Financial Econometrics II – Cross Section and Panel Data

Difference-in-differences

Andreas Fuster
Swiss Finance Institute @ EPFL

SFI Léman PhD program – 2025, Lecture 3

Overview

- **Motivation**
- Difference-in-differences – the basics
- Difference-in-differences – implementation
- Extensions: continuous treatment & staggered diff-in-diff

The setting

- Recap – if we have a model like

$$y = \beta_0 + \beta_1 x + u$$

  and $cov(x, u) \neq 0$, then one of the key assumptions for **causal inference is violated**.

- Another way to think about this is that the distribution of x (also after controlling for other covariates) is **not random**.

- For instance: firms with low leverage may have higher profits because low leverage is more likely for firms with some omitted variable contained in u that is also associated with high profits.

Quasi-natural experiments

- Ideally, the researcher could simply **run experiments** to achieve the necessary randomness.
  — In medicine, researchers randomly give a new drug to patients to determine the effect of this new drug.

- In corporate finance, we generally **cannot** do this **random assignment**.
  — We cannot randomly assign a firm's leverage to determine its effect on profits.

- Therefore, researchers in corporate finance rely on so-called "**quasi-natural experiments.**"

Quasi-natural experiments

- Quasi-natural experiments are events that cause random assignment or a change in a variable of interest x, typically in a panel setting.

  — Some regulatory change (often at state or country level) affects the leverage of a subset of firms.

  — A macro-level financial crisis differentially affects different firms depending on their debt maturity structure pre-crisis.

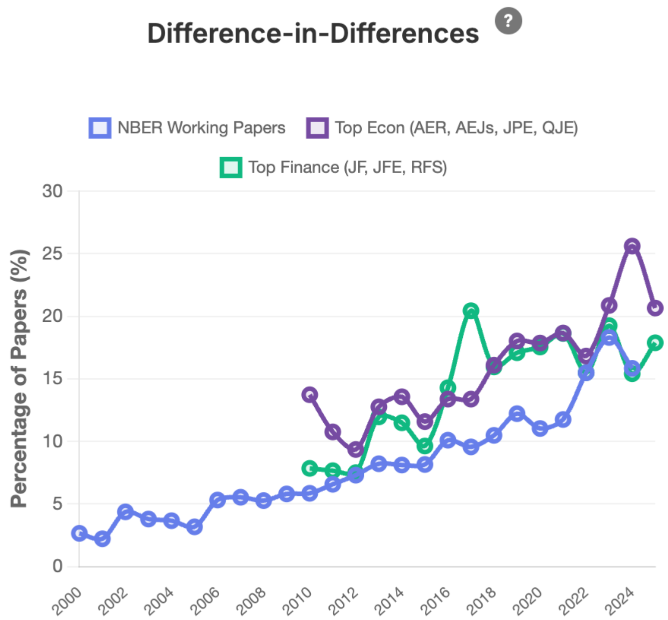
- Using this type of differential quasi-random treatment to draw **causal inferences** in a panel setting is very popular, and usually referred to as a **difference-in-differences** approach.

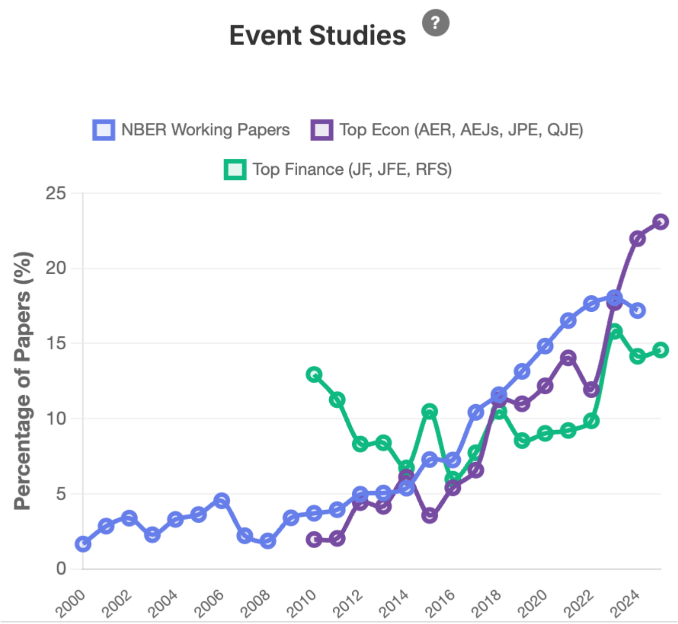
Difference-in-differences

- Intuition: DiD compares how outcomes change over time for a treated group versus a control group, and attributes the *extra* change in the treated group to the treatment.
- Key assumption: absent treatment, the treated group would have experienced the same change as the control group → the control group's change is a credible counterfactual for what would have happened to treated units.

Difference-in-differences

- Difference-in-differences is an extremely popular approach – probably the main “identification technology” used in recent years
  — also in finance – from https://paulgp.com/econlit-pipeline/dashboard.html:

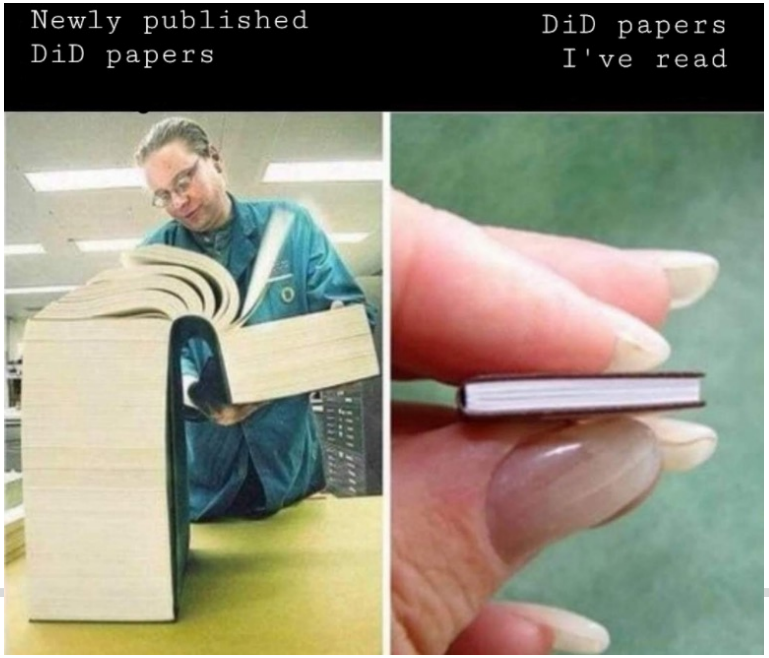




- Currie et al. (2020) note: “Over time, event studies have become almost synonymous with difference-in-differences: It is now rare to use difference-in-differences without showing an event study graph, and conversely it is rare to show event studies without a control group.”

Recent developments in DiD

- The popularity of DiD is likely due to the fact that it is very intuitive. And for a while, applied econ/finance researchers could follow a “standard playbook”.
- But: over the last ~7 years, **many** new methodological developments in this area. Will mention some of them, although cannot possibly cover them all.



Recent developments in DiD

- Some recent surveys (will refer to the "*" ones later):
  - * Baker, Andrew, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon and Pedro H. C. Sant'Anna. 2025. Difference-in-Differences Designs: A Practitioner's Guide. *Journal of Economic Literature.*
  - de Chaisemartin, Clément and Xavier D'Haultfoeuille. 2023. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *Econometrics Journal*.
    - These authors are also working on a textbook on DiD – draft available at https://ssrn.com/abstract=4487202
  - Miller, Douglas L. 2023. An introductory guide to event study models. *Journal of Economic Perspectives*.
  - * Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics.*
- For a great 2-hour overview, see the 2023 NBER Methods Lecture by Jesse Shapiro and Liyang Sun:

  https://www.nber.org/conferences/si-2023-methods-lectures-linear-panel-event-studies

Overview

- Motivation
- **Difference-in-differences – the basics**
- Difference-in-differences – implementation
- Extensions: continuous treatment & staggered diff-in-diff

Notation (following Roth et al. 2023)

- Will start with the DiD estimator in the simplest, canonical case – binary treatment, two periods (pre/post)
  - example: the famous Card-Krueger (1994) minimum wage study comparing fast-food employment in NJ vs. PA before vs. after NJ increased the minimum wage

- Notation: D_i = 0/1: treated no/yes; t = 1 before treatm., t = 2 after

- Last lecture, briefly introduced "potential outcomes" – useful here:
  - Denote $Y_{i,t}(0)$ unit *i*'s potential outcome if untreated; $Y_{i,t}(1)$ if treated. Would ideally like Average Treatment Effect: $\mathrm{E}[Y_{i,2}(1) - Y_{i,2}(0)]$
  - Observed outcome is $Y_{i,t} = D_i Y_{i,t}(1) + (1 - D_i) Y_{i,t}(0)$
  - DiD can uncover **Average Treatment Effect on the Treated (ATT):**

$$\tau_2 = \mathrm{E}[Y_{i,2}(1) - Y_{i,2}(0)\,|D_i = 1]$$

Two key assumptions

- The challenge in estimating τ_2 is that the untreated potential outcome $Y_{i,2}(0)$ for the treatment group is never observed
- The central idea behind DiD is that we can use the untreated group to construct this counterfactual outcome

- This requires the **"parallel trends" (PT) assumption**:

$$\mathrm{E}[Y_{i,2}(0) - Y_{i,1}(0)\,|D_i = 1] = E[Y_{i,2}(0) - \;Y_{i,1}(0)\,|D_i = 0]$$

- In words: without treatment, the average outcome for the treated and untreated groups would have evolved in parallel
  — Note: this is about the *change* across the two periods, not the *level*

- DGP where this holds: $Y_{i,t}(0) = \alpha_i + \phi_t + \varepsilon_{it}$ with $E(\varepsilon_{it}|D_i) = 0$
  — Treatment can be related to α_i, but not to the trend ε_{it}

Two key assumptions

- Second key assumption is **“no anticipation”**:

$$Y_{i,1}(0) = Y_{i,1}(1) \text{ for all } i \text{ with } D_i = 1$$

- In words: in the pre-treatment period, getting *subsequently* treated does not affect outcome yet.

- What does this get us? Rearrange PT assumption from last slide:

$$\begin{aligned}
\mathrm{E}[Y_{i,2}(0)|D_i = 1] &= \mathrm{E}[Y_{i,1}(0)|D_i = 1] + \mathrm{E}[Y_{i,2}(0) - Y_{i,1}(0)\,|D_i = 0] \\
\text{(by no anticip.)}\quad &= \mathrm{E}[Y_{i,1}(1)|D_i = 1] + \mathrm{E}[Y_{i,2}(0) - Y_{i,1}(0)\,|D_i = 0] \\
&\equiv \mathrm{E}[Y_{i,1}|D_i = 1] \quad + \mathrm{E}[Y_{i,2} - Y_{i,1}\,|D_i = 0]
\end{aligned}$$

- So then we can identify

$$\begin{aligned}
\tau_2 &= \mathrm{E}[Y_{i,2}(1) - Y_{i,2}(0)\,|D_i = 1] \\
&= \underbrace{\mathrm{E}\left[Y_{i,2} - Y_{i,1}\,\middle|D_i = 1\right]}_{\textbf{change for treated}} - \underbrace{\mathrm{E}[Y_{i,2} - Y_{i,1}\,|D_i = 0]}_{\textbf{change for untreated}}
\end{aligned}$$

The difference-in-differences estimator

- To estimate

$$\tau_2 = \mathrm{E}\left[Y_{i,2} - Y_{i,1} \,\middle|\, D_i = 1\right] - \mathrm{E}[Y_{i,2} - Y_{i,1} \mid D_i = 0]$$

we use the sample analogue:

$$\widehat{\tau_2} = (\bar{Y}_{t=2,D=1} - \bar{Y}_{t=1,D=1}) - (\bar{Y}_{t=2,D=0} - \bar{Y}_{t=1,D=0})$$

- Example from Card and Krueger (1994), where NJ was “treated” with an increase in minimum wage:

	Stores by state		
Variable	PA (i)	NJ (ii)	Difference, NJ – PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

Intuition – single differences

- Consider observing the treated group only, and attempting to estimate the treatment effect as

$$\bar{Y}_{t=2,D=1} - \bar{Y}_{t=1,D=1}$$

  — This would "work" only if without the treatment, the expected outcome would have remained unchanged between the two periods.

- Conversely, consider observing only the post-period, and attempting to estimate the treatment effect as

$$\bar{Y}_{t=2,D=1} - \bar{Y}_{t=2,D=0}$$

  — This would "work" only if without the treatment, the expected outcome would have been identical across the two groups – implausible unless treatment was fully randomly assigned.

- (See Roberts-Whited for more formal discussion of single-diff cases)

Difference-in-differences

- Rather than calculating means manually, we commonly use the regression version of the difference-in-differences estimator:

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 (d_i \times p_t) + u_{i,t}$$

- p_t equals 1 if period t is after treatment, and zero otherwise
- d_i equals 1 if unit i is in treated group, and zero otherwise.

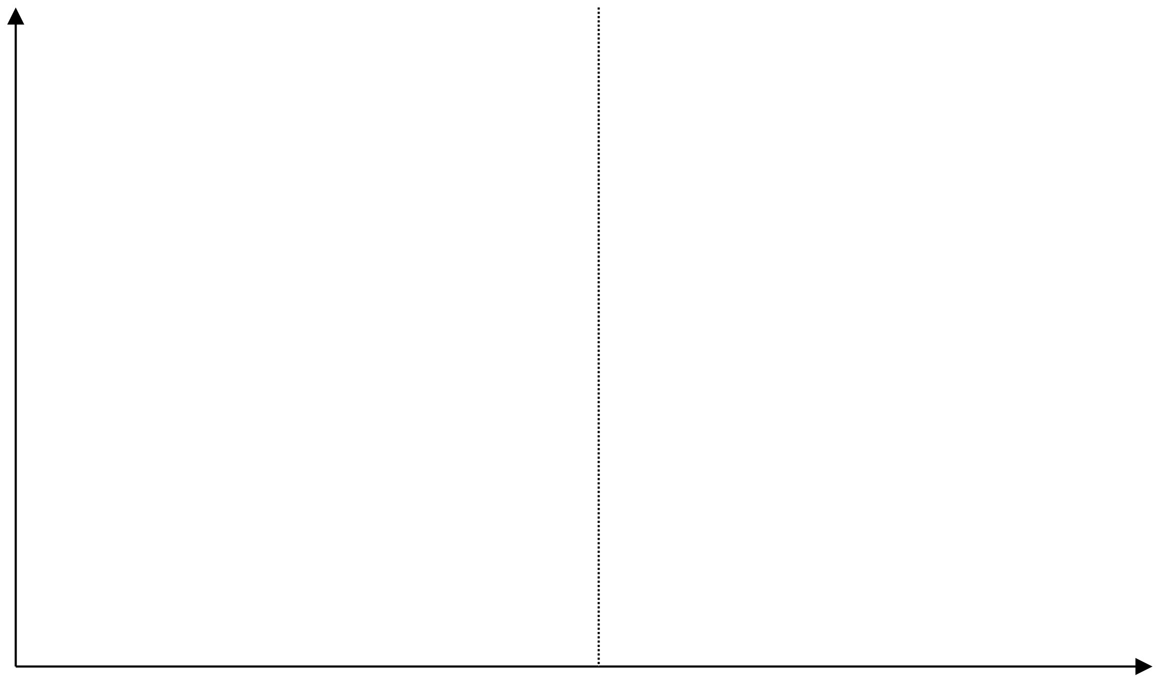
  — $\boldsymbol{\beta_1}$**:** measures average change in y due to trends common to both treated and control units

  — $\boldsymbol{\beta_2}$**:** measures average difference in level of y between treated and control units in the pre-treatment period.

  — $\boldsymbol{\beta_3}$**:** measures the average differential change in y from the pre- to post-treatment period for the treatment group **relative** to the change in y for the control group → can easily be shown to equal τ_2

Difference-in-differences

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 (d_i \times p_t) + u_{i,t}$$



Difference-in-differences

- The difference-in-differences estimator takes care of two major **identification threats**:

1. Any permanent, **time-invariant, difference between the treatment and control groups** is differenced away by inclusion of the d indicator variable.
2. Any **common trend affecting both the treatment and control group** is also differenced away by inclusion of the p indicator variable.

- Threats to the validity of the diff-in-diff estimator cannot come from permanent differences between the treatment and control groups, or shared trends.

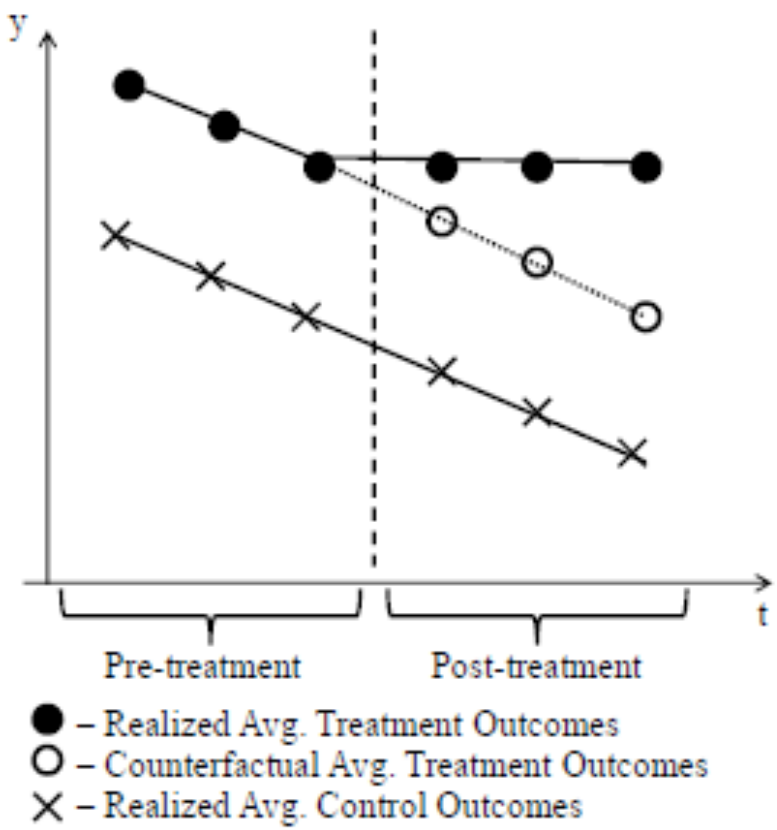
More on parallel trends

- As seen above, the crucial assumption with diff-in-diff is the **“parallel trends” assumption:** In the absence of treatment, the average change in the outcome variable would have been the same for both the treatment and control groups.
  - Formally in the regression version:
    $cov(d, u) = cov(p, u) = cov(dp, u) = 0$
  - Inherently untestable but can provide supportive evidence (later)

- If have multiple pre-treatment periods: requires trends in the outcomes for the treatment and control groups prior to the treatment to be the same.

Parallel trends assumption

Figure 1: **Difference-in-Differences Intuition**

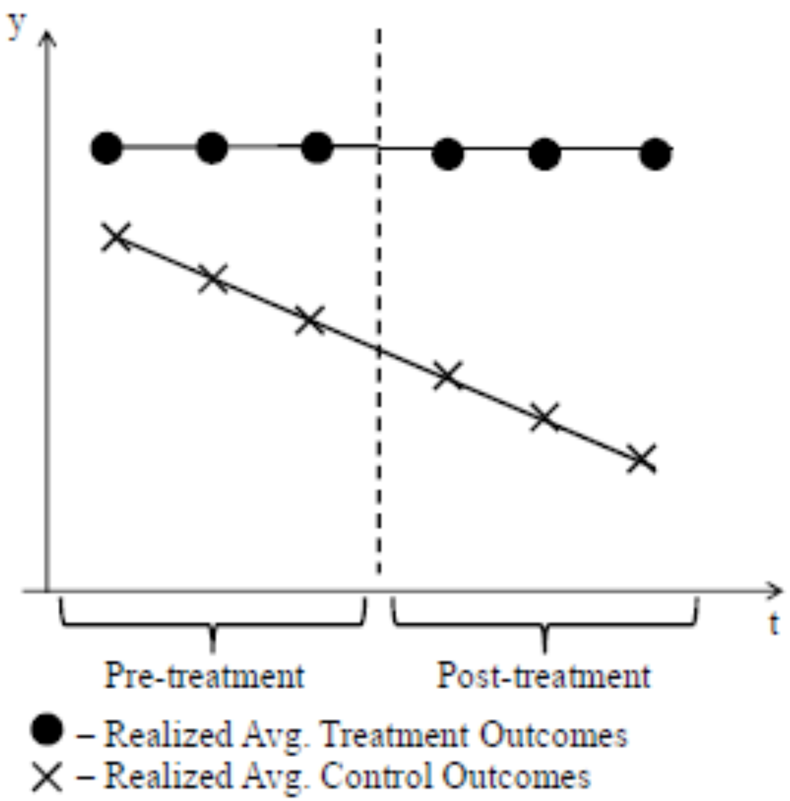


Source: Roberts and Whited (2012)

- Here, both series are trending, but that's not a problem, as long as trends in pre-treatment period are parallel
    - "kink" after treatment is what identifies the effect
- Level differences in pre-period are also fine but mean that the estimation may be sensitive to functional form assumptions (e.g. log(y) vs. y – relative vs. absolute changes)
    - also, with trends as depicted, could potentially estimate the model in first differences as well

Parallel trends assumption

Figure 2: **Violation of the Parallel Trends Assumption**



Source: Roberts and Whited (2012)

- Pattern depicted here is problematic for standard DiD estimator
  - Will estimate a large treatment effect just due to differential trends
- Absence of parallel pre-trends makes simple DiD estimator essentially "unusable"
  - But will return to this below
- Particular worry: treatment that happens **in response to** evolution before the treatment
  - "Squeaky wheel gets the grease"

SUTVA

- Another, “hidden”, key assumption is the so-called “Stable Unit Treatment Value Assumption”, or **SUTVA**

- Essentially, it means that a unit’s potential outcome is unaffected by **the treatment assignment of other units**
  - aka “no interference”
  - $Y_{i,t}(D_i)$ does not depend on $\{D_j, j \neq i\}$

- In corporate finance / banking settings, **often not realistic**, since firms interact with each other (directly or via market)
  - e.g. law change in one state may affect firms in other states if their products are substitutes

- Violations will lead to biased estimated treatment effects, although can often argue that sign of effect not affected

SUTVA

- Berg, Reisinger and Streitz (2021) contain thorough discussion of this issue, and advice for researchers
- So far the issue has been mostly ignored in finance settings, but this will likely change going forward
- Always worth keeping in the back of your mind when considering diff-in-diff designs – your life is much easier when you can argue that SUTVA holds
- (Related: discussion of general eqm effects – see e.g. Nakamura and Steinsson, JEP 2018, for discussion in macro context.)

Overview

- Motivation
- Difference-in-differences – the basics
- **Difference-in-differences – implementation**
- Extensions: continuous treatment & staggered diff-in-diff

General diff-in-diff estimator

- In a panel data set, we can add **firm and time fixed effects**:

$$y_{i,t} = \beta_0 + \beta_3(d_i \times p_t) + \alpha_i + \delta_t + u_{i,t}$$

Firm fixed effects control for the treatment assignment

Time fixed effects control for trends and subsume p_t

- This model may improve precision and fit
  - Intercept is allowed to vary by firm – may matter especially if have unbalanced panel
  - Allows common change in y to vary by time period (e.g. year)

Inference in diff-in-diff

- Classic reference is Bertrand, Duflo, Mullainathan (2004). They discuss three approaches:

1. Block bootstrap
2. Collapse into pre/post
3. **Clustering at the group level** – most common

— With small number of clusters, even wild bootstrap may no longer work (e.g. famous Card-Krueger minimum wage study had just two groups: NJ and PA).

— In such a case, may need to use randomization inference (see Cunningham section 4.2; Hagemann, JoE 2019, and other papers by same author)

The event study chart

- An extension that is very commonly done, if there are multiple time periods pre- and post-treatment, is the **"event study" / "dynamic diff-in-diff":**

$$y_{i,t} = \alpha_i + \delta_t + \sum_{\ell=-q}^{-2} \gamma_\ell \, \mathbb{I}(t - T_{i0} = \ell) + \sum_{\ell=0}^{m} \lambda_\ell \mathbb{I}(t - T_{i0} = \ell) + u_{i,t}$$

  - treatment starts at time T_{i0} (may differ across *i*)
  - the γ_ℓ coefficients show differential evolution prior to the treatment. Would like those to be close to zero – **"no differential pre-trends"**
  - the λ_ℓ coefficients show differential evolution after treatment – ideally "monotonic" and statistically significant (at least over some period). If effect only happens "late", sheds doubt on validity.
  - some flexibility as to which period is omitted (most common: -1, but also see 0 or start of pre-period)

Event study chart – examples

- Cunningham book, Section 9.4.3 discusses study by Miller et al. (QJE 2021) on how expansion of Medicaid in some states affected mortality. Time unit: years.
  - X-axis = “event time” because policy adoption staggered across states: 21 states expanded Medicaid in 2014, 3 states in 2015, 2 states in 2016, and 1 state in 2017. Cf. later

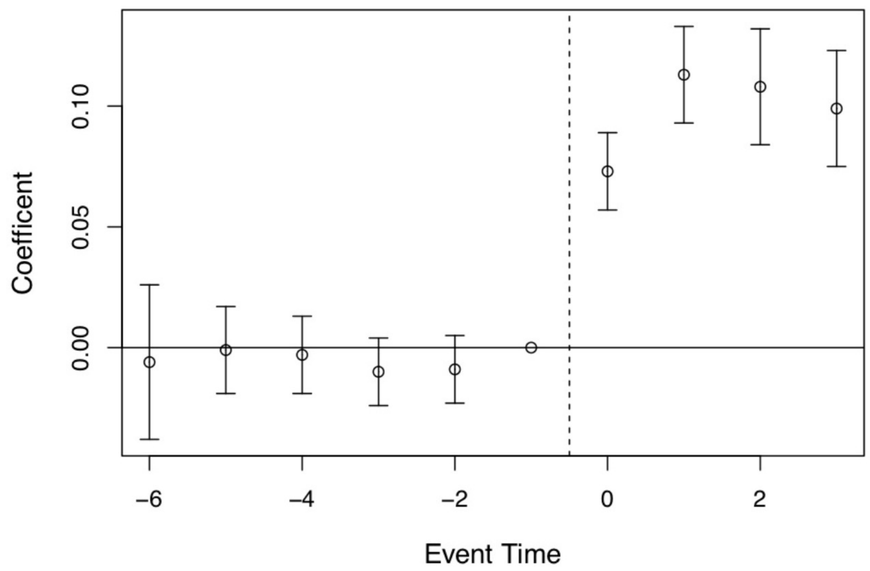


Figure 9.5: Estimates of Medicaid expansion’s effects on **coverage** using leads and lags in an event study model. Reprint from Miller et al. (2019).

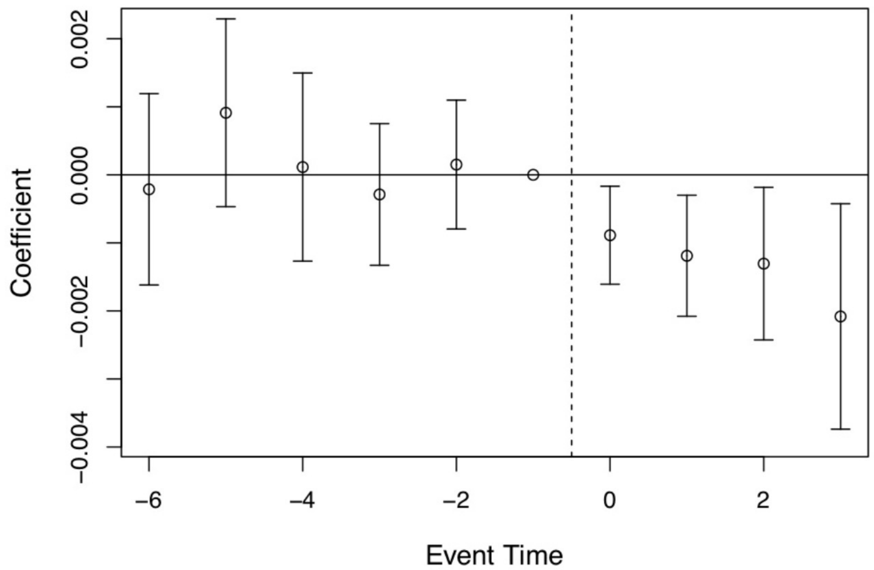


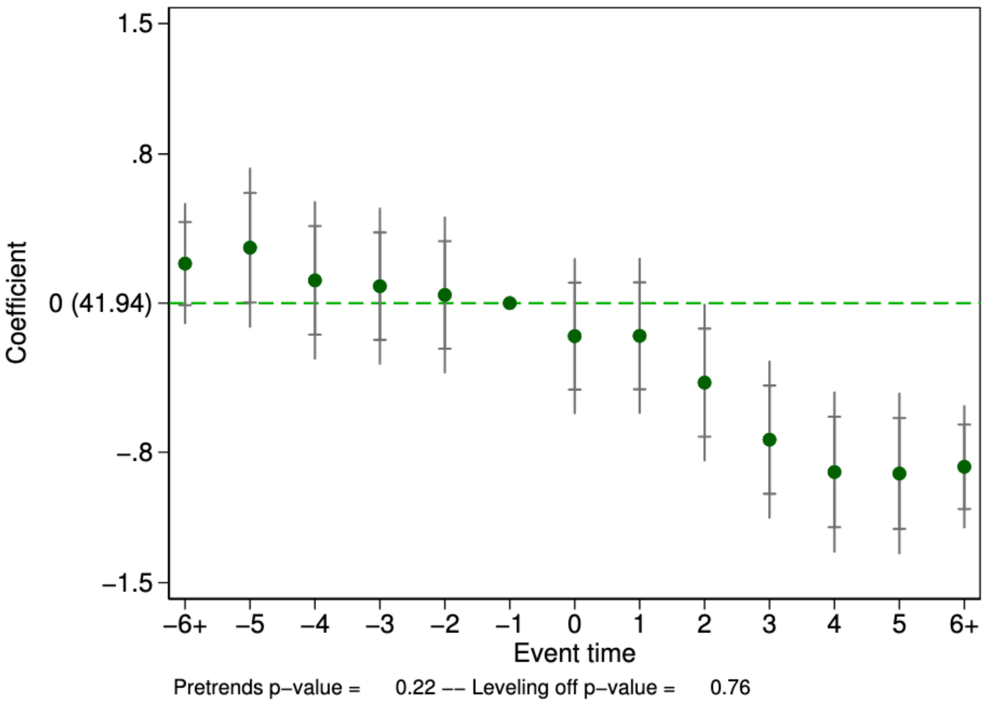
Figure 9.7: Miller et al. (2019) estimates of Medicaid expansion’s effects on on **annual mortality** using leads and lags in an event study model

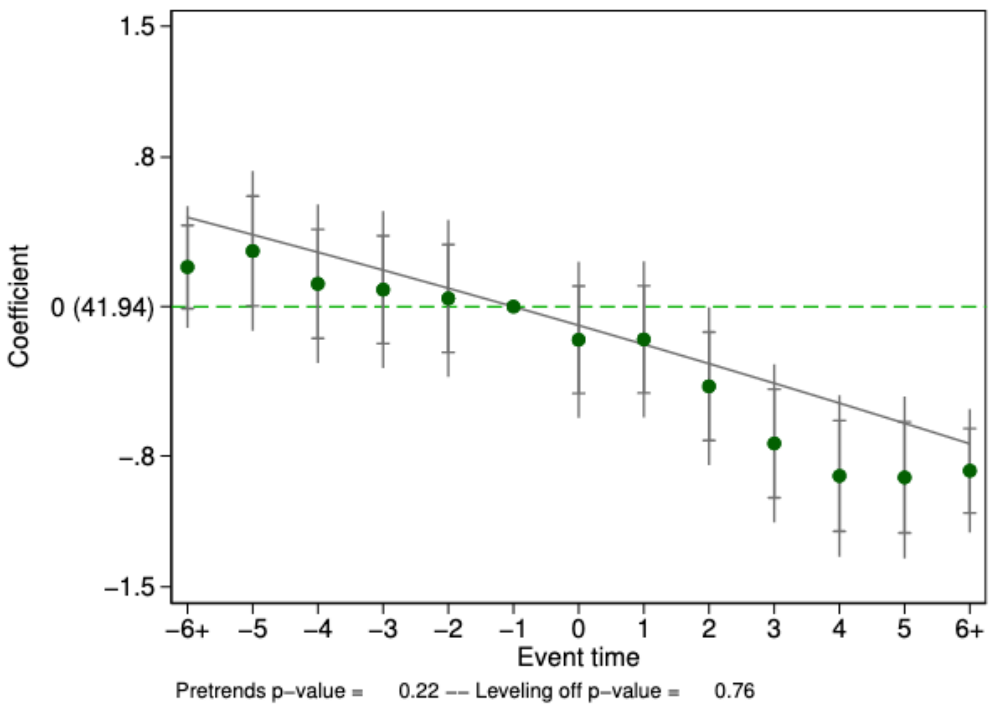
Event study charts (and estimation) – “state of the art”

- Freyaldenhoven, Hansen, Pérez and Shapiro (2021) (https://jorgeperezperez.com/files/EventStudy.pdf) provide various recommendations for making event study plots more informative
  - incl. video series on Youtube and package (“xtevent” in Stata or “EventStudyR”)
- For instance:
  - include a label for the mean of the pre-period
  - plot “uniform sup-t” confidence bands for the path of the effect, in addition to pointwise confidence intervals
    - see also https://ryanedmundkessler.github.io/software/
  - add *p*-values for “no pre-trends” and “dynamics levelling off”
  - plot path of the “least wiggly” confound whose path cannot be rejected
- Note: they focus on cumulative effects of a policy, which may or may not be what we want to show

Event study charts (and estimation) – “state of the art”

- Examples from Freyaldenhoven et al.:





Charts illustrate an important issue with pre-trend testing: Limited power against alternatives (often do not reject the null of no pre-trend, but would also not reject the null of *some* pre-trend)

Notes on parallel trends / pre-trends

- Until recently, the consensus was: “no significant pre-trends” = good; “significant pre-trends” = you can’t run your DiD

- This consensus is starting to shift – see Freyaldenhoven et al. (AER 2019) & work by Jonathan Roth (https://jonathandroth.github.io/papers/)

- In particular “A More Credible Approach to Parallel Trends” (Rambachan & Roth, 2023) – propose tools for robust inference in DiD settings where parallel trends may be violated.
  - E.g. consider restriction that the magnitude of the post-treatment violation of parallel trends can be no larger than a constant $\bar{M}$ times the maximal pre-treatment violation
  - Then, could report e.g. that the conclusion of a positive treatment effect is robust up to the value $\bar{M} = 2$.
  - Packages for Stata and R: https://github.com/mcaceresb/stata-honestdid#honestdid

The role of additional covariates (based on Baker et al., JEL 2025, Section 4)

- The outcomes $Y_{i,t}$ that we study are usually also affected by other variables $X_{i,t}$ aside from the treatment
- Therefore, one would ideally like the relevant $X_{i,t}$ to be balanced across treated and control group
  - — if they are not, this casts doubt on the PT assumption

- Common diagnostic: Normalized difference in pre-treatment means:

$$Norm.Diff. = \frac{\overline{X_T} - \overline{X_C}}{\sqrt{S_T^2 + S_C^2}}$$

- Rule of thumb: $|Norm.Diff.| > 0.25$ is potentially problematic
- Can also check balance in *changes* in covariates from pre to post, although these may themselves be affected by treatment
- If you have balance, then adding covariates shouldn't matter much for your estimates (but you may gain some precision)

The role of additional covariates (based on Baker et al., JEL 2025, Section 4)

- What to do if your groups are *not* balanced?
- Appeal to **conditional parallel trends** assumption :

$$\mathrm{E}[Y_{i,2}(0) - Y_{i,1}(0) \mid X_i, D_i = 1] = E[Y_{i,2}(0) - Y_{i,1}(0) \mid X_i, D_i = 0]$$

- This requires "common support" or "strong overlap", i.e. for any X_i, there are some treated units and some untreated ones

- How to estimate? Most commonly, researchers just add the **pre-treatment** covariates $X_{i,1}$ to the TWFE regression from above
  - sometimes interacted with the post dummy (p_t)
  - **! Don't control for variables that the treatment can affect !**

- **BUT!** This is not innocuous – with covariates, TWFE regression only recovers ATT under the (arguably restrictive) assumption that the treatment effect is identical across covariate groups

The role of additional covariates (based on Baker et al., JEL 2025, Section 4)

- There exist estimators that have better properties in such a situation:
  1. "Regression adjustment" (RA): Using the *control units only*, estimate how outcome changes relate to X, then impute this for treated units
  2. "Inverse probability weighting" (IPW)
     - estimate propensity score p_i (= probability of being treated) as a function of $X_{i,1}$, e.g. via a logit
     - then give more weight to control units with higher p_i ("closer" to treated)
     - need to check that p_i not too close to 1, otherwise can be unstable; in that case may need to trim these observations
  3. "Doubly robust" estimator: ~combines RA and IPW, and gives a consistent estimator if at least one of the auxiliary models is "right"
     - e.g. "dsdid" or "csdid" in Stata

Rarely seen in (published) finance papers so far, but this may change

Checking validity

- Other common validity checks that researchers often perform:

1. **Placebo tests #1:** Repeat the diff-in-diff analysis on pre-event years. That is, falsely assume that the onset of the treatment occurs one, two, three years before it actually does. The treatment effect should be statistically indistinguishable from zero.
    - event study chart essentially does that visually

2. **Placebo tests #2:** Make sure that variables that **should be unaffected by the event are unaffected by the event**. Replace the outcome variable of interest in the empirical model with these other variables.

Checking validity

3. **Diff-in-diff-in-diff (aka triple-differences):** can be seen as either another placebo (if there are subgroups that should not be affected by treatment) or as a test of mechanisms/ channels (e.g. some firms should be more affected than others)
   — as triple interactions can be hard to interpret, some authors prefer to do sample splits (and run DiD in subsamples)

4. **Treatment reversal:** If there is a reversal of the treatment, it should cause a return to the pre-treatment behavior.

Overview

- Motivation
- Difference-in-differences – the basics
- Difference-in-differences – implementation
- **Extensions: continuous treatment & staggered diff-in-diff**

Extensions of basic framework

- Two types of extensions of the basic DiD framework are commonly considered – traditionally without much discussion of underlying assumptions, but this has changed recently:

  1. Continuous treatment/exposure instead of binary treatment

  2. Staggered treatment (rather than single pre/post period)

- In both cases, recent literature has emphasized that our standard TWFE estimators can be problematic **if treatment effects are heterogeneous** across units or over time (which is often plausible)

Continuous treatment

- Rather than having treated/untreated, in many settings the treatment is continuous – or different units get varying “doses” of the treatment
  - — common when studying outcomes across locations with different shares of firms/households affected by some policy change

- Related to the Bartik IV design we discussed last time, but here we effectively run the “reduced form” only (and the “shifter” is pre/post rather than some aggregate variable)

- Common to just run same DiD regression with d_i continuous:

$$y_{i,t} = \alpha_i + \delta_t + \beta(d_i \times p_t) + u_{i,t}$$

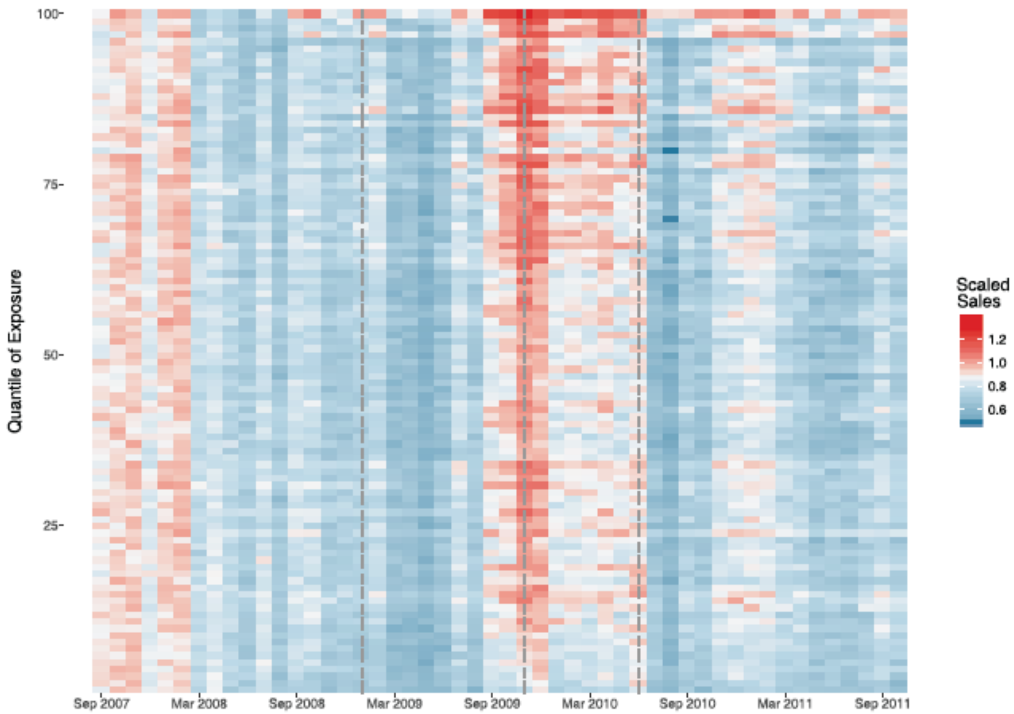
Continuous treatment

- Could also turn into dummies for above/below median exposure – that's typically done for charts
- But for estimation, appears more efficient to use all the variation
- However, also effectively assume that effect of "treatment dosage" is linear
- Useful to check that it is at least monotonic – and approximately linear
  - e.g. if have sufficient data, form dummies for quintiles/deciles and estimate effects for those
- Example: Berger et al. (JF 2020) on effects of first-time homebuyer credit in 2008-10 on home sales across zip codes

Berger et al. pre-trend & monotonicity “test”

- Exposure: “We define program exposure based on the number of potential first-time homebuyers in a ZIP code (…) [measured] as the year-2000 share of people in a ZIP code who are first-time homebuyers.”
- Effects on home sales over time by centile (b/c 9k ZIP codes):

Panel A. Difference-in-Differences Calendar Time Heatmap



Continuous treatment – caveat

- Recent work has shown that the TWFE estimator may not perform well in settings with **heterogeneous treatment effects.**

- In particular, Sun and Shapiro (2022) illustrate that when the treatment effects β_i are unit-specific, the regression

$$y_{i,t} = \alpha_i + \delta_t + \beta(d_i \times p_t) + u_{i,t}$$

may fail to recover a (weighted) average of these unit-specific effects and get a β estimate outside the range of all β_i (!)

  — they discuss that having some totally untreated units (i.e. $d_i = 0$) can help obtain a better estimator (intuitive – this helps to “anchor” the counterfactual)

  — see also Callaway, Goodman-Bacon and Sant'Anna (2024) for related discussion of issues with continuous treatments

Staggered diff-in-diff

- Often in finance, treatment doesn't happen for all units (e.g. states) at the same time – adoption over several years
- Then, "post" varies across units – a **"staggered" DiD**
  - — may have "always treated" or "never treated" units
- Very common – from Baker, Larcker and Wang (2022, "BLW"):

Table 1
Use of DiD and Staggered DiD in Finance and Accounting: 2000–2019.

	(1) DiD	(2) Staggered DiD
Journal of Finance	52	30
Journal of Financial Economics	163	85
Review of Financial Studies	138	75
Review of Finance	27	14
Journal of Financial and Quantitative Analysis	51	32
Finance	**431**	**236** (55%)

Staggered diff-in-diff

- While researchers approached this essentially in the same way as standard diff-in-diff – with TWFE regressions – a very active recent literature has pointed out potential issues with such designs

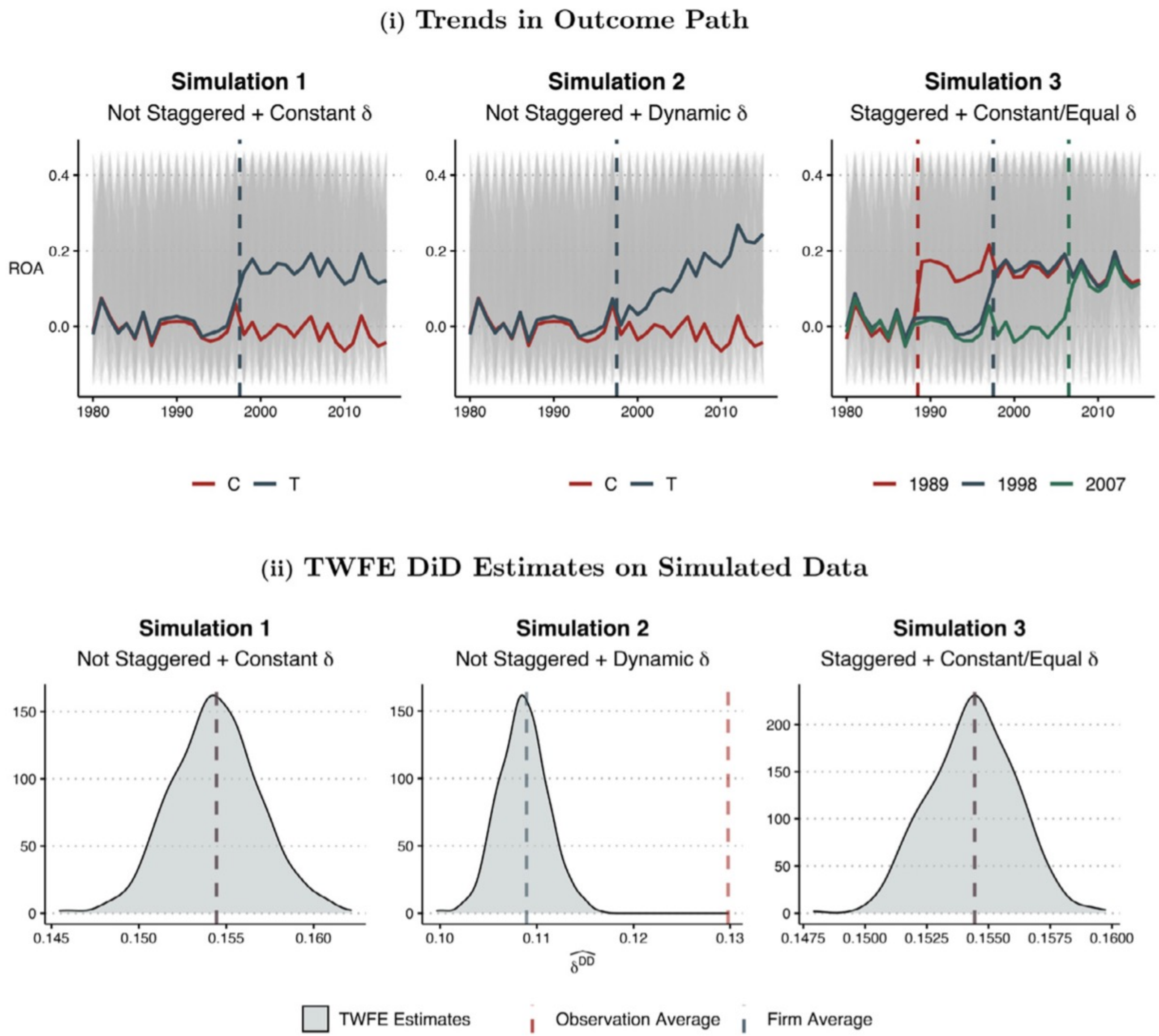
- These occur if treatment effects are **heterogeneous** – either across units or over time.

- Good entry point to the rapidly growing literature: BLW (2022), "How much should we trust staggered DiD?" (plus survey papers listed earlier, esp. Roth et al.)
  - https://asjadnaqvi.github.io/DiD/ provides links to various packages in Stata and R

BLW simulations – illustrating when there is a problem

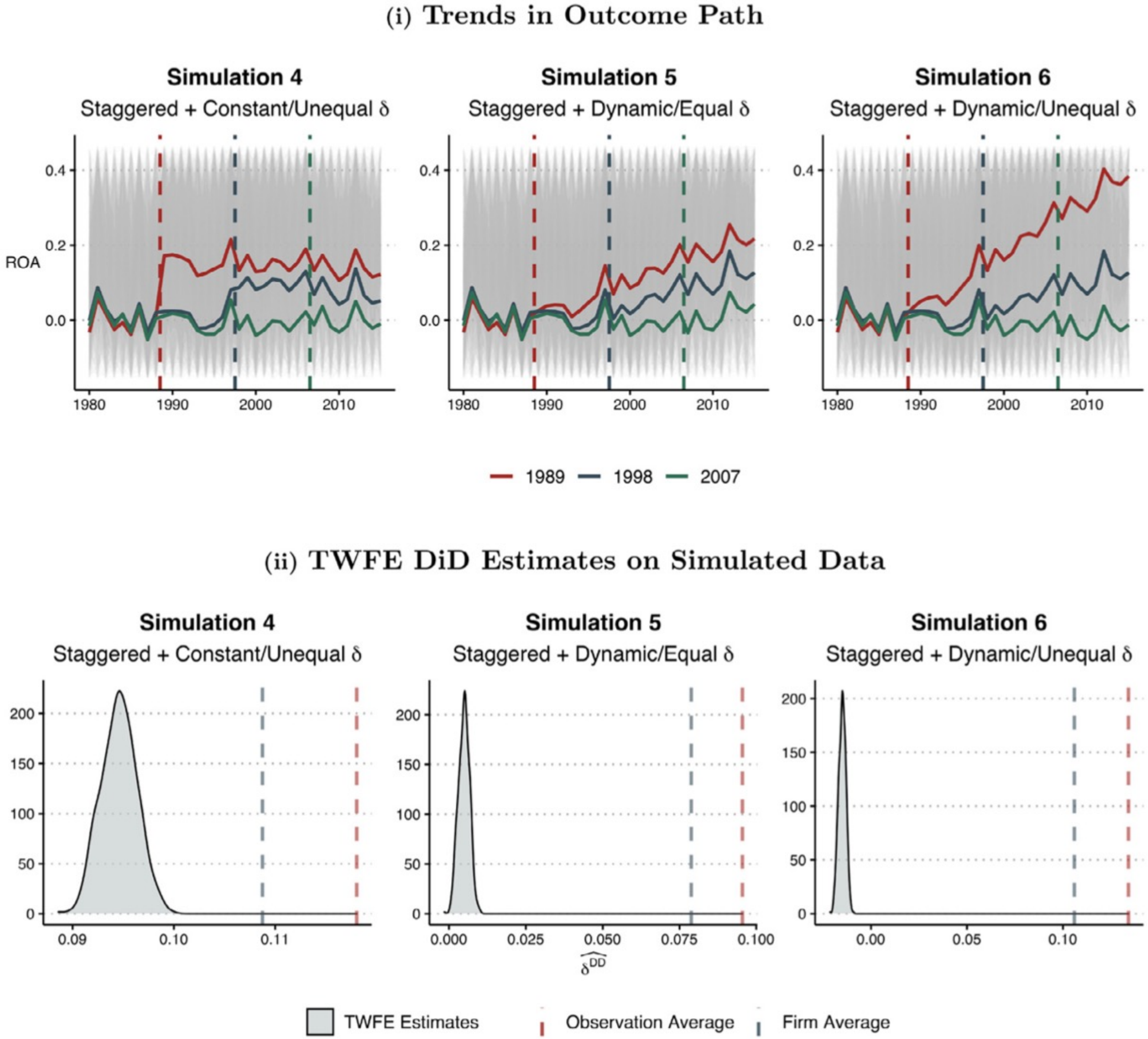
In these cases, TWFE DiD does “fine”:

(see paper for details about the simulations)



BLW simulations – illustrating when there is a problem

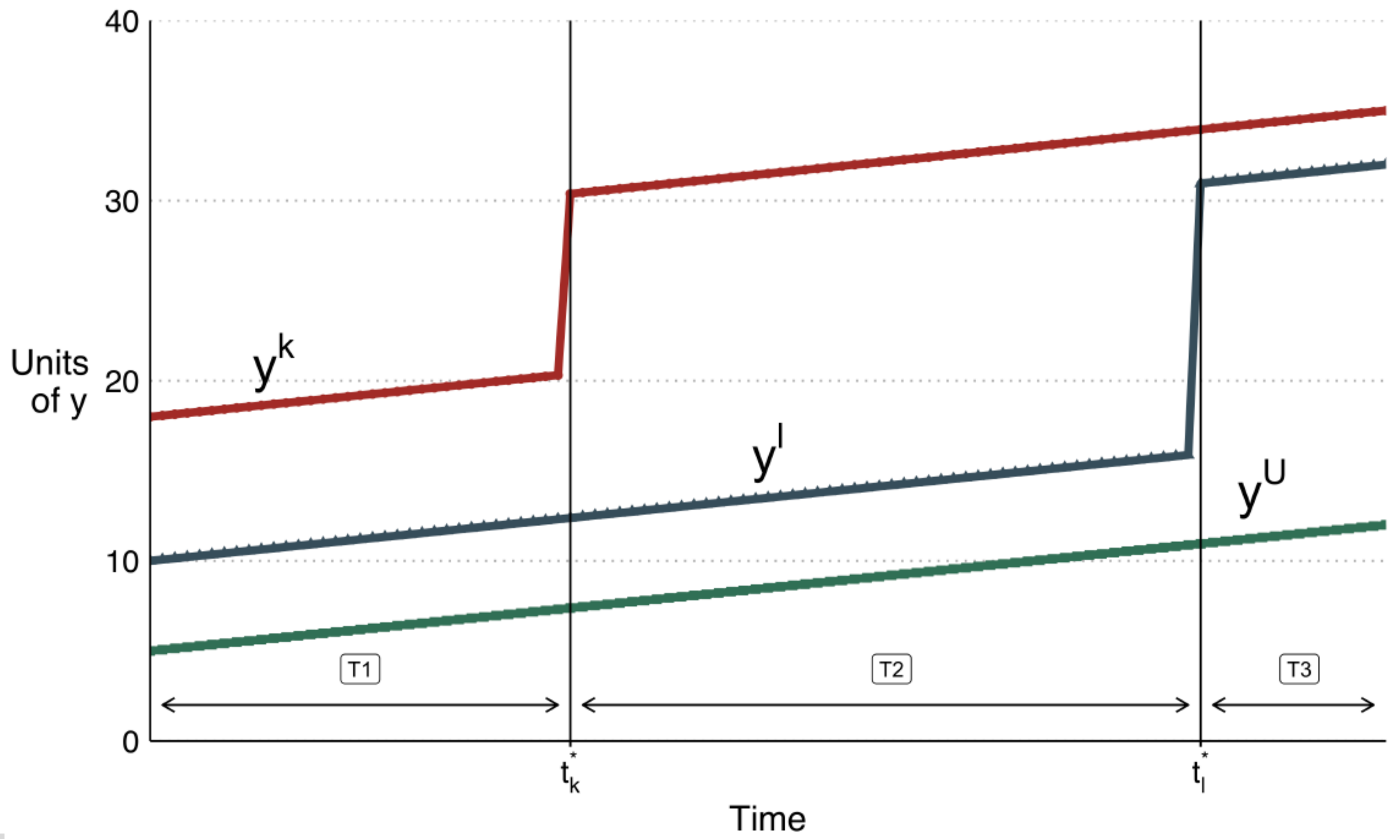
In these cases, TWFE DiD does much less fine – esp. Sim. 6, where the sign flips!



What's the problem? Goodman-Bacon (2021) decomposition

- Stylized setting with three groups – TWFE DiD is a weighted average of all possible two-group/two-period DiD estimators

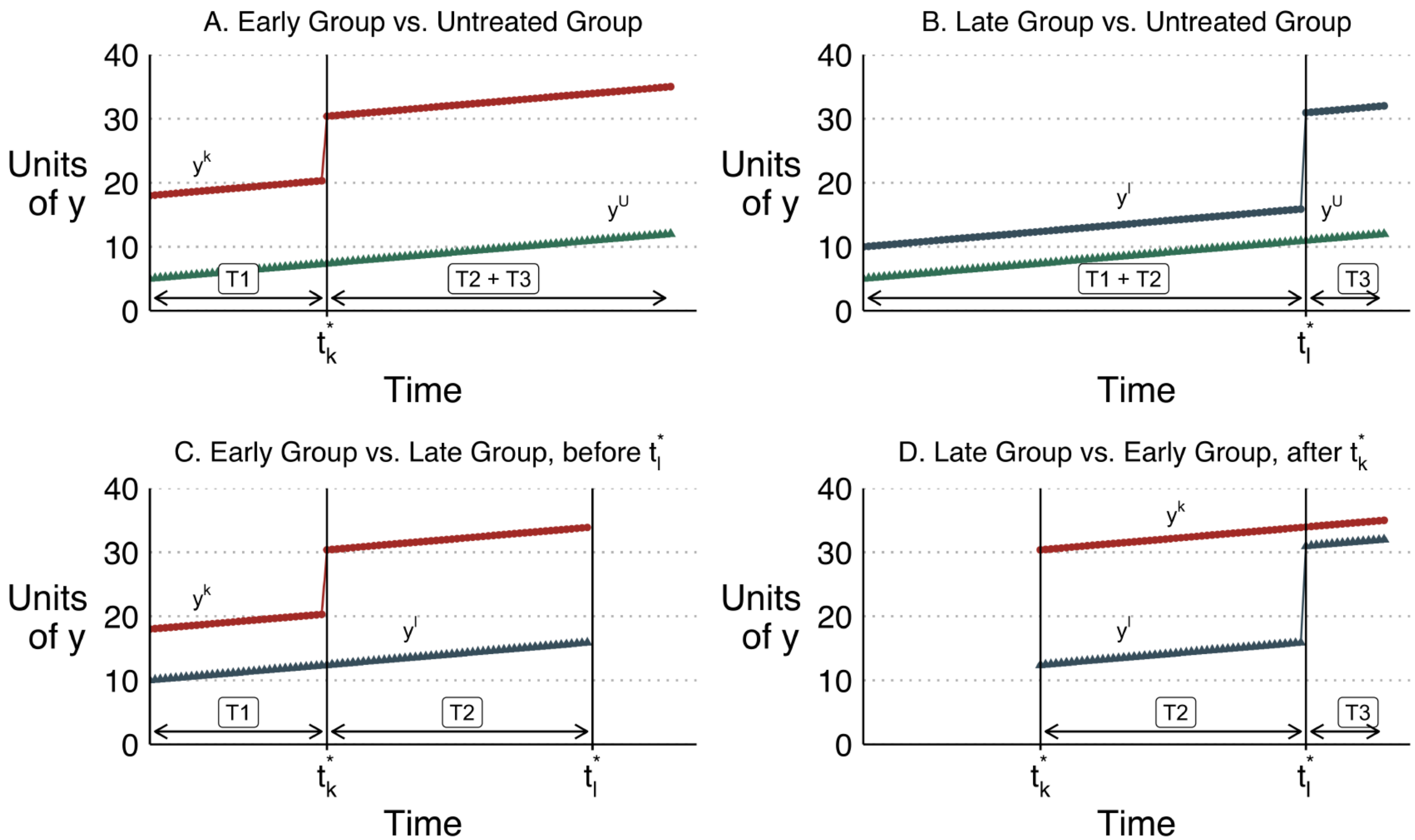
(i) Staggered treatment setting with three treatment groups.



What’s the problem? Goodman-Bacon (2021) decomposition

- Stylized setting with three groups – TWFE DiD is a weighted average of all possible two-group/two-period DiD estimators

(ii) Four constituent 2x2 designs.



Late vs. early

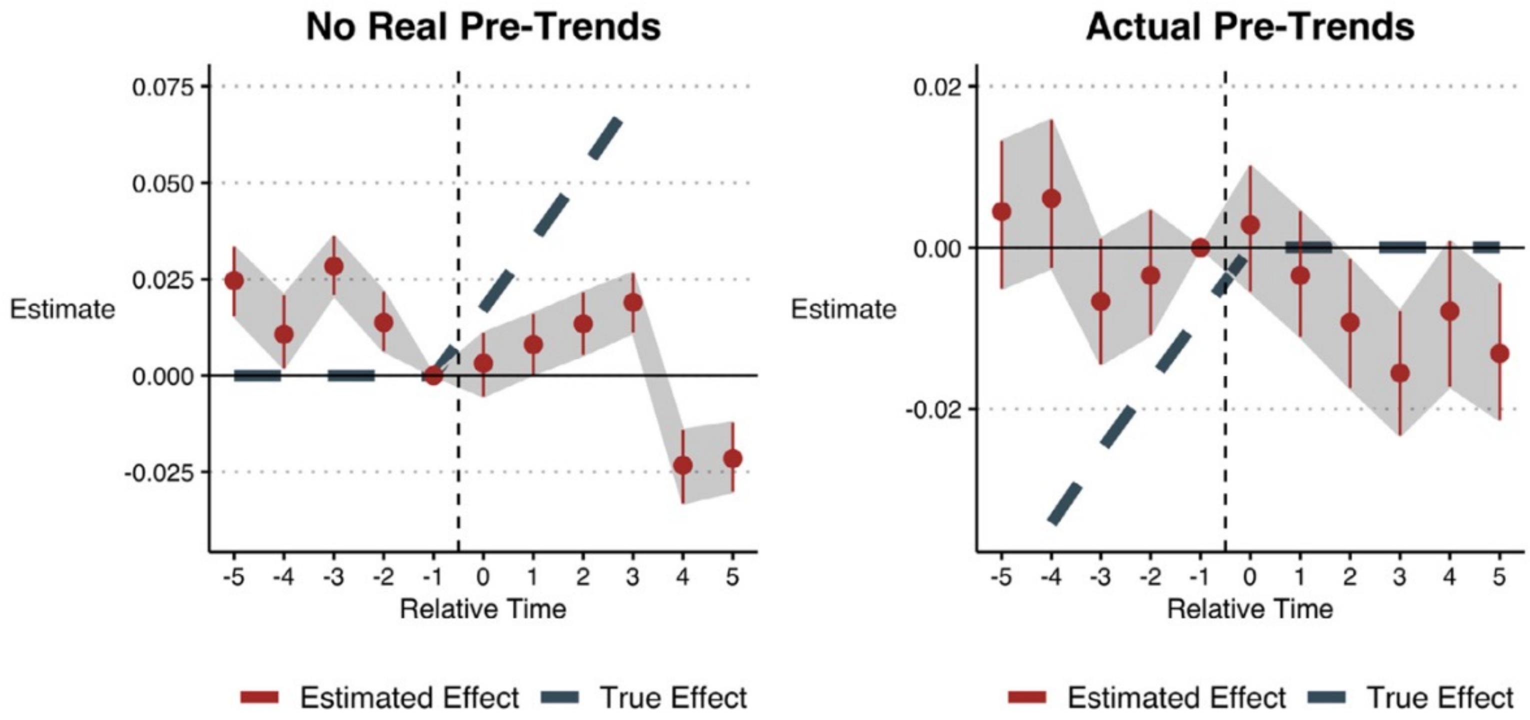
- What's the problem with Simulation 6? No untreated; "late vs. early" (constituent D) gets most weight, and "looks negative" due to weaker (but positive!) treatment effect
  - BLW: early-treated units as effective controls = "potentially problematic"

- What determines weights?
  - absolute size of the subsample
  - relative size of the treatment and effective comparison group in the subsample
  - timing of the treatment in the subsample
  - magnitude of treatment variance in the subsample

- Changing the panel length alone can change the staggered DiD estimate, even when each 2x2 DiD estimate is held constant.

Event study plots

- With staggering and treatment effect heterogeneity, standard event study plots can also be dramatically “off” – BLW Section 3.2
- Simulation evidence:



- **Binning** (e.g. <=−5, >=5) can have major effects on estimated path

Diagnostics & alternatives

- Goodman-Bacon (2021) proposes a useful diagnostic, namely the weight & average DiD of each of the 4 comparison types
- Then, different new estimators have been proposed, which effectively don't use early-treated units as controls for late-treated units – BLW section 4:
  - Callaway and Sant'Anna (2021) – BLW's recommendation
  - Sun and Abraham (2021)
  - Stacked Regression (used e.g. in Cengiz et al., 2019)
  - There are other alternatives – see next slide.
- BLW apply these methods to two published finance papers, finding in each case that results are not robust to the alternative estimators. (Though not clear how common that is.)

Imputation estimators

- Another type of alternative estimators are so-called “regression imputation” estimators – e.g. Borusyak et al. (2024) and Gardner et al. (2024)
- Basic idea very intuitive – 2-stage approach:
  1. Regress outcomes on unit and time fixed effects using *only the subsample of untreated observations* (incl. not-yet-treated)
  2. Based on this regression, impute a counterfactual $\hat{Y}_{i,t}(0)$ and estimate treatment effects relative to that

  — this yields unit-specific treatment effects that can then be aggregated to the ATT. See e.g. “did2s” package in Stata and R.

- Roth et al. (2023, sect. 3.3) compare assumptions underlying this estimator vs. Callaway-Sant’Anna and provide some guidance on which to use; see also Harmon (2024).

  — relative efficiency depends on serial correlation in errors

Event study plots with new estimators

s:fi

- A final thing to note is that the event study estimators from these new estimators are constructed differently from the TWFE event studies we are used to
  - illustrated for a particularly stark case in https://www.jonathandroth.com/assets/files/HetEventStudies.pdf

- What is particularly “confusing” is that the imputation estimators à la Borusyak et al. (2024) do not have an omitted base period
  - and in fact the pre- and post-treatment paths should not directly be compared

- Make sure to understand the plots generated by the method(s) you use (often depend on options chosen in the relevant software packages)

Summary and conclusion

- DiD is an extremely popular and intuitive methodology
- But methodological details have come under increased scrutiny in recent years, and this will certainly continue for a while – “new standards” are forming.
  - see e.g. BLW Section 6, Roth et al. (2023) and Baker et al. (2025) for a set of recommendations, but of course not the final word.
- Life is much easier with “standard” DiD than staggered DiD
  - on the other hand, there are many solutions available to the issues pointed out in the literature – by now expected to apply correctly
- In practice, also combinations of different “complications” – e.g. continuous and staggered. Often less guidance available so far.

Summary and conclusion

- A central issue in DiD is the parallel trends assumption – untreated (or later-treated) units provide the counterfactual for treated units
  - Another very active area of research

- If you don't seem to have parallel trends between treated and control units, one option is to re-weight control units to match pre-trends and/or characteristics of treated units – this is the **synthetic control approach** (see Abadie, JEL 2021 and NBER methods lectures 2021 for a recent overview)
  - useful in particular if only have one treated unit
  - won't cover here but worth learning about! Example of a published paper that uses it is Zevelev (RFS 2021)