FNCE 9260: Empirical Methods in Corporate Finance

Class 3: Regression Part 3 (DiD Extensions)

Daniel Garrett

Outline

Outline

Reading and Extra Materials

- Today's primary material is derived from the following:
  - *Mostly Harmless* Ch. 5, 7
  - Roberts and Whited (2013) Sections 4.3+ and 6
  - Bruce Hansen "Econometrics" Chapter 24 (has many estimators/proofs)
  - "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature" by Roth et al.
  - "How much should we trust staggered difference-in-differences estimates?" by Baker et al.

Identification

- We observe data that is the result of some data generating process (DGP) that maps from parameters to the observed data (lecture 1, slide 8)
- Assuming that we have the right model (i.e., parameters exist to index the data's distribution), we are interested in estimating the parameters of the data generating process
- Identification is the mapping from theory into data and answering "Can one logically deduce the unknown value of the parameter from the distribution of the observed data?"
  - In *reduced-form* empirical work, our model is ultimately our set of "identifying **assumptions**"
  - The best identification arguments are constructive, coming from sources like plausible identifying variation (i.e., exogenous changes in a variable that is usually endogenous)

Average Treatment Effects

- For a discrete treatment, we are often interested in how that treatment D_i changes outcomes Y_i on average
- The primary object of interest is the **Average Treatment Effect** defined as $E[Y_{1i}] - E[Y_{0i}]$, or the difference in expected outcomes upon receiving the treatment
- The *fundamental problem of causal inference* is that only one of these two outcomes is ever observed (no observance of parallel universes)
- So, we are often left observing the difference in means $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] =$

$$\underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{ATT} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{selection effect}}$$

Selection/OVB

- In OLS (what we will use in my portion of the course), we often estimate the model

$$Y_i = X_i\beta + e_i$$

and the estimand identifies the following:

$$\beta^{OLS} = \beta + E[X_iX_i']^{-1}E[X_ie_i]$$

- If e_i is conditionally independent of X_i and centered at zero, this is simply β
- But, if the true model were $Y_i = X_i\beta + U_i\gamma + \varepsilon_i$ (single unobservable), then $e_i = U_i\gamma + \varepsilon_i$ and the estimand identifies

$$\beta^{OLS} = \beta + \frac{Cov(X_i, U_i)\gamma}{Var(X_i)}$$

Fixed Effects: The 'Within' Transformation

- The true model is:

$$y_{it} = \alpha + \beta x_{it} + \delta f_i + \nu_{it}$$

We can subtract the average outcome for each unit from the true data generating process:

$$y_{it} - \bar{y}_i = \alpha - \alpha + \beta x_{it} - \beta \bar{x}_i + \delta f_i - \delta f_i + \nu_{it} - \bar{\nu}_i$$

where bars indicate means across time within unit.

- We are left with the following transformed regression equation where we've accounted for f_i but it drops out!

$$(y_{it} - \bar{y}_i) = \beta(x_{it} - \bar{x}_i) + (\nu_{it} - \bar{\nu}_i)$$

- Can run this version (adjusting DoF) or run pooled OLS with dummy variables for FE

Generalized Difference-in-Differences

- Generalized DiD models include time and unit FE:

$$y_{i,t} = \delta_t + \gamma_i + \beta_3(d_i \times p_t) + \nu_{i,t}$$

- Treatment doesn't have to be discrete, however, continuous treatment has some costs and benefits:
  - **Advantages:** (1) better use of all variation available in the data (2) interpretable magnitudes
  - **Disadvantages:** (1) makes strict functional form and measurement assumptions (2) influenced by treatment outliers

- Generalized DiD is often used with staggered treatments, which is where we will start today

Extraneous Note #1

- Standard errors should generally be clustered at the cross sectional unit level
- Draws of the same firm over time are *certainly correlated* in terms of the residual in most imaginable model
- See discussion in Bertrand, Duflo, and Mullainathan (2004) ("How much should we trust difference-in-differences estimates?")
- Modern solution is cluster-robust SE with clusters defined at unit of treatment (and sometimes unit and time) level
  - See Cameron, Gelbach, and Miller (2011) for multi-way clustering discussion

Extraneous Note #2

- Unbalanced panels and missing data are a feature of most corporate finance problems (e.g., firms in compustat change over time)
- Some potential solutions depend on why the data are missing:
  - Run a version of the analysis in a dataset that aggregates if data are missing due to entry/exit (Curtis, Garrett, Ohrn, Roberts, and Serrato 2023; Oberfield and Raval 2021)
  - FE for units takes out level impacts of changing composition (Garrett 2021; Giroud and Mueller 2010), need to check entry/exit, firm-cohort in Gormley and Matsa (2011)
  - DFL weighting to force composition to be constant wrt observables (Zwick and Mahon 2017)
  - Ignore the existence of missing data (many older Compustat papers)
  - Run a version of the analysis on a balanced panel (robustness/sensitivity)
- No perfect solution to my knowledge, be careful as **each of these can change the interpretation** *[why?]*

Extraneous Note #3: Different Common Panel Models

- There are a bunch of different panel estimators (many for computational tractability)
  - FE and 'within' estimator mean the same thing
  - Random effects (RE) assumes that unobserved heterogeneity is uncorrelated with observables/controls/treatments [this is not realistic, but could increase efficiency if it was]
  - Between effects (BE) regresses average outcomes on average covariates using OLS
  - The RE estimator is a matrix-weighted average of FE and BE
- FE is the only one I've seen in a paper published since 2010

Extraneous Note #4

- DiD treatment may not be immediate, or symmetric for treatment being added and taken away
- These can be very interesting parts of the economic story
- First differences is very different than other FE methods in this sense: assumes full treatment effect happens on first period
- It is now computationally easy to show these effects by year, and this should be done and shown graphically

$$y_{i,t} = \delta_t + \gamma_i + \sum_{k=0}^{T} \beta_k (treat_i \times \mathbb{1}(t = k)) + \nu_{i,t}$$

Outline

DiD in Most Modern Uses

- Many generalized DiD papers use "staggered implementation," treatment doesn't happen to all units at the same time
- **The benefits?** One can more credibly argue that
  1. Any unobserved, time varying factor correlated with treatment would need to be similarly correlated with a bunch of treatments over time
  2. Repeated treatments of differing sizes can give some ideas about the missing intercept problem (Not widely done)
- **The costs?** You will not get ATT with OLS, but new estimators offer a good solution

DiD in Most Modern Uses

- Many generalized DiD papers use "staggered implementation," treatment doesn't happen to all units at the same time
- **The benefits?** One can more credibly argue that
  1. Any unobserved, time varying factor correlated with treatment would need to be similarly correlated with a bunch of treatments over time
  2. Repeated treatments of differing sizes can give some ideas about the missing intercept problem (Not widely done)
- **The costs?** You will not get ATT with OLS, but new estimators offer a good solution

DiD in Bertrand and Mullainathan (2003): Business Combination (BC) laws

- It is notoriously hard to measure manager agency problems in the wild
- Managers differ in many observable (tenure, capital structure choices, etc) and unobservable (skill, preferences) ways that makes comparison tricky
- **One strategy**: find quasi-exogenous changes in the incentive set for managers to see how behavior changes
- The natural experiment in Bertrand and Mullainathan (2003) comes from the creation of "Business Combination" (BC) laws
  - States began passing laws to limit outside takeovers after the 1968 Williams Act
  - SCOTUS ruled these laws unconstitutional, so a second set of laws slowly passed in the 80s-90s
  - BC laws impose a 3-5 year moratorium on certain transactions between a target and an acquirer unless the target board specifically allows

DiD in Bertrand and Mullainathan (2003): Business Combination (BC) laws

- How do managers change behavior when there is a sudden drop in takeover threat?
- Bertrand and Mullainathan (2003) get plant level data from the Census and use a generalized DiD looking at a bunch of different outcomes

$$y_{jklt} = \alpha_t + \alpha_j + \gamma X_{jklt} + \delta BC_{kt} + \epsilon_{jklt}$$

where j indexes firm, k indexes state of incorporation, l is state of location, and t indexes time

- Interesting part of law changes: can control for local state economic outcomes, potentially different than incorporation state economic outcomes
- Basic idea to compare two plants in PA, one gets the BC law shock to incorporation state while other doesn't
  - Doesn't include all location-time FE for computation, adds outcomes for other plants in the same state-year $y_{lt,-i}$ *[is that a problem?]*

DiD in Bertrand and Mullainathan (2003): Business Combination (BC) laws

- TL;DR of the regression results:
  - Wages and employment seem to increase
  - Plant deaths decrease
  - Plant births also decrease
  - Investment doesn't really increase (not empire building on average)
  - TFP and profitability both decrease
- Seems like managers prefer "the quiet life" when available and overpay professional workers in their firms

DiD in Bertrand and Mullainathan (2003): Business Combination (BC) laws

- Threat to identification: there is a time-varying unobserved factor correlated with BC law passage across the states that implement such rules
- (Discussion) What could such an example be?

DiD in Giroud and Mueller (2010): More BC laws

- An example of a "triple difference" shows up in Giroud and Mueller (2010), which extends Bertrand and Mullainathan (2003) with an interaction with industry Herfindahl-Hirschman Index (HHI)

$$\text{HHI}_j = \sum_{i \in J} \left(\frac{y_i}{\sum_{k \in J} y_k} \right)^2$$

  For industry j (firms in set J), HHI is the squared sum of market shares
  - **NOTE:** HHI calculated *from Compustat* is **not a meaningful object** (Keil 2017). However, we will use it in the problem set since it has been a common thing to do

- Very simple approach **(replicating this in PS3)**:
  1. Using Compustat, first replicate Bertrand and Mullainathan (2003) ROA result
  2. Then add HHI control and interaction, find all impact is on interaction with HHI
  3. Finally, do triple diff in thirds for high, medium, and low HHI industries

DiD in Giroud and Mueller (2010): More BC laws

- An example of a "triple difference" shows up in Giroud and Mueller (2010), which extends Bertrand and Mullainathan (2003) with an interaction with industry Herfindahl-Hirschman Index (HHI)

$$\text{HHI}_j = \sum_{i \in J} \left(\frac{y_i}{\sum_{k \in J} y_k} \right)^2$$

  For industry j (firms in set J), HHI is the squared sum of market shares
  - **NOTE:** HHI calculated *from Compustat* is **not a meaningful object** (Keil 2017). However, we will use it in the problem set since it has been a common thing to do

- Very simple approach **(replicating this in PS3)**:
  1. Using Compustat, first replicate Bertrand and Mullainathan (2003) ROA result
  2. Then add HHI control and interaction, find all impact is on interaction with HHI
  3. Finally, do triple diff in thirds for high, medium, and low HHI industries

DiD in Karpoff and Wittry (2018): Even More BC laws

- There are now **many** more papers looking at BC laws and various corporate outcomes
- Karpoff and Wittry (2018) take a step back and challenge the basic identifying assumption behind these DiDs for a few reasons
  - Heterogeneous treatment effects:
    1. Other state takeover laws
    2. Pre-existing firm level defenses
    3. Legal regime due to other important court cases
  - Endogenous passage of BC laws
- They take the approach of adding controls to make the treatment *more random*
- Previous DiD papers following Bertrand and Mullainathan (2003) have **substantial changes in estimated impacts** with new controls

What's the problem with staggered DiD?

- There are 3 major problems with staggered DiD with a generalized DiD model:
  1. Treatment may be slow evolving and treated units are part of the control for later treatment cohorts *(how fast should ROA change if firm incentives change?)*
  2. Treatment impacts may be a function of time so cohorts treated at different times have different treatments and β_3 is a *weighted average* of these (legal regimes in BC law literature)
  3. Treatment may be slow and differently evolving across different units in different contexts

What's the problem with staggered DiD?

- There are 3 major problems with staggered DiD with a generalized DiD model:
  1. Treatment may be slow evolving and treated units are part of the control for later treatment cohorts *(how fast should ROA change if firm incentives change?)*
  2. Treatment impacts may be a function of time so cohorts treated at different times have different treatments and β_3 is a *weighted average* of these (legal regimes in BC law literature)
  3. Treatment may be slow and differently evolving across different units in different contexts

What's the problem with staggered DiD?

- There are 3 major problems with staggered DiD with a generalized DiD model:
  1. Treatment may be slow evolving and treated units are part of the control for later treatment cohorts *(how fast should ROA change if firm incentives change?)*
  2. Treatment impacts may be a function of time so cohorts treated at different times have different treatments and β_3 is a *weighted average* of these (legal regimes in BC law literature)
  3. Treatment may be slow and differently evolving across different units in different contexts

What's the problem with staggered DiD?

- There are 3 major problems with staggered DiD with a generalized DiD model:
  1. Treatment may be slow evolving and treated units are part of the control for later treatment cohorts *(how fast should ROA change if firm incentives change?)*
  2. Treatment impacts may be a function of time so cohorts treated at different times have different treatments and β_3 is a *weighted average* of these (legal regimes in BC law literature)
  3. Treatment may be slow and differently evolving across different units in different contexts

Staggered DiD Model with Multiple Events

- Using the following setup for the generalized DiD model ("Quiet Life" setup without firm FE or controls):

$$y_{ict} = \beta d_{ict} + p_t + m_c + \nu_{ict}$$

  - y_{ict} is outcome for unit i (e.g. firm)
  - in period t (e.g. year)
  - cohort c, where ***cohort*** indexes the different sets of firms treated by each separate event
- From Bertrand and Mullainathan (2003), different firms are affected by a change in regulation at different points in time
- Each group of firms that are treated at a given point in time are a cohort

Staggered DiD Model with Multiple Events

- Using the following setup for the generalized DiD model ("Quiet Life" setup without firm FE or controls):

$$y_{ict} = \beta d_{ict} + p_t + m_c + \nu_{ict}$$

  - y_{ict} is outcome for unit i (e.g. firm)
  - in period t (e.g. year)
  - cohort c, where ***cohort*** indexes the different sets of firms treated by each separate event
- From Bertrand and Mullainathan (2003), different firms are affected by a change in regulation at different points in time
- Each group of firms that are treated at a given point in time are a cohort

Staggered DiD Model with Multiple Events

- Model:

$$y_{ict} = \beta d_{ict} + p_t + m_c + \nu_{ict}$$

- Intuition of this approach:
  - Every untreated observation at a particular point in time acts as control for treated observations in that time period
  - A firm treated in 1999 by some event will act as a control for a firm treated in 1994 until itself becomes treated in 1999
  - When changing to treated, a given cohort jumps β to the new outcome level

- β will capture ***a** weighted average* treatment effect across the multiple events

- Goodman-Bacon (2021) showed that the **implicit weights may be negative**!

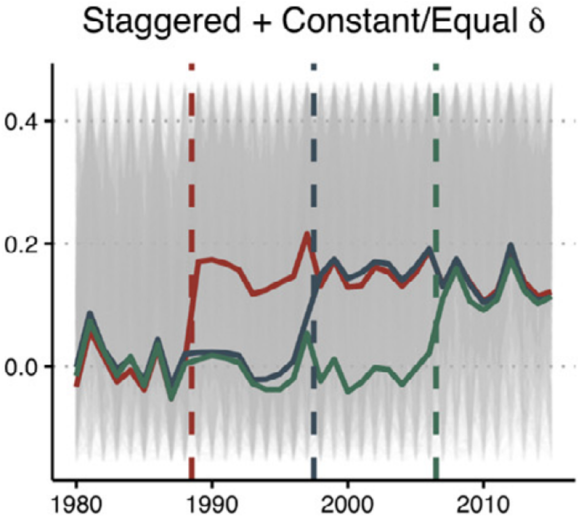
Problem #1: Using Treated as Control

- Model:

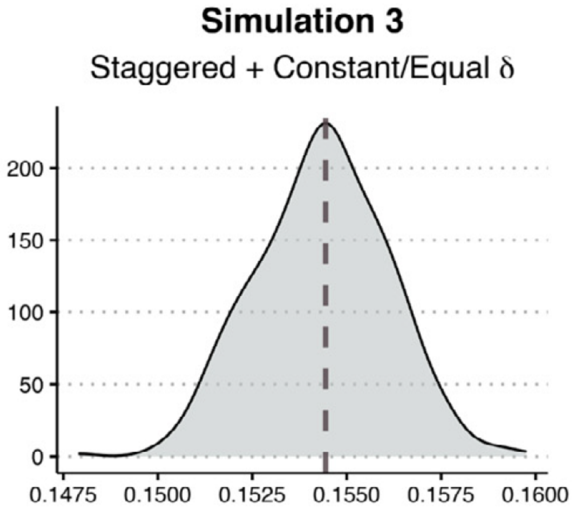
$$y_{ict} = \beta d_{ict} + p_t + m_c + \nu_{ict}$$

- The first problem is that treatment may be **slow evolving**
- Model assumes that the level jump with treatment happens immediately
  - Another assumption is that only one unit is treated, so treated are never in control for future cohort
- I lean on simulations from Baker, Larcker, and Wang (2022) to characterize this problem through examples

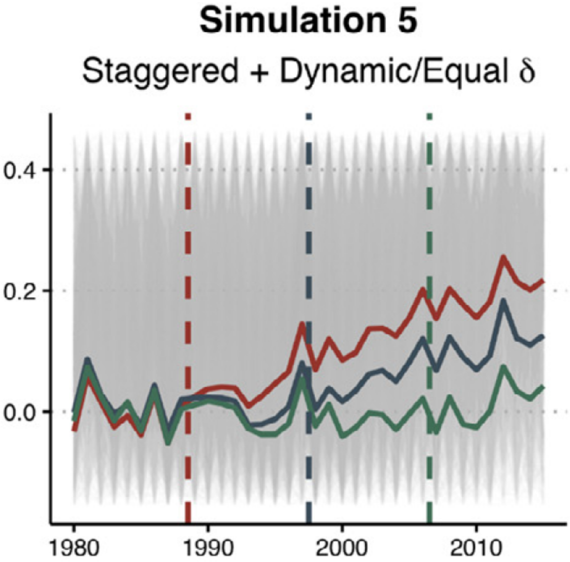
Prob # 1 Simulations: Baker, Larcker, and Wang (2022)



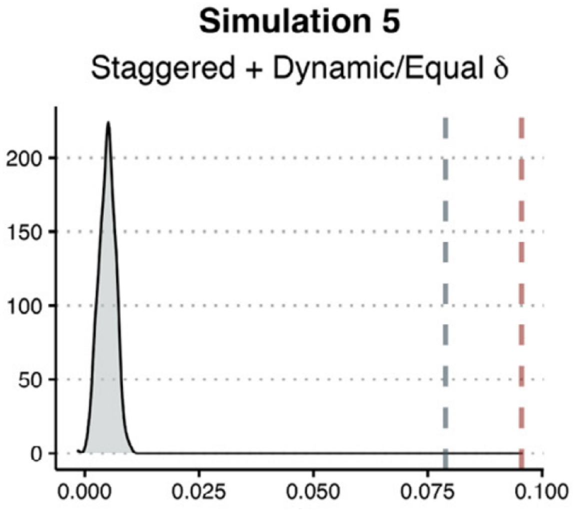
Prob # 1 Simulations: Baker, Larcker, and Wang (2022)



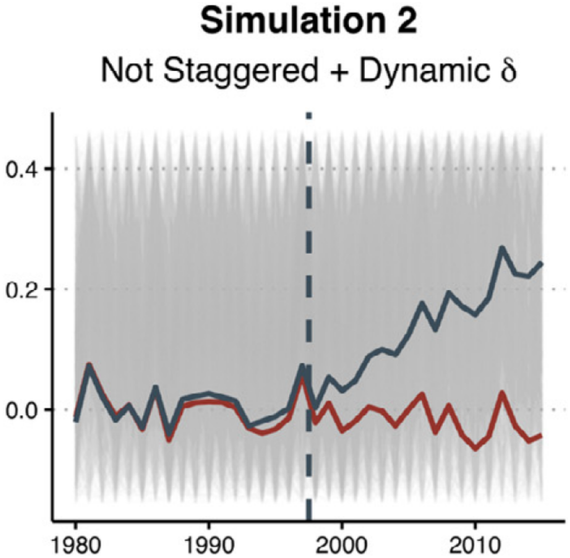
Prob # 1 Simulations: Baker, Larcker, and Wang (2022)



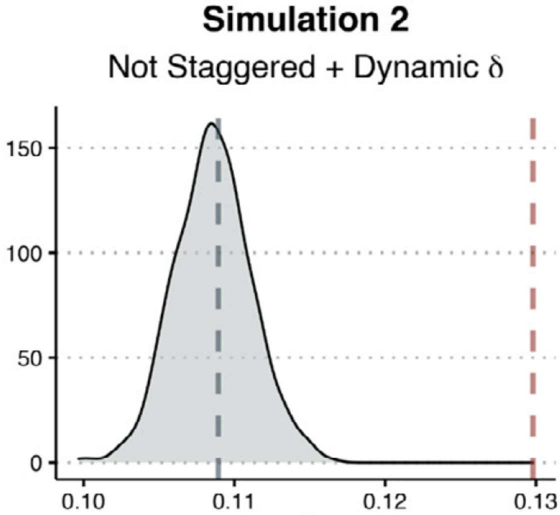
Prob # 1 Simulations: Baker, Larcker, and Wang (2022)



Prob # 1 Simulations (NOTE): Baker, Larcker, and Wang (2022)



Prob # 1 Simulations (NOTE): Baker, Larcker, and Wang (2022)



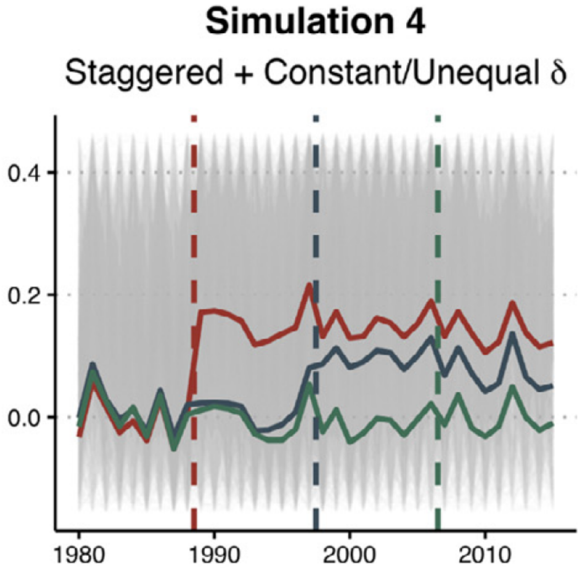
Problem #2: Different Treatment by Cohort

- Model:

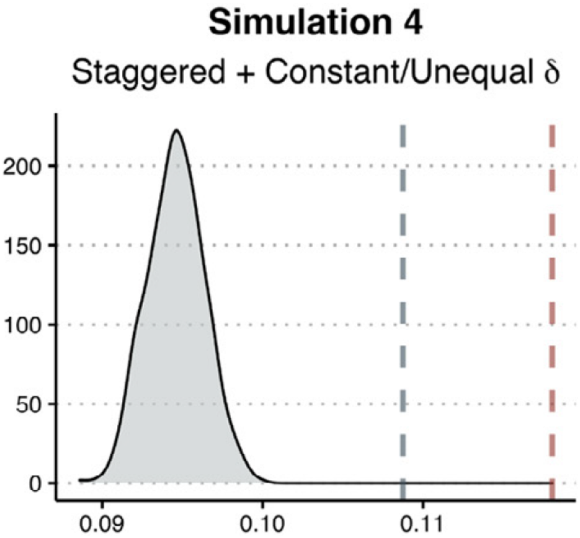
$$y_{ict} = \beta d_{ict} + p_t + m_c + \nu_{ict}$$

- The second problem is that each unit/cohort has a unique treatment not equal to the ATE
- As time goes on, treatment may be fundamentally different due to context
- Impact may also be driven by how many units are treated
- Tax competition is an area where treatment effects are a function of order:
  - Tax revenue expected from marijuana legalization depends heavily on how many other states pass similar laws
  - Tax revenue from legalizing different sorts of gambling is a function of neighboring states

Prob # 2 Simulations: Baker, Larcker, and Wang (2022)



Prob # 2 Simulations: Baker, Larcker, and Wang (2022)



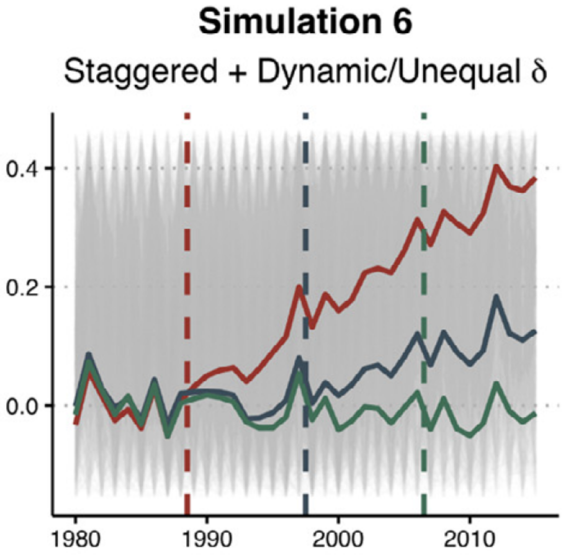
Problem #3: Differing Slopes Treatment by Cohort

- Model:

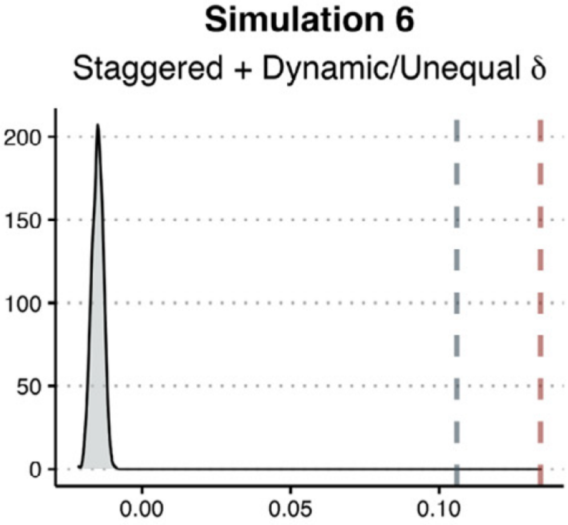
$$y_{ict} = \beta d_{ict} + p_t + m_c + \nu_{ict}$$

- If the treatment is slow evolving and has unequal impacts across cohorts, the **sign of the estimate from this regression may be wrong**

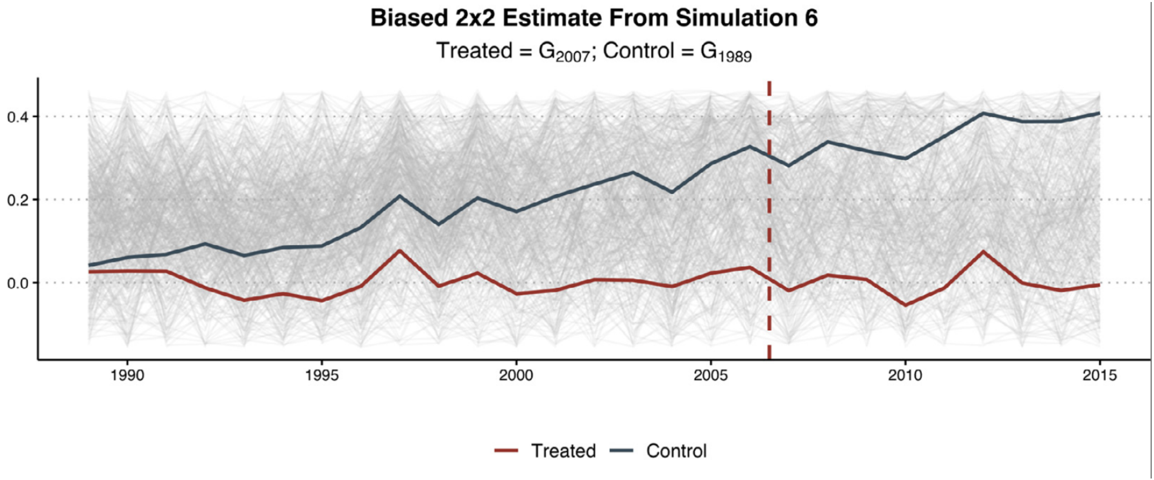
Prob # 3 Simulations: Baker, Larcker, and Wang (2022)



Prob # 3 Simulations: Baker, Larcker, and Wang (2022)



Prob # 3 Simulations: Baker, Larcker, and Wang (2022)



Note on Graphing the Aggregated Event Study for TWFE

- **Best practice:** show the event study separately for each cohort
- But, I include an aggregated event study for PS3 to give you all an example of creating this since it is a classic diagnostic
- **Intuition:** stack the cohorts with the same dummy variable for a few years before and after treatment
  - For each cohort c, define the time at which treatment happens, $\tilde{t}_c$
  - Create a running variable of time since treatment $D_{ct} = t - \tilde{t}_c$
  - Create a series of dummy variables for the time periods before and after treatment, defined jointly for all cohorts, $treat_k = \mathbb{1}(D_{ct} = k)$ (will probably need pooling state)
  - Regress outcome on dummy variables and normal FE

Note on Graphing the Aggregated Event Study for TWFE

- Here is a bit of stata code that implements this method:

```
gen treat = 0
replace treat = 1 if year >= year(treatment_date) & treatment_date != .

gen years_since_treat = year - year(treatment_date) + 5

forvalues x=0(1)11 {
gen dummy_`x' = years_since_treat==`x'
}
gen dummy_12 = years_since_treat>=12

reghdfe y dummy* X ..., absorb(gvkey year)
```

- The estimates are for outcome difference for each year pooled for all treatment cohorts (this uses implicit weights, more on this later)

Recapping Problems with Staggered DiD

- Model:

$$y_{ict} = \beta d_{ict} + p_t + m_c + \nu_{ict}$$

- Goodman-Bacon (2021) formally shows the following identification for the β from this regression

$$\underset{N\rightarrow \inf}{\text{plim}}\ \hat{\beta} = VWATT + VWCT - \Delta ATT$$

  - $VWATT$ is the variance weighted ATT estimate *[interpretation?]*
  - $VWCT$ is the variance weighted common trend, which is an extension of the parallel trends assumption
  - ΔATT is the weighted sum of the change in ATT around a latter treated unit's treatment window

- It is far from trivial to get the ATT from a staggered treatment for these reasons

- But many new estimators/methods have been proposed in the last few years!!

Callaway and Sant'Anna (2021) Solution

- Model:

$$y_{ict} = \beta d_{ict} + p_t + m_c + \nu_{ict}$$

- So, we understand staggered DiD is not the silver bullet that the literature of the last 20 years hoped
- A simple intuition for what we need to do to solve these issues:
  1. Allow treatment to vary across cohorts
  2. Stop treated cohorts with evolving treatments from affecting future cohorts
  3. Re-aggregate treatment effects with weights we think are more appropriate
- The Callaway and Sant'Anna (2021) solution (CS) is going got do exactly that

Callaway and Sant'Anna (2021) Solution

- csdid is the stata implementation (needs drdid), but it's also in R (package did)
- Program estimates **all** 2x2 DiD estimates of $ATT(g, t)$ (group time treatment effects)
- Unlike OLS that minimizes MSE, the CS solution focus on reweighting all of these comparisons
  1. Primary: simple weighted average of all $ATT(g, t)$'s
  2. Other possibilities like averages based on how long treatment lasts (treatment effect for each length of exposure, "dynamic" in csdid), etc.
- NOTE 1: built-in commands like csdid_plot and staggered will make our lives much easier
- NOTE 2: allows controls, but time-varying controls are held as fixed at their treatment period levels within unit
- ****NOTE 3:** These programs are not uniform in how they graph pre-trends!! (See "Interpreting Event-Studies from Recent Difference-in-Differences Methods" by J Roth)

Callaway and Sant'Anna (2021) Solution

- csdid is the stata implementation (needs drdid), but it's also in R (package did)
- Program estimates **all** 2x2 DiD estimates of $ATT(g, t)$ (group time treatment effects)
- Unlike OLS that minimizes MSE, the CS solution focus on reweighting all of these comparisons
  1. Primary: simple weighted average of all $ATT(g, t)$'s
  2. Other possibilities like averages based on how long treatment lasts (treatment effect for each length of exposure, "dynamic" in csdid), etc.
- NOTE 1: built-in commands like csdid_plot and staggered will make our lives much easier
- NOTE 2: allows controls, but time-varying controls are held as fixed at their treatment period levels within unit
- ****NOTE 3:** These programs are not uniform in how they graph pre-trends!! (See "Interpreting Event-Studies from Recent Difference-in-Differences Methods" by J Roth)

Sun and Abraham (2021) Solution

- The Sun and Abraham (2021) solution (SA) is very similar to the CS solution
- Implemented in stata using eventstudyinteract package
  - Calculates separate treatment effects for each cohort
  - Weights each treatment effect according to the share of the sample in that cohort
  - Limits control to never treated and not-yet-treated observations
- This method is very much in the spirit of Gibbons, Suárez Serrato, and Urbancic (2018) we will talk about shortly

Gormley and Matsa (2011) Solution

- The solution most common in finance before the last year was shown in Gormley and Matsa (2011)
- Unfortunately, doesn't have a special stata package to my knowledge
- But the method has the benefit of being very intuitive and easy to code up by hand
- This is the **"stacked regression"** method

Gormley and Matsa (2011) Solution

- Think of running generalized diff-in-diff for just one of the multiple events

$$y_{it} = \beta d_{it} + p_t + m_i + \nu_{it}$$

- But, contrary to standard difference-indifference, your sample is
  1. Restricted to a small window around the event
  2. Control restricted to not yet (or never) treated observations

- Much like the CS and SA solutions, you want to get separate effects for each cohort. Difference: you create this sample separately for each cohort

- With samples for each cohort, "stack" the samples in a one dataset and create a factor variable that separately identifies each sample
  - **Note:** it is ok for some observation units to appear multiple times in the stacked data [e.g. GOOG might be a control in event year 1999 but a treated firm in a later event in 2005]

Gormley and Matsa (2011) Solution

- Now we estimate the following pooled regression with all of the cohorts

$$y_{ict} = \beta d_{ict} + p_{tc} + m_{ic} + \nu_{ict}$$

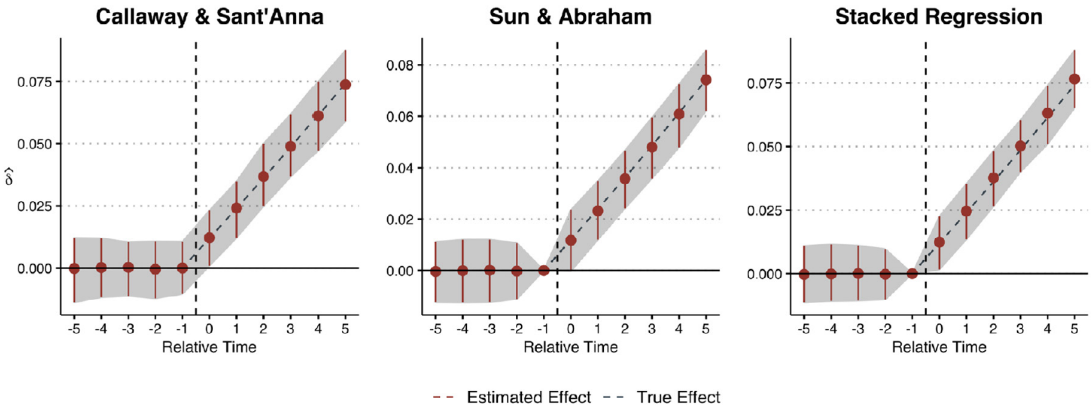
1. d_{ict} is now an indicator if unit i in cohort c is treated in t
2. p_{tc} gives cohort-time fixed effects
3. m_{ic} gives unit cohort fixed effects
    - GOOG from the last slide will get a separate FE for each time it shows up

- This has the OLS implicit weighting problem still, but the variance weighting across cohorts is solved

Gormley and Matsa (2011) Solution

- The stacked regression solution is very simple and mostly solves the problems of stacked DiD
- Easy to isolate a specific window for each cohort
- Very easily extends to triple differences and higher order differences
- Tends to go over well at finance seminars already

Back to Simulations from Baker, Larcker, and Wang (2022)



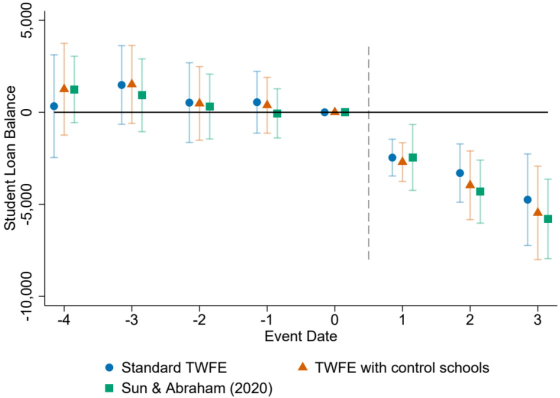
Simulation 6 (unequal treatment impacts that evolve) is pretty good in the stacked version relative to TWFE that had wrong sign

How to Choose a Solution?

- We've only covered a relatively superficial overview of these solutions, and you will have to try multiple in PS3
- Ultimately, you will wind up showing at least 2 versions in any paper with staggered DiD any time soon
  - CS allows controls and flexible weights (this also means more defense in interpreting)
  - SA is more parametric and takes a bit more setup in stata, gives more structure for comparison across papers
  - Stacked regressions are easy to parse, but may not fix the same problems as explicitly outside of specifically simulated scenarios
- In PS3, I only require that you do 2, but I strongly encourage attempting all 3 if you have time

Example from Menaka Hampole's Seminar in 2023

Figure 5: Effect of UNLPs on amount of student debt



Hampole (2022) showed how we will all be doing ES graphs going forward

Treatment Heterogeneity with One Treatment

- Even if there is only a single treatment, FE estimators still may not give good estimates of the ATT
- This comes down to interpretation a bit: if we are to draw a treated observation at random and treat them, how do we expect their outcome to change?
- FE estimators allow for separate intercepts, but the impact on the slope of each observation depends on variance, which means we mechanically will not get the "average" impact

Broken or Fixed Effects?

- This is the insight made clear in Gibbons, Suárez Serrato, and Urbancic (2018): OLS with FE **will not yield the ATE** with heterogeneous treatments because units are implicitly weighted by variance
- Starting with an empirically tractable definition of ATE:

$$\beta^{ATE} = \sum_g \pi_g \beta_g$$

- For each of g groups with a heterogeneous conditional expectation, the average treatment effect is the expected impact when drawing a unit on average (multiply expectations by population shares π_g)
- OLS with FE will not return this object (proposed solution: allow heterogeneous β_g)

Gibbons, Suárez Serrato, and Urbancic (2018) Example

- Gibbons, Suárez Serrato, and Urbancic (2018) turn to the example of Karlan and Zinman (2008), not 2009, to examine empirical importance
- Karlan and Zinman (2008) asks how credit demand changes with changes in interest rates in development scenarios
- Empirical design: interest rate randomization in panel regression with FE

$$y_i = \alpha + \beta r_i + \delta X_i + \varepsilon_{ib}$$

where X_i contains bank branch, demographic group, risk group, and mailer wave FE among other controls

Gibbons, Suárez Serrato, and Urbancic (2018) Example

- Simple result: 1 pp increase in loan rate decreases size by 4.4 units
- Gibbons, Suárez Serrato, and Urbancic (2018) replicates the estimation and is able to get the same impact
- But the impacts are entirely different by risk group!!!
- Can calculate the implied weights by the relative variances of treatment by group multiplied by the sample frequency of that group

Gibbons, Suárez Serrato, and Urbancic (2018) Example

Table 1: Karlan and Zinman (2008) treatment effect weighting.

		Weight	
Risk group	**Effect**	**FE**	**Sample**
Low	−32.4	0.044	0.125
Medium	−9.9	0.058	0.092
High	−2.7	0.898	0.783
Average		−4.393	−7.047
Std. error		(1.129)	(1.917)

Note: Goodman-Bacon (2021) shows these FE weights can be **negative** in staggered DiD

Average Partial Effect Estimate (GSSUtest in stata)

- Solution 1: Interaction Weighted Estimator (IWE)
  - Run normal fixed effects regression
  - Include interactions of FE with the treatment of interest (interest rate)
  - Reweight estimates according to sample shares
- Solution 2: Regression Weighted Estimator (RWE)
  - Estimate weights inversely proportional to the standard deviation of the conditional treatment values within each group
  - Run normal fixed effects regression with inverse variance weights
- Table A.17 of Garrett, Ordin, Roberts, and Suárez Serrato (2022) gives an example of applying the IWE estimator

Does Treatment Heterogeneity Matter? (Gibbons, Suárez Serrato, and Urbancic 2018)

- In Karlan and Zinman (2008), accounting for heterogeneity changes the prediction problem for a program manager

  How much will our revenue/necessary subsidy change with a change in interest rate?

  - Answer is potentially misleading without heterogeneity
- But the heterogeneity is important in its own right!! (and provides basis for future Karlan and Zinman papers)

Does Treatment Heterogeneity Matter? (Gibbons, Suárez Serrato, and Urbancic 2018)

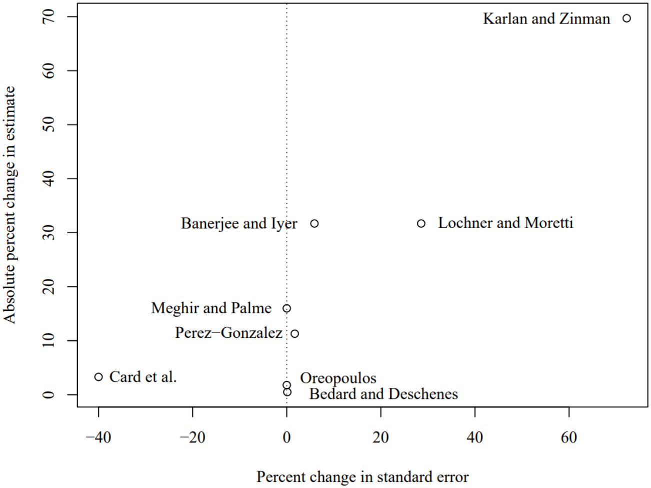
- In Karlan and Zinman (2008), accounting for heterogeneity changes the prediction problem for a program manager

  How much will our revenue/necessary subsidy change with a change in interest rate?

  - Answer is potentially misleading without heterogeneity

- But the heterogeneity is important in its own right!! (and provides basis for future Karlan and Zinman papers)

Does Treatment Heterogeneity Matter?

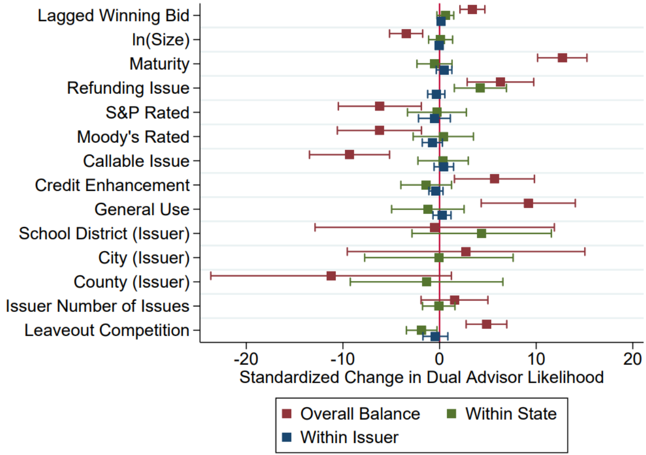


Note: Falsification Tests for DiD and Panel Models

- One can never directly test underlying identification assumption, but can do some falsification tests
  1. Compare balance of pre-treatment observables *[conditional on other controls/FE?]*
  2. Check that timing of observed change in outcome coincides with timing of event
  3. Check for treatment reversal if treatment is removed
  4. Check variables that should not be affected (structures investment in Zwick and Mahon (2017))
  5. Add a triple-difference

Test #1: Balance from Garrett (2021) *(Discussion)*

Figure 1: Linear Probability Estimates Explaining Dual Advisor Choice, 2008-2011



Can we do better than conditional expectations for heterogeneous treatments?

- Instead of looking at average treatment effects, sometimes we care more about the distributional impacts
- A very useful tool in this case is a *quantile regression*, which we introduce a crash course of in the next few slides
- The basic idea: instead of CEF, estimate the *Conditional Quantile Function* using the following minimization problem

$$Q_\tau(y_i|X_i) = \arg\min_{q(X_i)} E[\rho_\tau(y_i - q(X_i))]$$

where $\rho_\tau(u) = (\tau - \mathbb{1}(u \leq 0))u$

Quantile Regression

- Writing it out all at once may add some clarity

$$Q_\tau(y_i|X_i) = \arg\min_{q(X_i)} E[(\tau - \mathbb{1}(y_i - q(X_i) \leq 0))(y_i - q(X_i))]$$

- τ is the quantile that you are describing
- This function will minimize absolute deviations with a (potentially asymmetric) loss function
- The ρ_τ function is the "check function," which looks a whole lot like a check mark
- For overview, see Koenker and Hallock (2001)

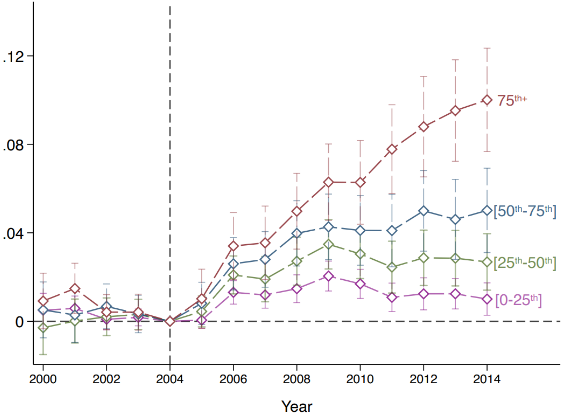
Quantile Regression

- Instead of the conditional expectation, quantile regression tells us how much a change in X will change the expected τ^{th} percentile of the outcome
- It's not widely used in finance, but sometimes we do explicitly care about what the new percentile will be for a certain moment (maybe median)
- Many studies such as Fonseca and Matray (2022) study how banking access affects the distribution of outcomes

Quantile Regression: Potential Example

- Fonseca and Matray (2022) ask how extending public banking services to unbanked communities impacts the distribution of wages
- **Design:** DiD comparing places that receive a bank to those that either already received or do not receive a public bank
- **Several outcomes:** gini coefficients, average wages within quartiles
- Heterogeneity by supply of different worker types
- **Finding?** Upon getting banking access wages go up, but much faster for high-wage and scare worker types

Quantile Regression: Fonseca and Matray (2022)



- Fonseca and Matray (2022) focus on sorting workers into quartiles each year, then linear regression
  - Yields conditional expectations within group
- Another option would be to show the evolution of the quantiles themselves using quantile regressions
  - Benefits and costs? *(discussion)*

Outline

Matching (Econometrics)

- The matching approach can be used to estimate any of our previous regressions
- It's been quite popular outside of finance (medicine, labor econ, etc.), but I've seen a bunch of matching finance papers this year
- Gaining popularity due to rhetorical simplicity:
  - For each treated observation, you find a "matching" untreated observation that serves as the counterfactual
  - Compare outcome of treated observations to outcome of matched observations
  - Very easy in any regression with pair FE for each match

Matching (Econometrics)

- The matching approach can be used to estimate any of our previous regressions
- It's been quite popular outside of finance (medicine, labor econ, etc.), but I've seen a bunch of matching finance papers this year
- Gaining popularity due to rhetorical simplicity:
  - For each treated observation, you find a "matching" untreated observation that serves as the counterfactual
  - Compare outcome of treated observations to outcome of matched observations
  - Very easy in any regression with pair FE for each match

Why Match?

- Matching solves a different sort of non-random assignment than our DiD and general panel methods
  - Treatment is not random. If it were, would not need to match each treated obs. to a control before taking average difference in outcomes
- The idea is that we can find a specific control where, between the treatment and the matched control, treatment is conditionally random
- In fact, *Mostly Harmless* Section 3.3.1 goes through the similarities of matching as an estimator when compared to OLS with controls
- The difference is that matching gives more weight to unit types based on count while OLS gives more weight to certain types when there is *more variation in treatment*

Why Match?

- Matching solves a different sort of non-random assignment than our DiD and general panel methods
  - Treatment is not random. If it were, would not need to match each treated obs. to a control before taking average difference in outcomes
- The idea is that we can find a specific control where, between the treatment and the matched control, treatment is conditionally random
- In fact, *Mostly Harmless* Section 3.3.1 goes through the similarities of matching as an estimator when compared to OLS with controls
- The difference is that matching gives more weight to unit types based on count while OLS gives more weight to certain types when there is *more variation in treatment*

Why not Match?

- Matching has two weaknesses that are worth thinking through before implementing
  - Simple matching** throws out data and variation in practice
  - The identifying assumptions are almost identical to OLS, so the same biases will apply
- If the original treatment isn't random (exogenous), it is often difficult to believe that controlling for some controls will restore randomness...

Why not Match?

- Matching is ***not*** an effective solution for the following:
  - To fix simultaneity bias
  - To eliminate measurement error bias
  - To fix omitted variable bias from unobservable variables

A Generalized Matching Model

- We want to know the treatment effect of treatment d
- The outcome is given by
  - $y(1)$ if $d = 1$
  - $y(0)$ if $d = 0$
- There also can be k observable covariates $X = (x_1, ..., x_k)$

A Generalized Matching Model

- Assumption #1 – **Unconfoundedness**
  - Outcomes $y(0)$ and $y(1)$ are statistically independent of treatment, d, conditional on the observable covariates, X
  - Stronger than *conditional mean independence*, needed for propensity score in particular
- Assumption #2 – **Overlap**
  - For each value of covariates, there is a positive probability of being in the treatment group and in the control group (a match exists conditional on X)
  - In practice, perfect matches on covariates don't usually exist (there is only one GOOG), which causes solvable bias

A Generalized Matching Model

- A natural estimator is to find the treatment effect for each X and then integrate over the distribution of X
- Roberts and Whited (2013), page 68, goes through the identification arguments for this estimator
- In practice, k may be large and exact matches may never exist (X may have continuous elements...)
- **At least two solutions:**
  1. Matching on covariates with arbitrary distance metric (not covering for time)
  2. Propensity score matching

Propensity Score Matching

- Propensity Score Matching is an approach in which we distill the covariates X into a probability of treatment, $Pr(d=1|X)$
- Propensity scores have been popular because we can think of the conditional expectation of treatment as the probability, so we can use all of our regression models
- If *unconfoundedness* (assumption #1) is true, then you can just include a control for the propensity score in a regression to get ATE
  - **Note:** do not confuse this with a "control function," which controls for residuals from a first-stage IV regression
  - Can make selection worse with endogeneity, not better

Propensity Score Matching

- Propensity Score Matching is an approach in which we distill the covariates X into a probability of treatment, $Pr(d = 1|X)$
- Propensity scores have been popular because we can think of the conditional expectation of treatment as the probability, so we can use all of our regression models
- If *unconfoundedness* (assumption $\#1$) is true, then you can just include a control for the propensity score in a regression to get ATE
  - **Note:** do not confuse this with a "control function," which controls for residuals from a first-stage IV regression
  - Can make selection worse with endogeneity, not better

Propensity Score Matching

- Propensity Score Matching is an approach in which we distill the covariates X into a probability of treatment, $Pr(d = 1|X)$
- Propensity scores have been popular because we can think of the conditional expectation of treatment as the probability, so we can use all of our regression models
- If *unconfoundedness* (assumption $\#1$) is true, then you can just include a control for the propensity score in a regression to get ATE
  - **Note:** do not confuse this with a "control function," which controls for residuals from a first-stage IV regression
  - Can make selection worse with endogeneity, not better

Propensity Score Matching

- To implement propensity score matching:
  1. Estimate a regression describing d as a function of covariates X (Logit, Probit, or OLS all can be appropriate)
     - Note: using less than full controls on first stage can improve finite sample properties
  2. Predicted value for observation i is its propensity score, $ps(X_i)$
  3. Select the M closest observations not receiving treatment based on distance from propensity score
     - If $M = 1$, this is referred to as *nearest neighbor matching*

Propensity Score Matching: Pros and Cons

Pros:

- Can relax functional form assumptions of OLS, so a good robustness check
- Matching often used to screen sample to make sure distributions are similar (treatment and control are "balanced")
- "The goal of matching is to reduce imbalance in the empirical distribution" of confounders (King and Nielsen 2019)

Cons:

- Doubles down on the identifying assumptions in OLS in requiring conditional independence
- Often presented as if it fixes an underlying selection/endogeneity problem (needs a design)
- Huge number of researcher choices (M, replacement, first stage model, etc.), and **instability of results to these choices**
- ORDER OFTEN MATTERS, 'set seed' and 'sort' in stata

Matching in the Wild

- Given relative instability, we don't see much matching in empirical corporate finance
- Examples where matching has become weirdly popular:
  1. Climate finance and green bond studies
  2. Learning about large multinational corporation behavior

Matching in the Wild: Green Bonds

- There is a **massive vacuum of writing** about green securities, who issues them, how behavior changes, and how those securities are priced
- An early paper, Baker, Bergstresser, Serafeim, and Wurgler (2018), sought to use municipal bonds (frequent, small issuance relative to corporates) to see how the market prices green bonds
- **Design:** panel comparison of green and non-green bonds with OLS with FE *(not issuer-day)*
- **Finding:** green bonds seem to be issued at a material premium (lower yield)

Matching in the Wild: Green Bonds

- Larcker and Watts (2020) takes issue with the research design in Baker, Bergstresser, Serafeim, and Wurgler (2018), stating

  Baker et al. (2018) use a pooled fixed-effects model in their analyses. ... [T]his approach is insufficient to adequately control for nonlinearities and issuer-specific time variation, which ultimately leads to spurious inferences. ...

  It is easy to imagine a situation where the fixed effects will be inadequate. For example, green issuers (which tend to be significantly larger) may outperform non-green issuers over the sample period. Even when controlling for rating-maturity-issuance month fixed effects and issuer fixed effects, a greenium would be observed in this setting when it does not actually exist.

- Thoughts on this critique?

Matching in the Wild: Green Bonds

- Larcker and Watts (2020) take advantage of the fact that munis for different issues may be issued by the same issuer on the same day

  We avoid these problems by taking advantage of the unique institutional features of the municipal securities market that give us a nearly perfect counterfactual security.

- **Design:** Match green bonds to the nearest non-green bond issued by the same issuer on the same day

- **Finding:** Bonds with the same maturity issued on the same day by the same issuer have the same yield regardless of green status

Matching in the Wild: Green Bonds

- Now the green bond literature is full of matching papers to try to make the control and treatment match perfectly on controls
- Outside of munis, this is being used in corporates, sovereigns, and other alternative assets (Flammer 2021; Kölbel and Lambillon 2022)
- (Discussion) What makes matching appropriate or not for these papers?

Matching in the Wild: International Tax Avoidance

- In my ongoing work on tax avoidance, we have been trying to use matching methods
- **The idea:** Large multinational firms (MNCs) who are impacted by changes in international tax considerations for financing (check the box, repatriation holidays, etc.) are not like most firms
- Matching will restrict the sample of control firms to those that are most observably similar to impacted multinationals, but they will never be identical
- **Our working insight:** while MNCs are very hard to match, the locations in which they exist are easier to match, so we *measure the impact of international capital flow constraints through the reflections of large firms in local labor markets*
- Going further, the instability of propensity score matching has pushed us in the direction of a generalization...

Matching in the Wild: International Tax Avoidance

- In my ongoing work on tax avoidance, we have been trying to use matching methods
- **The idea:** Large multinational firms (MNCs) who are impacted by changes in international tax considerations for financing (check the box, repatriation holidays, etc.) are not like most firms
- Matching will restrict the sample of control firms to those that are most observably similar to impacted multinationals, but they will never be identical
- **Our working insight:** while MNCs are very hard to match, the locations in which they exist are easier to match, so we *measure the impact of international capital flow constraints through the reflections of large firms in local labor markets*
- Going further, the instability of propensity score matching has pushed us in the direction of a generalization...

Matching in the Wild: International Tax Avoidance

- In my ongoing work on tax avoidance, we have been trying to use matching methods
- **The idea:** Large multinational firms (MNCs) who are impacted by changes in international tax considerations for financing (check the box, repatriation holidays, etc.) are not like most firms
- Matching will restrict the sample of control firms to those that are most observably similar to impacted multinationals, but they will never be identical
- **Our working insight:** while MNCs are very hard to match, the locations in which they exist are easier to match, so we *measure the impact of international capital flow constraints through the reflections of large firms in local labor markets*
- Going further, the instability of propensity score matching has pushed us in the direction of a generalization...

Inverse Probability of Treatment Weights (IPW)

- *Note: Adair Morse walked by while I was writing this slide in 2022 and said she prefers PSM to IPW, but agrees with me that both are used to obfuscate more than solve issues*
- Inverse Probability Weighting (IPW) is a type of matching that uses all data instead of restricting to "matches"
- The idea is to reweight the data so that the treatment and control have the same observable distributions
- I generally cite Hirano, Imbens, and Ridder (2003), but others cite Abadie (2005) as the first IPW paper
  - We also mentioned DiNardo, Fortin, and Lemieux (1996) weights, and IPW is also nested by DFL weights that *can* be used to observably balance an unbalanced panel

Implementing IPW

- Some intuition can be found by writing down the difference in conditional means that we observe
$$E[Y|X, D = 1] - E[Y|X, D = 0]$$
- We know this difference does not give us the ATE outside of randomization, but let's say we estimated $P(X) = Pr(D = 1|X)$
- Scale both sides by a 1 (need overlap assumption for non-zero propensities):
$$E\left[\frac{Y}{P(X)}|X, D = 1\right] P(X) - E\left[\frac{Y}{(1 - P(X))}|X, D = 0\right](1 - P(X))$$

Implementing IPW

- Simplifying terms and integrating gives us the IPW estimator (sometimes called Horvitz-Thompson (HT) estimator in survey literatures)

$$E\left[\frac{DY}{P(X)} - \frac{Y(1-D)}{(1-P(X))}\right]$$

- Again, note the overlap assumption makes the denominator non-zero, but it can empirically get pretty close to zero
- It also may be the case that estimates of $\hat{P}(X)$ are imprecise. That's fine
- Ideally, include controls that both predict treatment **and affect outcome** (not just predicting treatment)

Implementing IPW

- Implementing IPW requires three steps:
  - First stage estimation of the likelihood of a firm falling in the treatment or control groups (logit, probit, OLS, or non-parametric are all fine)
  - Create inverse probability weights of treatment according to the equation:

$$weight_i = \frac{treat_i}{P(treat_i = 1)} + \frac{1 - treat_i}{P(treat_i = 0)},$$

    where $P(treat_i = 1)$ is the likelihood of treatment from the first stage.
  - Run weighted least squares using IPW in a second stage (bootstrap to get appropriate SE, or use teffects)
  - Note: There are problems with these weights 'blowing up' or becoming too large, it's worth graphing these estimates to check

Outline

Synthetic Control

- Abadie (2021) reviews the literature on synthetic control in an empiricist-friendly way
- Synthetic control has primarily been used in *comparative case studies* where an aggregate unit faces a large change
- Famous examples are
  - CA with a major cigarette tax increase (Abadie, Diamond, and Hainmueller 2010)
  - The Mariel Boatlift and immigration (Peri and Yasenov 2019)
  - Corporate political connections and firm returns (Acemoglu, Johnson, Kermani, Kwak, and Mitton 2016)

Synthetic Control

- Since Abadie (2021) is already a solid resource, I only replicate the bare-bones model intuition here

The synthetic control method is based on the idea that, when the units of observation are a small number of aggregate entities, a combination of unaffected units often provides a more appropriate comparison than any single unaffected unit alone.

Synthetic Control Model

- Data are available for $J+1$ units, assume that unit 1 is treated, J units in the "donor pool" untreated

- Panel covers T periods with first T_0 periods happening before any treatment

- For each unit j, observe
  - The outcome at each period Y_{jt}
  - k predictors of the outcome $X_{1j}, ..., X_{kj}$

- For each period and unit, let Y_{jt}^N be the potential response without any intervention, Y_{jt}^I with intervention

- The treatment we care about is then $\tau_{1t} = Y_{1t}^I - Y_{1t}^N$

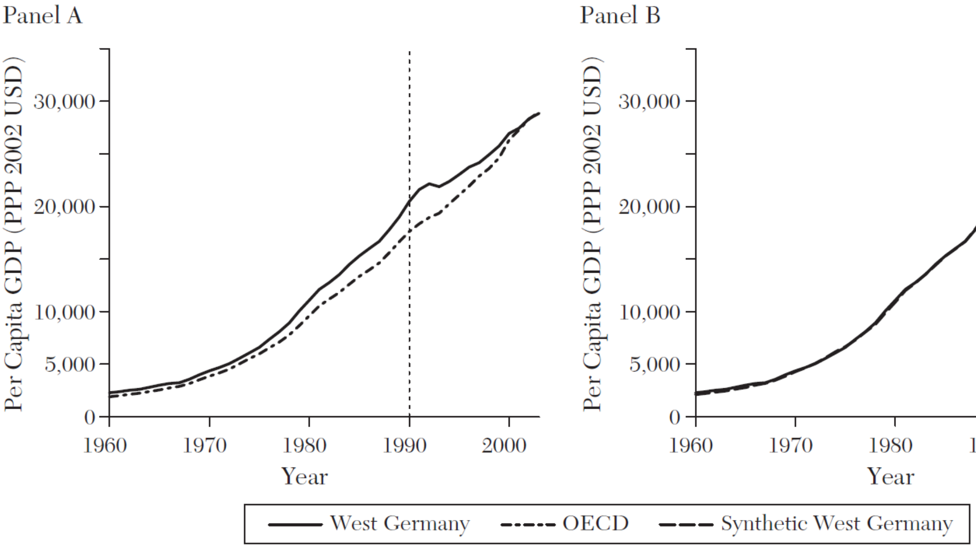
Synthetic Control Model

- The treatment we care about is $\tau_{1t} = Y_{1t}^I - Y_{1t}^N$
- And finding Y_{1t}^I is trivial since it is observed for $t > T_0$
- Then the problem of estimating Y_{1t}^N is the problem, but we are going to do something a lot *like* DiD
- Using the donor pool, want to identify a weighted average group of controls that match Y_{1t}^N for $t \leq T_0$
- Easier said than done, as we still need some way of finding some "optimal" weights, luckily Abadie, Diamond, and Hainmueller (2010) already did

Synthetic Control Notes

- There are many ways to weight, but the solution in Abadie, Diamond, and Hainmueller (2010) weights each X_{jt} to minimize mean squared prediction error (MSPE)
- Another option is to weight according to the inverse of the variance of the controls
  - Given how many times we've mentioned that today, I think it's a pretty safe bet
- Other papers have suggested choosing weights to match an out-of-sample prediction (discussed extensively in Abadie, Diamond, and Hainmueller (2015))

Synthetic Control Example, See Abadie (2021) for Extension



Concluding Thoughts

- DiD is still an amazing identification strategy, especially with staggered treatment and modern methods
- Quantile regression can be used to easily show impacts on inequality
- Matching methods provide a good robustness to traditional OLS with FE and DiD
- But, we can probably do a little better than propensity score matching with IPW with weaker assumptions
- **The treatment of all of these methods is woefully incomplete in this lecture. Please go back to the cited primary sources before you stake your future on a given method here.**

Concluding Thoughts

- DiD is still an amazing identification strategy, especially with staggered treatment and modern methods
- Quantile regression can be used to easily show impacts on inequality
- Matching methods provide a good robustness to traditional OLS with FE and DiD
- But, we can probably do a little better than propensity score matching with IPW with weaker assumptions
- **The treatment of all of these methods is woefully incomplete in this lecture. Please go back to the cited primary sources before you stake your future on a given method here.**

Abadie, Alberto, 2005, Semiparametric difference-in-differences estimators, *The review of economic studies* 72, 1–19.

——— , 2021, Using synthetic controls: Feasibility, data requirements, and methodological aspects, *Journal of Economic Literature* 59, 391–425.

——— , Alexis Diamond, and Jens Hainmueller, 2010, Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program, *Journal of the American statistical Association* 105, 493–505.

——— , 2015, Comparative politics and the synthetic control method, *American Journal of Political Science* 59, 495–510.

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton, 2016, The value of connections in turbulent times: Evidence from the united states, *Journal of Financial Economics* 121, 368–391.

Baker, Andrew C, David F Larcker, and Charles CY Wang, 2022, How much should we trust staggered difference-in-differences estimates?, *Journal of Financial Economics* 144, 370–395.

Baker, Malcolm, Daniel Bergstresser, George Serafeim, and Jeffrey Wurgler, 2018, Financing the response to climate change: The pricing and ownership of us green bonds, Discussion paper National Bureau of Economic Research.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, 2004, How much should we trust differences-in-differences estimates?, *The Quarterly journal of economics* 119, 249–275.

Bertrand, Marianne, and Sendhil Mullainathan, 2003, Enjoying the quiet life? corporate governance and managerial preferences, *Journal of political Economy* 111, 1043–1075.

Callaway, Brantly, and Pedro HC Sant'Anna, 2021, Difference-in-differences with multiple time periods, *Journal of Econometrics* 225, 200–230.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, 2011, Robust inference with multiway clustering, *Journal of Business & Economic Statistics* 29, 238–249.

Curtis, E Mark, Daniel G Garrett, Eric C Ohrn, Kevin A Roberts, and Juan Carlos Suárez Serrato, 2023, Capital investment and labor demand, Discussion paper National Bureau of Economic Research.

DiNardo, John, Nicole M Fortin, and Thomas Lemieux, 1996, Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach, *Econometrica* 64, 1001–1044.

Flammer, Caroline, 2021, Corporate green bonds, *Journal of Financial Economics* 142, 499–516.

Fonseca, Julia, and Adrien Matray, 2022, Financial inclusion, economic development, and inequality: Evidence from brazil, .

Garrett, Daniel, 2021, Conflicts of interest in municipal bond advising and underwriting, Working paper.

———, Andrey Ordin, James W Roberts, and Juan Carlos Suárez Serrato, 2022, Tax Advantages and Imperfect Competition in Auctions for Municipal Bonds, *The Review of Economic Studies* rdac035.

Gibbons, Charles E, Juan Carlos Suárez Serrato, and Michael B Urbancic, 2018, Broken or fixed effects?, *Journal of Econometric Methods* 8, 20170002.

Giroud, Xavier, and Holger M Mueller, 2010, Does corporate governance matter in competitive industries?, *Journal of financial economics* 95, 312–331.

Goodman-Bacon, Andrew, 2021, Difference-in-differences with variation in treatment timing, *Journal of Econometrics* 225, 254–277.

Gormley, Todd A, and David A Matsa, 2011, Growing out of trouble? corporate responses to liability risk, *The Review of Financial Studies* 24, 2781–2821.

Hampole, Menaka V, 2022, Financial frictions and human capital investments, .

Hirano, Keisuke, Guido W Imbens, and Geert Ridder, 2003, Efficient estimation of average treatment effects using the estimated propensity score, *Econometrica* 71, 1161–1189.

Karlan, Dean S, and Jonathan Zinman, 2008, Credit elasticities in less-developed economies: Implications for microfinance, *American Economic Review* 98, 1040–1068.

Karpoff, Jonathan M, and Michael D Wittry, 2018, Institutional and legal context in natural experiments: The case of state antitakeover laws, *The Journal of Finance* 73, 657–714.

Keil, Jan, 2017, The trouble with approximating industry concentration from compustat, *Journal of Corporate Finance* 45, 467–479.

King, Gary, and Richard Nielsen, 2019, Why propensity scores should not be used for matching, *Political analysis* 27, 435–454.

Koenker, Roger, and Kevin F Hallock, 2001, Quantile regression, *Journal of economic perspectives* 15, 143–156.

Kölbel, Julian F, and Adrien-Paul Lambillon, 2022, Who pays for sustainability? an analysis of sustainability-linked bonds, *An Analysis of Sustainability-Linked Bonds (January 12, 2022)*.

Larcker, David F, and Edward M Watts, 2020, Where's the greenium?, *Journal of Accounting and Economics* 69, 101312.

Oberfield, Ezra, and Devesh Raval, 2021, Micro data and macro technology, *Econometrica* 89, 703–732.

Peri, Giovanni, and Vasil Yasenov, 2019, The labor market effects of a refugee wave synthetic control method meets the mariel boatlift, *Journal of Human Resources* 54, 267–309.

Roberts, Michael R, and Toni M Whited, 2013, Endogeneity in empirical corporate finance1, in *Handbook of the Economics of Finance*, vol. 2 . pp. 493–572 (Elsevier).

Sun, Liyang, and Sarah Abraham, 2021, Estimating dynamic treatment effects in event studies with heterogeneous treatment effects, *Journal of Econometrics* 225, 175–199.

Zwick, Eric, and James Mahon, 2017, Tax policy and heterogeneous investment behavior, *American Economic Review* 107, 217–48.