# Causal Inference Workshop

## Week 5 - Difference-in-Differences, Triple Differences, TWFE, ...

Causal Inference Workshop

February 16, 2024

# Workshop outline

A. Causal inference fundamentals
  - Modeling assumptions matter too
  - Conceptual framework (potential outcomes framework)
B. Design stage: common identification strategies
  - IV + RDD [coding]
  - Event Studies, DiD, DiDiD, New TWFE Lit [coding]
  - Synthetic Control / Synthetic DiD [coding]
C. Analysis stage: strengthening inferences
  - Limitations of identification strategies, pre-estimation steps
  - Estimation [controls] and post-estimation steps [supporting assumptions]
D. Other topics in causal inference and sustainable development
  - Inference (randomization inference, bootstrapping)
  - Weather data regressions, other common/fun SDev topics [coding]
  - Remote sensing data, other common/fun SDev topics

# Causal inference roadmap

- *Potential outcomes* [framework]
  - Causal effect is the difference between two potential outcomes
  - We can't observe this difference, but can see differences in average observed outcomes
  - If **(conditional) independence assumption** holds, can estimate unbiased ATT

- *Identification* [application/implementation] [last week, and today, ... and next week!]
  - In most empirical settings, IA and CIA do not hold, which is why we need an **identification strategy**
  - Want to eliminate selection bias (identification problem)

- *Estimation* [application/implementation]
  - (Usually) use linear regression model
  - $\hat{\beta}_{OLS}$ unbiased estimator for ATT if $e$ is uncorrelated with treatment (regression problem)

# Outline

Workshop outline

Overview

Event studies

Difference-in-differences (DiD)

Triple differences (DiDiD)

Two-Way Fixed Effects (TWFE)

# Overview of event studies, DiD+, and TWFEs

- General setup:
  - **Objective:** estimate the impact of some treatment at a certain time
  - **Data**: panel data (multiple units, multiple time periods)

$\rightarrow$ to estimate **causal** effect of event, we want the counterfactual of it occurring (this is the unit's outcome had the event not occurred)

# Overview of event studies, DiD+, and TWFEs

- General setup:
  - **Objective:** estimate the impact of some treatment at a certain time
  - **Data**: panel data (multiple units, multiple time periods)

  $\rightarrow$ to estimate **causal** effect of event, we want the counterfactual of it occurring (this is the unit's outcome had the event not occurred)

- Event study, DiD+, and synthetic control methods all use some counterfactual
  - **Event study**: All units treated, assume group's *past* value is CF (*within* variation)

# Overview of event studies, DiD+, and TWFEs

- General setup:
  - **Objective:** estimate the impact of some treatment at a certain time
  - **Data**: panel data (multiple units, multiple time periods)

  $\rightarrow$ to estimate **causal** effect of event, we want the counterfactual of it occurring (this is the unit's outcome had the event not occurred)

- Event study, DiD+, and synthetic control methods all use some counterfactual
  - **Event study**: All units treated, assume group's *past* value is CF (*within* variation)
  - **DiD, DiDiD, TWFE**: Some units never get treated, assume CF *trends* of treated and untreated are parallel, use control units to remove trends in $Y$ in treated (*within* and *between* variation)

# Overview of event studies, DiD+, and TWFEs

- General setup:
  - **Objective:** estimate the impact of some treatment at a certain time
  - **Data**: panel data (multiple units, multiple time periods)

  $\rightarrow$ to estimate **causal** effect of event, we want the counterfactual of it occurring (this is the unit's outcome had the event not occurred)

- Event study, DiD+, and synthetic control methods all use some counterfactual
  - **Event study**: All units treated, assume group's *past* value is CF (*within* variation)
  - **DiD, DiDiD, TWFE**: Some units never get treated, assume CF *trends* of treated and untreated are parallel, use control units to remove trends in $Y$ in treated (*within* and *between* variation)
  - **Synthetic control, synthetic DiD**: Units large aggregates, most (some) units never get treated, combine untreated units to construct optimal synthetic unit that captures CF trajectory of treated unit(s)

# Outline

Workshop outline

Overview

Event studies

Difference-in-differences (DiD)

Triple differences (DiDiD)

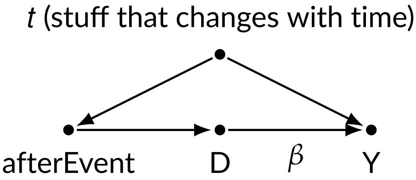
Two-Way Fixed Effects (TWFE)

# Event study, DGP and identifying assumptions

- Want to estimate the causal effect of an *event*, which occurs at time $\tau$ and affects *all units* in the population, on some outcome $Y$
  (treatment assignment is a function of the period)

- Identifying assumptions

| | |
|---|---|
| A1. exogeneity (random timing) | event unpredictable, not result of $Y$ |
| A2. sample composition | does not vary over time |



# Event study, some economics examples

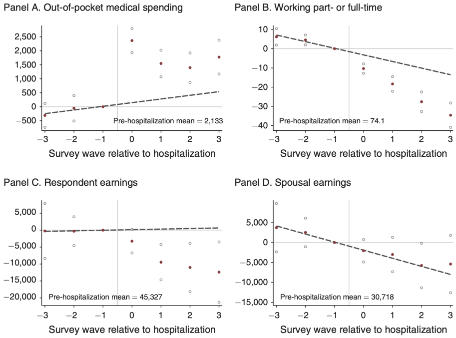
- Many finance studies with very high frequency stock market data
- Economic effects of hospitalization (Dobkin et al. 2018, AER)
  - Look at the effects of hospitalization on various metrics (unpaid bills, earnings, credit)
  - *Identifying assumption*: timing of the admission is uncorrelated with the outcome
    $\rightarrow$ deteriorating health or adverse health effects of job loss would violate this assumption
  - Look at nonparametric event study, then main parametric specification has **linear trend** in event time (based on nonparametric results)

$$y_{it} = \gamma_t + \alpha X_{it} + \delta r + \sum_{r=0}^{3} \mu_r \mathbb{1}\{r\} + \epsilon_{it}$$

$\rightarrow$ timing of the admission is uncorrelated with deviations of outcome from linear trend

# Event study, some economics examples

- Many finance studies with very high frequency stock market data
- Economic effects of hospitalization (Dobkin et al. 2018, AER)



# Event study, estimand and estimator

- Estimand

$$
\begin{aligned}
\beta_{ES} &= \mathbb{E}[Y_{it} | t = \tau] - \mathbb{E}[Y_{it} | t = \tau - 1] \\
&= \mathbb{E}[Y^1_{i,\tau}] - \mathbb{E}[Y_{i,\tau-1}] \\
&= \mathbb{E}[Y^1_{i,\tau}] - \mathbb{E}[Y^0_{i,\tau}] \\
&= \underbrace{\mathbb{E}[Y^1_{i,\tau} - Y^0_{i,\tau}]}_{ATT}
\end{aligned}
$$

- Estimator
  - $\hat{\beta}_{OLS}$ of the regression on a set of time dummies (omitting the period before the event to normalize it as 0) consistently estimates $\beta_{ES}$

$$
Y_{it} = \sum_{t=-K}^{\tau-2} [\beta_t \mathbb{1}\{t\}] + \beta \mathbb{1}\{\tau\} + \sum_{t=\tau+1}^{L} [\beta_t \mathbb{1}\{t\}] + e_{it}
$$

# Event study, best practices, strengths, and weaknesses

- Best practices
  - Plot and report all $\beta_t$'s, to check that they are not changing up to the event (change would suggest presence of pre-trends, which makes it hard to interpret the event in the absence of a sharp trend discontinuity)
  - As in any observational study, adjust for all other relevant pre-treatment variables

- Strengths & weaknesses
  - + Flexible, allows looking at whether effects are dynamic
    - Often empirical papers look at: 1) effect on average, 2) pre-trends, and 3) dynamics
  - May be difficult to rule out other things changing at the same time (unobserved confounders... the rooster concluding the sun rises because of his crowing?)

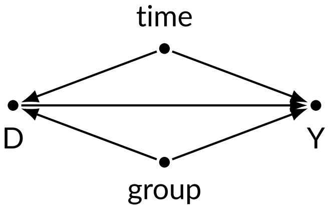
# Outline

# DiD, DGP and identifying assumptions

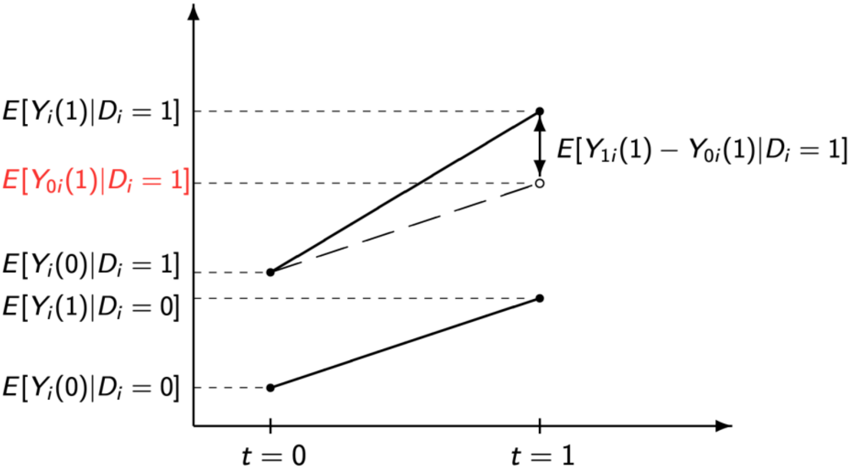
- Treatment assignment/exposure is now a function of *two* dimensions: group (treatment/control) and most commonly time (pre/post) exposure
  - $G_i = \mathbb{1}\{i \in \text{treatment group}\}$ and $P_t = \mathbb{1}\{t \in \text{post period}\} = \mathbb{1}\{t \geqslant \tau\}$
    $\rightarrow$ compare *change over time* in $Y$ in treated to *change over time* in $Y$ in control

- Identifying assumptions

| | |
|---|---|
| A1. "parallel" counterfactual trends | $\mathbb{E}[Y^0_{i,1} - Y_{i,0} \mid G_i = 1] = \mathbb{E}[Y^0_{i,1} - Y_{i,0} \mid G_i = 0]$ |
| A2. group composition | does not vary over time |



# DiD, DGP and identifying assumptions



# DiD, some classic examples

- 1850s cholera death rates (Snow 1855, table from The Effect Book)

| Region Supplier | Death Rates 1849 | Death Rates 1854 |
|---|---|---|
| Non-Lambeth Only (Dirty) | 134.9 | 146.6 |
| Lambeth + Others (Mix Dirty and Clean) | 130.1 | 84.9 |

Death rates are deaths per 10,000 1851 population.

- Fast food industry minimum wages (Card and Krueger 1994, AER)

# DiD, some classic examples

- 1850s cholera death rates (Snow 1855, table from The Effect Book)


- Fast food industry minimum wages (Card and Krueger 1994, AER)

# DiD, some classic examples

- 1850s cholera death rates (Snow 1855, table from The Effect Book)
- Fast food industry minimum wages (Card and Krueger 1994, AER)
  - Examine impact of minimum wage on employment (NJ vs. PA, fast food)

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

| | Stores by state | | | Stores in New Jersey[a] | | | Differences within NJ[b] | |
|---|---|---|---|---|---|---|---|---|
| Variable | PA (i) | NJ (ii) | Difference, NJ – PA (iii) | Wage = \$4.25 (iv) | Wage = \$4.26–\$4.99 (v) | Wage ≥ \$5.00 (vi) | Low–high (vii) | Midrange–high (viii) |
| 1. FTE employment before, all available observations | 23.33 (1.35) | 20.44 (0.51) | –2.89 (1.44) | 19.56 (0.77) | 20.08 (0.84) | 22.25 (1.14) | –2.69 (1.37) | –2.17 (1.41) |
| 2. FTE employment after, all available observations | 21.17 (0.94) | 21.03 (0.52) | –0.14 (1.07) | 20.88 (1.01) | 20.96 (0.76) | 20.21 (1.03) | 0.67 (1.44) | 0.75 (1.27) |
| 3. Change in mean FTE employment | –2.16 (1.25) | 0.59 (0.54) | 2.76 (1.36) | 1.32 (0.95) | 0.87 (0.84) | –2.04 (1.14) | 3.36 (1.48) | 2.91 (1.41) |
| 4. Change in mean FTE employment, balanced sample of stores[c] | –2.28 (1.25) | 0.47 (0.48) | 2.75 (1.34) | 1.21 (0.82) | 0.71 (0.69) | –2.16 (1.01) | 3.36 (1.30) | 2.87 (1.22) |
| 5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d] | –2.28 (1.25) | 0.23 (0.49) | 2.51 (1.35) | 0.90 (0.87) | 0.49 (0.69) | –2.39 (1.02) | 3.29 (1.34) | 2.88 (1.23) |

# DiD, estimand and estimator

- Estimand

$$\beta_{DID} = \Big(\mathbb{E}[Y_{i,1}|G_i = 1] - \mathbb{E}[Y_{i,0}|G_i = 1]\Big) - \Big(\mathbb{E}[Y_{i,1}|G_i = 0] - \mathbb{E}[Y_{i,0}|G_i = 0]\Big)$$
$$= \ldots$$
$$= \underbrace{\mathbb{E}[Y_{i,1}^1 - Y_{i,1}^0|G_i = 1]}_{ATT \text{ (in post period)}}$$

- Estimator

$$Y_{it} = \alpha + \beta_G G_i + \beta_P P_t + \beta_{DID} G_i P_t + e_{it}$$

With two groups and two periods, $\beta_{DID} = \beta_{FE}$ for the below (to be continued...!)

$$Y_{it} = \lambda_G + \delta_P + \beta_{FE} G_i P_T + e_{it}$$

# DiD, best practices, strengths, and weaknesses

- Best practices
  - Support assumption of parallel counterfactual trends by showing that pre-treatment trends coincide (and estimate “dynamic” version of model and check that coefficients for time periods before treatment are zero)
  - Adjust for other relevant pre-treatment confounders

- Strengths & weaknesses
  + Repeated observations get rid of unobserved time-invariant confounders
  + Pre-trends not a problem (unlike event studies) as long as trends of the groups are parallel
  + Repeated cross-sectional ok, as long as sample composition doesn’t vary over time
  - Anything beyond simple binary DiD has potential issues that need to be addressed, as we’ll soon discuss

# Outline

# DiDiD, DGP and identifying assumptions

- Treatment also varies along a third dimension or "subgroup" (e.g., gender or space)
  - $S_i = \mathbb{1}\{i \in \text{treatment in dimension 3}\}$

- Identifying assumptions

| | |
|---|---|
| A1. "parallel" counterfactual trends | $\mathbb{E}[Y^0_{i,1} - Y_{i,0} \mid G_1, S_1] - \mathbb{E}[Y^0_{i,1} - Y_{i,0} \mid G_1, S_0] =$ |
| across subgroups | $\mathbb{E}[Y^0_{i,1} - Y_{i,0} \mid G_0, S_1] - \mathbb{E}[Y^0_{i,1} - Y_{i,0} \mid G_0, S_0]$ |
| A2. subgroup composition | does not vary over time |

# DiDiD, an economics example

- Student loan information and academic choices (Schmeiser et al. 2016, AER:P&P)
  - Intervention at Montana State University: letter sent to students with high loan amounts
  - Only sent to students with high loans at Montana State University (reasonable that parallel trends assumption for high loan vs. no high loan would be violated)
  - Use University of Montana students (could just compare high loan students from two schools, but maybe parallel trends violated, e.g., different rules about switching majors)
  - Compare across 1) university ($G_i$), 2) loan amount ($S_i$), and 3) timing of the letter ($P_t$)

→ students who are sent letter 2ppt more likely to switch majors

# DiDiD, estimand and estimator

- Estimand

$$\begin{aligned}
\beta_{DIDID} &= \Big[ \big(\mathbb{E}[Y_{i,1}|G_1, S_1] - \mathbb{E}[Y_{i,0}|G_1, S_1]\big) - \big(\mathbb{E}[Y_{i,1}|G_0, S_1] - \mathbb{E}[Y_{i,0}|G_0, S_1]\big) \Big] - \\
&= \Big[ \big(\mathbb{E}[Y_{i,1}|G_1, S_1] - \mathbb{E}[Y_{i,0}|G_1, S_1]\big) - \big(\mathbb{E}[Y_{i,1}|G_0, S_1] - \mathbb{E}[Y_{i,0}|G_0, S_1]\big) \Big] \\
&= \dots \\
&= \underbrace{\mathbb{E}[Y^1_{i,1} - Y^0_{i,1}|G_1, S_1]}_{ATT \text{ (in post period)}}
\end{aligned}$$

- Estimator

$$\begin{aligned}
Y_{it} = \alpha &+ \beta_G G_i + \beta_S S_i + \beta_P P_t \\
&\beta_{GS} G_i S_i + \beta_{GP} G_i P_t + \beta_{SP} S_i P_t + \\
&\beta_{DIDID} G_i S_i P_t + e_{it}
\end{aligned}$$

# DiDiD, best practices, strengths, and weaknesses

- Best practices
  - Triple difference makes for a very specific control group; so justify why a double difference isn't satisfactory
  - As in any observational study, adjust for all other relevant pre-treatment variables

- Strengths & weaknesses
  + Triple difference can difference out more confounding elements, so harder to find confounders
  - Requires more data and variation

# Outline

# So, what’s the deal with TWFE?...

# TWFE problem overview

- Canonical 2x2 DiD (2 periods and 2 groups, treated groups experience treatment in period 2) **great** – easy to understand what is estimated and assumptions needed

- However, most actual research these days is **not** using 2x2 DiD, but rather TWFE:

$$y_{it} = \lambda_i + \delta_t + \beta_{TWFE} D_{it} + e_{it}$$

- Problems begin because we do not (or did not) actually understand how this estimator is comparing mean outcomes across groups

- Big picture takeaway: TWFE regressions estimate *weighted sums* of the ATE in each group and period, with weights that sum to one but may make TWFE estimator biased (and *may* even be negative)

# TWFE problem overview

- Big picture takeaway: TWFE regressions estimate *weighted sums* of the ATE in each group and period, with weights that sum to one but may make TWFE estimator biased (and *may* even be negative)

$$y_{it} = \lambda_{g[i]} + \delta_t + \beta_{TWFE} D_{g[i]t} + e_{g[i]t}$$

$\rightarrow \hat{\beta}_{TWFE}$ is a specific weighted sum of the ATE in each treated (g,t) cell (Chaisemartin and D'Haultfœuille 2020)

$$\beta_{ATT} = \mathbb{E}\left[\sum_{(gt):D_{gt}=1} \frac{N_{gt}}{N_1} ATE_{gt}\right], \qquad \mathbb{E}[\hat{\beta}_{TWFE}] = \mathbb{E}\left[\sum_{(gt):D_{gt}=1} \frac{N_{gt}}{N_1} w_{gt} ATE_{gt}\right]$$

$\rightarrow$ weights may vary and can even be **negative!** (Goodman-Bacon 2021)
negative weights huge potential issue, as it may be $\hat{\beta}_{TWFE} < 0$ even if all $ATE_{gt} > 0$ (e.g., $1.5 \times 1 - 0.5 \times 4$)...)

# TWFE problem overview

- When do we need to be worried about these f***ing weights?
- How do we start digesting all the TWFE literature?
- What should we do if we want to use TWFE?
- Should we even use TWFE????
- ???????????????????????
- AHHHHHHH!

# TWFE problem overview

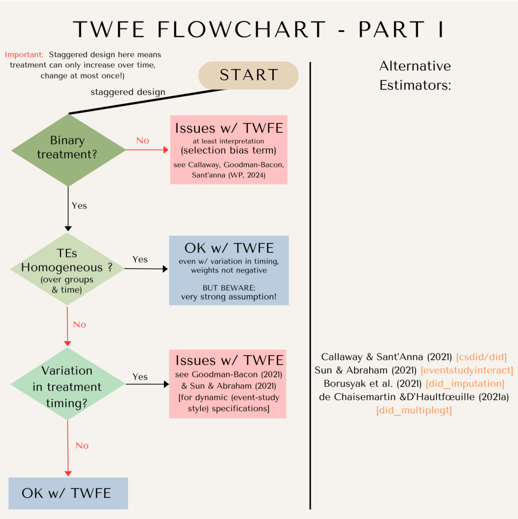
- When do we need to be worried about these f***ing weights?
- How do we start digesting all the TWFE literature?
- What should we do if we want to use TWFE?
- Should we even use TWFE????
- ???????????????????????
- AHHHHHHH!
- $\uparrow$ Pretty much the TWFE literature over the past $x$ years...?

# TWFE and when it's a problem

1. You should worry when **treatment effects are heterogeneous** (not constant across groups or over time) – this is *most* studies!
2. You should also understand stronger assumptions with **continuous** treatments
3. Definitional point: often people refer to "staggered" designs (technically means treatment can only increase, BUT people usually refer to binary treatments that can only increase, but can increase at any time as "staggered")

- Suggested reading order:
  - Most recent de Chaisemartin and D'Haultfoeuille review: *Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey*
  - Chaisemartin and D'Haultfœuille 2020 (AER, overview of problem & alternate estimator)
  - Goodman-Bacon (2021) helpful for understanding intuition behind negative weights and "forbidden comparisons" (using early-treated group as control for late-treated group)
    → then decide what else you need to learn about for your specific set-up!

# TWFE and when it's a problem



# Questions? Comments?

Thank you!

# References

Heavily based on Claire Palandri's 2022 version of the Causal Inference Workshop.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84 (4): 772–793. ISSN: 00028282, accessed February 15, 2024. http://www.jstor.org/stable/2118030.

Chaisemartin, Clément de, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *The American Economic Review* 110 (9): pp. 2964–2996. ISSN: 00028282, 19447981, accessed February 15, 2024. https://www.jstor.org/stable/26966322.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo. 2018. "The Economic Consequences of Hospital Admissions." *American Economic Review* 108 (2): 308–52. https://doi.org/10.1257/aer.20161038. https://www.aeaweb.org/articles?id=10.1257/aer.20161038.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." Themed Issue: Treatment Effect 1, *Journal of Econometrics* 225 (2): 254–277. ISSN: 0304-4076. https://doi.org/https://doi.org/10.1016/j.jeconom.2021.03.014. https://www.sciencedirect.com/science/article/pii/S0304407621001445.

Schmeiser, Maximilian, Christiana Stoddard, and Carly Urban. 2016. "Student Loan Information Provision and Academic Choices." *The American Economic Review* 106 (5): 324–328. ISSN: 00028282, accessed February 15, 2024. http://www.jstor.org/stable/43861037.