FNCE 9260: Empirical Methods in Corporate Finance

Class 2: Regression Part 2 (Panel Data and DiD)

Daniel Garrett

Outline

1. Logistics
   - Introduction and Unsolicited Advice
   - Quick Review
2. Panel Data
   - Fixed Effects
   - Lagged y and First Differences
3. Differences-in-Differences
4. Examples
   - Bonus Depreciation

Introduction

- **Education:** PhD in Economics from Duke
- **Fields:** Corporate finance, public finance, financial intermediation, public economics
- **Current working projects:**
  - Studying how political fights over **ESG policies in banks** are affecting public finance markets (with Ivanov). Following up with changes in **banking behavior and disclosure**
  - Exploring how firms change use of inputs (labor v. physical capital, foreign v. domestic production) when facing different types of **corporate tax wedges** (with Curtis, Roberts, Ohrn, and Suárez Serrato)
  - **Do cities and states make mistakes** when they hire a banker? Very weird patterns in underwriter/advisor switching (with Guenzel and Shahrabi)

Unsolicited Advice #1

- Life is short and academic researchers are not always great at optimizing their own happiness
- From Eric Zwick's "12 step program," I find that step 13 is most important:

  Enjoy yourself. Ask yourself continually, ***"Am I having fun or will I soon?"***

- Even if you are having fun in graduate school overall, this question helps focus attention on projects, data, coauthors, etc. that are going to keep one from burning out

Unsolicited Advice #2

- Academic research is done under uncertainty
- As a student, you are regularly given the results of academic research, but you don't necessarily get a sense for the result generation process
- As you start independent projects, know that (1) many projects fail at their original aim and (2) you are not expected to know the answers right away
- Manski's 2010 essay "Unlearning and Discovery" provides a discussion of dealing with the uncertainty of a research project/career

  > "I have repeatedly found that I am able to make new discoveries only after I unlearn orthodoxies and go back to basics with an open mind."

Reading and Extra Materials

- The amount of material covered in this course is astronomical, you probably will not fully internalize all of it in the next 14 weeks (and we will not teach every possible aspect)
- You will, however, want to keep the slides, syllabus, and a list of the readings to refer to later
- Today's primary material is derived from the following:
  - *Mostly Harmless* Ch. 5
  - Roberts and Whited (2013) Sections 4 and 7
  - Wooldridge "Introductory Econometrics" Chapter 10
  - Bruce Hansen "Econometrics" Chapters 17 and 18 (has many estimators/proofs)

Coding Language

- I sometimes will use Stata examples, but that's not the only language you can use
- *Causal Inference for the Brave and True* has good examples taking the same techniques and implementing them in python (and nice discussion of how causal inference complements data science)
- *Real Econometrics: The Right Tools to Answer Important Questions* has a very nice R guide if that's your jam
- My goal is more to expose you to the ideas and identification assumptions than to critique code or offer specific solutions since those change over time

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?"
  - One can calculate all sorts of statistics to describe the data, but the mechanism that generates the data requires a model
  - 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

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

Outline

Panel Data Overview

- “Panel data” is the common term used to refer to data where individual units are observed over time
  - A set of countries for the last 20 years and measures of investment and employment
  - Publicly traded firms with financial outcomes listed in Compustat
  - A group of households who are surveyed for consumption over time
- There are at least **3 huge benefits** to panel data over purely cross sectional or time series data
  - Controlling for time-invariant heterogeneity without an instrument or measurement (this class)
  - Allowing for broad forms of heterogeneity (next class)
  - Modeling dynamic relationships and effects (later in the semester)

Panel Data Overview

- "Panel data" is the common term used to refer to data where individual units are observed over time
  - A set of countries for the last 20 years and measures of investment and employment
  - Publicly traded firms with financial outcomes listed in Compustat
  - A group of households who are surveyed for consumption over time
- There are at least **3 huge benefits** to panel data over purely cross sectional or time series data
  - Controlling for time-invariant heterogeneity without an instrument or measurement (this class)
  - Allowing for broad forms of heterogeneity (next class)
  - Modeling dynamic relationships and effects (later in the semester)

Omitted Variable Motivation

- A huge concern for empirical work in finance is the presence of omitted variables
- Variables are omitted for many reasons, but insufficient data (i.e., firms fail to report segment-level assets) and a mechanical inability to measure (i.e., worker productivity in an non-production role) are common
- For both types of omissions, sometimes panel data can help...

Omitted Variable Motivation

- A huge concern for empirical work in finance is the presence of omitted variables
- Variables are omitted for many reasons, but insufficient data (i.e., firms fail to report segment-level assets) and a mechanical inability to measure (i.e., worker productivity in an non-production role) are common
- For both types of omissions, sometimes panel data can help...

Omitted Variable Example, Part 1

- Consider the firm-level relationship between investment and leverage (Ahn, Denis, and Denis 2006):

$$investment_{i,t} = \beta_0 + \beta_1 leverage_{i,t-1} + \nu_{i,t}$$

where $investment$ is capital expenditures/assets for firm i in year t. $leverage_{i,t-1}$ is long-term debt/assets last period.

- What are some potential variables that have been omitted from this model?

Omitted Variable Example, Part 2

- What if we expand this relationship to countries, thinking about the level of government indebtedness and total economic output (Reinhart and Rogoff 2010)

$$growth_{i,t} = \beta_0 + \beta_1 leverage_{i,t-1} + \nu_{i,t},$$

where $growth$ is total output in country i in year t scaled by the output in period $t-1$. $leverage_{i,t-1}$ is the outstanding public (governmental) debt as a share of GDP.

- What are some potential variables that have been omitted from this model?

Omitted Variable Example, Part 3

- It's relatively easy to think of other factors related to investment or GDP growth, but potentially much harder to measure **all** of these factors!!
- Many of these factors influencing the oucome of interest are related to the amount of leverage (either at firm or country level) as well
- So, we should know we will get biased estimates just running these regressions
- **Note:** you are *allowed and sometimes encouraged* to run bad regressions, just know they aren't causal

Benefit #1 of Panel Data

- What panel data primarily solves in many current papers is eliminating problems from **time-invariant omitted variables**
- In the first example, management capacity may be a **fixed characteristic of the firm** affecting leverage and investment
- Panel data will allow us to remove such **"fixed effects"** simply and easily, leaving only variation that comes from the relationship of interest

Setting up the model

- Panel data are data where each unit is observed multiple times (there are exceptions to this definition)
- Assume there are N units, indexed according to i
- Assume there are T periods over which units are observed, indexed by t
- A panel is "balanced" if all N units are observed for all T periods, otherwise it is "unbalanced"

Setting up the model

- Consider the following model:

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

where

$$\begin{aligned} &E(\nu_{it}) = 0 \\ &corr(x_{it}, f_i) \neq 0 \\ &corr(f_i, \nu_{it}) = 0 \\ &corr(x_{it}, \nu_{is}) = 0 \ \forall s, t \in T \quad \text{(strict exogeneity)} \end{aligned}$$

- I think in terms of y_{it} being investment, x_{it} being a tax rate, and f_i being firm technology, personally, but you can think of anything
  - e.g., crime y_{it}, police x_{it}, and local criminal activity f_i

Pooled OLS

- The true model is:

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

- We don't observe f_i (firm technology is not trivial to measure), but we decide to put all of our data into stata and run the **pooled regression**:

```
reg y x, cl(unit)
```

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

- From last class, we know this gives us a specific bias from the OVB formula!!!

$$\beta^{pooled} = \beta + \frac{Cov(x_{it}, f_i)\delta}{Var(x_{it})}$$

Pooled OLS Example

- Crime as a function of police is a famous example ($Cov(x_{it}, f_i) > 0$)

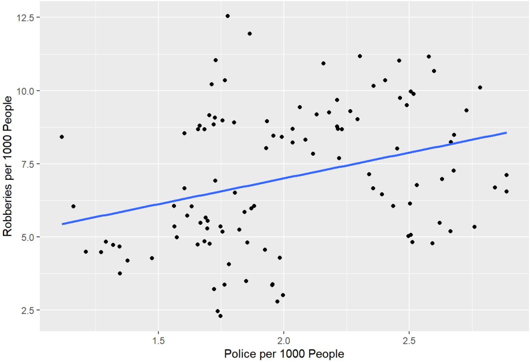


Figure 8.1: Robberies and Police for Large Cities in California

(source: *Real Econometrics: The Right Tools to Answer Important Questions* R Companion by Tony Carilli)

The 'Within' Transformation

- The true model is:

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

- We know the pooled OLS will yield a biased estimate, so what can we do instead?
- The 'within' transformation restricts to variation *within* each unit i to identify the parameter of interest
- Note that the average outcome for each unit is the following:

$$\bar{y}_i = \alpha + \beta \bar{x}_i + \delta f_i + \bar{\nu}_i$$

where bars indicate means across time within unit.

The 'Within' Transformation

- The true model is:

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

  and now 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!

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

The 'Within' Transformation

- The transformed model is:

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

- So, the assumption we need for identification is now that $(x_{it} - \bar{x}_i)$ and $(\nu_{it} - \bar{\nu}_i)$ are not correlated
- Since $\bar{x}_i = \frac{1}{T}\sum x_{it}$ and $\bar{\nu}_i = \frac{1}{T}\sum \nu_{it}$ contain all of their observations, the **exogeneity restriction is stronger** than required for OLS (strict exogeneity)

Fixed Effect Models in General

- The within transformed regression is a **fixed effect** model—it takes into account everything that is fixed at the unit i level and differences those averages away!!
- This fixed variable f_i is *extremely* general
  - Captures all things, observable ***and unobservable***, that are don't change within an observation
  - The panel allows the researcher to work around the level impact of time-invariant characteristics of units
  - Often referred to as "unobserved heterogeneity" in empirical work

Fixed Effect Example

- Allowing β to vary according to i is also an option, from our crime example

```
areg crime c.police#i.city, a(city) cl(city)
```

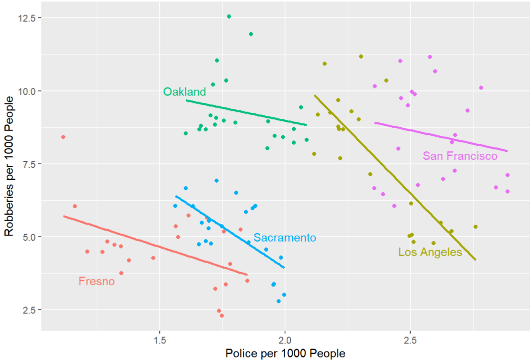


Figure 8.3: Robberies and Police for Specified Cities in California with City-Specific Regression Lines

(source: *Real Econometrics: The Right Tools to Answer Important Questions* R Companion by Tony Carilli)

Estimating Fixed Effect Models

- The most common Stata commands I see include **xtreg**, **areg**, and **reghdfe**, all of which do the within transformation
  - reghdfe only does this transformation on the first listed FE to be "absorbed," so it's best to make that the most granular FE
- The within-transformed estimator leads to decreasing the degrees of freedom, so I suggest using a premade tool or checking the Hansen or Wooldridge texts to get a correction formula if you need to manually code
  - The correction additionally depends on the error structure, so this is not necessarily trivial

Pooled Regression with Dummy Variables

- Computation power used to be a bigger deal than it is today (unless you're using truly 'big data'), so lots of approaches to similar estimators exist
- A nearly identical approach is to simply plug a dummy variable D_i for each unit i into the regression equation:

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

- The relative scales of f_i and δ generally don't matter**, so this rescales $f_i = 1$ and we can estimate the combined δf_i simply using dummies, yielding same estimates

```
reg y x i.unit, cl(unit)
```

- **Sometimes the magnitude of the FE can matter: common to regress dummy estimate on observables

Notes on Pooled Regression with Dummy Variables

- Some texts refer to this object with special names, but I will call it OLS with a lot of parameters
- These dummy variables are perfectly colinear with the intercept α, therefore Stata and R will automatically drop one unit's dummy variable
- The reported intercept is interpreted as the average value for the unit whose dummy variable was dropped
  - For "xtreg, fe", the reported intercept is the average of the individual FE instead of dropping a certain FE
  - You can replicate this by omitting the intercept, saving the FE, and finding the average in reghdfe (assuming a balanced panel):

```
reghdfe y x, a(A=unit) cl(unit) nocon
```

Notes on Pooled Regression with Dummy Variables

- There are two available R^2 measures for a FE regression: within and total
  - Pooled OLS with FE (reg) and areg will give the total
  - xtreg will only give the within
- reghdfe produces both (and is what I use, and everything I have seen recently in replication packages)
- R^2 is not usually relevant, but is being commonly used in "coefficient stability" tests like Oster (2019)

  Garrett, Ordin, Roberts, and Suárez Serrato (2022) has a detailed appendix documenting how to use/interpret such a test

Benefits and costs of using OLS with FE

Benefits

- Can be used to group any type of "unit"
- Year FE (or industry-by-year FE) are common to control for unobserved heterogeneity within time groups
- Allows for arbitrary correlation between each fixed effect f_i and each control x_i variable within group i
- Don't need to model specific group-level differences with parametric functions**

Costs

- Cannot separately identify impact of variables that don't vary within group (will need something like IV)
- Changes the interpretation of coefficients (parametric structure of model)
- **Biases of measurement error/selection can be amplified**
- (Historically) computationally intensive to estimate a bunch of FE

Benefits and costs of using OLS with FE

Benefits

- Can be used to group any type of "unit"
- Year FE (or industry-by-year FE) are common to control for unobserved heterogeneity within time groups
- Allows for arbitrary correlation between each fixed effect f_i and each control x_i variable within group i
- Don't need to model specific group-level differences with parametric functions**

Costs

- Cannot separately identify impact of variables that don't vary within group (will need something like IV)
- Changes the interpretation of coefficients (parametric structure of model)
- **Biases of measurement error/selection can be amplified**
- (Historically) computationally intensive to estimate a bunch of FE

Benefits and costs of using OLS with FE

- Revealed preference would indicate that researchers think the benefits are bigger than costs

(source: Khoa Vu)

A Few Notes on the Costs #1

- One should be careful about what variation is perfectly correlated with the FE
  - Take a stylized model of firm investment as a function of firm size:

    $$\ln(investment)_{it} = \alpha + \beta \ln(firmsize)_{it} + \gamma mature_i + \delta_t + f_i + \varepsilon_{it}$$

    where $\ln(investment)_{it}$ is for firm i in year t and $mature_i$ is an indicator equal to one if the firm was founded before 1980.
  - Here, γ is not identified because $mature_i$ is perfectly collinear with f_i
- **The place to be careful is to figure out if Stata correctly omits estimating γ instead of maybe omitting an additional f_i**

A Few Notes on the Costs #2

- Larcker and Watts (2020) estimate a model exploring the greenium in municipal bonds
- The empirical specification looks at repeated issues from the same issuer
  - Model explaining yields

$$yield_{ijt} = \alpha + \beta green_i + \delta_t + \xi_j + \varepsilon_{ijt}$$

    where $\ln(yield)_{ijt}$ is for bond i, from issuer j, at time t

  - $green_i$ is an indicator equal to 1 for a green bond
- One should be careful (as the authors are) in interpreting the estimate of β as the within-issuer greenium, and not the total greenium
- Consistent with a signaling story with positive greenium, even though $\hat{\beta} = 0$

A Few Notes on the Costs #3

- Computation power can matter at this point, but much less than historically
- From Todd Gormley's FNCE9260 slides in 2016: "Estimating 4-digit SIC×year and firm FE in Compustat requires ≈40 GB memory. No one has this; hence, no one does it"
- Now everyone is using much more memory than this, and the regressions run fine (multicore stata 17 also helps)
- If using many FE in reghdfe like the following

  ```
  reghdfe y x, a(fe1 fe2 fe3 ...) cl(fe_cluster)
  ```

  one should make **fe1 the most granular FE**, reghdfe does the within transformation only on that variable (person in the LEHD, county-industry in bonus paper #2 later, etc.)

A Few Notes on the Costs #4

- FE are identified by reorganizing the within transformation:

$$\hat{f}_i = \bar{y}_i - \hat{\beta}\bar{x}_i \quad \forall i = 1, ..., N$$

- This estimate of the FE is unbiased, but **inconsistent**, meaning sample size increases in N doesn't lead to necessary asymptotics

- this is the "Incidental parameters problem" in that increasing N also adds more parameters

- This also means that non-linear models like Logit, Tobit, and Probit don't work well with a bunch of FE
  - This is one reason Jeffrey Wooldridge is a huge defender of poisson regression, detailed analysis in Wooldridge (1999)

Lagged dependent variables with FE

- Supposed we think the model of the world is recursive but also includes a fixed effect:

$$y_{i,t} = \alpha + \rho y_{i,t-1} + \beta x_{i,t} + f_i + \nu_{i,t}, \quad |\rho| < 1$$

- Same as before, but now the true model includes the lagged dependent variable
- Even if $x_{i,t}$ and f_i are independent, this model **cannot** be estimated with OLS or with FE
- Analytical statement of the bias can be tedious (17.37 in Hansen), but the basic idea is that the within transformation puts all values of $y_{i,t}$ on the RHS with the error, so **strict exogeneity** is not possible
- Will need an instrument for an unbiased estimate

Lagged dependent variables with FE

- Supposed we think the model of the world is recursive but also includes a fixed effect:

$$y_{i,t} = \alpha + \rho y_{i,t-1} + \beta x_{i,t} + f_i + \nu_{i,t}, \quad |\rho| < 1$$

- Same as before, but now the true model includes the lagged dependent variable
- Even if $x_{i,t}$ and f_i are independent, this model **cannot** be estimated with OLS or with FE
- Analytical statement of the bias can be tedious (17.37 in Hansen), but the basic idea is that the within transformation puts all values of $y_{i,t}$ on the RHS with the error, so **strict exogeneity** is not possible
- Will need an instrument for an unbiased estimate

Lagged y or FE?

- Assume we don't know which model is the true model

$$y_{i,t} = \alpha + \rho y_{i,t-1} + \beta x_{i,t} + \nu_{i,t}, \quad |\rho| < 1$$

$$y_{i,t} = \alpha + \beta x_{i,t} + f_i + \nu_{i,t}$$

- If we know the sign of β, we can estimate both models and do a bounding exercise
- For $\beta > 0$,
  - Estimate will be too high if you incorrectly use FE while lagged y model is correct
  - Estimate will be too low if you incorrectly use lagged y while FE model is correct
- 'Bounding' or 'Bracketing' (See *Mostly Harmless*, page 246)

Back to FE models! First Differences

- The within transformation (subtracting unit means) is only one of many ways to estimate a FE model
- Another is to subtract the previous (or any) observation
- Let the true model be

$$y_{i,t} = \alpha + \beta x_{i,t} + f_i + \nu_{i,t}$$

  This is also the true model the period before

$$y_{i,t-1} = \alpha + \beta x_{i,t-1} + f_i + \nu_{i,t-1}$$

- Subtracting the latter statement from the former yields a different FE estimator

$$(y_{i,t} - y_{i,t-1}) = \beta(x_{i,t} - x_{i,t-1}) + (\nu_{i,t} - \nu_{i,t-1})$$

First Differences or Within Estimator?

- With two observations per unit, these methods are identical
- With more than two observations, both estimators are consistent, but efficiency differs
  - If $\nu_{i,t}$ are serially uncorrelated, within is more efficient (or Pooled OLS with FE)
  - If $\nu_{i,t}$ follow a random walk, then FD is more efficient
- If *strict exogeneity is violated* (i.e., $correl(x_{i,t}, \nu_{i,s}) \neq 0$ for some $s \neq t$), within may be better than FD, since inconsistency shrinks with larger T, while FD inconsistency is constant in T
- Good idea? Try both and if they aren't the same, investigate error correlations

Outline

How does panel data relate to ATE estimation?

- Reminder, we generally care about estimating ATE: $E[Y_{1i}] - E[Y_{0i}]$
- But, we don't observe that! We usually have $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] =$

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

- If we have randomization of treatment (so $E[Y_{0i}|D_i = 0] = E[Y_{0i}|D_i = 1]$), then this is already ATE, but randomization a very strong assumption
  - Absent treatment the treated group would have the same outcome y as the untreated group
  - And the untreated group would have had the same expected outcomes as the treated group if they were instead treated

ATE in a Simple Regression Model

- Everything can be re-expressed in the format of a regression model:

$$y_i = \beta_0 + \beta_1(\mathbb{1}\{d_i = 1\}) + \nu_i$$

- The data only show us $E[y_{1i}|d_i = 1] - E[y_{0i}|d_i = 0]$, which in this formulation is is equal to

$$\beta_1 + \underbrace{E[y_{0i}|d_i = 1] - E[y_{0i}|d_i = 0]}_{\text{selection effect}}$$

- So, if we just estimate this regression model using OLS, we will get a **biased estimate**

ATE in a Simple Regression Model

- We can also easily add controls the the regression model
- With controls X_i, our OLS estimator still may include a bias if the treatment is not random conditional on controls

$$\beta_1 + \underbrace{E[y_{0i}|d_i = 1, X_i] - E[y_{0i}|d_i = 0, X_i]}_{\text{selection effect}}$$

- The big problems remain if the financial phenomenon of interest is selected into by observations
  - a firm has to choose to start a subsidiary in a tax haven
  - an entrepreneur chooses to take out a loan
  - a homebuyer chooses whether to have a conforming mortgage or not
- If observably similar entities make different choices, this should indicate **something unobservable matters**!!!

ATE in the Cross-Section (Single Difference)

- Without panel data, we only observe each observation once
- Some observations are treated, $d_i = 1$, and others are not
- We write down the following model to estimate

$$y_i = \beta_0 + \beta_1(\mathbb{1}\{d_i = 1\}) + \nu_i$$

- We know from a few slides ago that an OLS estimate of this regression is biased unless $E(\nu_i | d_i) = 0$, or that treatment is uncorrelated with the error

ATE in the Time Series (Single Difference)

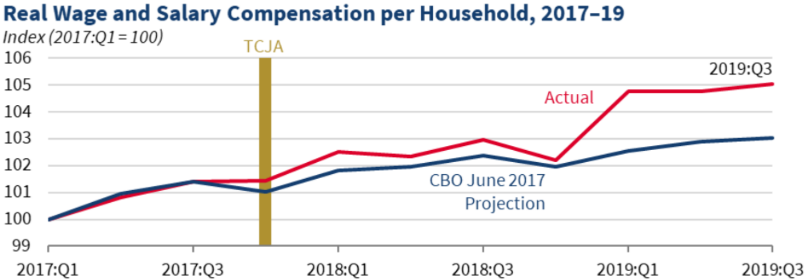
- Without panel data, we only observe the treated unit, but we observe them **before and after treatment**!!
- The unit is untreated for $t \in 0, ..., s-1$ and then treated in period $t = s$ and future periods
- We write down the following model to estimate

$$y_t = \beta_0 + \beta_1(\mathbb{1}\{d_t = 1\}) + \nu_t$$

- We know from a few slides ago that an OLS estimate of this regression is biased unless $E(\nu_t|d_t) = 0$, or that treatment timing is uncorrelated with the error

ATE in the Time Series (Single Difference), Example

- Big questions often have this kind of treatment (one treatment, hits everyone)



Sources: Bureau of Economic Analysis; Census Bureau; Congressional Budget Office; CEA calculations.
Note: Values are adjusted to real terms using the PCE chain price index. Values are indexed such that 2017:Q1 is equal to 100 to account for BEA annual revisions.

- If the CEA wants to estimate national wage impacts of TCJA17, they want to come up with a counterfactual!!!

Natural Panel Solution: Difference-in-Differences (DiD)

- Diffence-in-differences is just the combination of the two single-difference models
- Treatment happens in the cross-section ***and*** over time
- Plain language description: compare the pre to post change in the outcome for the treatment group (1st difference) to the pre to post change in the outcome for the control group (2nd difference)

Difference-in-Differences (DiD) Description

- A DiD regression estimator:

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

  - $p_t = 1$ if treatment occurred before period t
  - $d_i = 1$ if unit i will ever be treated

(Discussion) How can we interpret estimates of β_1, β_2, and β_3?

Difference-in-Differences (DiD) Description

- A DiD regression estimator:

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

	Post-Treatment, (1)	Pre-Treatment, (2)	Difference, (1)-(2)
Treatment, (a)	$\beta_0+\beta_1+\beta_2+\beta_3$	$\beta_0+\beta_2$	$\beta_1+\beta_3$
Control, (b)	$\beta_0+\beta_1$	β_0	β_1
Difference, (a)-(b)	$\beta_2+\beta_3$	β_2	β_3

Source: Michael Wittry

Difference-in-Differences (DiD) Description

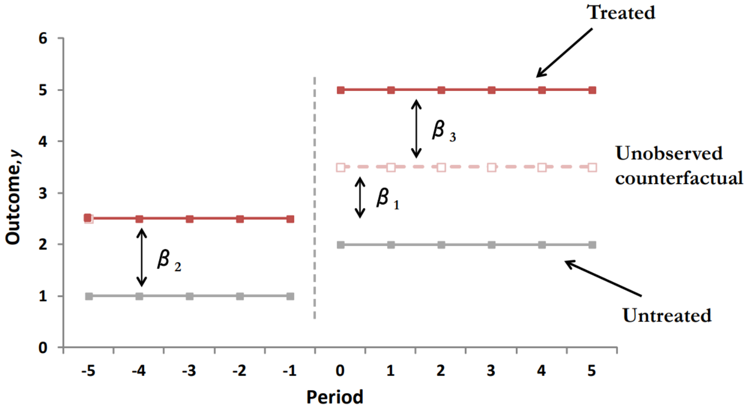
- The identifying assumption behind a causal interpretation to this estimator is often called **the parallel trends assumption**:

  > Absent the treatment, the change in outcome y for the treated units would not have been different than the change in outcome y for untreated units

- It's relatively more simple to observe this assumption in graphical form, so I lean on Todd Gormley's graphs

Difference-in-Differences (DiD) Description

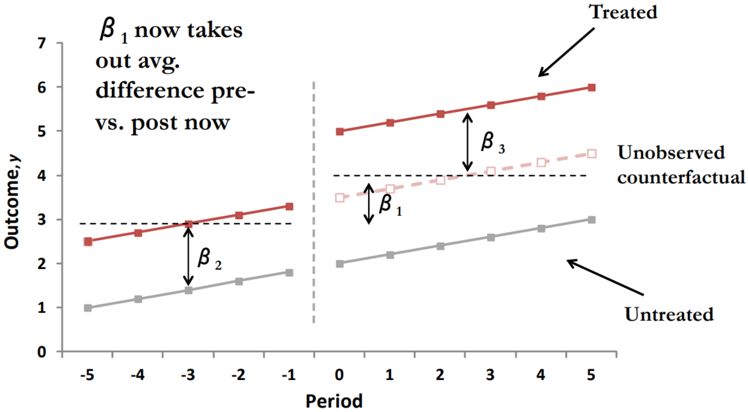
$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 (d_i \times p_t) + \nu_{i,t}$$



Source: Todd Gormley

Difference-in-Differences (DiD) Description

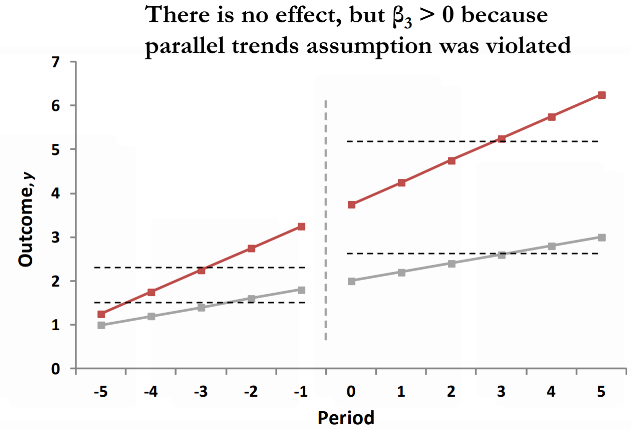
$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 (d_i \times p_t) + \nu_{i,t}$$



Source: Todd Gormley

Difference-in-Differences (DiD) Description

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



Source: Michael Wittry Garrett

Parallel Trends Cautionary Note!!

- Sometimes people look to the trends of treated and control outcomes in the preperiod to argue that "parallel trends holds," or variations of that argument
- Parallel trends in the preperiod is **neither necessary nor sufficient** to establish the parallel trends assumption that identifies a DiD model
- I will expand on this argument later, but the **assumption** is inherently untestable since it only matters ***when treatment happens***, not necessarily before.
- The key threat to identification of a DiD model is a **time-varying omitted variable**–something that affects the treatment group differently around the same time as the treatment

Miscellaneous DiD Comments

- 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, i.e., *constant linear dose response* and (2) influenced by treatment outliers
- Generalized DiD models include time and unit FE, which generally absorbs β_0, β_1, and β_2:

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

- Generalized DiD is often used with staggered treatments, which we will get to briefly by the end of class

Outline

Bonus Depreciation as a Plausibly Exogeneous Shock

- I begin with a discussion of 4 different DiDs resulting from a policy called *bonus depreciation* in the US
- A single shock to the after-tax cost of physical capital can be used in many ways
- This example should highlight (1) different ways of measuring the response to the same variation and (2) the amount of latitude that a researcher has in defining a regression model
- The DiDs of interest
  - Discrete bonus depreciation and changes in investment
  - Continuous bonus depreciation and impact on investment
  - Continuous exposure to discrete bonus depreciation and local labor markets
  - Discrete bonus depreciation and changes in firm inputs and production

Bonus Depreciation, Zwick and Mahon (2017)

- How does business investment respond to tax incentives?
  - Foundational question in corporate tax literature from Hall and Jorgenson (1967)
- How do financial constraints affect business investment?
  - Foundational question in corporate finance going back to Fazzari, Hubbard, and Petersen (1988), ***extremely*** hard to measure constraints
- What models of firm behavior are consistent with these responses?
  - Is the investment response consistent with maximizing firm value? A capital structure pecking order? Something else? (short termism vs. salience vs. agency problems)

Zwick and Mahon (2017) use a (1) simple DiD model with discrete treatment and (2) generalized DiD model with continuous treatment to address the intersection of these questions

Bonus Depreciation, Background

First, some background on bonus depreciation

- In the US, firms recoup investment costs through depreciation expenses for tax purposes
- The schedule of depreciation expenses is set by IRS Pub. 946, which describes asset classes according to the Modified Accelerated Cost Recovery System (MACRS)
- Somewhat serendipitously for research, in the US "the cost recovery periods were not intended to reflect actual useful lives, or even some percentage of the useful lives" (Brazell, Dworin, and Walsh 1989)
- In 2001, the US implemented "bonus depreciation" allowing firms to deduct 30%-100% of capex in the first year of purchase

Bonus Depreciation, Background

First, some background on bonus depreciation

- In the US, firms recoup investment costs through depreciation expenses for tax purposes
- The schedule of depreciation expenses is set by IRS Pub. 946, which describes asset classes according to the Modified Accelerated Cost Recovery System (MACRS)
- Somewhat serendipitously for research, in the US "the cost recovery periods were not intended to reflect actual useful lives, or even some percentage of the useful lives" (Brazell, Dworin, and Walsh 1989)
- In 2001, the US implemented "bonus depreciation" allowing firms to deduct 30%-100% of capex in the first year of purchase

Bonus Depreciation, Background

- Present value of depreciation deductions for a $1 investment

$$z_0 = \sum_{t=0}^{T} \frac{D_t}{(1+r)^t}$$

- T is the recovery period
- D_t are the deductions allowed in year t such that $\sum_t D_t = 1$
- r is the discount rate

- $z_0 \in (0,1]$ is smaller for long tax-life assets
- Bonus increases PV of deductions by:

$$\underbrace{(b+(1-b)z_0)}_{\text{Bonus Depreciation}} - \underbrace{z_0}_{\text{MACRS}} = b(1-z_0)$$

- Value of bonus is higher when z_0 is small–i.e., for long tax-life assets

Bonus Depreciation, Background

- Present value of depreciation deductions for a \$1 investment

$$z_0 = \sum_{t=0}^{T} \frac{D_t}{(1+r)^t}$$

- T is the recovery period
- D_t are the deductions allowed in year t such that $\sum_t D_t = 1$
- r is the discount rate

- $z_0 \in (0, 1]$ is smaller for long tax-life assets
- Bonus increases PV of deductions by:

$$\underbrace{(b + (1-b)z_0)}_{\text{Bonus Depreciation}} - \underbrace{z_0}_{\text{MACRS}} = b(1 - z_0)$$

- Value of bonus is higher when z_0 is small–i.e., for long tax-life assets

Bonus Depreciation, Background

A firm purchases a computer with a 5-year usable life that costs $1,000

Year	0	1	2	3	4	5	Total
MACRS							
Deduction	200	320	192	115	115	58	1000
Tax Benefit	70	112	67.2	40.3	40.3	20.2	350
50% Bonus							
Deduction	600	160	96	57.5	57.5	29	1000
Tax Benefit	210	56	33.6	20.2	20.2	10	350

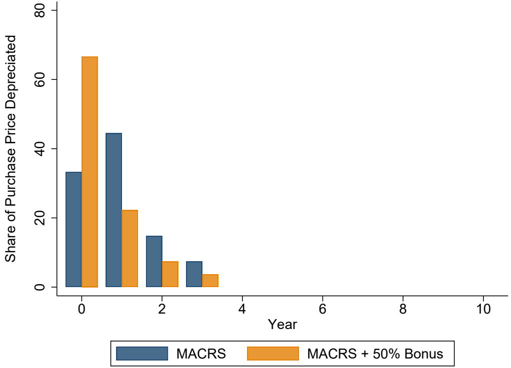
With a 7% discount rate, bonus increases PV but not total benefit

- PV of tax benefits without Bonus: $311 ($z_0 = 0.889$)
- PV of tax benefits with Bonus: $331 ($z_{bonus} = 0.946$)
- $20, or 2% of purchase price, savings from Bonus
- $140, or 14% of purchase price, **new cash flow in year 0**

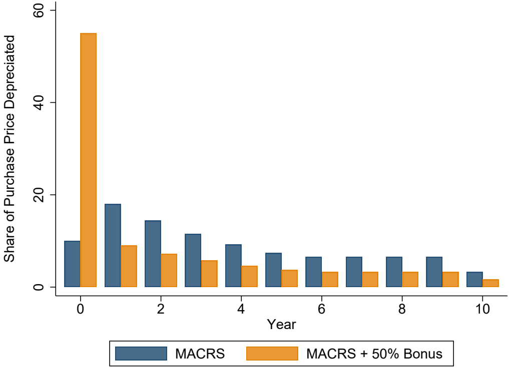
Bonus Depreciation, Background

- Tax rules specify timeline of depreciation deductions D_t
- Bonus depreciation: $\begin{cases} \text{Immediate depreciation} & b\% \\ \text{Remaining deductions} & (1-b)D_t \end{cases}$
- Consider effects of $b = 50\%$ on two types of assets:

A. 3-year MACRS Assets



B. 10-year MACRS Assets



Bonus Depreciation, Zwick and Mahon (2017)

- Bonus allowance is larger for longer lived items
- Industries differ in relative intensity of longer lived assets
- Use tax filings to compute z_0 at the 4-digit NAICS level

Short Duration (NAICS)	Long Duration (NAICS)
Rental and Leasing (532)	Utilities (221)
Publishing (511)	Pipeline Transport (486)
Data Processing (518)	Railroads (482)
Ground Transit (485)	Accommodations (721)
Professional Services (541)	Food Manufacturing (311)

- Cross-sectional variation in bonus generosity to identify effect of bonus in a DiD

Bonus Depreciation, Zwick and Mahon (2017)

- Data: US Corporate tax returns, 1993-2010
  - Size-stratified samples of ≈100,000 returns per year (unbalanced panel)
  - Total tax return outcomes: investment, income, expenses, assets, payouts, employment, industry, location
  - Initially restrict sample to firms with average eligible investment to greater than $100k *[why?]*
- DiD models:
  - Model 1 comparing longest ($d_i = 1$) to shortest ($d_i = 0$) by year

$$y_{i,t} = \alpha_i + \delta_t + \beta_t d_i + \gamma X_{i,t} + \varepsilon_{i,t}$$

  - Model 2 estimating a generalized DiD ($d_{i,t} = z_{N,t}$)

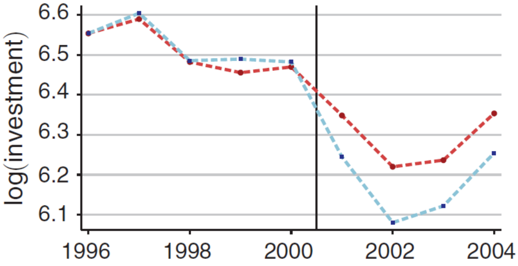
$$y_{i,t} = \alpha_i + \delta_t + \beta d_{i,t} + \gamma X_{i,t} + \varepsilon_{i,t}$$

Bonus Depreciation, Zwick and Mahon (2017)

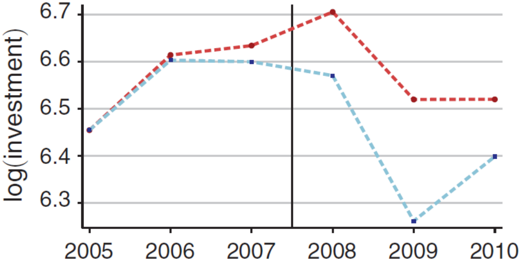
- Model 1: Discrete long vs. short tax-life comparison
  - "Donut Hole" approach dropping most of the data
  - Includes controls for 10-piece splines in assets, sales, profit margin, and age *[why?]*
  - **Non-parametrically reweights the data** according to DiNardo, Fortin, and Lemieux (1996) in 10 bins of assets and 10 bins of sales *[why? argument similar to Oberfield and Raval (2021)]*
  - Normalizes $\beta_{1996} = 0$ (omit that treatment indicator) then adds the δ_t estimates
  - Separates intensive and extensive margins (logs are a big problem in finance, never just add $+1$ (Cohn, Liu, and Wardlaw 2022))
- Identifying assumption: **Parallel Trends*****
  - If no bonus, average outcome paths similar across industries.
  - Concern: time-varying industry shocks coinciding with bonus. (e.g., durables investment more resilient in downturns)

Bonus Depreciation, Zwick and Mahon (2017)

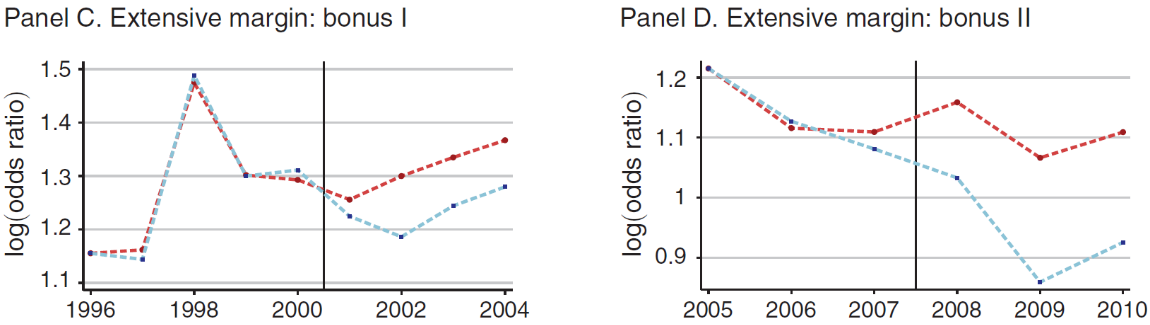
Panel A. Intensive margin: bonus I



Panel B. Intensive margin: bonus II



Bonus Depreciation, Zwick and Mahon (2017)



	(1)	(2)	(3)	(4)	(5)	(6)
	Intensive margin: LHS variable is log(investment)					
$z_{N,t}$	3.69	3.78	3.07	3.02	3.73	4.69
	(0.53)	(0.57)	(0.69)	(0.81)	(0.70)	(0.62)
$CF_{it}/K_{i,t-1}$		0.44				
		(0.016)				
Observations	735,341	580,422	514,035	221,306	585,914	722,262
Clusters (firms)	128,001	100,883	109,678	63,699	107,985	124,962
R^2	0.71	0.74	0.73	0.80	0.72	0.71
	Extensive margin: LHS variable is $\log(P(\text{investment} > 0))$					
$z_{N,t}$	3.79	3.87	3.12	3.59	3.99	4.00
	(1.24)	(1.21)	(2.00)	(1.14)	(1.69)	(1.13)
$CF_{it}/K_{i,t-1}$		0.029				
		(0.0100)				
Observations	803,659	641,173	556,011	247,648	643,913	803,659
Clusters (industries)	314	314	314	274	277	314
R^2	0.87	0.88	0.88	0.93	0.90	0.90
	Tax term: LHS variable is investment/lagged capital					
$\frac{1-t_c z}{1-t_c}$	−1.60	−1.53	−2.00	−1.42	−2.27	−1.50
	(0.096)	(0.095)	(0.16)	(0.13)	(0.14)	(0.10)
$CF_{it}/K_{i,t-1}$		0.043				
		(0.0023)				
Observations	637,243	633,598	426,214	211,029	510,653	631,295
Clusters (firms)	103,890	103,220	87,939	57,343	90,145	103,565
R^2	0.43	0.43	0.48	0.54	0.45	0.44
Controls	No	No	No	No	Yes	No
Industry trends	No	No	No	No	No	Yes

Bonus Depreciation, Zwick and Mahon (2017)

- Interpretation of the intensive margin DiD in column (3), the 2001 episode:
  - One unit increase in $z_{N,t}$ leads to 307% more investment
  - Average change in $z_{N,t}$ with 30% bonus is increase of 0.036 (mean $z_0 = 0.88$ in table 2)
  - Combination of change in treatment due to bonus and the point estimate implies $0.036 \times 3.07 = 0.11$, or an 11% increase in investment on average
  - Implies average impact is the same magnitude as the gap between least and most treated (about 12%)**
- In this context, generalized DiD has the benefit that it has an more interpretation for the coefficient based on the parameterization of the treatment (we know exactly how much bonus changes the NPV of depreciation deductions)

Bonus Depreciation, Zwick and Mahon (2017)

- TL,DR for the rest of the paper:
  - This effect is massive, how can it be true?
  - Small firms have a larger response (heterogeneous responses missed in earlier studies on public firms)
  - Non-profitable firms have no response despite having (delayed) benefit
  - Firms take on a bit of leverage, increase overall activity
  - Effects **much larger** for firms low on cash
- (Discussion) Did anything jump out as weird in the DiDs?

Bonus Depreciation, Garrett, Ohrn, and Suárez Serrato (2020)

- While Zwick and Mahon (2017) and others show beneficial impact on investment, what about overall firm behavior that we care about?
- In particular, growing concern that bonus depreciation would lead to firms replacing workers with machines (Acemoglu, Manera, and Restrepo 2020)
- Garrett, Ohrn, and Suárez Serrato (2020) uses a local labor market approach to assess whether workers, on the whole, are beneficiaries of bonus or if they are net replaced
  - **Note:** This is a shift-share type measurement that you will see with more rigor in Lecture 4.

Bonus Depreciation, Garrett, Ohrn, and Suárez Serrato (2020)

Long Duration Exposure

- 4-digit NAICS industry duration measurement from Zwick & Mahon (2017): z_i
  - stable over time; MACRS equipment types specified by industry definitions
- QCEW Annual Employment data at the 4-digit NAICS × county level

Local Employment and Earnings Data

- QCEW Annual Employment at the 3-digit NAICS × county levels

"Local" Capital Stock

- Current-Cost Net Capital Stock of Private Nonresidential Fixed Assets from BEA
  - Broken out into 3-digit NAICS industries at the national level
- QCEW Annual Employment at the 3-digit NAICS × county levels

Bonus Depreciation, Garrett, Ohrn, and Suárez Serrato (2020)

Definition of Exposure to Long Duration Industries

- Define Emp_{ict} to be employment of industry i in county c in year t
- Define $long_i = \begin{cases} 1 & \text{bottom tercile of } z_i \text{ distribution} \\ 0 & \text{otherwise} \end{cases}$

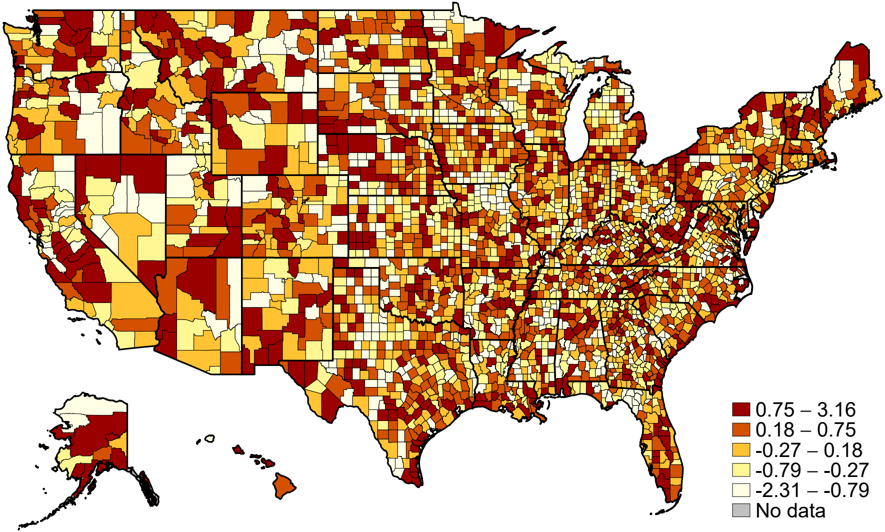
Local bonus **Exposure** for a county is defined as the following:

$$\text{Exposure}_c = \frac{\sum_i Emp_{ic2001} \mathbb{1}(long_i = 1)}{\sum_i Emp_{ic2001}}$$

Discretization

- eliminates the effects of outliers in z_i distribution
- eliminates concerns about discounting assumption
- Results robust to $long_i$ defined at the 25^{th} and 40^{th} percentiles

Bonus Depreciation, Garrett, Ohrn, and Suárez Serrato (2020)



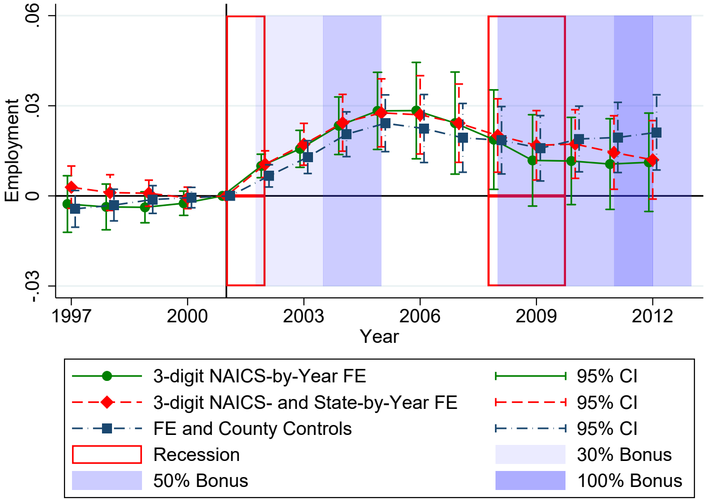
Bonus Depreciation, Garrett, Ohrn, and Suárez Serrato (2020)

- Diff-in-diff: compare changes in employment across otherwise similar county-industries
- Let $\Delta Emp_{ict} \equiv \frac{Emp_{ict} - Emp_{ic2001}}{Emp_{ic2001}}$ be cumulative employment growth (computation burden trick to eliminate ≈150,000 county-industry FE)
- Estimating equation:

$$\Delta Emp_{ict} = \sum_{y=1997}^{2012} \beta_y \left[\text{Exposure}_c \times \mathbb{1}(t = y) \right] + \mathbf{X'}_{ct}\gamma_t + \mu_{st} + \nu_{it} + \epsilon_{ict}$$

- State×year (750), NAICS 3-digit industry×year FE (1,515)
- $\mathbf{X}_{ct}$ controls include
  - other county economic shocks (trade, routine labor, other tax policies),
  - existing capital per person,
  - demographics (race, education)

Bonus Depreciation, Garrett, Ohrn, and Suárez Serrato (2020)



Bonus Depreciation, Garrett, Ohrn, and Suárez Serrato (2020)

New fact

- An IQR increase in Bonus $\implies$ 2.1% employment growth

Jobs Created

- Jobs created = average exposure (2.72 IQR units) × estimated effect (2.1%, assuming no exposure to long-duration industries has no spillover)
- Relative to the 109.3 million workers in the US in 2001 (QCEW)
  $\rightarrow$ increase in employment of **6.24 million jobs**

This calculation is based only on partial equilibrium effects. If bonus depreciation increases or decreases aggregate employment, that is a "missing intercept problem"

Bonus Depreciation, Garrett, Ohrn, and Suárez Serrato (2020)

- **Parallel trends impossible hard to verify,** so we leaned on a placebo
- Structures, land, drilling equipment, and intellectual property were not eligible for bonus
- Identify industries with 5x more stock in non-eligible categories
- NAICS codes:
  - Oil and gas extraction (2111)
  - Rail transportation (4821)
  - Performing arts and sports (7111, 7112)
  - Travel accommodation and RV parks (7211, 7212)
  - Other Services (81)
- Exposure to long-duration industries with mostly non-eligible capital does not have an effect on employment

Bonus Depreciation, Garrett, Ohrn, and Suárez Serrato (2020)

Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)

- While the number of workers in local markets was increasing, that says very little about worker welfare
- It may be that good paying manufacturing jobs are replaced, more people move into support roles and other industries
- Using plant-level data matching workers with capital, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023) uses a similar discretized DiD to test plant impacts, heterogeneity, and to identify worker subsitutability in response to bonus

Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)

- Marshall (1890) & Hicks (1932) note labor demand depends on:
  - Scale effect: firm expands production and hires more workers
  - Substitution/complementarity between labor and capital
- To separate these effects, the model assumes:

1. Bonus lowers cost of capital — $\phi = \frac{\partial \ln(\text{Cost of Capital})}{\partial \text{Bonus}} < 0$
2. Product demand elasticity (CES) — $\eta > 1$
3. Production function has CRTS with inputs K, L, J — s_K, s_L, s_J: Cost shares; σ_{KL}: Allen elasticity of substitution; Substitutes $\sigma_{KL} > 0$; Complements $\sigma_{KL} < 0$

Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)

- Marshall (1890) & Hicks (1932) note labor demand depends on:
  - Scale effect: firm expands production and hires more workers
  - Substitution/complementarity between labor and capital
- To separate these effects, the model assumes:

1. Bonus lowers cost of capital — $\phi = \frac{\partial \ln(\text{Cost of Capital})}{\partial \text{Bonus}} < 0$
2. Product demand elasticity (CES) — $\eta > 1$
3. Production function has CRTS with inputs K, L, J —
   s_K, s_L, s_J: Cost shares
   σ_{KL}: Allen elasticity of substitution
   Substitutes $\sigma_{KL} > 0$
   Complements $\sigma_{KL} < 0$

Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)

- The firm FOCs of this model make a prediction for $\frac{\partial \ln K}{\partial \text{Bonus}}$—how much capital will change with respect to a capital subsidy
- Let $\beta^K = \frac{\partial \ln K}{\partial \text{Bonus}}$ and $\beta^L = \frac{\partial \ln L}{\partial \text{Bonus}}$
- A DiD explaining $\ln K$ and $\ln L$ (and all other factors) will pin down this model and give elasticity estimates

Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)

- Model predictions for reduced-form effects:

$$\text{Capital}: \quad \beta^K = (\underbrace{s_K\eta}_{\text{Scale}} + \underbrace{s_J\sigma_{KJ} + s_L\sigma_{KL}}_{\text{Substitution}}) \times \underbrace{-\phi}_{\text{Cost of Capital} > 0}$$

$$\text{Labor}: \quad \beta^L = s_K(\eta - \sigma_{KL}) \times -\phi$$

- Labor demand increases if:
  1. K-L are complements: $\sigma_{KL} < 0$
  2. Scale effect dominates: $\eta > \sigma_{KL} > 0$

Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)

- Model predictions for reduced-form effects:

$$\text{Capital}: \quad \beta^K = (\underbrace{s_K\eta}_{\text{Scale}} + \underbrace{s_J\sigma_{KJ} + s_L\sigma_{KL}}_{\text{Substitution}}) \times \underbrace{-\phi}_{\text{Cost of Capital} > 0}$$

$$\text{Labor}: \quad \beta^L = s_K(\eta - \sigma_{KL}) \times -\phi$$

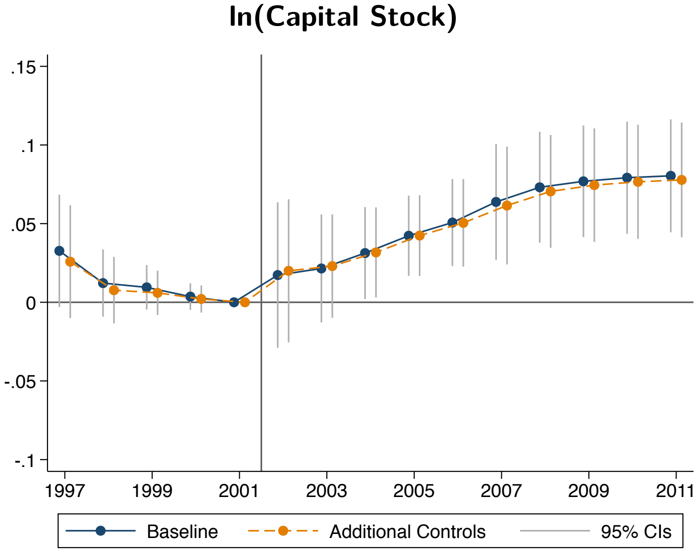
- Labor demand increases if:
  1. K-L are complements: $\sigma_{KL} < 0$
  2. Scale effect dominates: $\eta > \sigma_{KL} > 0$

Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)

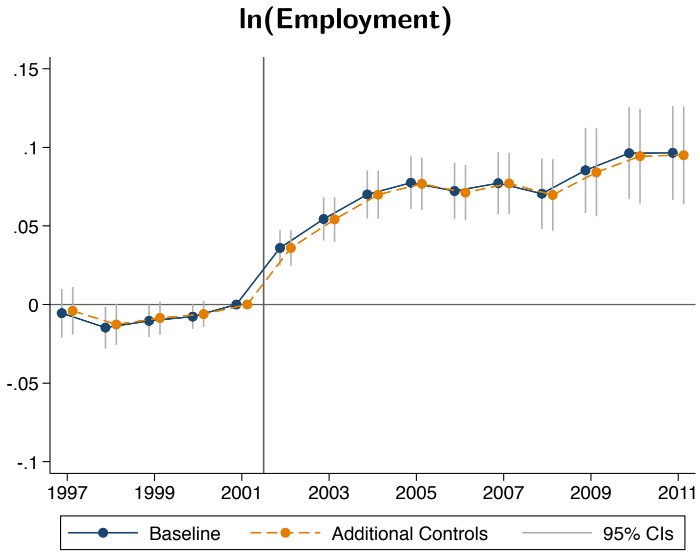
$$Y_{it} = \alpha_i + \mu_{\mathbf{it}} + \sum_{y=1997}^{2011} \beta_y \mathbb{I}[\text{Long Duration}]_i \times \mathbb{I}[y = t]_t + \varepsilon_{it}$$

- Outcome Y_{it} for plants i in year t (log capital, log employment)
- $\mu_{\mathbf{it}}$: plant characteristics interacted with year fixed effects control for time-varying determinants of outcomes
- $\mathbb{I}[\text{Long Duration}]_i = 1$ for long-duration plants (2001 primary industry)
- $\beta_{1997} - \beta_{2011}$: differences in outcomes between long-duration and short-duration plants in each year

Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)



Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)



Identification and Estimation

- Compare labor and scale (weighted average of all inputs) effects:

$$\left.\begin{aligned} \text{Labor}: \quad \beta^L &= s_K(\eta - \sigma_{KL}) \times -\phi \\ \text{Scale effect}: \quad \bar{\beta} &= s_K \eta \times -\phi \end{aligned}\right\} \Longrightarrow \begin{aligned} \frac{\beta^L}{\bar{\beta}} &= 1 - \frac{\sigma_{KL}}{\eta} \\ \sigma_{KL} &= \eta\left(1 - \frac{\beta^L}{\bar{\beta}}\right) \end{aligned}$$

- Labor and scale effects identify ϕ and η
- Estimate model parameters $(\sigma_{KL}, \sigma_{KJ}, \phi, \eta)$ via Classical Minimum Distance
  - Combines information from effects on inputs and revenue
  - Parameters consistent with cost minimization $s_J\sigma_{KJ} + s_L\sigma_{KL} > 0$ (Allen, 1938)
  - Estimate or calibrate $\eta \in [2, 5]$

Identification and Estimation

- Compare labor and scale (weighted average of all inputs) effects:

$$\left.\begin{array}{lrcl} \text{Labor}: & \beta^L & = & s_K(\eta - \sigma_{KL}) \times -\phi \\ \text{Scale effect}: & \bar{\beta} & = & s_K \eta \times -\phi \end{array}\right\} \Longrightarrow \begin{array}{c} \frac{\beta^L}{\bar{\beta}} = 1 - \frac{\sigma_{KL}}{\eta} \\ \sigma_{KL} = \eta\left(1 - \frac{\beta^L}{\bar{\beta}}\right) \end{array}$$

- Labor and scale effects identify ϕ and η
- Estimate model parameters $(\sigma_{KL}, \sigma_{KJ}, \phi, \eta)$ via Classical Minimum Distance
  - Combines information from effects on inputs and revenue
  - Parameters consistent with cost minimization $s_J \sigma_{KJ} + s_L \sigma_{KL} > 0$ (Allen, 1938)
  - Estimate or calibrate $\eta \in [2, 5]$

Identification and Estimation

- Compare labor and scale (weighted average of all inputs) effects:

$$\left.\begin{array}{lcc} \text{Labor}: & \beta^L = & s_K(\eta - \sigma_{KL}) \times -\phi \\ \text{Scale effect}: & \bar{\beta} = & s_K \eta \times -\phi \end{array}\right\} \implies \begin{array}{c} \frac{\beta^L}{\bar{\beta}} = 1 - \frac{\sigma_{KL}}{\eta} \\ \sigma_{KL} = \eta \left(1 - \frac{\beta^L}{\bar{\beta}}\right) \end{array}$$

- Labor and scale effects identify ϕ and η
- Estimate model parameters $(\sigma_{KL}, \sigma_{KJ}, \phi, \eta)$ via Classical Minimum Distance
  - Combines information from effects on inputs and revenue
  - Parameters consistent with cost minimization $s_J \sigma_{KJ} + s_L \sigma_{KL} > 0$ (Allen, 1938)
  - Estimate or calibrate $\eta \in [2, 5]$

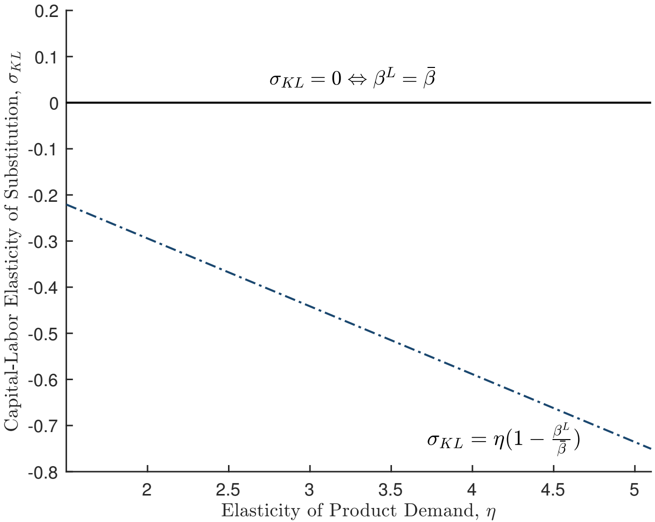
Identification and Estimation

- Compare labor and scale (weighted average of all inputs) effects:

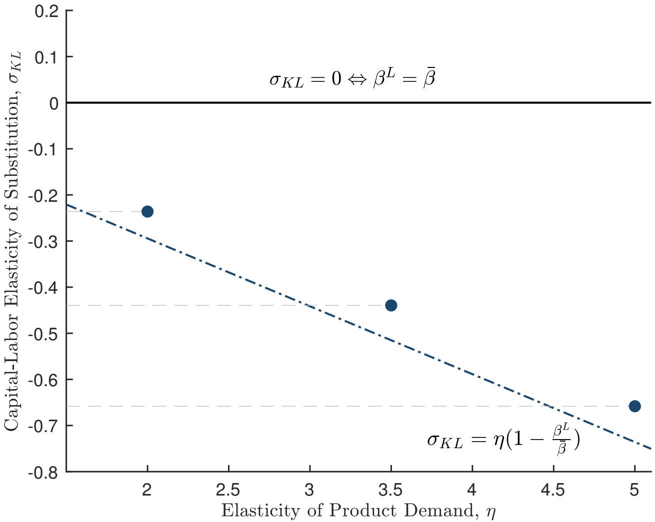
$$\left.\begin{aligned} \text{Labor :} \quad \beta^L &= s_K(\eta - \sigma_{KL}) \times -\phi \\ \text{Scale effect :} \quad \bar{\beta} &= s_K \eta \times -\phi \end{aligned}\right\} \implies \begin{aligned} \frac{\beta^L}{\bar{\beta}} &= 1 - \frac{\sigma_{KL}}{\eta} \\ \sigma_{KL} &= \eta\left(1 - \frac{\beta^L}{\bar{\beta}}\right) \end{aligned}$$

- Labor and scale effects identify ϕ and η
- Estimate model parameters $(\sigma_{KL}, \sigma_{KJ}, \phi, \eta)$ via Classical Minimum Distance
  - Combines information from effects on inputs and revenue
  - Parameters consistent with cost minimization $s_J \sigma_{KJ} + s_L \sigma_{KL} > 0$ (Allen, 1938)
  - Estimate or calibrate $\eta \in [2, 5]$

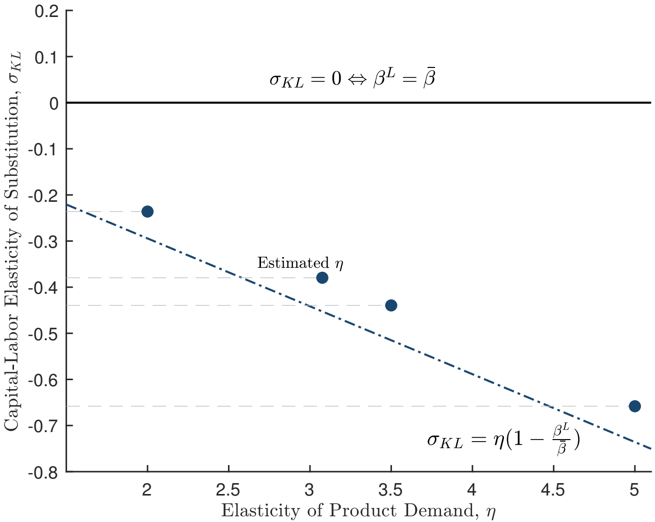
Classical Minimum Distance Estimates of σ_{KL}



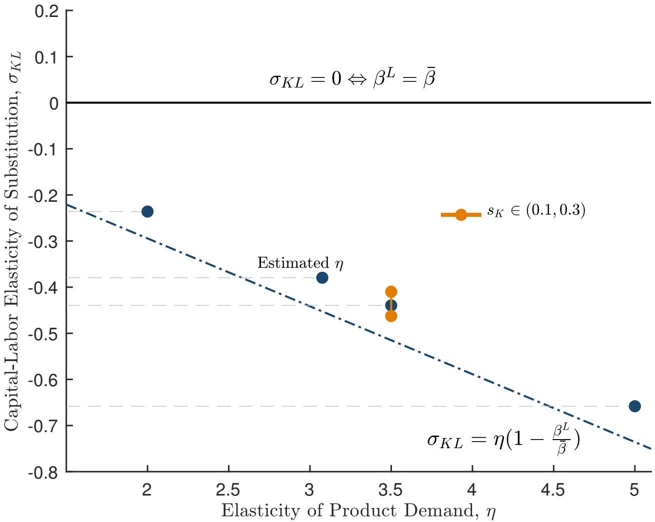
Classical Minimum Distance Estimates of σ_{KL}



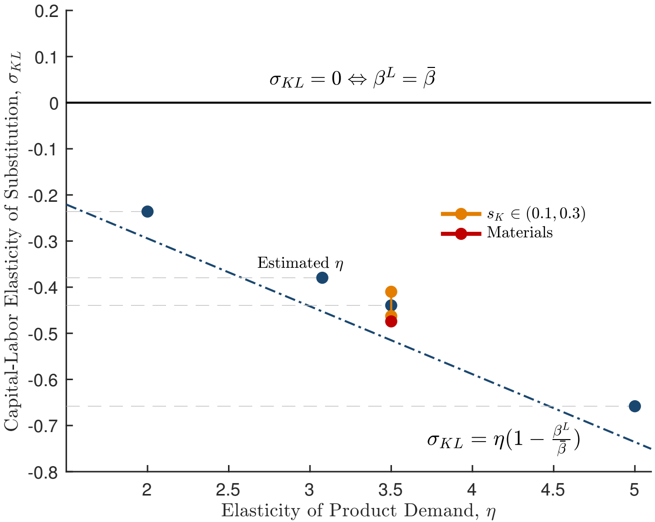
Classical Minimum Distance Estimates of σ_{KL}



Classical Minimum Distance Estimates of σ_{KL}



Classical Minimum Distance Estimates of σ_{KL}



Bonus Depreciation, Curtis, Garrett, Ohrn, Roberts, and Serrato (2023)

$$Y_{it} = \alpha_i + \mu_{\mathbf{it}} + \sum_{y=1997}^{2011} \beta_y \mathbb{I}[\text{Long Duration}]_i \times \mathbb{I}[y = t]_t + \varepsilon_{it}$$

- Estimates of β_{2011} on 3 different outcomes pins down a simple model of capital-labor substitution
- Estimates suggest that bonus did not lead to replacing workers with machines, even at the plant level
- How does the definition of treatment as long-duration being the top 3rd of plants matter?

Bonus Depreciation, Overall

- Giant threat to identification: there is a time-varying unobserved factor correlated with tax-lives of assets at an industry level that is changing around 2001
- (Discussion) What could such an example be?

Bonus Depreciation, Overall

- There is a second fundamental concern with these papers–the estimates all come from comparing slightly less treated to slightly more treated observations
- A manifestation of the "**missing intercept problem**": aggregate impacts are perfectly collinear with FE when we try to learn about aggregates from cross-sectional variation
- See "Moll_2021_lecture.pdf" on Canvas for a short lecture by Ben Moll on this issue if interested
  - Classic problem in many recent, impactful papers (Mian and Sufi 2009; Autor, Dorn, and Hanson 2016)
  - "Good methodological question to think about. High return from any progress."

Bonus Depreciation, Note #1

- One does not have to stop at 2 differences (DiD)
- A very natural thing to do is to extend to more differences (diff-in-diff-in-diff-in-diff...)
- Curtis, Garrett, Ohrn, Roberts, and Serrato (2023) showcases one example of triple difference, which is the difference between DiD estimators for two different groups. Generally:

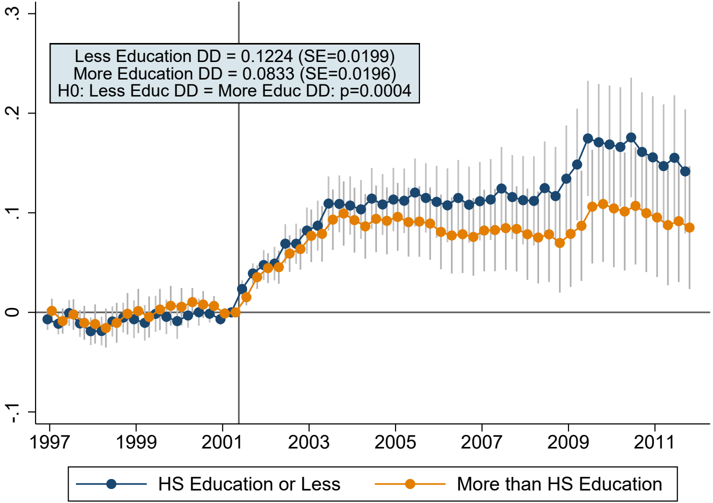
$$y_{i,j,t} = \delta_t + \gamma_i + \xi_j + \beta_3(d_i \times p_t) + \beta_4(d_i \times k_j) + \beta_5(p_t \times k_j) + \beta_7(d_i \times p_t \times k_j) + \nu_{i,t}$$

- What is the interpretation of each of the βs now?

Bonus Depreciation, Note #1

- A weird result in the paper is that average wages for treated plants go down
- The paper shows that this is due to a changing composition of workers: relatively lower wage workers make up more of the employment gains
- We formalize this with a Triple Diff with education level of the worker
  - Define new observation at state-industry-education type
  - New indicator and interactions for high- and low-education workers
  - Can either (presented) show separate betas by estimating separately by j or (not presented other than p-value) show difference term by estimating stacked (being careful with clusters for errors)

Bonus Depreciation, Note #1



Bonus Depreciation, Note #1

- Running each subsample model separately (note this implicitly includes other interactions)

```
foreach demo in hs_educ morethan_hs {
gen beta_`demo' = 0
gen se_high_`demo' = 0
gen se_low_`demo' = 0

eststo r`demo': reghdfe ln_emp treat_q1-treat_q17 treat_q19-treat_q60 ///
if educ_stratum=="`demo'" [aweight=emp2001], ///
absorb(state_naics_fe state_qtr_fe i.pre_growth_`demo'#i.qtr) ///
vce(cluster state_naics_fe)
}
```

Bonus Depreciation, Note #1

- When pooling into stacked regression for statistical tests, **cluster variable across subsamples should be the same**

```
foreach var of varlist state_naics_fe state_qtr_fe year post_treat {
gen `var'_more = `var'*(e_stratum=="morethan_hs")
gen `var'_hs   = `var'*(e_stratum=="hs_educ")
}

* handling the joint cluster variable
gen state_naics_fe = state_naics_fe_hs
replace state_naics_fe = state_naics_fe_more if e_stratum=="morethan_hs"

reghdfe ln_emp post_treat_more post_treat_hs [aweight=emp2001], ///
absorb(state_naics_fe_more state_qtr_fe_more ///
state_naics_fe_hs state_qtr_fe_hs) ///
cluster(state_naics_fe) nocons
```

Bonus Depreciation, Note #2

- The logic of taking differences across the estimates for different populations is very powerful
- Kitagawa (1955), Oaxaca (1973), and Blinder (1973) introduce a very intuitive way of decomposing any change in outcome variables
- Intuitively, the predicted change in y is the sum of the change in covariates multiplied by original $\hat{\beta}$s ("quantities") plus the change in $\hat{\beta}$s multiplied by the original covariates ("prices")
- I also present a version of this in Curtis, Garrett, Ohrn, Roberts, and Serrato (2023) to decompose the decline in wages into different sources

Bonus Depreciation, Note #2

- We can write down four equations that describe the relative wage in pre vs. post, treated vs. untreated

$$
\begin{aligned}
wage_{jst}^{\text{treat, pre}} &= \alpha_{js}^{\text{treat, pre}} + \gamma_{st}^{\text{treat, pre}} + \beta^{\text{treat, pre}} X_{jst}^{\text{treat, pre}} + \varepsilon_{jst} \\
wage_{jst}^{\text{treat, post}} &= \alpha_{js}^{\text{treat, post}} + \gamma_{st}^{\text{treat, post}} + \beta^{\text{treat, post}} X_{jst}^{\text{treat, post}} + \varepsilon_{jst} \\
wage_{jst}^{\text{control, pre}} &= \alpha_{js}^{\text{control, pre}} + \gamma_{st}^{\text{control, pre}} + \beta^{\text{control, pre}} X_{jst}^{\text{control, pre}} + \varepsilon_{jst} \\
wage_{jst}^{\text{control, post}} &= \alpha_{js}^{\text{control, post}} + \gamma_{st}^{\text{control, post}} + \beta^{\text{control, post}} X_{jst}^{\text{control, post}} + \varepsilon_{jst}.
\end{aligned}
$$

- X_{jst} in each regression include the share of young employees, share of employees with less than a high school education, share of non-white employees, and share of employees that are female

Bonus Depreciation, Note #2

- We can estimate each of these models using OLS with FE, which yields FE and $\hat{\beta}$
- Taking a difference between the first two models yields

$$\Delta \ln(\hat{wage})^{\text{treat}} = \Delta \text{FE}^{\text{treat}} + \hat{\beta}^{\text{treat, post}} \bar{X}^{\text{treat, post}} - \hat{\beta}^{\text{treat, pre}} \bar{X}^{\text{treat, pre}}.$$

- The trick is to add and subtract $\hat{\beta}^{\text{treat,pre}} \bar{X}^{\text{treat,post}}$ on the RHS

$$\begin{aligned} \Delta \ln(\hat{wage})^{\text{treat}} =& \Delta \text{FE}^{\text{treat}} \\ &+ (\hat{\beta}^{\text{treat,post}} - \hat{\beta}^{\text{treat,pre}}) \overline{X}^{\text{treat,pre}} \\ &+ \hat{\beta}^{\text{treat,pre}} (\overline{X}^{\text{treat,post}} - \overline{X}^{\text{treat,pre}}) \end{aligned}$$

Bonus Depreciation, Note #2

- Change in wages for treated firms is:

$$\Delta \ln(\hat{wage})^{\text{treat}} = \underbrace{\Delta \text{FE}^{\text{treat}}}_{\text{Unobserved heterogeneity}} + \underbrace{\Delta \hat{\beta}^{\text{treat}} \overline{X}^{\text{treat, pre}}}_{\text{Wages conditional on observables}} + \underbrace{\hat{\beta}^{\text{treat, pre}} \Delta \overline{X}^{\text{treat}}}_{\text{Observable composition}}$$

- In the DiD context:

$$\Delta \ln(\hat{wage})^{\text{treat}} - \Delta \ln(\hat{wage})^{\text{control}} = \Delta \hat{FE}^{\text{treat}} - \Delta \hat{FE}^{\text{control}}$$
$$+ \Delta \hat{\beta}^{\text{treat}} \overline{X}^{\text{treat, pre}} - \Delta \hat{\beta}^{\text{control}} \overline{X}^{\text{control, pre}}$$
$$+ \hat{\beta}^{\text{treat, pre}} \Delta \overline{X}^{\text{treat}} - \hat{\beta}^{\text{control, pre}} \Delta \overline{X}^{\text{control}}$$

Note: Easiest to do statistical tests on these objects by estimating all in one regression with FE interactions for subsamples

Bonus Depreciation, Note #2

- Change in wages for treated firms is:

$$\Delta \ln(\hat{wage})^{\text{treat}} = \underbrace{\Delta \text{FE}^{\text{treat}}}_{\text{Unobserved heterogeneity}} + \underbrace{\Delta \hat{\beta}^{\text{treat}} \overline{X}^{\text{treat, pre}}}_{\text{Wages conditional on observables}} + \underbrace{\hat{\beta}^{\text{treat, pre}} \Delta \overline{X}^{\text{treat}}}_{\text{Observable composition}}$$

- In the DiD context:

$$\begin{aligned} \Delta \ln(\hat{wage})^{\text{treat}} - \Delta \ln(\hat{wage})^{\text{control}} &= \Delta \hat{FE}^{\text{treat}} - \Delta \hat{FE}^{\text{control}} \\ &+ \Delta \hat{\beta}^{\text{treat}} \overline{X}^{\text{treat, pre}} - \Delta \hat{\beta}^{\text{control}} \overline{X}^{\text{control, pre}} \\ &+ \hat{\beta}^{\text{treat, pre}} \Delta \overline{X}^{\text{treat}} - \hat{\beta}^{\text{control, pre}} \Delta \overline{X}^{\text{control}} \end{aligned}$$

Note: Easiest to do statistical tests on these objects by estimating all in one regression with FE interactions for subsamples

Bonus Depreciation, Note #2

- Change in wages for treated firms is:

$$\Delta \ln(\hat{wage})^{\text{treat}} = \underbrace{\Delta \text{FE}^{\text{treat}}}_{\text{Unobserved heterogeneity}} + \underbrace{\Delta \hat{\beta}^{\text{treat}} \overline{X}^{\text{treat, pre}}}_{\text{Wages conditional on observables}} + \underbrace{\hat{\beta}^{\text{treat, pre}} \Delta \overline{X}^{\text{treat}}}_{\text{Observable composition}}$$

- In the DiD context:

$$\begin{aligned} \Delta \ln(\hat{wage})^{\text{treat}} - \Delta \ln(\hat{wage})^{\text{control}} &= \Delta \hat{FE}^{\text{treat}} - \Delta \hat{FE}^{\text{control}} \\ &+ \Delta \hat{\beta}^{\text{treat}} \overline{X}^{\text{treat, pre}} - \Delta \hat{\beta}^{\text{control}} \overline{X}^{\text{control, pre}} \\ &+ \hat{\beta}^{\text{treat, pre}} \Delta \overline{X}^{\text{treat}} - \hat{\beta}^{\text{control, pre}} \Delta \overline{X}^{\text{control}} \end{aligned}$$

Note: Easiest to do statistical tests on these objects by estimating all in one regression with FE interactions for subsamples

DiD Beyond Bonus Depreciation

- Bonus depreciation is exciting because exogenous shocks to capital prices are few and far between, but the single treatment limits interpretation
- Many generalized DiD papers use "staggered implementation," treatment doesn't happen to all units at the same time
  - The closest to this in bonus depreciation is Ohrn (2019) using the state-level rollout
- **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 Beyond Bonus Depreciation

- Bonus depreciation is exciting because exogenous shocks to capital prices are few and far between, but the single treatment limits interpretation

- Many generalized DiD papers use "staggered implementation," treatment doesn't happen to all units at the same time
  - The closest to this in bonus depreciation is Ohrn (2019) using the state-level rollout

- **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 (we will talk more about this next class)
- (Discussion) What could such an example be?

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 any given model
- See discussion in Bertrand, Duflo, and Mullainathan (2004)
- Modern solution is cluster-robust SE with clusters defined at unit (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), need to check entry/exit
  - DFL weighting to force composition to be constant wrt observables (Zwick and Mahon 2017)
  - Ignore the existence of missing data (many Compustat papers, not a good solution)
  - 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**

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

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

In Next Class

- Starting with presentations from "problem set" 2
- Heterogeneous treatment effects with panel data
  - Across characteristics (with some quantile regression)
  - Across cohorts (starting from De Chaisemartin and d'Haultfoeuille (2020))
- Matching, synthetic control, and semiparametrics
- Readings:
  - Chapters 5.2 and 7.1 of *Mostly Harmless*
  - Gibbons et al. (2019) and Abadie (2021) from the syllabus are particularly practitioner friendly

Acemoglu, Daron, Andrea Manera, and Pascual Restrepo, 2020, Does the us tax code favor automation?, Discussion paper National Bureau of Economic Research.

Ahn, Seoungpil, David J Denis, and Diane K Denis, 2006, Leverage and investment in diversified firms, *Journal of financial Economics* 79, 317–337.

Autor, David H, David Dorn, and Gordon H Hanson, 2016, The china shock: Learning from labor-market adjustment to large changes in trade, *Annual review of economics* 8, 205–240.

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.

Blinder, Alan S, 1973, Wage discrimination: reduced form and structural estimates, *Journal of Human resources* pp. 436–455.

Brazell, David W, Lowell Dworin, and Michael Walsh, 1989, *A history of federal tax depreciation policy* . , vol. 64 (Office of Tax Analysis, US Treasury Department).

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

Cohn, Jonathan B, Zack Liu, and Malcolm I Wardlaw, 2022, Count (and count-like) data in finance, *Journal of Financial Economics* 146, 529–551.

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.

Fazzari, Steven, R Glenn Hubbard, and Bruce Petersen, 1988, Investment, financing decisions, and tax policy, *The American economic review* 78, 200–205.

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.

Garrett, Daniel G, Eric Ohrn, and Juan Carlos Suárez Serrato, 2020, Tax policy and local labor market behavior, *American Economic Review: Insights* 2, 83–100.

Hall, Robert E, and Dale W Jorgenson, 1967, Tax policy and investment behavior, *The American Economic Review* 57, 391–414.

Kitagawa, Evelyn M, 1955, Components of a difference between two rates, *Journal of the american statistical association* 50, 1168–1194.

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

Mian, Atif, and Amir Sufi, 2009, The consequences of mortgage credit expansion: Evidence from the us mortgage default crisis, *The Quarterly journal of economics* 124, 1449–1496.

Oaxaca, Ronald, 1973, Male-female wage differentials in urban labor markets, *International economic review* pp. 693–709.

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

Ohrn, Eric, 2019, The effect of tax incentives on us manufacturing: Evidence from state accelerated depreciation policies, *Journal of Public Economics* 180, 104084.

Oster, Emily, 2019, Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics* 37, 187–204.

Reinhart, Carmen M, and Kenneth S Rogoff, 2010, Growth in a time of debt, *American economic review* 100, 573–78.

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).

Wooldridge, Jeffrey M, 1999, Distribution-free estimation of some nonlinear panel data models, *Journal of Econometrics* 90, 77–97.

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