What's Trending in Difference-in-Differences?
A Synthesis of the Recent Econometrics Literature

Jonathan Roth
Brown University

Pedro H. C. Sant'Anna
Vanderbilt University

Alyssa Bilinski
Brown University

John Poe
University of Michigan

Seminar at NY FED, May 2023

Popularity of Difference-in-Differences methods

Currie, Kleven and Zwiers (2020) at AEA P&P

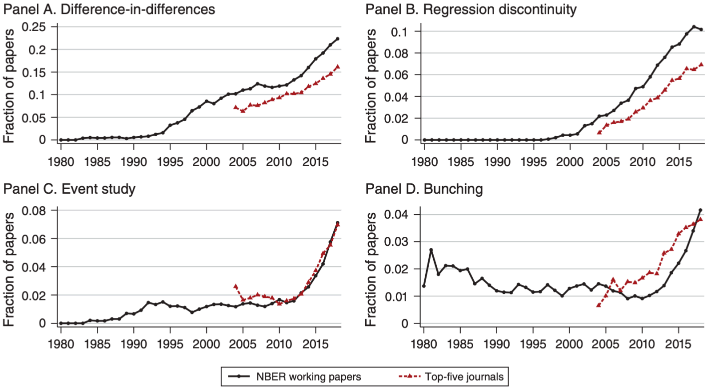


FIGURE 4. QUASI-EXPERIMENTAL METHODS

Notes: This figure shows the fraction of papers referring to each type of quasi-experimental approach. See Table A.I for a list of terms. The series show five-year moving averages.

Introduction

- The last few years have seen an explosion of econometrics on DiD, making it hard to keep up
- In Roth, Sant'Anna, Bilinski and Poe (2023), we attempt to synthesize the recent literature and provide concrete recommendations for practitioners

The simplest case

- We will start a description of DiD in the simplest "canonical" case
- Why? Because recent DiD lit can be viewed as relaxing various components of the canonical model while preserving others

In the canonical DiD model, we have:

- 2 periods: treatment occurs (for some units) in period 2
- Identification of the ATT from parallel trends and no anticipation
- Estimation using sample analogs, equivalent to OLS with TWFE
- A large number of independent observations (or clusters)

Canonical DiD – with math

- Panel data on $Y_{i,t}$ for $t = 1, 2$ and $i = 1, ..., N$
- **Treatment timing:** Some units ($D_i = 1$) are treated in period 2; everyone else is untreated ($D_i = 0$)
- **Potential outcomes:** Observe $Y_{i,t}(1) \equiv Y_{i,t}(0, 1)$ for treated units; and $Y_{i,t}(0) \equiv Y_{i,t}(0, 0)$ for comparison

Key identifying assumptions

- **Parallel trends:**

$$\mathbb{E}\left[Y_{i,t=2}(0) - Y_{i,t=1}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i,t=2}(0) - Y_{i,t=1}(0) \mid D_i = 0\right]. \quad (1)$$

- **No anticipation:** $Y_{i,t=1}(1) = Y_{i,t=1}(0)$
  - Intuitively, the outcome in period 1 isn't affected by treatment status in period 2
  - Often left implicit in notation, but important for interpreting DiD estimand as a causal effect in period 2

Identification

- **Target parameter:** Average treatment effect on the treated (ATT) in period 2

$$\tau_{ATT} = E[Y_{i,t=2}(1) - Y_{i,t=2}(0) | D_i = 1]$$

- Under parallel trends and no anticipation, can show that

$$\tau_{ATT} = \underbrace{(E[Y_{i,t=2}|D_i = 1] - E[Y_{i,t=1}|D_i = 1])}_{\text{Change for treated group}} - \underbrace{(E[Y_{i,t=2}|D_i = 0] - E[Y_{i,t=1}|D_i = 0])}_{\text{Change for comparison group}},$$

a "difference-in-differences" of population means

Estimation and Inference

- The most conceptually simple estimator replaces population means with sample analogs:

$$\hat{\tau}_{DiD} = (\overline{Y}_{12} - \overline{Y}_{11}) - (\overline{Y}_{02} - \overline{Y}_{01})$$

  where $\overline{Y}_{dt}$ is sample mean for group d in period t

- Conveniently, $\hat{\tau}_{DID}$ is algebraically equal to OLS coefficient $\hat{\beta}$ from the TWFE regression

$$Y_{i,t} = \alpha_i + \phi_t + D_{i,t}\beta + \epsilon_{i,t}, \tag{2}$$

  where $D_{i,t} = D_i * 1[t = 2]$.

- **Inference:** And clustered standard errors are valid as the number of clusters grows large

Characterizing the recent literature

We can group the recent innovations in DiD lit by which elements of the canonical model they relax:

- **Multiple periods and staggered treatment timing**
- **Relaxing or allowing PT to be violated**
- **Inference with a small number of clusters**

Today, we will focus on the first topic due to the time constraints.

Structure of the rest of the talk

- We will talk about:

  1. Problems with “static” TWFE - Goodman-Bacon (2021).
  2. Problems with TWFE event-study - Sun and Abraham (2021).
  3. Resolving these problems - Callaway and Sant’Anna (2021).

- Many new papers keep coming out but we won’t have time to dig deeper into them.

- Many DiD empirical applications, however, deviate from the canonical 2x2 DiD setup
  - More than two time periods
  - Variation in treatment timing

Does TWFE “work” in setups with variation in treatment timing?

Traditional methods: TWFE regressions

- We know that, in the 2x2 case,

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

- It is tempting to "extrapolate" from this setup and use variations of the following TWFE specification to estimate causal effects:

$$Y_{i,t} = \alpha_i + \alpha_t + \beta \cdot D_{i,t} + \varepsilon_{i,t}$$

  where dummies $D_{i,t} = 1\{t - G_i \geq 0\}$, where G_i indicates the period unit i is first treated (Group).

- $D_{i,t}$ is an indicator for unit i being treated by period t.
- For simplicity, let's assume that treatment is "irreversible": once a unit is treated, it is forever treated - aka **staggered design**

Does TWFE “work” in setups with variation in treatment timing?

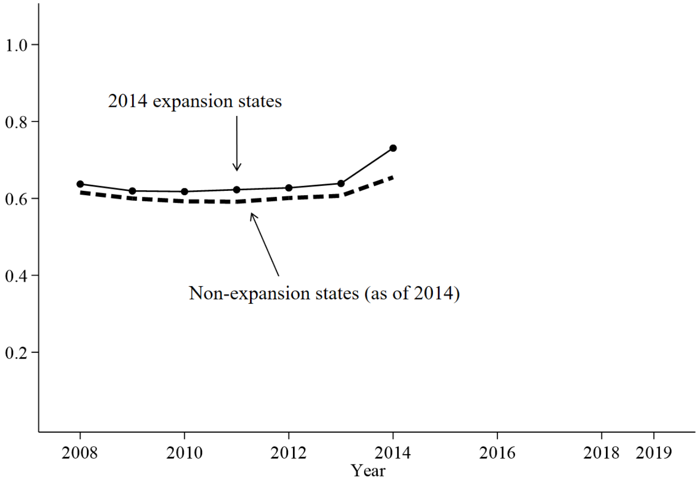
Example: Effect of ACA Medicaid Expansion on Health Insurance rate

Empirical Example: Medicaid Expansion

- To motivate our problem, let's look at a classical example: Medicaid Expansion
- We want to analyze its effect on health insurance rate among low-income, childless adults aged 25-64.

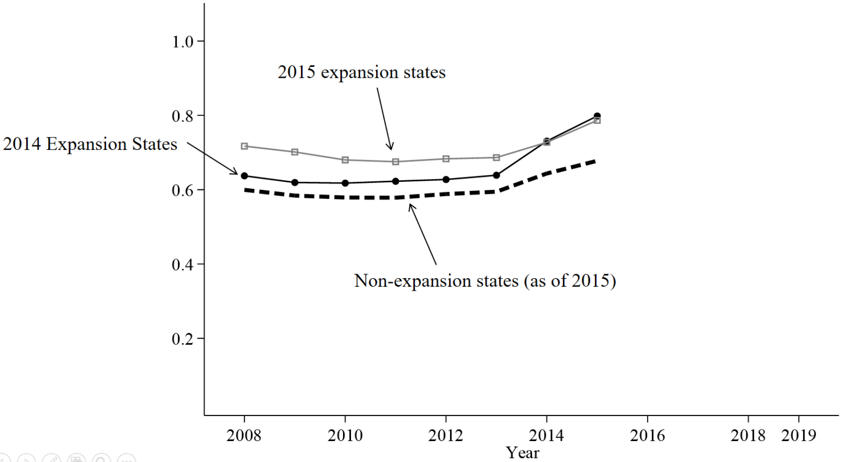
ACA Medicaid Expansion Circa 2014

Figure 1: Health Insurance Rate (low-income Childless Adults Aged 25-64)



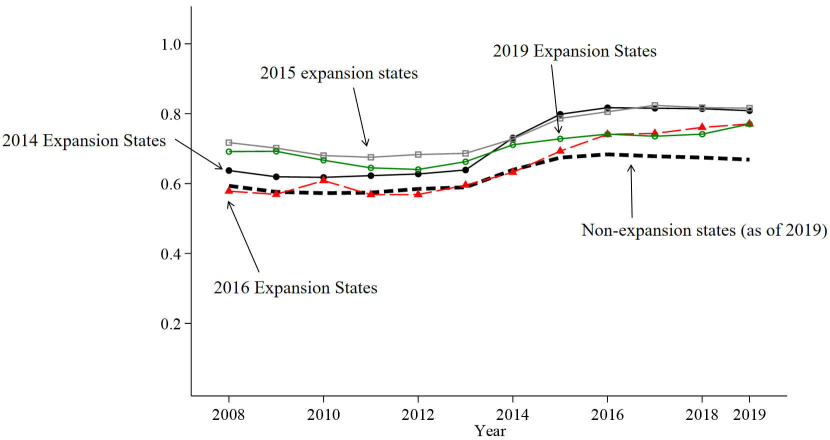
ACA Medicaid Expansion Circa 2015

Figure 2: Health Insurance Rate (low-income Childless Adults Aged 25-64)



ACA Medicaid Expansion Circa 2019

Figure 3: Health Insurance Rate (low-income Childless Adults Aged 25-64)



ACA Medicaid Expansion Circa 2019

- 23 states expanded circa 2014 - 4 did it earlier (ACA is effectively relabeled), we drop them.
- 3 states expanded circa 2015
- 2 states expanded circa 2016
- 1 states expanded circa 2017
- 2 states expanded circa 2019
- 16 states haven't expanded by 2019

OLS estimate of β

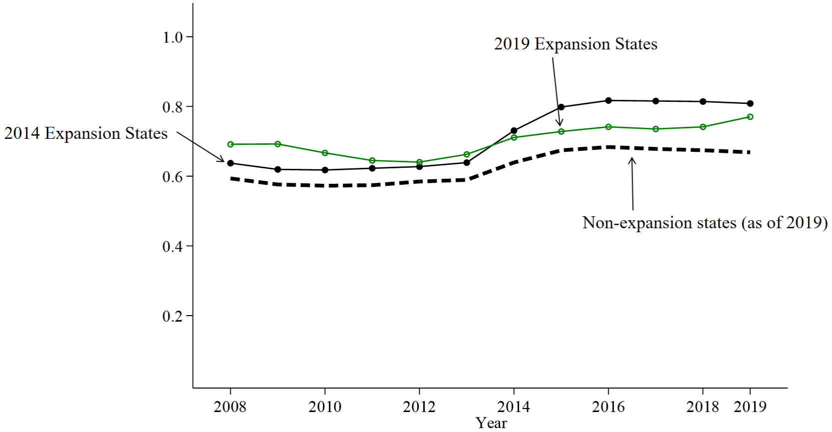
- Let $\widehat{\beta}$ be the OLS estimator of the following TWFE regression specification:

$$Y_{i,t} = \alpha_i + \alpha_t + \beta \cdot D_{i,t} + \varepsilon_{i,t}$$

- What is $\widehat{\beta}$?
- Goodman-Bacon (2021) shows that we can answer this question following these three steps:
  1. Remove unit means
  $$D_{i,t} - \overline{D}_i$$
  2. Remove time means of $(D_{i,t} - \overline{D}_i)$:
  $$\widetilde{D}_{i,t} = (D_{i,t} - \overline{D}_i) - (\overline{D}_t - \overline{\overline{D}})$$
  3. Calculate univariate regression of $Y_{i,t}$ on $\widetilde{D}_{i,t}$:
  $$\widehat{\beta} = \frac{(nT)^{-1} \sum_{i,t} Y_{i,t} \cdot \widetilde{D}_{i,t}}{(nT)^{-1} \sum_{i,t} \widetilde{D}_{i,t}^2}$$

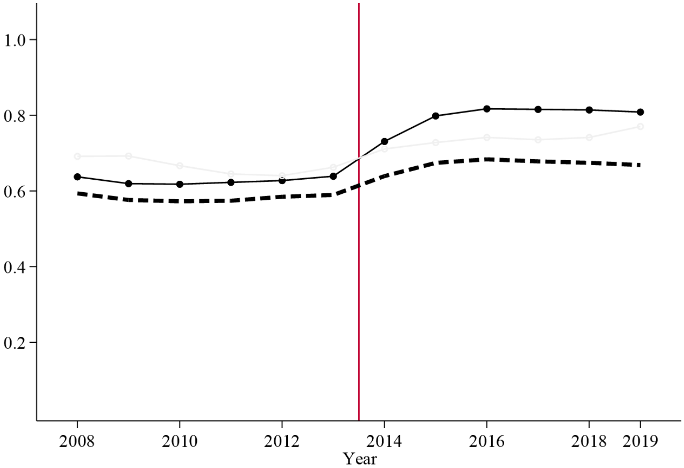
Three Groups Example

Figure 4: Health Insurance Rate (low-income Childless Adults Aged 25-64)



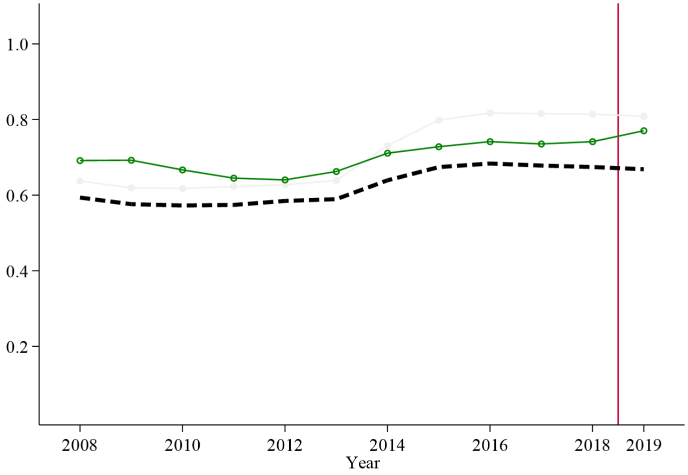
Treated in 2014 vs. Never-Treated

Figure 5: Health Insurance Rate (low-income Childless Adults Aged 25-64)



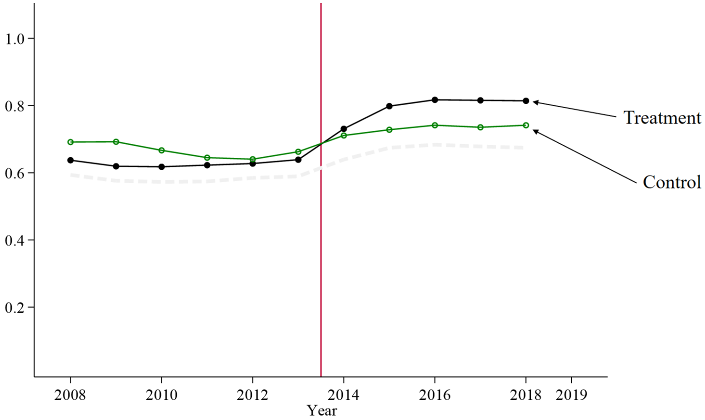
Treated in 2019 vs. Never-Treated

Figure 6: Health Insurance Rate (low-income Childless Adults Aged 25-64)



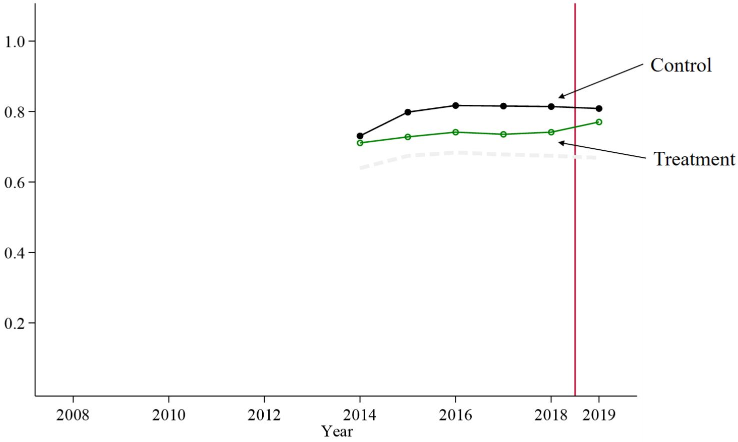
Treated in 2014 vs. Treated in 2019 ($t < 2019$)

Figure 7: Health Insurance Rate (low-income Childless Adults Aged 25-64)



Treated in 2019 vs. Treated in 2014 ($t \geq 2014$)

Figure 8: Health Insurance Rate (low-income Childless Adults Aged 25-64)



OLS estimate of β

- OLS is “variational hungry” and exploit all these 2x2 comparisons.
- But how does OLS aggregate them?
- Main result of Goodman-Bacon (2021) is the Bacon-Decomposition:

$$\widehat{\beta} = s_{k,U} \cdot \widehat{\beta}_{k,U} + s_{\ell,U} \cdot \widehat{\beta}_{\ell,U} + \left[s_{k,\ell} \cdot \widehat{\beta}_{k,\ell} + s_{\ell,k} \cdot \widehat{\beta}_{\ell,k} \right]$$

- In our example:
  - $k = 2014$
  - $\ell = 2019$
  - $U =$ never-treated

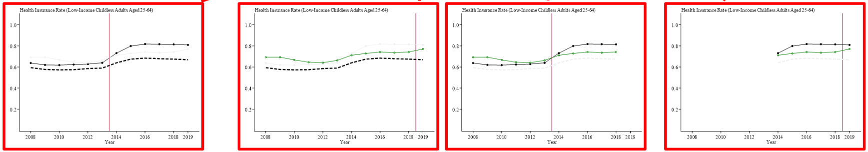
Does TWFE “work” in setups with variation in treatment timing?

Bacon Decomposition

- Main result of Goodman-Bacon (2021) is the Bacon-Decomposition:

Figure 9: Bacon-Decomposition: The 2x2 $\widehat{\beta}$

$$\hat{\beta}^{DD} = s_{kU}\hat{\beta}^{DD}_{kU} + s_{\ell U}\hat{\beta}^{DD}_{\ell U} + \left[s^{k}_{k\ell}\hat{\beta}^{DD,k}_{k\ell} + s^{\ell}_{k\ell}\hat{\beta}^{DD,\ell}_{k\ell}\right]$$



- Main result of Goodman-Bacon (2021) is the Bacon-Decomposition:

Figure 10: Bacon-Decomposition: The weights

$$\hat{\beta}^{DD} = s_{kU}\hat{\beta}^{DD}_{kU} + s_{\ell U}\hat{\beta}^{DD}_{\ell U} + \left[s^{k}_{k\ell}\hat{\beta}^{DD,k}_{k\ell} + s^{\ell}_{k\ell}\hat{\beta}^{DD,\ell}_{k\ell}\right]$$

$$s_{kU} = \frac{(n_k + n_U)^2 n_{kU}(1 - n_{kU})\bar{D}_k(1 - \bar{D}_k)}{V(\tilde{D}_{it})}$$

$$s^{k}_{k\ell} = \frac{\left((n_k + n_\ell)(1 - \bar{D}_\ell)\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell}\frac{1 - \bar{D}_k}{1 - \bar{D}_\ell}}{V(\tilde{D}_{it})}$$

$$s^{\ell}_{k\ell} = \frac{\left((n_k + n_\ell)\bar{D}_k\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k}\frac{\bar{D}_\ell}{\bar{D}_k}}{V(\tilde{D}_{it})}$$

Sample size^2

- Main result of Goodman-Bacon (2021) is the Bacon-Decomposition:

Figure 11: Bacon-Decomposition: The weights

$$\hat{\beta}^{DD} = s_{kU}\hat{\beta}^{DD}_{kU} + s_{\ell U}\hat{\beta}^{DD}_{\ell U} + \left[s^{k}_{k\ell}\hat{\beta}^{DD,k}_{k\ell} + s^{\ell}_{k\ell}\hat{\beta}^{DD,\ell}_{k\ell}\right]$$

$$s_{kU} = \frac{(n_k + n_U)^2 n_{kU}(1 - n_{kU})\bar{D}_k(1 - \bar{D}_k)}{\hat{V}(\widetilde{D}_{it})}$$

$$s^{k}_{k\ell} = \frac{\left((n_k + n_\ell)(1 - \bar{D}_\ell)\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell}\frac{1 - \bar{D}_k}{1 - \bar{D}_\ell}}{\hat{V}(\widetilde{D}_{it})}$$

$$s^{\ell}_{k\ell} = \frac{\left((n_k + n_\ell)\bar{D}_k\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k}\frac{\bar{D}_\ell}{\bar{D}_k}}{\hat{V}(\widetilde{D}_{it})}$$

If you did TWFE on this subsample, what would the variance of $\widetilde{D}_{it}$ be?

Bacon-Decomposition: General case

Theorem (Goodman-Bacon (2021) decomposition)

Assume that there are $k = 1, \ldots, K$ groups of treated units ordered by treatment time t_k^* and one "never-treated" group, U, which does not receive treatment in the data. The share of units in group k is n_k, and the share of periods that group k spends under treatment is $\overline{D}_k$. The regression estimate from a two-way fixed effects model is a weighted average all two-group DiD estimators:

$$\widehat{\beta} = \sum_{k \neq U} \left(s_{k,U} \cdot \widehat{\beta}_{k,U} \right) + \sum_{k \neq U} \sum_{\ell > k} \left(s_{k,\ell} \cdot \widehat{\beta}_{k,\ell} + s_{\ell,k} \cdot \widehat{\beta}_{\ell,k} \right),$$

where the weights are given by

$$s_{k,U} = \frac{(n_k + n_U)^2 \widehat{V}_{k,U}}{\widehat{V}\left(\widetilde{D}_{i,t}\right)}, \quad s_{k,\ell} = \frac{\left((n_k + n_\ell)\left(1 - \overline{D}_\ell\right)\right)^2 \widehat{V}_{k,\ell}}{\widehat{V}\left(\widetilde{D}_{i,t}\right)}, \quad s_{\ell,k} = \frac{\left((n_k + n_\ell)\,\overline{D}_k\right)^2 \widehat{V}_{\ell,k}}{\widehat{V}\left(\widetilde{D}_{i,t}\right)},$$

such that $\sum_{k \neq U} s_{k,U} + \sum_{k \neq U} \sum_{\ell > k} \left(s_{k,\ell} + s_{\ell,k} \right) = 1.$

What does this mean to TWFE regressions?

TWFE computes weighted-averages of 2x2 DiD's

- $\widehat{\beta} = 0.074$ in the empirical application.

- **OLS weights use sample size and variance**

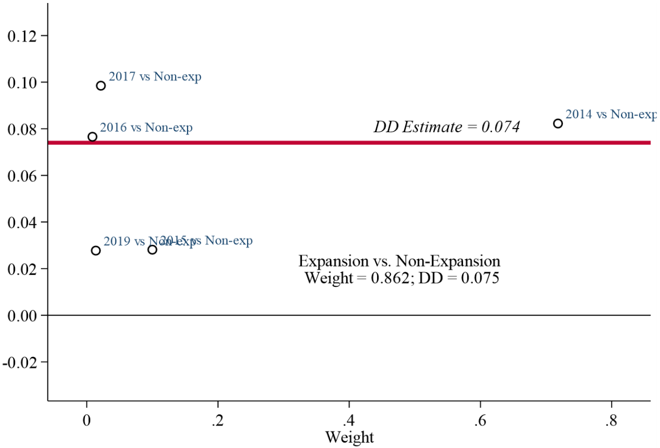
- Is that what you really want?

- **TWFE exploits all 2x2 DiD comparisons**
  - Treated vs. "Never-treated"
  - Early-treated vs. Later-treated
  - Later-treated vs. Already-treated

- Are all these comparisons "reasonable" to attach a causal interpretation to $\widehat{\beta}$?

Bacon-Decomposition: Treated vs. Never-Treated

Figure 12: Bacon-Decomposition: The weights



TWFE regressions, in general,

do not recover an easy-to-interpret

causal parameter of interest,

unless we rule out TE heterogeneity/dynamics

What happens when we consider a TWFE event-study specification?

Event-Study via TWFE specifications

Event-Study via TWFE specifications

- One of the main attractive features of observing multiple time periods is that we can attempt to "learn" about treatment effect dynamics.

- Status-quo in the literature is to consider variants of the TWFE event-study regression

$$Y_{i,t} = \alpha_i + \alpha_t + \gamma_k^{-K} D_{i,t}^{<-K} + \sum_{k=-K}^{-2} \gamma_k^{lead} D_{i,t}^k + \sum_{k=0}^{L} \gamma_k^{lags} D_{i,t}^k + \gamma_k^{L+} D_{i,t}^{>L} + \varepsilon_{i,t}$$

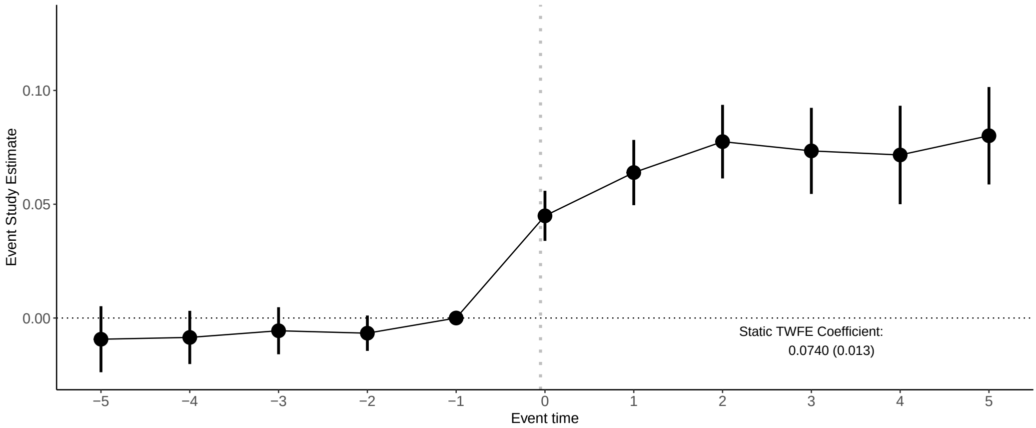
  with the event study dummies $D_{i,t}^k = 1\{t - G_i = k\}$, where G_i indicates the period unit i is first treated (Group).

- $D_{i,t}^k$ is an indicator for unit i being k periods away from initial treatment at time t.

Does this strategy “work”?

ACA Medicaid Expansion: TWFE Event-study specification

Figure 13: Health Insurance Rate (low-income Childless Adults Aged 25-64



Event-Study via TWFE specifications

- Can we (a priori) “trust” these results?
- What type of treatment effect parameter is being reported in this event-study?
- What kind of assumptions are we implicitly relying on?
- What kind of comparisons are being made “behind the scenes”?
- These are important questions!

Event-Study via TWFE specifications

Sun and Abraham (2021)

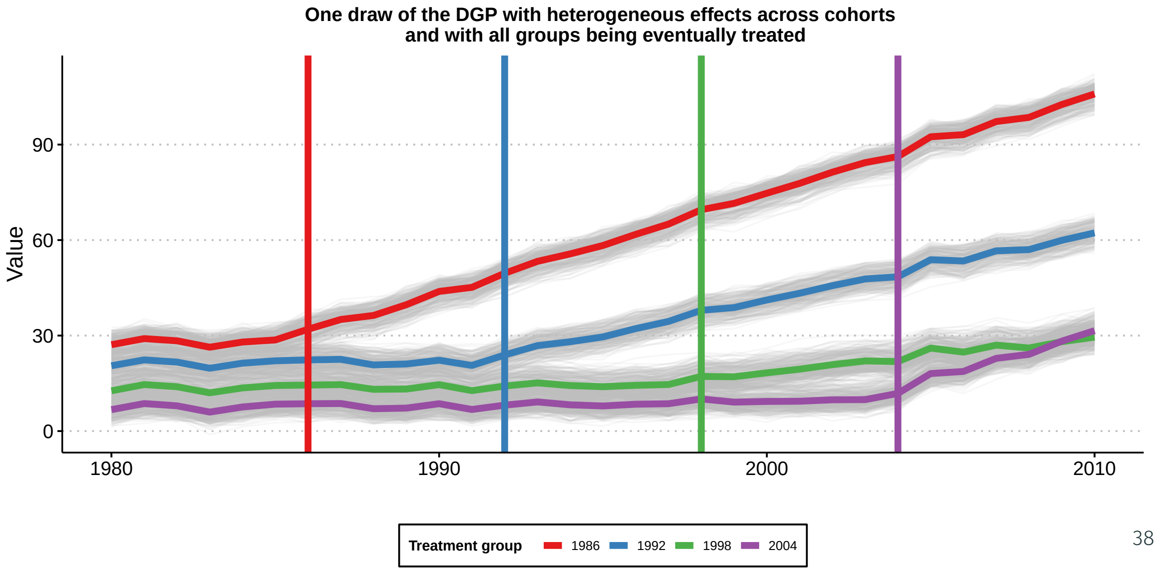
Problem with Event-Study via TWFE specifications: Sun and Abraham (2021)

- Sun and Abraham (2021) bring "bad" news, once again!
- Even when we impose the Strong unconditional parallel trends and the no-anticipation assumption, the OLS coefficients of the TWFE ES specification are, in general, very hard to interpret.
- Coefficient on a given lead or lag can be contaminated by effects from other periods
- Pre-trends can arise solely from treatment effects heterogeneity!
- Even under treatment effect homogeneity across cohorts (they all share same dynamics in event-time), the OLS coefficients can still be contaminated by treatment effects from the excluded periods.

Event-Study via TWFE specifications

Stylized example using simulated data

Stylized example using simulated data



Stylized example using simulated data

- 1000 units ($i = 1, 2, \ldots, 1000$) from 40 states ($state = 1, 2, \ldots, 40$).
- Data from 1980 to 2010 (31 years).
- 4 different groups based on year that treatment starts: $g = 1986, 1992, 1998, 2004$.
- Randomly assign each state to a group.
- Outcome:

$$Y_{i,t} = \underbrace{(2010 - g)}_{\text{cohort-specific intercept}} + \underbrace{\alpha_i}_{N\left(\frac{state}{5},1\right)} + \underbrace{\alpha_t}_{\frac{(t-g)}{10}+N(0,1)} + \underbrace{\tau_{i,t}}_{\mu_g \cdot (t-g+1)\cdot 1\{t \geq g\}} + \underbrace{\varepsilon_{i,t}}_{N\left(0,\left(\frac{1}{2}\right)^2\right)}$$

- $\mu_{1986} = \mu_{2004} = 3,\ \mu_{1992} = 2,\ \mu_{1998} = 1$
- ATT for group g at the first treatment period is μ_g, at the second period since treatment is $2 \cdot \mu_g$, etc.

Traditional methods: TWFE event-study regression

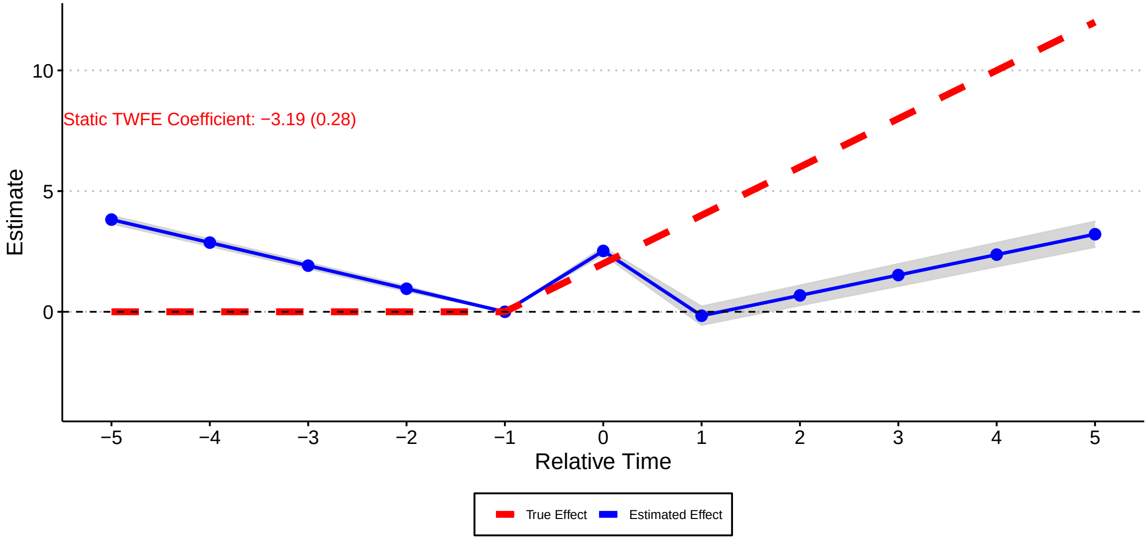
- What if we tried to estimate the treatment effects using traditional TWFE event-study regressions,

$$Y_{i,t} = \alpha_i + \alpha_t + \gamma_k^{-K} D_{i,t}^{<-K} + \sum_{k=-K}^{-2} \gamma_k^{lead} D_{i,t}^{k} + \sum_{k=0}^{L} \gamma_k^{lags} D_{i,t}^{k} + \gamma_k^{L+} D_{i,t}^{>L} + \varepsilon_{i,t},$$

  with K and L to be equal to 5 ?

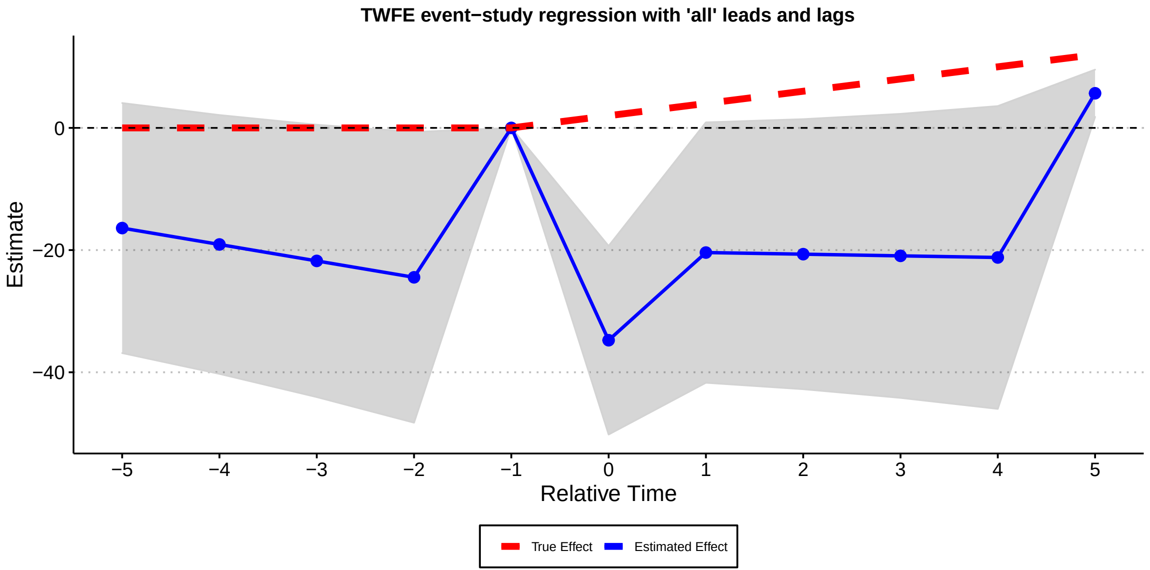
- Simulate data and repeat 1,000 times to compute bias and simulation standard deviations.

TWFE event−study regression with binned end−points

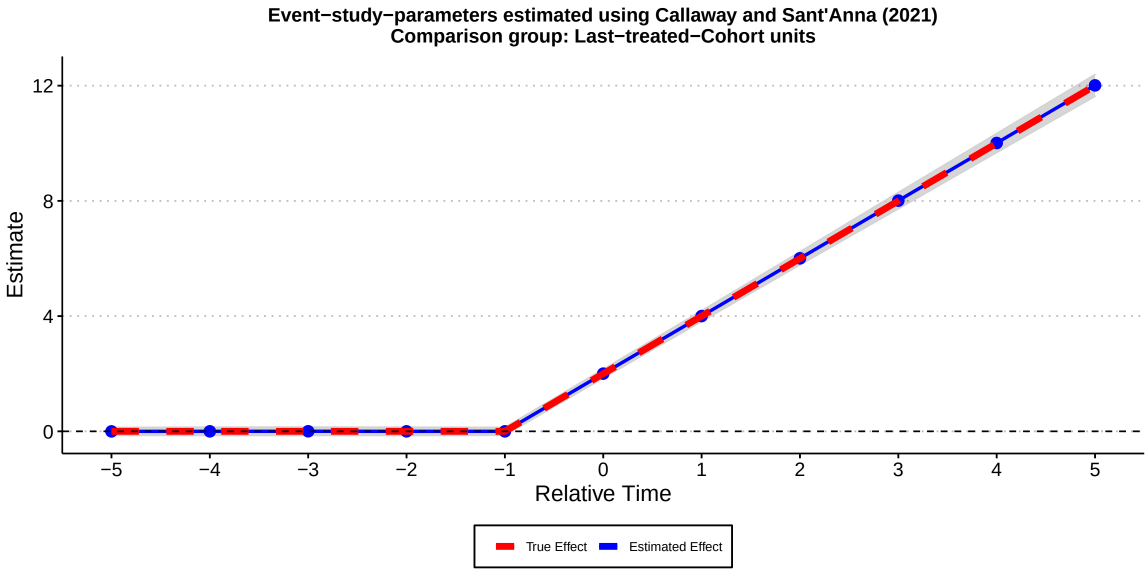


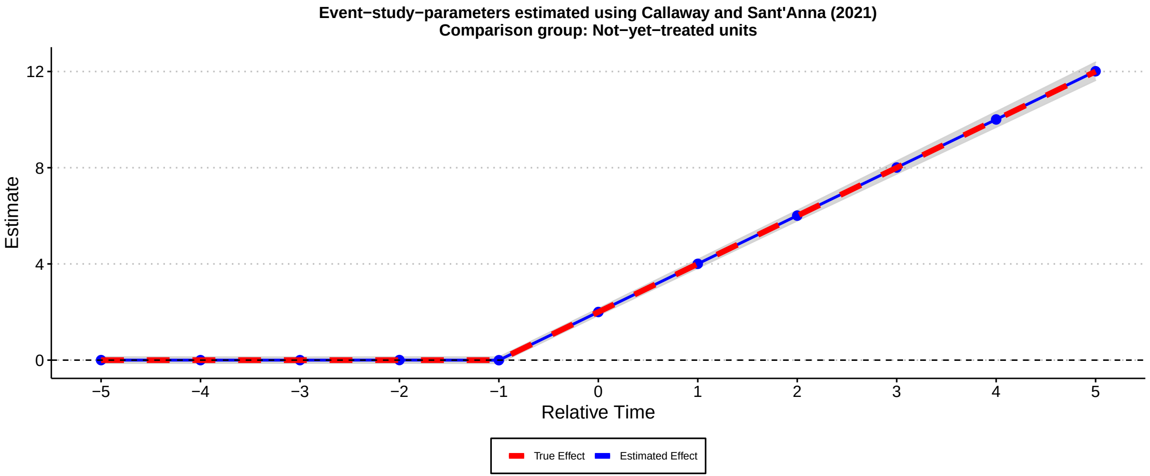
Traditional methods: TWFE event-study regression

- What if we include all possible leads and lags in the TWFE event study specification, i.e., to set K and L to the maximum allowable in the data, making inclusion of $D_{i,t}^{<-K}$ and of $D_{i,t}^{>L}$ unnecessary ?


TWFE event−study regression with 'all' leads and lags
Estimate
0
−20
−40
−5
−4
−3
−2
−1
0
1
2
3
4
5
Relative Time
True Effect
Estimated Effect

How can we proceed?


Event−study−parameters estimated using Callaway and Sant'Anna (2021)
Comparison group: Last−treated−Cohort units
Estimate
12
8
4
0
−5
−4
−3
−2
−1
0
1
2
3
4
5
Relative Time
True Effect
Estimated Effect


Event−study−parameters estimated using Callaway and Sant'Anna (2021)
Comparison group: Not−yet−treated units
Estimate
12
8
4
0
−5
−4
−3
−2
−1
0
1
2
3
4
5
Relative Time
True Effect
Estimated Effect

Callaway and Sant'Anna (2021)

Main take-way message

Clearly separate identification, aggregation, and estimation/inference steps!

Let’s talk about identification

Callaway and Sant'Anna (2021)

Identification

Building block of the analysis

- If sample size was not a limitation (we have all the data in the world), what kind of question we would like to answer?

- In staggered setups, a parameter that is interesting and has clear economic interpretation is the $ATT(g,t)$

$$ATT(g,t) = \mathbb{E}\left[Y_t(g) - Y_t(\infty) \,|\, G_g = 1\right], \ for\ t \geq g.$$

- Average Treatment Effect at time t of starting treatment at time g, among the units that indeed started treatment at time g.

Identifying Assumptions: No-Anticipation

- Given that we never observe $Y(\infty)$ in post-treatment periods among units that have been treated, we need to make assumptions to identify $ATT(g,t)$'s
- No-Anticipation Assumption: For all i,t and $t < g, g'$, $Y_{i,t}(g) = Y_{i,t}(g')$.
- Unit treatment effects are zero before treatment takes place.
- Exactly the same content as in the 2x2 case.

Parallel trend assumption based on a “never treated” group

Assumption (Parallel Trends based on a “never-treated”)

For each $t \in \{2, \ldots, T\}$, $g \in \mathcal{G}$ *such that* $t \geq g$,

$$\mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty) | G_g = 1] = \mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty) | C = 1]$$

Parallel Trends based on not-yet treated groups

Assumption (Parallel Trends based on "Not-Yet-Treated" Groups)

For each $(s, t) \in \{2, \ldots, T\} \times \{2, \ldots, T\}$, $g \in \mathcal{G}$ *such that* $t \geq g$, $s \geq t$

$$\mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty) | G_g = 1] = \mathbb{E}[Y_t(\infty) - Y_{t-1}(\infty) | D_s = 0, G_g = 0].$$

ATT(g,t) Estimand: “never-treated” as comparison group

- Under no-anticipation and PT based on “never-treated”, we have

$$ATT_{unc}^{nev}(g,t) = \mathbb{E}[Y_t - Y_{g-1}|G_g = 1] - \mathbb{E}[Y_t - Y_{g-1}|C = 1].$$

- This looks very similar to the two periods, two-groups DiD result without covariates.
- The difference is now we take a “long difference”.
- Same intuition carries, though!
- This result appears in Callaway and Sant’Anna (2021) and Sun and Abraham (2021).

ATT(g,t) Estimand: not-yet treated as comparison group

- If one wants to use an the units that have not-yet been exposed to treatment by time t, we have a different estimand:

$$ATT^{ny}_{unc}(g,t) = \mathbb{E}[Y_t - Y_{g-1}|G_g = 1] - \mathbb{E}[Y_t - Y_{g-1}|D_t = 0, G_g = 0].$$

- This looks similar to the two periods, two-groups DiD result without covariates, too.
- The difference is now we take a “long difference” , and that the comparison group changes over time.
- Same intuition carries, though!
- This result appears in Callaway and Sant’Anna (2021) and de Chaisemartin and D’Haultfœuille (2020), though de Chaisemartin and D’Haultfœuille (2020) focus exclusively in instantaneous treatment effects, i.e., the case with $g = t$.

Callaway and Sant'Anna (2021)

Aggregation

Second step: Aggregation

Summarizing ATT(g,t)

- $ATT(g, t)$ are very useful parameters that allow us to better understand treatment effect heterogeneity.

- We can also use these to summarize the treatment effects across groups, time since treatment, calendar time.

- Practioners routinely attempt to pursue this avenue:
  - Run a TWFE “static” regression and focus on the β associated with the treatment.
  - Run a TWFE event-study regression and focus on β associated with the treatment leads and lags.
  - Collapse data into a 2 x 2 Design (average pre and post treatment periods).

Summarizing ATT(g,t)

- We propose taking weighted averages of the $ATT(g,t)$ of the form:

$$\sum_{g=2}^{T}\sum_{t=2}^{T}\mathbf{1}\{g \leq t\}w_{gt}ATT(g,t)$$

- The two simplest ways of combining $ATT(g,t)$ across g and t are, assuming no-anticipation,

$$\theta_M^O := \frac{2}{T(T-1)}\sum_{g=2}^{T}\sum_{t=2}^{T}\mathbf{1}\{g \leq t\}ATT(g,t) \tag{3}$$

  and

$$\theta_W^O := \frac{1}{\kappa}\sum_{g=2}^{T}\sum_{t=2}^{T}\mathbf{1}\{g \leq t\}ATT(g,t)P(G=g|C\neq 1) \tag{4}$$

- Problem: They “overweight” units that have been treated earlier

Summarizing ATT(g,t): Cohort-heterogeneity

- More empirically motivated aggregations do exist!
- Average effect of participating in the treatment that units in group g experienced:

$$\theta_S(g) = \frac{1}{T - g + 1} \sum_{t=2}^{T} \mathbf{1}\{g \leq t\} ATT(g, t)$$

Summarizing ATT(g,t): Calendar time heterogeneity

- Average effect of participating in the treatment in time period t for groups that have participated in the treatment by time period t

$$\theta_C(t) = \sum_{g=2}^{T} \mathbf{1}\{g \leq t\} ATT(g,t) P(G = g | G \leq t, C \neq 1)$$

Summarizing ATT(g,t): Event-study / dynamic treatment effects

- The effect of a policy intervention may depend on the length of exposure to it.
- Average effect of participating in the treatment for the group of units that have been exposed to the treatment for exactly e time periods

$$\theta_D(e) = \sum_{g=2}^{T} \mathbf{1}\{g + e \leq T\} ATT(g, g+e) P(G = g | G + e \leq T, C \neq 1)$$

- This is perhaps the most popular summary measure currently adopted by empiricists.

Third step: Estimation and Inference

Callaway and Sant'Anna (2021)

Estimation and Inference

Estimation

- Identification results suggest a simple plug-in estimation procedure.
- Replace population expectations with their empirical analogues.
- Callaway and Sant'Anna (2021) allows for covariates and provides high-level conditions that first-step estimators have to satisfy.

Inference

- Under relatively weak regularity conditions,

$$\sqrt{n}\left(\widehat{ATT}(g,t) - ATT(g,t)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{gt}(\mathcal{W}_i) + o_p(1)$$

- From the above asymptotic linear representation and a CLT, we have

$$\sqrt{n}\left(\widehat{ATT}(g,t) - ATT(g,t)\right) \xrightarrow{d} N(0, \Sigma_{g,t})$$

  where $\Sigma_{gt} = \mathbb{E}[\psi_{gt}(\mathcal{W})\psi_{gt}(\mathcal{W})']$.
- Above result ignores the dependence across g and t, and "multiple-testing" problems.
- **Solution:** Use bootstrap to do simultaneous inference.
- Details are on the paper (and also on slides available on my webpage).

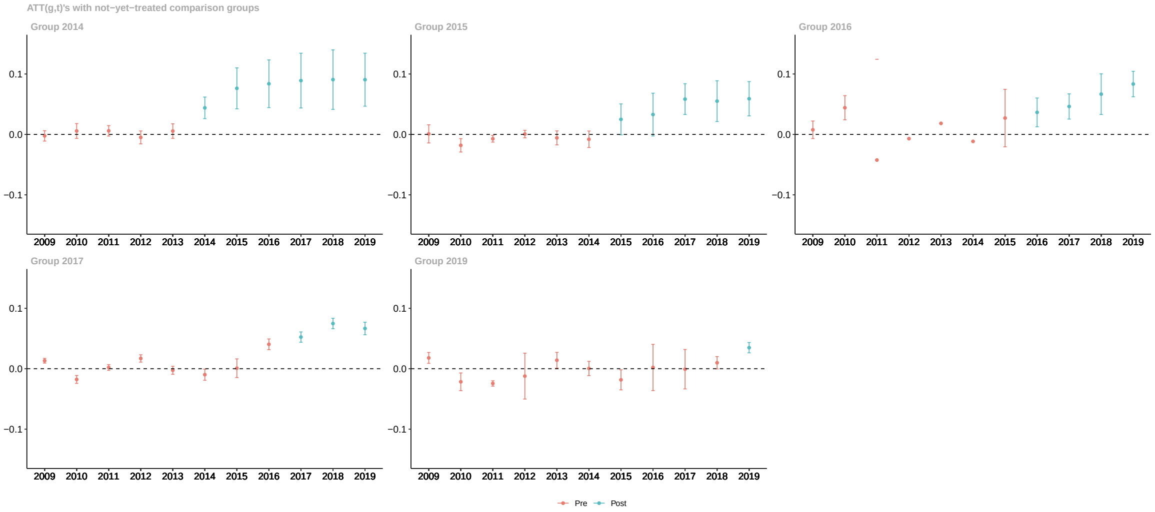
Let's go back to the ACA Medicaid Expansion Example

ACA Medicaid Expansion

- 23 states expanded circa 2014 - 4 did it earlier (ACA is effectively relabeled), we drop them.
- 3 states expanded circa 2015
- 2 states expanded circa 2016
- 1 states expanded circa 2017
- 2 states expanded circa 2019
- 16 states haven't expanded by 2019

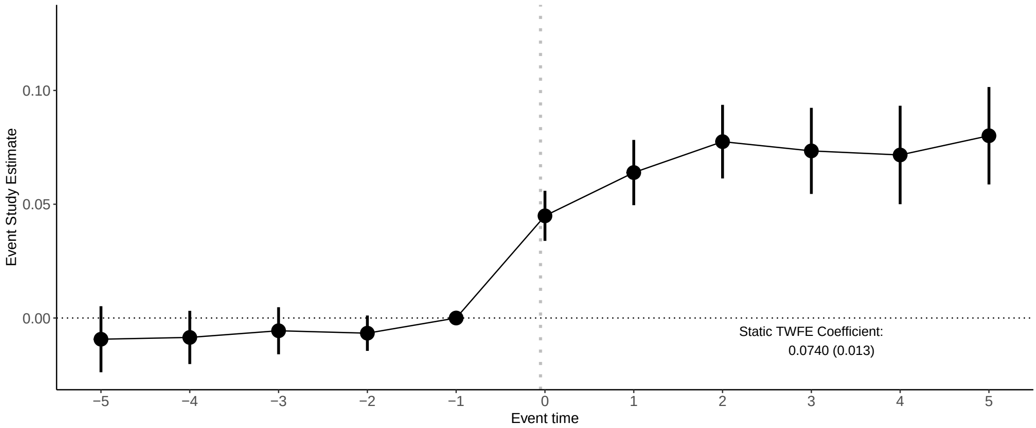
Challenge setup to make inference on ATT(g,t)'s per se

ACA Medicaid Expansion: Not-yet-treated as comparison group



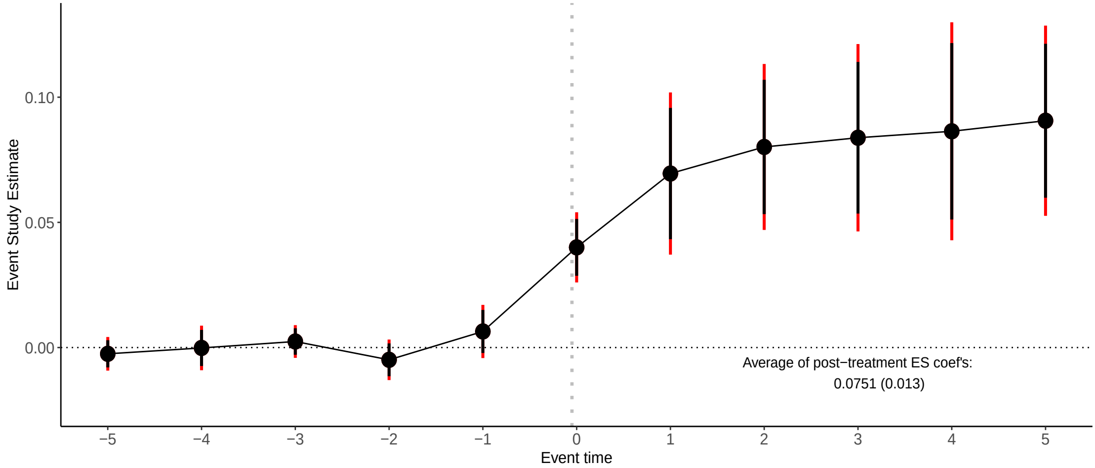
ACA Medicaid Expansion: TWFE Event-study specification

Figure 14: Health Insurance Rate (low-income Childless Adults Aged 25-64)



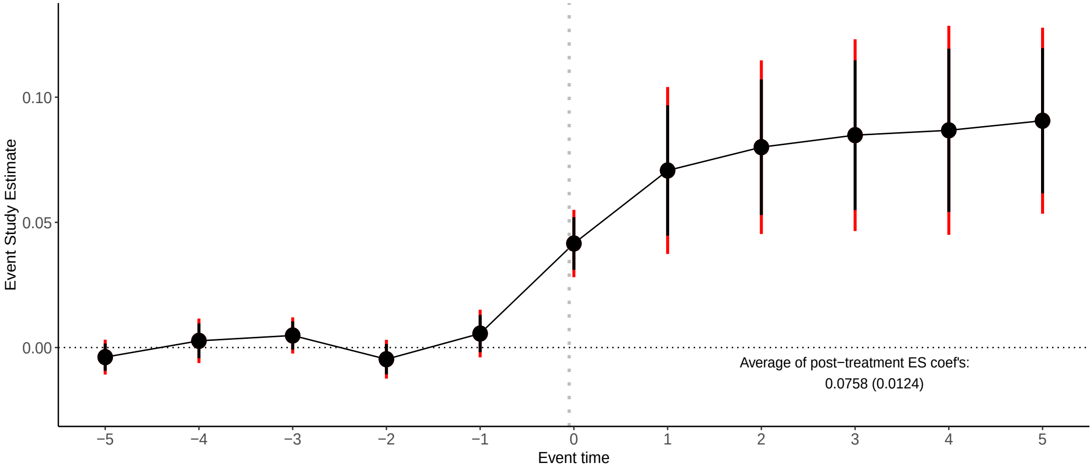
ACA Medicaid Expansion: CS Event-study specification

Figure 15: Results using "never-treated" as a comparison group



ACA Medicaid Expansion: CS Event-study specification

Figure 16: Results using “not-yet-treated” as comparison groups



Take-way messages

DiD procedures multiple time periods

- With multiple time periods and variation in treatment timing, TWFE does not respect our assumptions:
  - OLS is "variational hungry" and makes many comparisons of means
  - Some of these comparisons are bad: use already-treated units as a comparison group to "later-treated" groups
  - This can lead to "negative weighting" problems.
- Solution to the TWFE problem is simple
  - Separate the identification, aggregation and estimation/inference parts of the problem
- Use $ATT(g,t)$ as building block so we can transparently see how things are constructed
- Many different aggregation schemes are possible: they deliver different parameters!
- Can allow for covariates via regressions adjustments, IPW and DR.

References

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

Currie, Janet, Henrik Kleven, and Esmée Zwiers, "Technology and Big Data Are Changing Economics: Mining Text to Track Methods," *AEA Papers and Proceedings*, May 2020, *110*, 42–48.

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).

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *Journal of Econometrics*, 2023, *Forthcoming*.

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