_0(y_{i,1} - y_{i,0}|X))]$$

# 2x2 DID with Covariates

What to kind of controls can be used?

**Panel**

- They should not be unique to treatment or control group
- They should capture pre-treatment characteristics
  (only pre-treatment values are used)
- Time varying variables are allowed only if the changes are strictly exogenous. (bad control risk)

**Repeated Cross-section**

- Time varying controls are possible but SZ assumes stationary in the covariates (Abadie, 2005).
- They should not change much across time
- The changes should be strictly exogenous.
- Post-treatment units should represent their pre-treatment counterparts.

# 2x2 DID with Covariates: How?

How are ATT's Estimated?

SZ discusses the properties of three types of estimators, which accommodate to panel and repeated cross-sectional data.

- Outcome regression
- Reweighting approach (IPW and IPT)
- Doubly Robust Estimators: Combining OR and RW

For simplicity, I will focus on panel estimators.

# 2x2 DID with Covariates

$$ATT = E(\Delta y_i | D_i = 1) - \hat{E}(\gamma_i | D_i = 1)$$

**Outcome Regression**

$$S1 : \gamma_i = \Delta y_i = \gamma(x) + v_i \quad \forall i | D_i = 0$$
$$S2 : E(\gamma_i | x) = \hat{\gamma}(x)$$
$$S3 : \widehat{ATT}_{OR} = E(\Delta y_i | D_i = 1) - E(\hat{\gamma}(x) | D_i = 1)$$

Depends strongly on the correct specification for the outcome change.

# 2x2 DID with Covariates

$$ATT = E(\Delta y_i | D_i = 1) - \hat{E}(\gamma_i | D_i = 1)$$

**Re-weighted Approach**

$$S1 : P(D_i = 1|X) = F(X) \rightarrow \hat{\pi}(X)$$

$$S2 : \omega(x) = \frac{\hat{\pi}(X)}{1 - \hat{\pi}(X)}$$

$$S3 : \widehat{ATT}^{1}_{ipw} = E(\Delta y_i | D_i = 1) - \frac{E(\omega(x)\Delta y_i | D_i = 0)}{E(\omega(x)|D_i = 0)}$$

$$S3 : \widehat{ATT}^{2}_{ipw} = E(\Delta y_i | D_i = 1) - \frac{E(\omega(x)\Delta y_i | D_i = 0)}{E(D_i)/(1 - E(D_i))}$$

Depends strongly on the correct specification of the propensity score.

# 2x2 DID with Covariates

$$ATT = E(\Delta y_i | D_i = 1) - \hat{E}(\gamma_i | D_i = 1)$$

**Doubly Robust OPT 1**

$$S1 : P(D_i = 1|X) = \hat{\pi}(X) \rightarrow \omega(x) = \frac{\hat{\pi}(X)}{1 - \hat{\pi}(X)}$$

$$S2 : \gamma_i = \Delta y_i = \gamma_\omega(x) + v_i \quad \forall i | D_i = 0 \quad weighted \quad by \quad \omega(x)$$

$$S3 : \widehat{ATT}^1_{DR} = E(\Delta y_i | D_i = 1) - E(\hat{\gamma}_\omega(x) | D_i = 1)$$

Both Doubly Robust estimators are consistent if either the outcome model or the propensity score is correctly specified

# 2x2 DID with Covariates

$$ATT = E(\Delta y_i | D_i = 1) - \hat{E}(\gamma_i | D_i = 1)$$

**Doubly Robust OPT 2**

$$S1 : P(D_i = 1|X) = \hat{\pi}(X) \rightarrow \omega(x) = \frac{\hat{\pi}(X)}{1 - \hat{\pi}(X)}$$

$$S2 : \gamma_i = \Delta y_i = \gamma(x) + v_i \quad \forall i | D_i = 0$$

$$S3 : \widehat{ATT}^2_{DR} = E(\Delta y_i | D_i = 1)$$

$$- E(\hat{\gamma}(x)|D_i = 1) - \frac{E\Big(\omega(x)(\Delta y_i - \hat{\gamma}(x))|D_i = 0\Big)}{E\Big(\omega(x)|D_i = 0\Big)}$$

# Table of Contents

# drdid: Doubly Robust DID estimator

```
# Command is available from ssc
ssc install drdid, replace
# General Syntax
drdid y x1 x2 ... [if/in] [w=.], [ivar(pid)]  time(tmt) treatment(trt) //
        [method] [SE CI options]
# ivar -> Panel, Otherwise RC. tmt and trt Should be Binary
# "Method" Options:
drimp [rc1]   : ATT_DR  OPT 1: p(trt=1|X)~IPT
dripw [rc1]   : ATT_DR  OPT 2: p(trt=1|X)~Logit
stipw         : ATT_IPW OPT 1
ipw           : ATT_IPW OPT 2
reg           : ATT_OR
all           : Estimates all options
```

# drdid: Doubly Robust DID estimator

```
# "SE/CI" options: Default Asymptotic Standard Errors
cluster(cvar)    : Clustered Standard errors
wboot[(opt)]     : Wildbootstrap:
   reps()        : Repetitions, Default 999
   wbtype()      : WBtype: mammen / rademacher
rseed()          : Seed for replication
gmm              : Estimation via gmm.
level()          : Level of Confidence. Default 95
# Other options
noisily          : Shows all intermediate steps

## Additional utilities
# Predicts IPWeights or Pscores
drdid_predict {newvar}, [weight pscore]
# Displays Intermediate Steps
drdid_display, bmatrix(name) vmatrix(name)
```

# Example

```
use https://friosavila.github.io/playingwithstata/drdid/lalonde.dta, clear

. drdid re age educ black married nodegree hisp re74 if treated==0 | sample==2,
ivar(id) time(year) tr(experimental) dripw

Doubly robust difference-in-differences                  Number of obs = 32,834
Outcome model  : least squares
Treatment model: inverse probability
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
experimental |
   (1 vs 0)  |  -871.3271   396.0211    -2.20   0.028    -1647.514   -95.14007
------------------------------------------------------------------------------
```

# Example

```
drdid re age educ black married nodegree hisp re74 if
treated==0 | sample==2, cluster(id) time(year) tr(experimental) drimp
# output Omitted
drdid_predict wgt, weight

tabstat age educ black married nodegree hisp re74  if
treated==0 | sample==2 [w=wgt], by(experimental)

exper |       age      educ     black   married  nodegree      hisp      re74
------+------------------------------------------------------------------------
    0 |  24.44706  10.18824        .8  .1576471  .8141176  .1129412  3672.485
    1 |  24.44706  10.18824        .8  .1576471  .8141176  .1129412  3672.485
------+------------------------------------------------------------------------
Total |  24.44706  10.18824        .8  .1576471  .8141176  .1129412  3672.485
-------------------------------------------------------------------------------
```

# Table of Contents

# GxT and the TWFE

As mentioned before, empirical research usually differs from the canonical approach: Many periods (T) and treatment at different times (G).

The common approach is to estimate this model using a TWFE:

$$y_{it} = \alpha_i + \gamma_t + \theta^{TWFE} * p_{it} + u_{it}$$

$$p_{it} = 1 \quad if \ already \ treated$$

This would capture the ATT only if the effect is homogeneous across units.

# Limitations of TWFE

There are two ways of understanding the limitations of the simple TWFE estimator.

**Negative Weights** Assume a balanced panel and no controls:

$$\hat{\theta}^{TWFE} = \frac{\sum \tilde{P}_{it} y_{it}}{\sum \tilde{P}_{it}^2} \quad ; \tilde{P}_{it} = P_{it} + \bar{P} - (\bar{P}_i + \bar{P}_t)$$

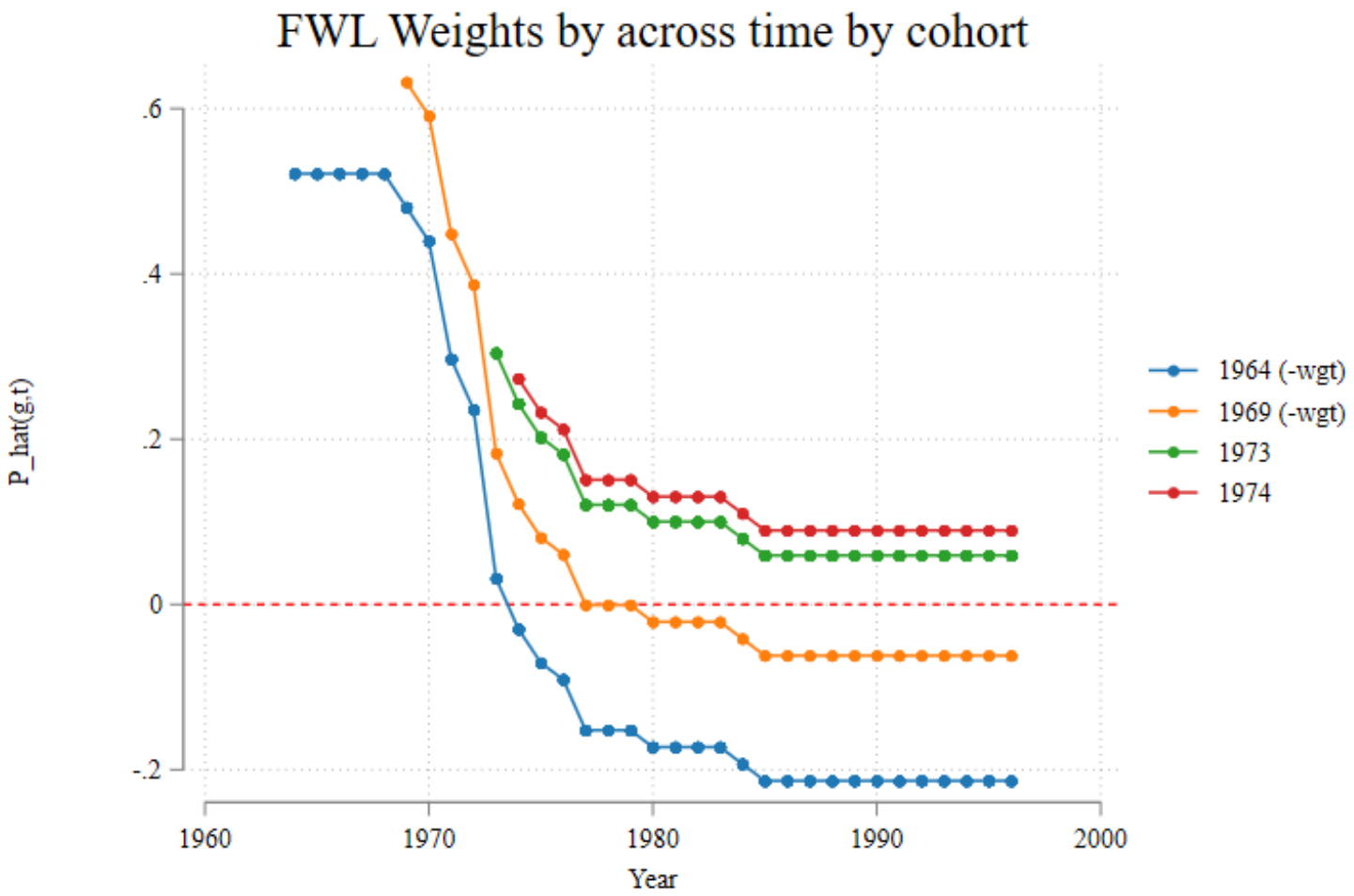
Treated units $P_{it} = 1$ should always receive a positive weight.
however, $\tilde{P}_{it}$ could be negative because:

- $\bar{P}_t$ is larger at later periods
- $\bar{P}_i$ is larger for units treated earlier

Thus, $\tilde{P}_{it}$ could be negative for units treated earlier, but seen at later periods.

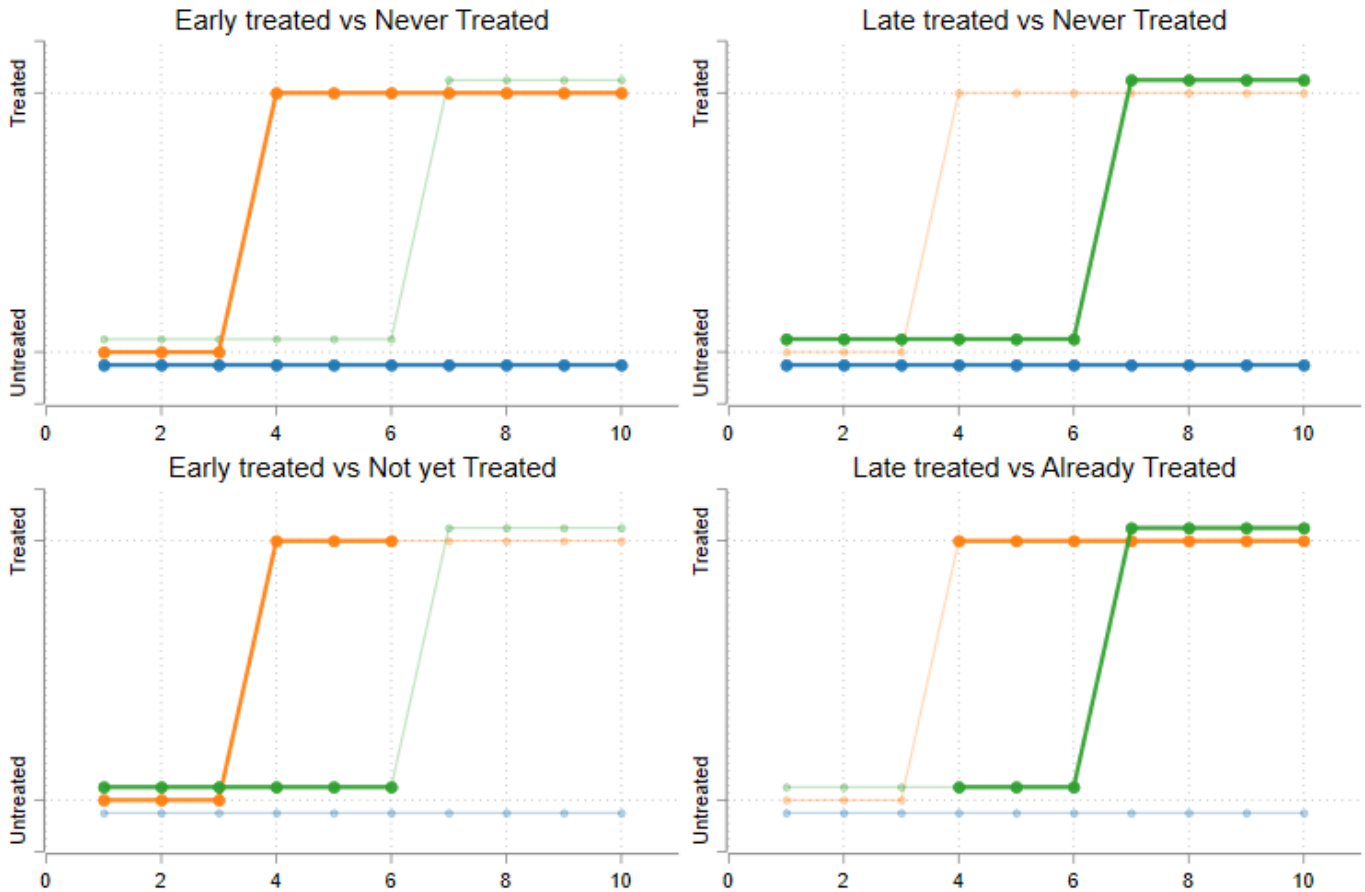
# Negative Weights

`Data:Bacon_example.dta`



# Good vs Bad Control groups

Goodman-Bacon (2021): TWFE estimator can be described as a weighted average of all 2x2 DID designs given the variation in the treatment status. But not all are good designs:



# Solution: a more flexible estimator

The "problem" with TWFE is how LR computes coefficients.

The alternative is to use a more flexible specification, or/and avoiding "bad" control groups.

There are many options, but two are the easiest to understand:
Opt1: Sun and Abraham (2021) and Wooldridge (2021)

$$y_{it} = \alpha_i + \gamma_t + \sum_{g=g_0}^{G} \sum_{s=g}^{T} \theta(g,s) * 1(G=g, t=s) + e$$

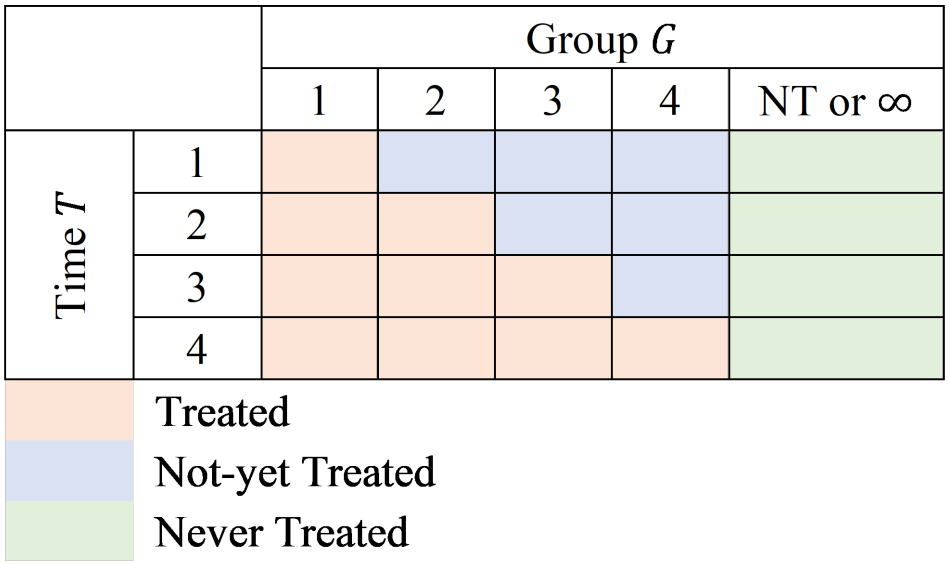
Opt2: Callaway and Sant'Anna (2021)

Breakdown a single GxT DID to multiple 2x2 DID.

One can aggregate ATT's as needed

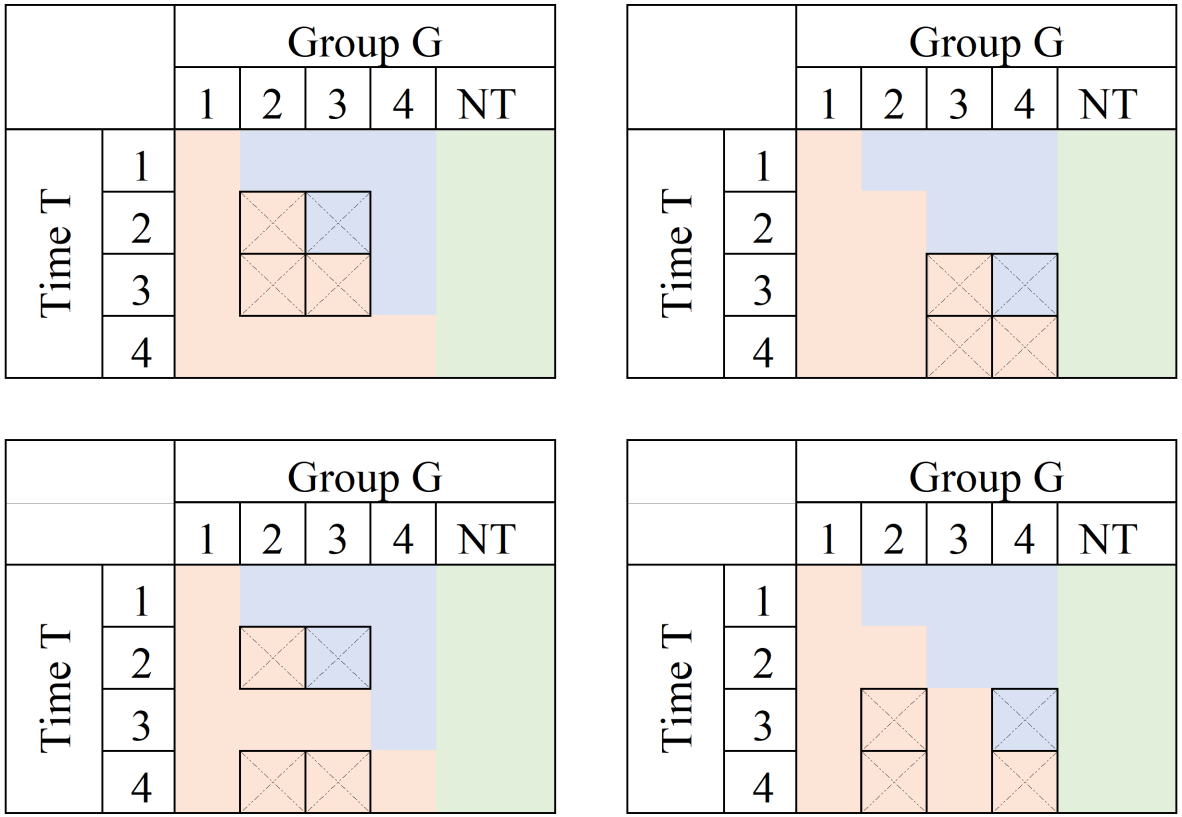
# DID with Multiple Time Periods

- CS breaks down the problem from one GxT DID to many 2x2 DID, using only good designs.
- Each design estimates a particular ATT(G,T).
- `drdid` can be applied on each design independently.



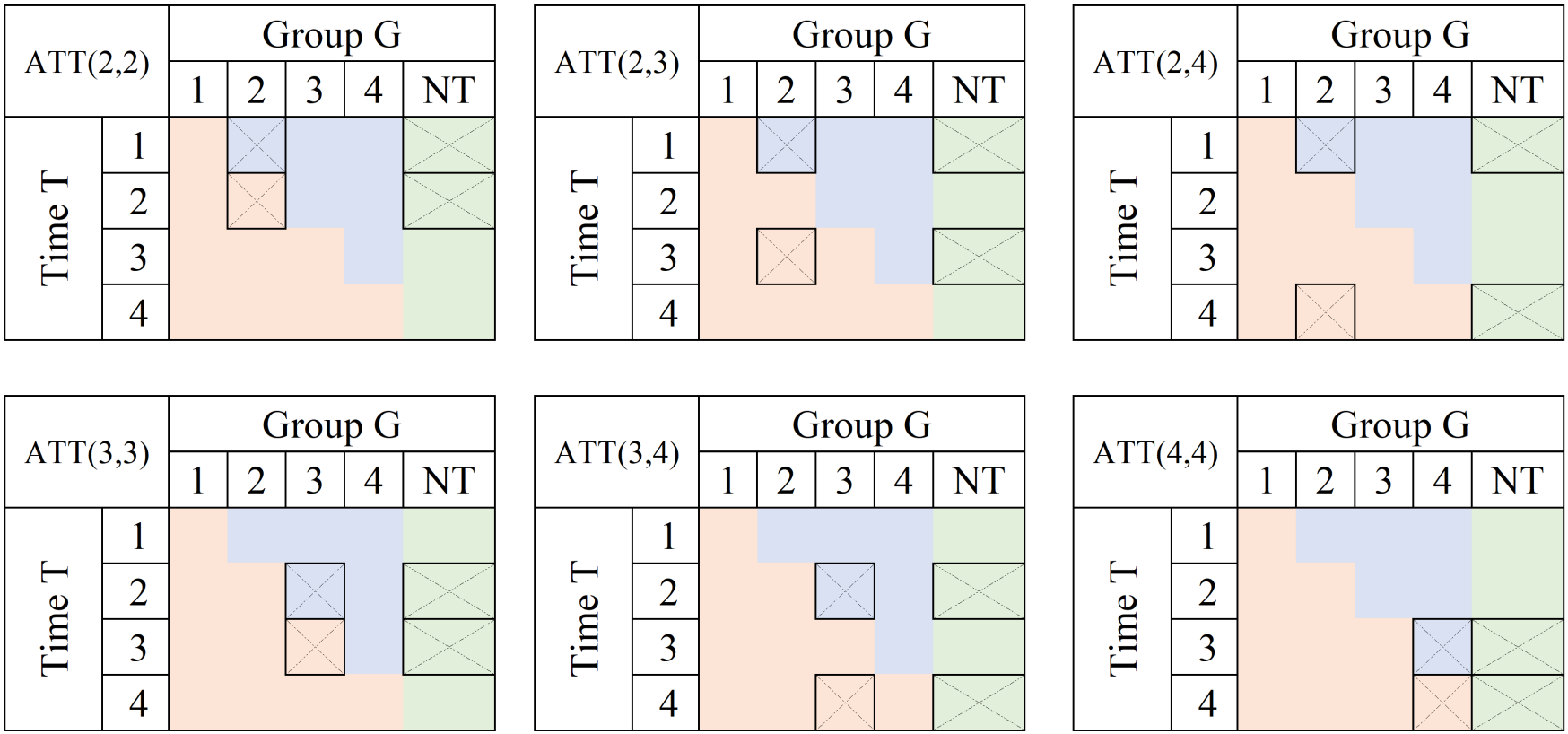
# DID with Multiple Time Periods

Figure: 2x2 to avoid



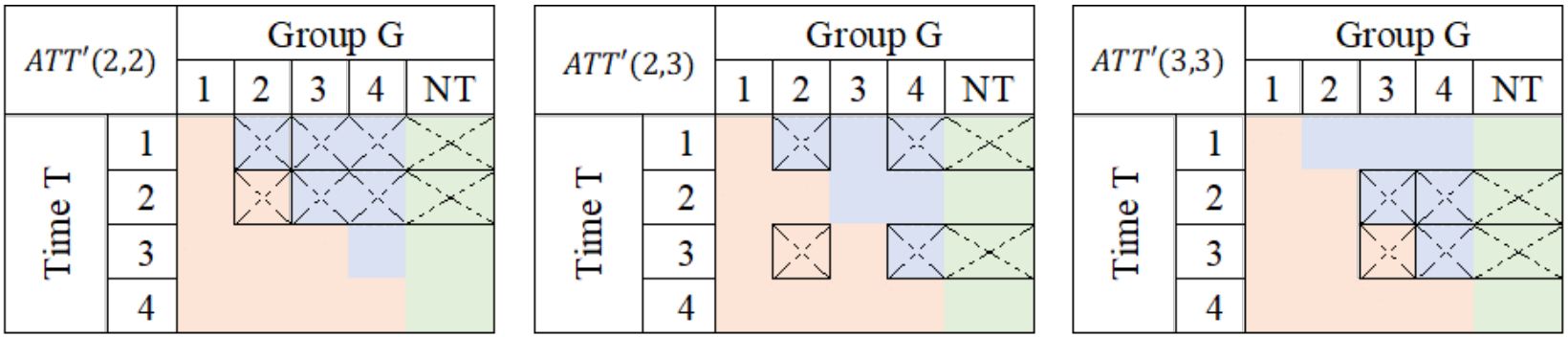
# DID with Multiple Time Periods

Figure: 2x2 Using Never Treated

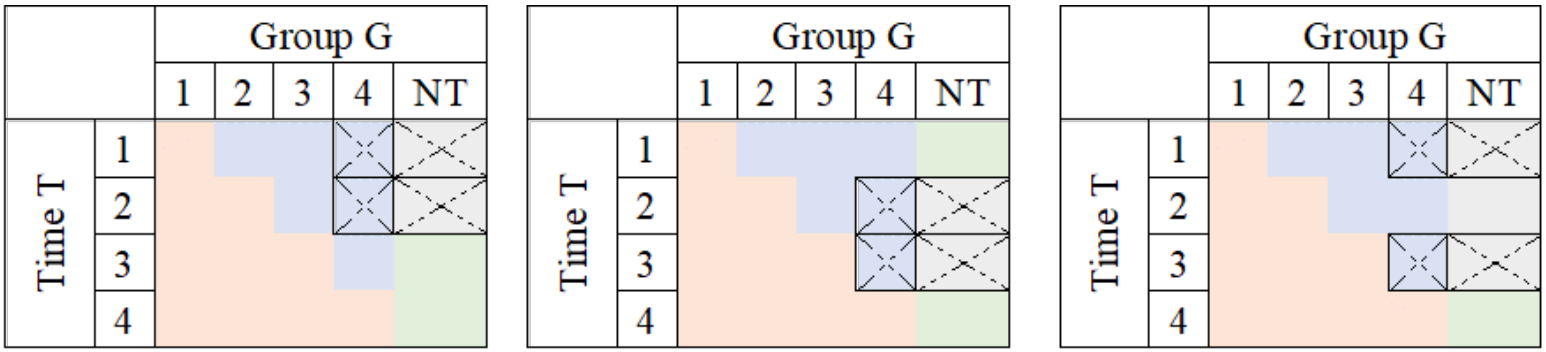


# DID with Multiple Time Periods

## 2x2 Using Not-yet Treated




## 2x2 Pre-treatment trends



# DID with Multiple Time Periods: ATT(g,t)

In CS, the main Building block is the ATT(g,t):

**ATT(g,t)**

The treatment effect for units treated at time $g$, measured at time $t$.

If $g \leq t$, it is useful to estimate treatment effects

$$ATT^{NT/NY}(g,t) = E(y_{i,t} - y_{i,g-1}|G_i = g) - E(y_{i,t} - y_{i,g-1}|G_i = NT \, or \, G_i > t)$$

If $g > t$, it is useful for pre-trend tests:

$$ATT^{S}(g,t) = E(y_{i,t} - y_{i,t-1}|G_i = g) - E(y_{i,t} - y_{i,t-1}|G_i = NT)$$

$$ATT^{L}(g,t) = E(y_{i,g-1} - y_{i,t-1}|G_i = g) - E(y_{i,g-1} - y_{i,t-1}|G_i = NT)$$

# Aggregating ATT(g,t)'s

CS identifies all feasible ATT(g,t)'s, but also suggests a series of possible aggregations.

General Structure:

$$ATT_{TYP} = \frac{\sum_{TYP} w_{g,t} ATT(g,t)}{\sum_{TYP} w_{g,t}}$$

where $w_{g,t}$ are based on the number of treated observations used in a particular ATT(g,t).

$$Simple : t \geq g$$
$$Group : t \geq g \& g = h$$
$$Calendar : t \geq g \& t = s$$
$$Event : t - g = e$$
$$CEvent : c_1 \leq t - g \leq c_2$$

# Table of Contents

# csdid: DID with Multiple Time periods

```
# Command is available from ssc, requires drdid
ssc install csdid, replace
# General Syntax
csdid y x1 x2 ... [if/in] [w=.], [ivar(pid)]  time(T) gvar(G) //
        [method(@@@)] [long] [SE CI options]
# G=0 Never treated, and G subset of T (Not a 0-1 dummy)
# method(@@) options:
drimp [rc1]  (default) ; dripw [rc1]
stdipw ; ipw ; reg
# Pretreatment ATT
long          : Long Gaps for Pretreatment ATT(g,t)
# SE option
pointwise     : If "wboot" were requested, one can call for pointwise CI
                Default is uniform CI (valid for joint tests)
# One can save the RIF's into a file to save results.
saverif(file) [replace] : Useful for estimation of "wboot"
                          SE on other Aggregations
```

# csdid: DID with Multiple Time periods

```
# Post-Estimation:
after csdid    : estat        [agg pretrend], [option]
using rif_file: csdid_stats [agg ], [option SE options]
# Possible aggregations
attgt    , [post estore(@@) esave(@@) ]
simple   , [post estore(@@) esave(@@) ]
calendar, [post estore(@@) esave(@@) ]
group    , [post estore(@@) esave(@@) ]
event    , [window(%1 %2)] [post estore(@@) esave(@@) ]
cevent   ,  window(%1 %2)  [post estore(@@) esave(@@) ]

# Visualization (after csdid, estat or csdid_stat),
csdid_plot, [options]: Plots previously estimated ATTGT's
```

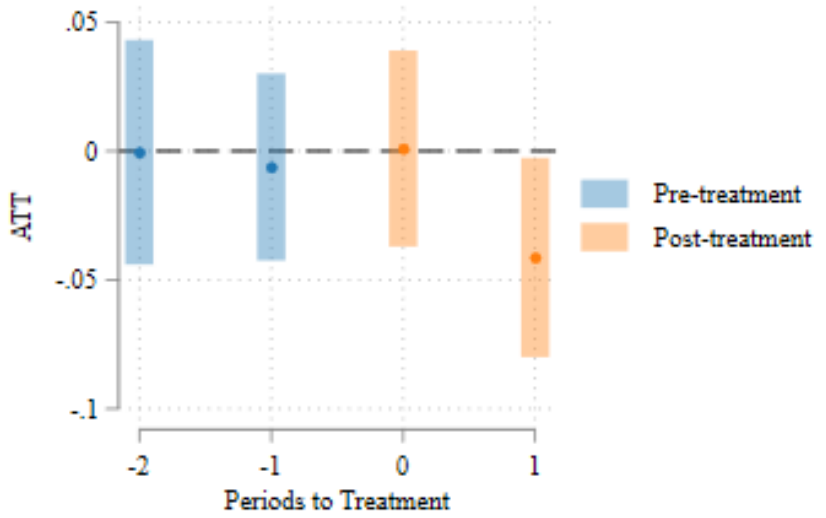
# csdid: Example

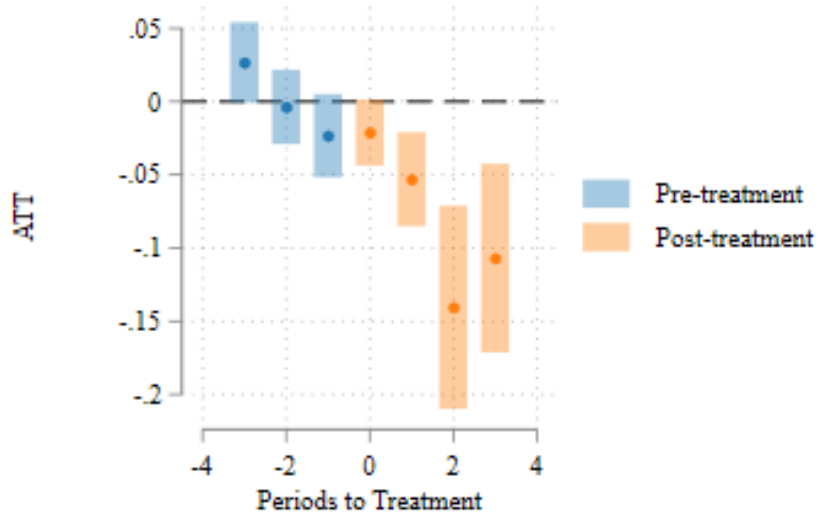
```
use https://friosavila.github.io/playingwithstata/drdid/mpdta.dta, clear
tab year first_treat
           |                first.treat
      year |         0      2004      2006      2007 |     Total
-----------+--------------------------------------------+----------
      2003 |       309        20        40       131 |       500
      2004 |       309        20        40       131 |       500
      2005 |       309        20        40       131 |       500
      2006 |       309        20        40       131 |       500
      2007 |       309        20        40       131 |       500
-----------+--------------------------------------------+----------
     Total |     1,545       100       200       655 |     2,500

# DID estimation, using Panel estimators, and DRIPW method.
# Asymptotic Standard errors
 csdid lemp lpop , ivar(countyreal) time(year) gvar(first_treat) method(dripw)
```

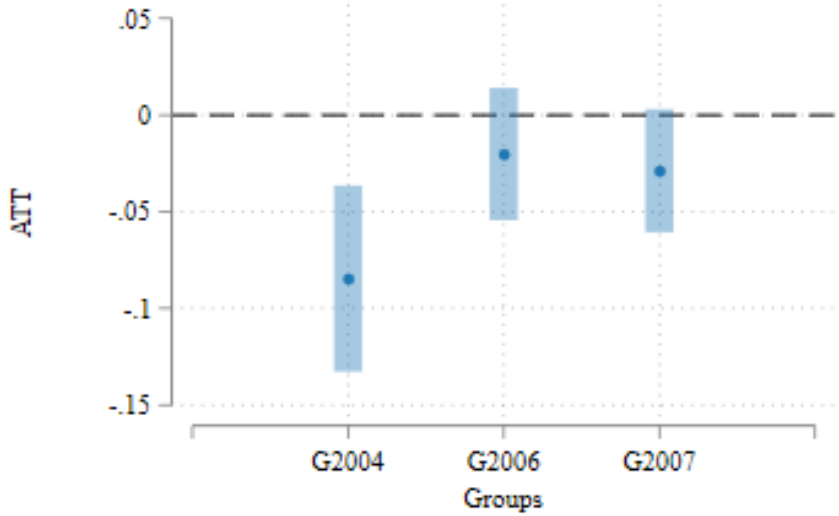
```
 csdid lemp lpop , ivar(countyreal) time(year) gvar(first_treat) method(dripw)
 saverif(rif)
File rif.dta will be used to save all RIFs
.......... # Each dot is a successful 2x2 estimation. X Otherwise
file rif.dta saved
Difference-in-difference with Multiple Time Periods # Type of Model
Outcome model  : least squares
Treatment model: inverse probability
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
g2004        | # Indicates the Cohort
 t_2003_2004 |  -.0145297   .0221292    -0.66   0.511     -.057902    .0288427
 t_2003_2005 |  -.0764219   .0286713    -2.67   0.008    -.1326166   -.0202271
 t_2003_2006 |  -.1404483   .0353782    -3.97   0.000    -.2097882   -.0711084
 t_2003_2007 |  -.1069039   .0328865    -3.25   0.001    -.1713602   -.0424476
-------------+----------------------------------------------------------------
# t_#1_#2: Indicate which period was used as Pre (#1) and  Post (#2) Periods
# Ommited Output for g2006 & g2007

 Control: Never Treated
See Callaway and Sant'Anna (2020) for details
```

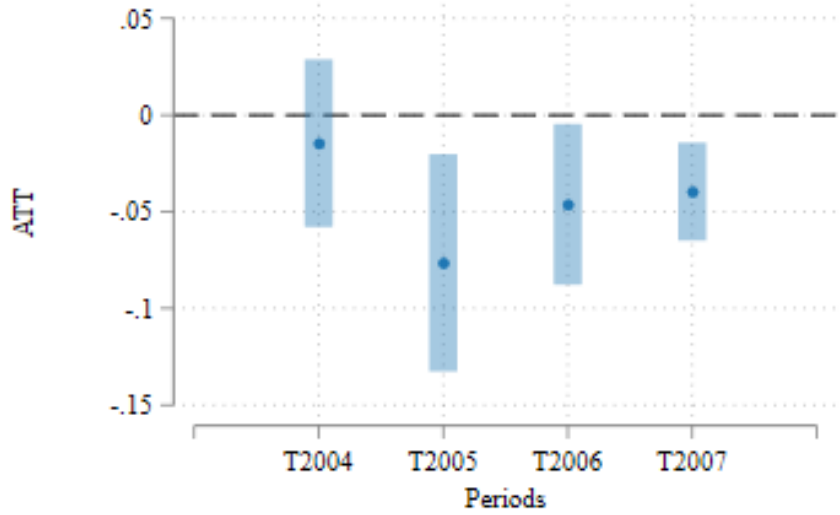
```
estat cevent, window(0 2)
ATT for events between 0 2 # Average of ATT from T+0 to T+2
Event Study:Aggregate effects
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ATTC |  -.0369435    .010938    -3.38   0.001    -.0583815   -.0155055
------------------------------------------------------------------------------

estat event
ATT by Periods Before and After treatment
Event Study:Dynamic effects
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         Tm3 |   .0267278   .0140657     1.90   0.057    -.0008404     .054296
         Tm2 |  -.0036165   .0129283    -0.28   0.780    -.0289555    .0217226
         Tm1 |   -.023244   .0144851    -1.60   0.109    -.0516343    .0051463
         Tp0 |  -.0210604   .0114942    -1.83   0.067    -.0435886    .0014679
         Tp1 |  -.0530032   .0163465    -3.24   0.001    -.0850417   -.0209647
         Tp2 |  -.1404483   .0353782    -3.97   0.000    -.2097882   -.0711084
         Tp3 |  -.1069039   .0328865    -3.25   0.001    -.1713602   -.0424476
------------------------------------------------------------------------------
```

## csdid_plot


.05
0
-.05
-.1
ATT
-2
-1
0
1
Periods to Treatment
Pre-treatment
Post-treatment


.05
0
-.05
-.1
-.15
-.2
ATT
-4
-2
0
2
4
Periods to Treatment
Pre-treatment
Post-treatment


.05
0
-.05
-.1
-.15
ATT
G2004
G2006
G2007
Groups


.05
0
-.05
-.1
-.15
ATT
T2004
T2005
T2006
T2007
Periods

# Table of Contents

# Conclusions

- The use of DID estimators of ATT's using DID has changed drastically over the last couple of years.
- Today I showed commands to implement two of these estimators:
- `drdid` implements Sant'Anna and Zhao (2020) estimator, which emphasizes the benefits of doubly robust DID estimators.
- `csdid` implements Callaway and Sant'Anna (2021), which proposes a strategy to identify and aggregate the treatment effects for GxT DID.

For other examples, discussions, and explanations check https://friosavila.github.io/playingwithstata/index.html

# Table of Contents

# References

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

Borusyak, Kirill and Xavier Jaravel, "Revisiting Event Study Designs," SSRN Scholarly Paper ID 2826228, Social Science Research Network, Rochester, NYAugust 2017.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," Journal of Econometrics, 2021, 225 (2), 200-230,

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," American Economic Review, 2020,110(9), 2964–2996.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," Journal of Econometrics, 2021, 225(2), 254-277.

# References

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.

Rios-Avila, Fernando, Pedro H. C. Sant'Anna, Brantly Callaway, and Asjad Naqvi "csdid and drdid: Doubly Robust Differences-in-Differences with Multiple Time Periods," Working Paper, 2021.

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

Sun, Liyan and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," Journal of Econometrics,2021, 225(2), 175-199.

Wooldridge, Jeffrey M., Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators (August 17, 2021). Available at SSRN: https://ssrn.com/abstract=3906345

# Table of Contents

# RC estimators

For next slides, I use the following notation:

$$E(y|D=d,T=t)=E_{d,t}(y)$$

where $y$ is the observed outcome for an unit.
$D=1$ if a unit is in the treated group.
$T=1$ if we are considering the Post-period.
Thus, $E_{1,0}(y)$ is the average observed outcome for the treated group, pre-treatment (T=0).
and the ATT estimator will be:

# RC Estimators: Regression Outcome

S1: Estimate 2 outcome models:

$$y = \theta_0(x) + e \; if \; D = 0 \; and \; T = 0$$

$$y = \theta_1(x) + e \; if \; D = 0 \; and \; T = 1$$

$$\hat{\gamma}(x) = \hat{\theta}_1(x) - \hat{\theta}_0(x)$$

S2: ATT:

$$ATT_{or} = E_{1,1}(y) - E_{1,0}(y) - E(\hat{\gamma}(x)|D = 1)$$

# RC Estimators: IPW1

S1: Estimate Pscore

$$P(D=1|X) = F(X) \rightarrow \hat{\pi}(x)$$

S2: Estimation of weights

$$\omega(x) = \frac{\hat{\pi}(x)}{1-\hat{\pi}(x)}$$

S3: ATT

$$ATT^1_{IPW} = E_{1,1}(y) - E_{1,0}(y) - \left( \frac{E_{0,1}(\omega(x)y)}{E_{0,1}(\omega(x))} - \frac{E_{0,0}(\omega(x)y)}{E_{0,0}(\omega(x))} \right)$$

# RC Estimators: IPW2

S1: Estimate Pscore $\hat{\pi}(x)$ and Weights $\omega(x)$
S2: Define $\pi = E(D)$ and $\tau = E(t)$
S3: Expected growth absent of treatment

$$\hat{E}(\gamma(x)|D=1) = \frac{E_{0,1}(\omega(x)y)}{\pi/(1-\pi)} \frac{E(t|D=0)}{\tau} - \frac{E_{0,0}(\omega(x)y)}{\pi/(1-\pi)} \frac{1-E(t|D=0)}{1-\tau}$$

S4: ATT

$$ATT^2_{IPW} = E_{1,1}(y)\frac{E(D\,t)}{\pi\tau} - E_{1,0}(y)\frac{E(D(1-t))}{\pi(1-\tau)} - E(\gamma(x)|D=1)$$

# RC Estimators: DR1

S1: Estimate Pscore $\pi(x)$ and weights $\omega(x)$
S2: Estimation of Outcome Regressions using WLS with $\omega(x)$ for the control group $(D = 0)$:

$$\theta^w_{0,1}(x)\, if\ t = 1\ and\ \theta^w_{0,0}(x)\, if\ t = 0$$

S3: Estimate ATT as:

$$ATT^1_{DR} = E_{1,1}(y - \hat{\theta}^w_{0,1}(x)) - E_{1,0}(y - \hat{\theta}^w_{0,0}(x)) -$$
$$\left( \frac{E_{0,1}(\omega(x)(y - \hat{\theta}^w_{0,1}(x))}{E_{0,1}(\omega(x))} - \frac{E_{0,0}(\omega(x)(y - \hat{\theta}^w_{0,0}(x))}{E_{0,0}(\omega(x))} \right)$$

# RC Estimators: DR2

S1: Estimate Pscore $\pi(x)$ and weights $\omega(x)$
S2: Estimation of Outcome Regressions using WLS with $\omega(x)$ for the control group ($D = 0$). And OR using OLS for the treated group ($D = 1$):

$$\theta^w_{0,1}(x); \theta^w_{0,0}(x)$$

$$\theta_{1,1}(x); \theta_{1,0}(x)$$

S3: Estimate ATT as

$$ATT^2_{DR} = E(\hat{\theta}_{1,1}(x) - \hat{\theta}_{1,0}(x) - (\hat{\theta}^w_{0,1}(x) - \hat{\theta}^w_{0,0}(x))|D = 1)$$
$$- \left( \frac{E_{0,1}(\omega(x)(y - \hat{\theta}^w_{0,1}(x))}{E_{0,1}(\omega(x))} - \frac{E_{0,0}(\omega(x)(y - \hat{\theta}^w_{0,0}(x))}{E_{0,0}(\omega(x))} \right)$$