Advanced DID

Introduction

MIXTAPE SESSIONS

Welcome!

Welcome to the Advanced Difference-in-Differences Mixtape Workshop!

- I am excited to learn with you all today.

Who Am I?

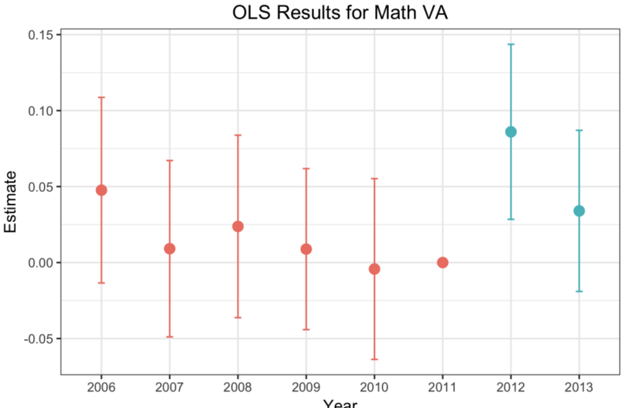
- Assistant Professor of Economics at Brown University.
- I consider myself an *applied econometrician*.
- The main goal of my research is to develop *usable* tools that improve the quality of empirical work.

My DiD Journey

- Early-on in graduate school, I was an aspiring labor economist running a lot of DiDs...

Figure 7: Event Study Results for the Effects of Retirements in 2011 on Math Value-Added

(a) OLS



My DiD Journey

- I realized I had a lot of questions about the methodology of what I was doing.
  - → Should I believe parallel trends holds in this context?
  - → Why do I have pre-trends in some of my specifications but not others?
  - → Is it okay if I focus only on the specifications without pre-trends...?

My DiD Journey

- I realized I had a lot of questions about the methodology of what I was doing.
  - → Should I believe parallel trends holds in this context?
  - → Why do I have pre-trends in some of my specifications but not others?
  - → Is it okay if I focus only on the specifications without pre-trends...?
- Pretty quickly I started writing methodological papers about these topics
  - → I never published the Act 10 paper – oops!
- But the goal of my research has always been to try to inform real-world analyses of economic topics

My DiD Journey

- I realized I had a lot of questions about the methodology of what I was doing.
  - → Should I believe parallel trends holds in this context?
  - → Why do I have pre-trends in some of my specifications but not others?
  - → Is it okay if I focus only on the specifications without pre-trends...?
- Pretty quickly I started writing methodological papers about these topics
  - → I never published the Act 10 paper – oops!
- But the goal of my research has always been to try to inform real-world analyses of economic topics
- Today I hope to share with you some of the insights that I and others have learned over the last few years, with the **goal of helping you improve your research.**
  - → Focus on both theory and applying it in practice!

(Approximate) Schedule for the day

- 10-11 Preliminaries & The Canonical DiD Model
- 11-11:15 Break
- 11:15-12:30 Staggered treatment timing and heterogeneous treatment effects
- 12:30-1 Lunch
- 1-2 Coding Exercise
- 2-3:15 Violations of Parallel Trends
- 3:15-4:15 Coding Exercise
- 4:15-5:00 Open ”Office Hour” for your DiD questions

Course logistics

- I strongly encourage you all to participate and ask questions!
  - → It's more fun for me and helps you learn better!
- There are several ways that you can ask questions:
  - → Raise hand on Zoom
  - → Text question on Discord
- I will pause periodically for you to ask live questions and to review messages on Discord

Introduction

- **Difference-in-differences** (DiD) is one of the most popular strategies for estimating causal effects in non-experimental contexts.
  - → Used in over 20% of NBER WPs (Currie et al., 2020)
- The last few years have seen an explosion of econometrics on DiD, making it hard to keep up (sorry!)
- In Roth, Sant'Anna, Bilinski, and Poe (JOE, 2023), we attempted to synthesize the recent literature and provide concrete recommendations for practitioners
- This course is loosely based on the structure in that paper, focusing on staggered timing (Section 3) and violations of parallel trends (Section 4)

The simplest case

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

The simplest case

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

In the canonical DiD model, we have:

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

Canonical DiD – with math

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

Canonical DiD – with math

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

Key Identifying Assumption - Parallel Trends

- The **parallel trends** assumption states that if the treatment hadn't occurred, average outcomes for the treatment and control groups would have evolved in parallel

$$\underbrace{E[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1]}_{\text{Counterfactual change for treated group}} = \underbrace{E[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0]}_{\text{Change for untreated group}}$$

Key Identifying Assumption - Parallel Trends

- The **parallel trends** assumption states that if the treatment hadn't occurred, average outcomes for the treatment and control groups would have evolved in parallel

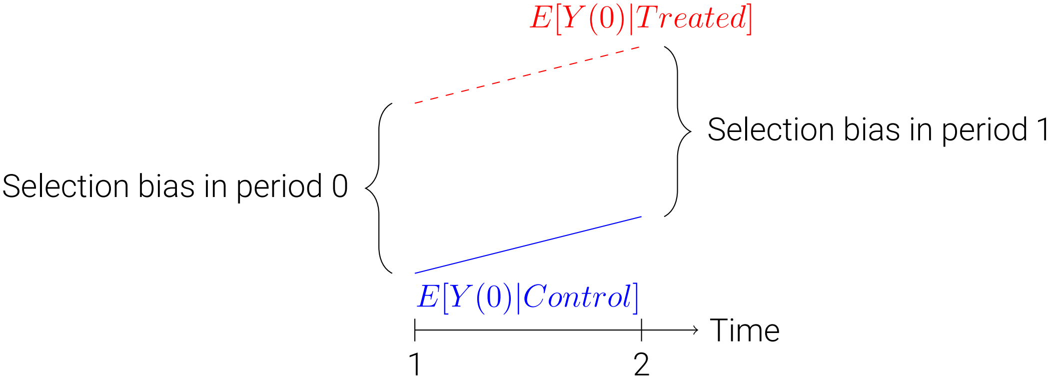
$$\underbrace{E[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1]}_{\text{Counterfactual change for treated group}} = \underbrace{E[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0]}_{\text{Change for untreated group}}$$

- The parallel trends assumption can also be viewed as a **selection bias stability** assumption:

$$\underbrace{E[Y_{i2}(0) \mid D_i = 1] - E[Y_{i2}(0) \mid D_i = 0]}_{\text{Selection bias in period 2}} = \underbrace{E[Y_{i1}(0) \mid D_i = 1] - E[Y_{i1}(0) \mid D_i = 0]}_{\text{Selection bias in period 1}}$$

- PT allows for there to be selection bias! But it must be stable over time

Visualizing PT



Key identifying assumptions

- **Parallel trends:**

$$\mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0\right]. \tag{1}$$

- **No anticipation:** $Y_{i1}(1) = Y_{i1}(0)$
  - → Intuitively, outcome in period 1 isn't affected by treatment status in period 2
  - → Often left implicit in notation, but important for interpreting DiD estimand as a causal effect in period 2

Identification

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

$$\tau_{ATT} = E[Y_{i2}(1) - Y_{i2}(0) | D_i = 1]$$

Identification

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

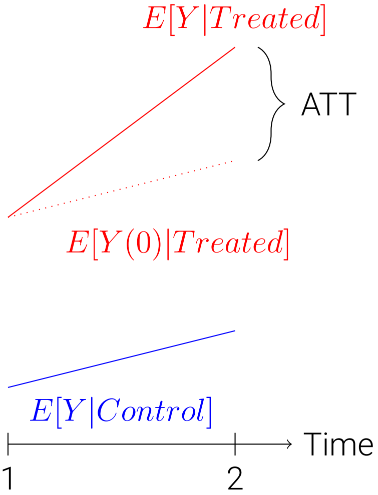
$$\tau_{ATT} = E[Y_{i2}(1) - Y_{i2}(0) | D_i = 1]$$

- Under parallel trends and no anticipation, can show that

$$\tau_{ATT} = \underbrace{(E[Y_{i2}|D_i = 1] - E[Y_{i1}|D_i = 1])}_{\text{Change for treated}} - \underbrace{(E[Y_{i2}|D_i = 0] - E[Y_{i1}|D_i = 0])}_{\text{Change for control}},$$

a “difference-in-differences” of population means

Visualizing Identification



Proof of Identification Argument

- Start with

$$E[Y_{i2} - Y_{i1}|D_i = 1] - E[Y_{i2} - Y_{i1}|D_i = 0]$$

Proof of Identification Argument

- Start with
$$E[Y_{i2} - Y_{i1}|D_i = 1] - E[Y_{i2} - Y_{i1}|D_i = 0]$$
- Apply definition of POs to obtain:

$$E[Y_{i2}(1) - Y_{i1}(1)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0]$$

Proof of Identification Argument

- Start with
$$E[Y_{i2} - Y_{i1} | D_i = 1] - E[Y_{i2} - Y_{i1} | D_i = 0]$$
- Apply definition of POs to obtain:
$$E[Y_{i2}(1) - Y_{i1}(1) | D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0) | D_i = 0]$$
- Use No Anticipation to substitute $Y_{i1}(0)$ for $Y_{i1}(1)$:
$$E[Y_{i2}(1) - Y_{i1}(0) | D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0) | D_i = 0]$$

Proof of Identification Argument

- Start with
$$E[Y_{i2} - Y_{i1} | D_i = 1] - E[Y_{i2} - Y_{i1} | D_i = 0]$$
- Apply definition of POs to obtain:
$$E[Y_{i2}(1) - Y_{i1}(1) | D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0) | D_i = 0]$$
- Use No Anticipation to substitute $Y_{i1}(0)$ for $Y_{i1}(1)$:
$$E[Y_{i2}(1) - Y_{i1}(0) | D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0) | D_i = 0]$$
- Add and subtract $E[Y_{i2}(0) | D_i = 1]$ to obtain:

$$E[Y_{i2}(1) - Y_{i2}(0) | D_i = 1] +$$
$$[(E[Y_{i2}(0) | D_i = 1] - E[Y_{i1}(0) | D_i = 1]) - (E[Y_{i2}(0) | D_i = 0] - E[Y_{i1}(0) | D_i = 0])$$

Proof of Identification Argument

- Start with
$$E[Y_{i2} - Y_{i1}|D_i = 1] - E[Y_{i2} - Y_{i1}|D_i = 0]$$
- Apply definition of POs to obtain:
$$E[Y_{i2}(1) - Y_{i1}(1)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0]$$
- Use No Anticipation to substitute $Y_{i1}(0)$ for $Y_{i1}(1)$:
$$E[Y_{i2}(1) - Y_{i1}(0)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0]$$
- Add and subtract $E[Y_{i2}(0)|D_i = 1]$ to obtain:
$$E[Y_{i2}(1) - Y_{i2}(0)|D_i = 1] +$$
$$[(E[Y_{i2}(0)|D_i = 1] - E[Y_{i1}(0)|D_i = 1]) - (E[Y_{i2}(0)|D_i = 0] - E[Y_{i1}(0)|D_i = 0])]$$
- Cancel the last terms using PT to get $E[Y_{i2}(1) - Y_{i2}(0)|D_i = 1] = \tau_{ATT}$

Estimation and Inference

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

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

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

Estimation and Inference

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

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

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

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

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

  where $D_{it} = D_i * 1[t = 2]$. Also equivalent to β from $\Delta Y_i = \alpha + \Delta D_i \beta + u_{it}$.

Estimation and Inference

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

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

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

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

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

  where $D_{it} = D_i * 1[t = 2]$. Also equivalent to β from $\Delta Y_i = \alpha + \Delta D_i \beta + u_{it}$.

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

Characterizing the recent literature

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

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

Will focus today on the first two

References I

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

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, August 2023, *235* (2), 2218–2244.