

# Module 4: Difference-in-Differences and Effects of Medicaid Expansion

## Part 4: Even More Difference-in-Differences

---

Ian McCarthy | Emory University  
Econ 470 & HLTH 470

# Problem with TWFE

---

# Problem in an Equation

Consider standard TWFE specification with a single treatment coefficient,

$$y_{it} = \alpha + \delta D_{it} + \gamma_i + \gamma_t + \varepsilon_{it}.$$

We can decompose  $\hat{\delta}$  into three things:

$$\hat{\delta}_{twfe} = VWATT + VWPT - \Delta ATT$$

1. A variance-weighted ATT
2. Violation of parallel trends
3. Heterogeneous effects over time

# Problem with words

- Best case: Variance-weighted ATT
- Differential timing **alone** can introduce bias because already treated act as controls for later treated groups (when seeking single regression coefficient)
- Heterogeneity and differential timing introduces "contamination" via negative weights assigned to some underlying 2x2 DDs

# Solution

Only consider "clean" comparisons:

- Separate event study for each treatment group vs never-treated or not-yet-treated
- Callaway and Sant'Anna (2020)
- Sun and Abraham (2020)
- de Chaisemartin and D'Haultfoeuille (2020)
- Stacking regression: Cengiz et al. (2019)
- Imputation: Gardner (2021), and Borusyak et al. (2021)

# Changing mindset for estimation

1. Define target parameter (e.g., ATT)...this is pretty new as a starting point
2. Identification
3. Estimation
4. Aggregation
5. Inference

# An aside on covariates and DD

---

# Incorporating covariates

- "Easy" to do in regression setting, but risks of using outcomes as controls
- Two general ways:
  1. Outcome regression (imputation-based)
  2. Propensity score



# Outcome regression (Heckman et al 1997)

$$\hat{\delta}^{reg} = E[Y_{t=1}|D = 1] - \left[ E[Y_{t=0}|D = 1] + \frac{1}{n^T} \sum_{i \in N_{d=1}} \left( \hat{\mu}_{d=0,t=1}(X_i) - \hat{\mu}_{d=0,t=0}(X_i) \right) \right],$$

where  $\hat{\mu}_{d,t}$  is the prediction from a regression among the untreated group using baseline covariates.

- Heckman forms prediction as regression of  $\Delta Y$  on  $X_i$  among untreated group, although could also consider separate regressions on levels
- Conceptually...take observed value among treatment group in post-period, subtract pre-period value and the predicted trend

# IPW (Abadie, 2005)

$$\hat{\delta}^{ipw} = E \left[ \frac{D - \hat{p}(D = 1|X)}{1 - \hat{p}(D = 1|X)} \frac{Y_{t=1} - Y_{t=0}}{P(D = 1)} \right]$$

- $Y_{t=1}$  is the observed outcome at time  $t = 1$ , and similarly for  $Y_{t=0}$
- $\hat{p}$  denotes the estimated propensity score from regression of  $D$  on  $X$  in pre-period
- Conceptually...upweight change among treated that look a lot like the control group, downweight change among treated that look different than controls

# DR (Sant'Anna and Zhou)

$$\hat{\delta}^{dr} = E \left[ \left( \frac{D}{P(D=1)} - \frac{\frac{\hat{p}(X)(1-D)}{1-\hat{p}(X)}}{E \left[ \frac{\hat{p}(X)(1-D)}{1-\hat{p}(X)} \right]} \right) (E[Y_{t=1}|D=1] - E[Y_{t=0}|D=1] - \Delta \hat{\mu}_0(X)) \right]$$

- Notice how this combines Heckman's outcome regression in the second part and Abadie's IPW in the first part

# The "New" DD

# The New DD

I'll organize this into three types of estimators:

1. GT
2. Stacked
3. Imputation

# GT1: Callaway and Sant'Anna

---

# CS Estimator (Conceptually)

- "Manually" estimate group-specific treatment effects for each period
- Each estimate is propensity-score weighted
- Aggregate the treatment effect estimates (by time, group, or both)

# CS Estimator (More Formally)

Group-specific treatment effects:

$$ATT(g, t) = E[Y_{1,t} - Y_{0,t} | G_g = 1],$$

where  $G$  denotes all feasible groups or cohorts (e.g.,  $g = 1$  could mean states expanding in 2014,  $g = 2$  denotes states expanding in 2015, etc.)



# CS Estimator (More Formally)

With this  $(g, t)$  notation, CS show:

$$ATT(g, t; \tau)^{DR} = E \left[ \left( \frac{G_g}{P(G_g = 1)} - \frac{\frac{\hat{p}_g(X)C}{1-\hat{p}_g(X)}}{E \left[ \frac{\hat{p}_g(X)C}{1-\hat{p}_g(X)} \right]} \right) \left( E[Y_t | G_g = 1] - E[Y_{g-\tau-1} | G_g = 1] - \Delta \hat{\mu}_{g,t,\tau}(X) \right) \right]$$

- $\tau$  denotes time from treatment, such that  $Y_{g-\tau-1}$  denotes the outcome for some reference time period,  $t = g - \tau - 1$
- $\Delta \hat{\mu}$  captures the predicted change from an outcome regression
- $\hat{p}_g$  denotes the predicted probability of being in the treatment cohort  $g$

# CS Estimator (More Formally)

- CS show a similar version of their estimator using a "not-yet-treated" control group rather than a never-treated.
- Different versions include...
  - "regression" based: drop the propensity score part
  - "IPW": drop  $\Delta\hat{\mu}$
- Only time-invariant covariates allowed

# CS Estimator (More Formally)

Finally, aggregate all of the  $(g, t)$  treatment effects:

$$\hat{\delta} = \sum_{g \in \mathcal{G}} \sum_{t=2}^{\mathcal{T}} w(g, t) \times ATT(g, t)$$

# CS Estimator (in Practice)

## Stata

```
ssc install csdid
ssc install event_plot
ssc install drdid

insheet using "data/acs_medicaid.txt", clear
gen perc_unins=uninsured/adult_pop
egen stategroup=group(state)
drop if expand_ever=="NA"
replace expand_year="0" if expand_year=="NA"
destring expand_year, replace

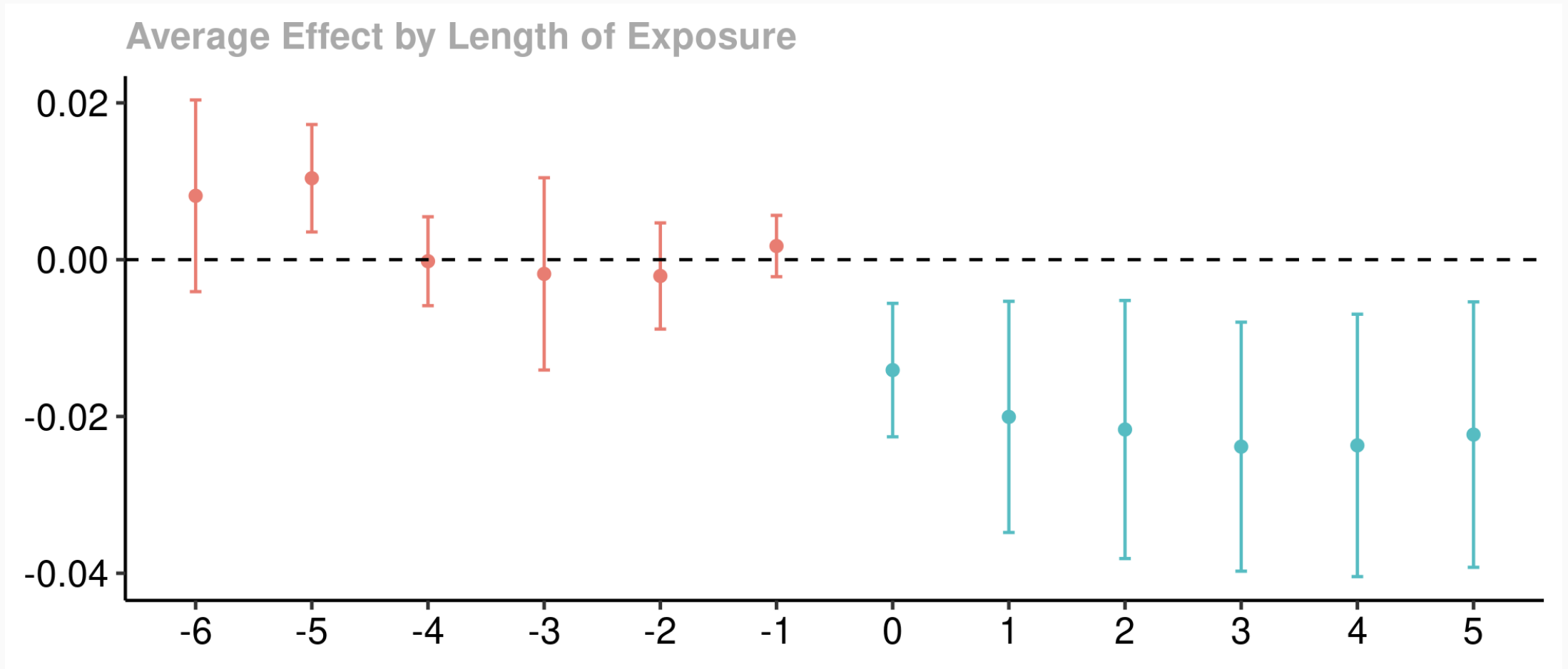
csdid perc_unins, ivar(stategroup) time(year) gvar(expan
estat event, estore(cs)
event_plot cs, default_look graph_opt(xtitle("Periods si
```

## R

```
library(tidyverse)
library(did)
library(DRDID)
mcaid.data <- read_tsv("../data/acs_medicaid.txt")
reg.dat <- mcaid.data %>%
  filter(!is.na(expand_ever)) %>%
  mutate(perc_unins=uninsured/adult_pop,
         post = (year ≥ 2014),
         treat=post*expand_ever,
         expand_year=ifelse(is.na(expand_year),0,expand_
  filter(!is.na(perc_unins)) %>%
  group_by(State) %>%
  mutate(stategroup=cur_group_id()) %>% ungroup()

mod.cs <- att_gt(yname="perc_unins", tname="year", idnam
               gname="expand_year",
               data=reg.dat, panel=TRUE, est_method="d
               allow_unbalanced_panel=TRUE)
mod.cs.event <- aggte(mod.cs, type="dynamic")
```

# CS in Practice



# GT2: Sun and Abraham

---

# Sun and Abraham

- Standard event study problems:
  - coefficient estimates are potentially biased due to treatment/control group construction
  - i.e., "contamination" of individual  $\delta_\tau$  from other leads/lags
- Solution: Estimate fully interacted model

$$y_{it} = \gamma_i + \gamma_t + \sum_g \sum_{\tau \neq -1} \delta_{g\tau} \times \mathbf{1}(i \in C_g) \times D_{it}^\tau + \beta x_{it} + \epsilon_{it}$$

# Sun and Abraham

$$y_{it} = \gamma_i + \gamma_t + \sum_g \sum_{\tau \neq -1} \delta_{g\tau} \times \mathbf{1}(i \in C_g) \times D_{it}^\tau + \beta x_{it} + \epsilon_{it}$$

- Intuition: Standard regression with different event study specifications for each treatment group
- Aggregate  $\delta_{g\tau}$  for standard event study coefficients and overall ATT



# Sun and Abraham in Practice

## Stata

```
ssc install eventstudyinteract
ssc install avar
ssc install event_plot

insheet using "data/acs_medicaid.txt", clear
gen perc_unins=uninsured/adult_pop
drop if expand_ever=="NA"
egen stategroup=group(state)
replace expand_year="." if expand_year=="NA"
destring expand_year, replace
gen event_time=year-expand_year
gen nevertreated=(event_time=.)

forvalues l = 0/4 {
    gen L`l'event = (event_time=`l')
}
forvalues l = 1/2 {
    gen F`l'event = (event_time=-`l')
}
gen F3event=(event_time<=-3)
eventstudyinteract perc_unins F3event F2event L0event L1
```

## R

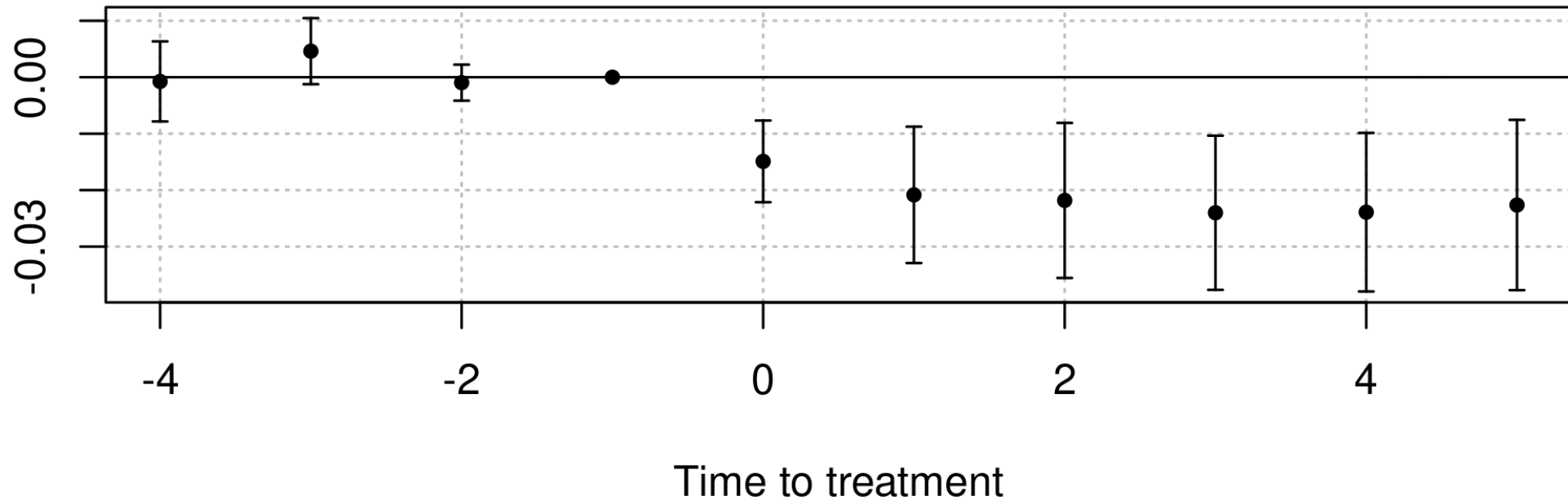
```
library(tidyverse)
library(modelsummary)
library(fixest)
mcaid.data <- read_tsv("../data/acs_medicaid.txt")
reg.dat <- mcaid.data %>%
  filter(!is.na(expand_ever)) %>%
  mutate(perc_unins=uninsured/adult_pop,
         post = (year >= 2014),
         treat=post*expand_ever,
         expand_year = ifelse(expand_ever==FALSE, 10000,
                              time_to_treat = ifelse(expand_ever==FALSE, -1,
                                                      time_to_treat = ifelse(time_to_treat < -4, -4,
                                                                              time_to_treat))
         )

mod.sa <- feols(perc_unins~sunab(expand_year, time_to_tr
                           cluster=~State,
                           data=reg.dat)
```

# Sun and Abraham in Practice

## SA Event study

Estimate and 95% Conf. Int.



# GT3: de Chaisemartin and D'Haultfoeuille (CH)

---

# CH

- More general than other approaches
- Considers "fuzzy" treatment (i.e., non-discrete treatment)
- Considers fixed effects and first-differencing
- Allows treatment to turn on and off (not allowed in CS or SA)

New paper from Callaway, Goodman-Bacon, and Sant'Anna also looks at DD with continuous treatment

# CH Approach

- Essentially a series of 2x2 comparisons
- Aggregates up to overall effects

# CH in Practice

## Stata

```
ssc install did_multiplgt
ssc install event_plot

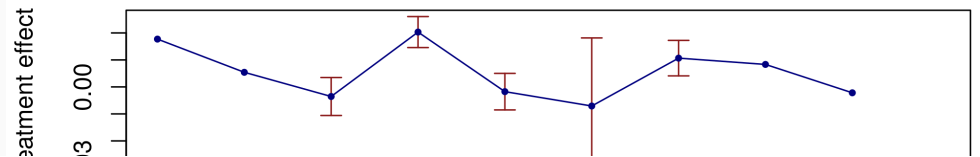
insheet using "data/acs_medicaid.txt", clear
gen perc_unins=uninsured/adult_pop
drop if expand_ever="NA"
egen stategroup=group(state)
replace expand_year="." if expand_year="NA"
destring expand_year, replace
gen event_time=year-expand_year
gen nevertreated=(event_time=.)
gen treat=(event_time ≥ 0 & event_time ≠ .)

did_multiplgt perc_unins stategroup year treat, robust_
event_plot e(estimates)#e(variances), default_look graph
title("de Chaisemartin and D'Haultfoeuille (2020)") xlab
```

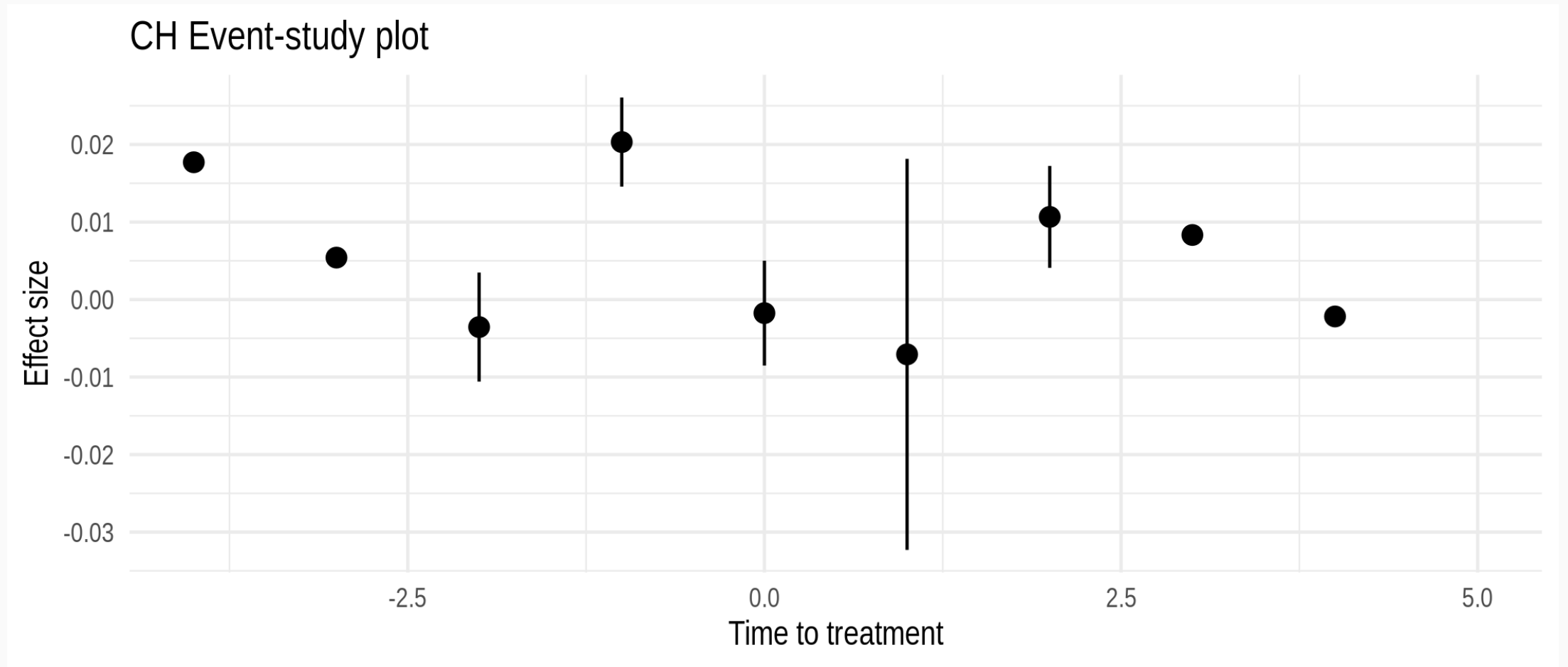
## R(not the same as in **Stata**)

```
library(DIDmultiplgt)
mcaid.data ← read_tsv("../data/acs_medicaid.txt")
reg.dat ← mcaid.data %>%
  filter(!is.na(expand_ever)) %>%
  mutate(perc_unins=uninsured/adult_pop,
         treat=case_when(
           expand_ever=FALSE ~ 0,
           expand_ever=TRUE & expand_year<year ~ 0,
           expand_ever=TRUE & expand_year ≥ year ~ 1))

mod.ch ← did_multiplgt(df=reg.dat, Y="perc_unins", G="
  placebo=4, dynamic=5, brep=50,
  parallel=TRUE, average_effect="
```



# CH in Practice



# CH in practice

Some barriers to this estimator in practice (at least, as implemented in `R` right now)

- Relatively slow
- Not user friendly
- Odd results



# Stacked regression

---

# Cengiz et al. (2019)

- "Stacked" event studies
- Estimate event study for every treatment group, using never-treated as controls
- Aggregate to overall average effects

# Cengiz et al. (2019)

1. Define event window,  $t \in [\kappa_a, \kappa_b]$  (e.g., 3 pre-periods and 5 post-periods,  $\kappa_a = 3$  and  $\kappa_b = 5$ )
2. Split the data into  $g = 1, \dots, G$  different "groups", as defined by treatment cohort, each with adoption date denoted by  $\omega_g$ 
  - observations outside of the  $[\omega_g - \kappa_a, \omega_g + \kappa_b]$  interval are dropped
3. Append (i.e., stack) each  $g$ th dataset
4. Run stacked event study allowing for different set of event study coefficients and fixed effects for every group  $g$

$$y_{itg} = \sum_{\tau=-\kappa_a}^{\kappa_b} \delta_{\tau} \times D_{ig} \times \mathbf{1}(t - \omega_g = \tau) + \gamma_{ig} + \gamma_{\tau g} + \varepsilon_{itg}$$

# Cengiz et al. (2019)

- Intuitively: run event study on every cohort,  $g$
- Control units (never treated or very late treated) will be duplicated over cohorts
- Need to cluster at the unit or unit/cohort level (probably unit level otherwise not accounting for duplication)
- **Alternative:** Among controls included in multiple cohorts, randomly assign them to one cohort

# Quick comparison

- Allows time-varying covariates
- Inference is less clear
- Likely estimating some variance-weighted ATT...not clear what those weights are anymore
- Seemingly stronger parallel trends assumptions for each cohort

# Imputation estimators

---

# Gardner (2021)

1. Estimate group and time fixed effects via first stage regression only among non-treated units
2. Predict outcome for all observations and residualize
3. Run standard event study specification on residualized outcome variable

Note: Estimate with GMM to account for first-stage prediction

# Gardner (2021) in Practice

## Stata

did2s

## R

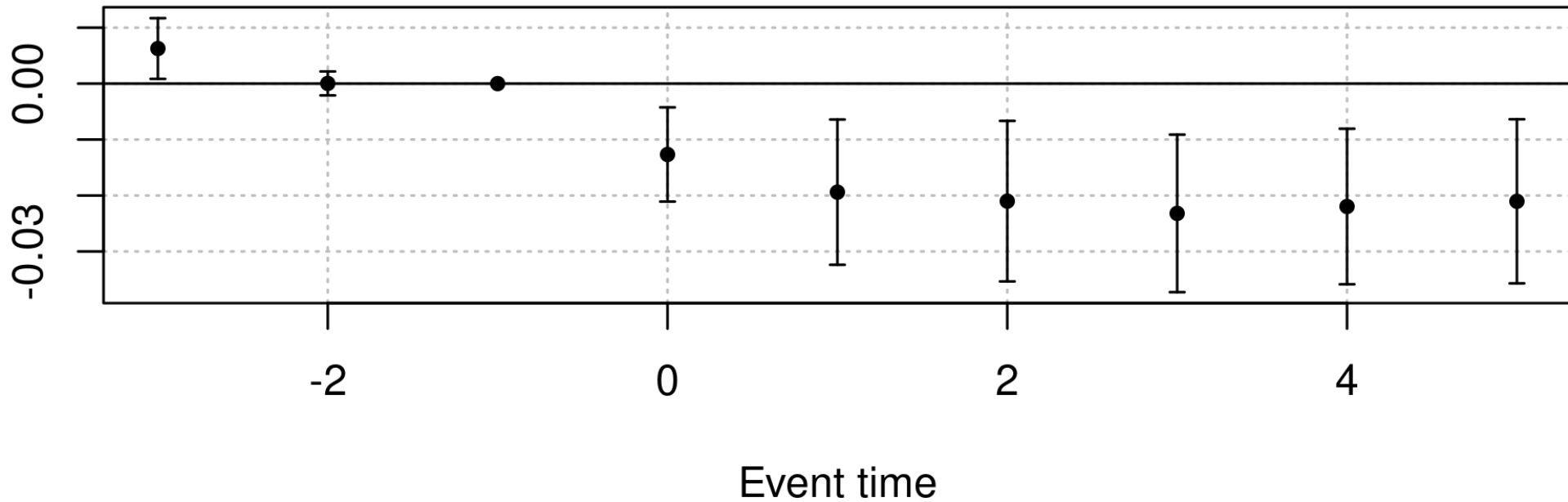
```
library(did2s)
mcaid.data ← read_tsv("../data/acs_medicaid.txt")
reg.dat ← mcaid.data %>%
  filter(!is.na(expand_ever)) %>%
  mutate(perc_unins=uninsured/adult_pop,
         post = (year ≥ 2014),
         treat=post*expand_ever,
         expand_year = ifelse(expand_ever==FALSE, 10000,
                              time_to_treat = ifelse(expand_ever==FALSE, -1,
                                                         time_to_treat = ifelse(time_to_treat < -3, -3,
                                                                                       0)))
mod.2s ← did2s(reg.dat, yname="perc_unins",
               treatment="treat",
               first_stage = ~ 0 | State + year,
               second_stage = ~i(time_to_treat, ref=-1)
               cluster_var="State")
```



# Gardner (2021) in Practice

Estimate and 95% Conf. Int.

## 2SDID Event Study



# Borusyak et al. (2021)

- Estimate regression only for untreated observations
- Predicted untreated outcome among the treated observations and take the difference
- Aggregate differences to form overall weighted average effect

# Borusyak et al. in practice

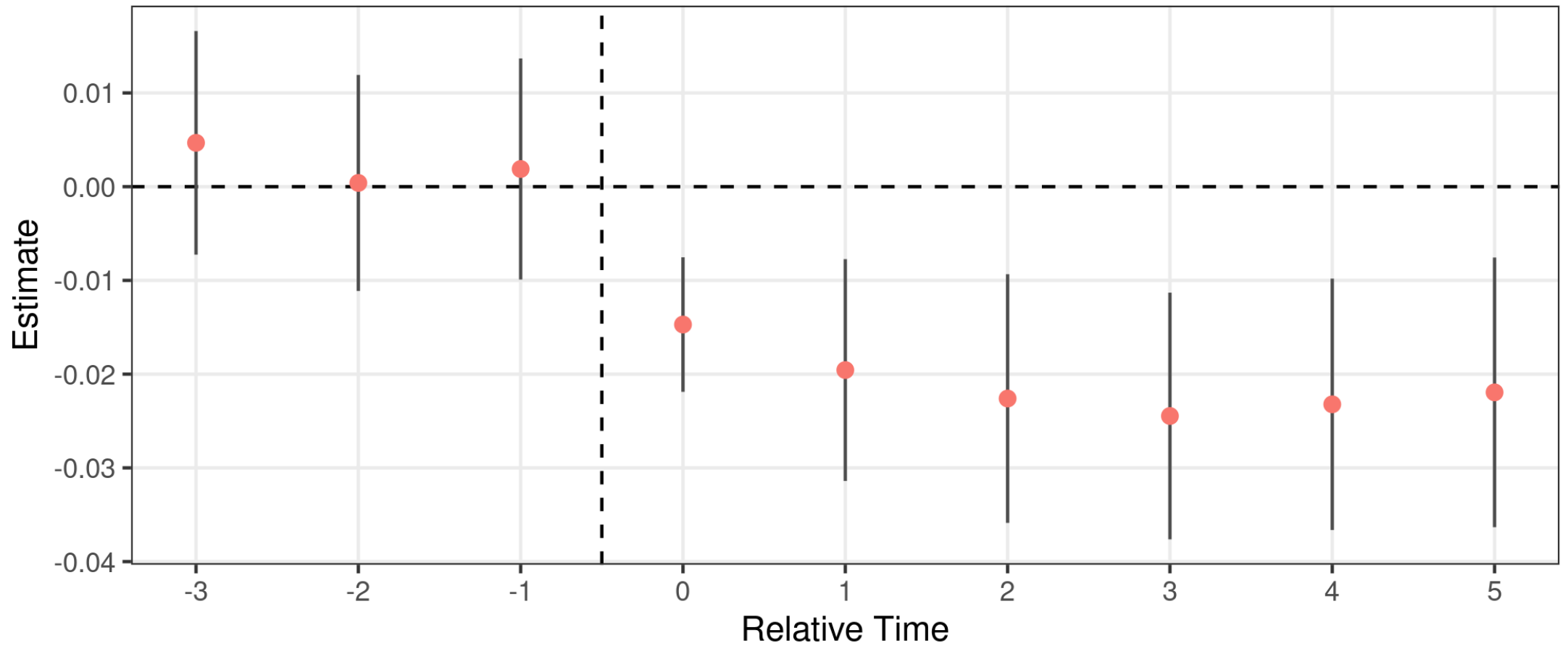
## Stata

```
did_imputation
```

## R

```
library(didimputation)
mcaid.data ← read_tsv("../data/acs_medicaid.txt")
reg.dat ← mcaid.data %>%
  filter(!is.na(expand_ever)) %>%
  mutate(perc_unins=uninsured/adult_pop,
         post = (year ≥ 2014),
         treat=post*expand_ever,
         expand_year = ifelse(expand_ever==FALSE, 0, exp
mod.bea ← did_imputation(reg.dat, yname="perc_unins",
                        gname="expand_year",
                        tname="year",
                        idname="State",
                        first_stage = ~ 0 | State + year,
                        cluster_var="State",
                        horizon=TRUE,
                        pretrends=-3:-1)
```

# Borusyak et al. in practice



# Putting things together

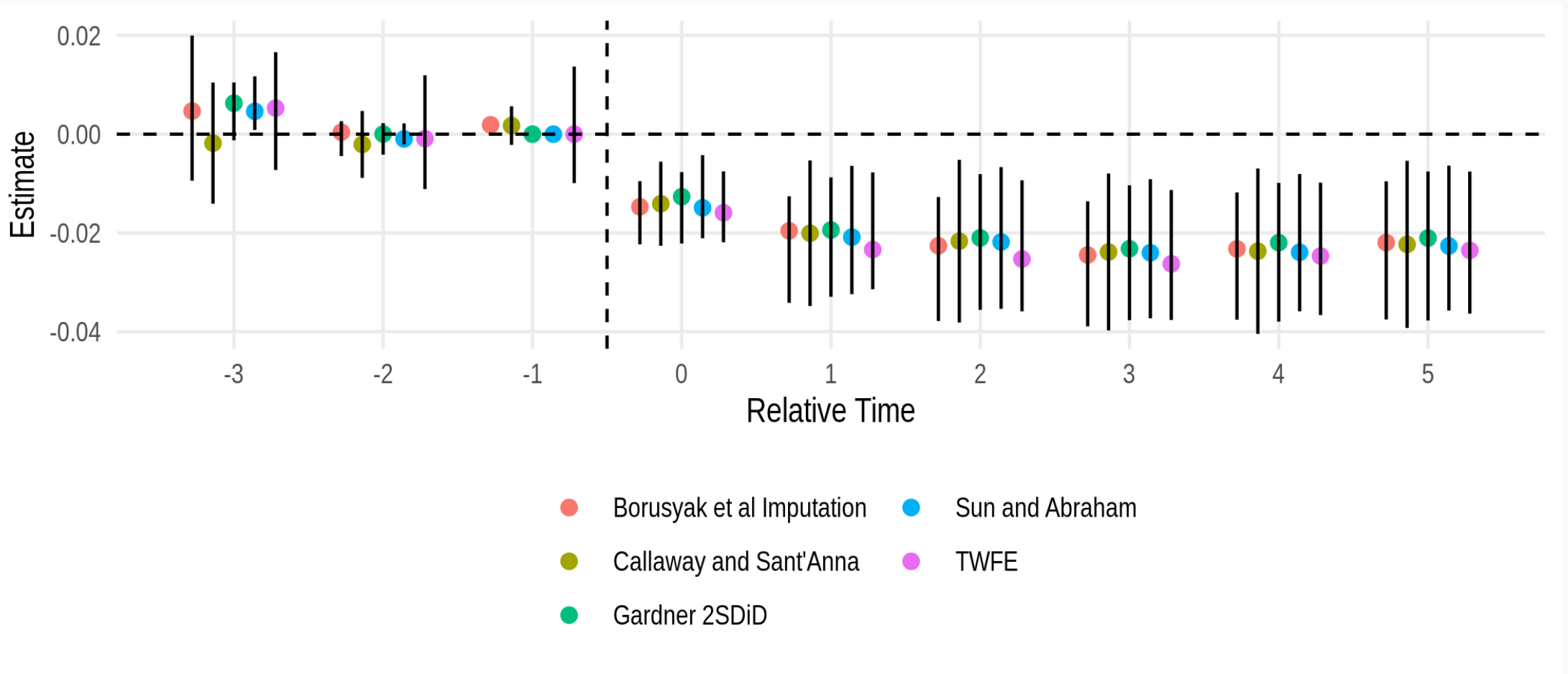
---

# Seems like lots of "solutions"

- Callaway and Sant'Anna (2020)
- Sun and Abraham (2020)
- de Chaisemartin and D'Haultfoeuille (2020)
- Cengiz et al (2019)
- Gardner (2021) and Borusyak et al. (2021)

Goodman-Bacon (2021) explores the problems but doesn't really propose a solution (still very important work though!)

# Comparison



# Comparison

## Similarities

- Focus on clean treatment/control
- Focus on event study framework (not a single overall effect)
- Impose some form of parallel trends assumption

## Differences

- Is there a "never treated" group?
- Can treatment turn on and off?
- How to include covariates?
- How to do inference?



# General advice

1. Do you have staggered treatment adoption? If so, will need to consider something beyond TWFE event study (even if it doesn't change results)
2. Do you need time-varying covariates? If so, consider Sun and Abraham or stacked regression (2SDD and imputation can only use pre-treatment covariates)
3. Is treatment "strict"? If not, CH is only option right now
4. Does treatment turn on and off again? If so, CH or perhaps focus on "clean" treatment adoptions
5. Inference? Stacked regression is harder here.

# Other topics

- Can you test for parallel pre-trends?
- Recent work says such tests are underpowered
- Consider potential violations of parallel trends and assess results
  - Intuitively "easy" to do in manual  $(g, t)$  or imputation setting, harder in pure regression setting