

# Causal Inference using Difference-in-Differences

Anzony Quispe<sup>1</sup>

July 18, 2025

---

<sup>1</sup>This material is based on Pedro Sant'Anna and Gemma Dipoppa's class.

# Overview

1. Motivation
2. Basic DiD Setup
3. Assumptions
4. Event Studies
5. TWFE

# Causal Effect

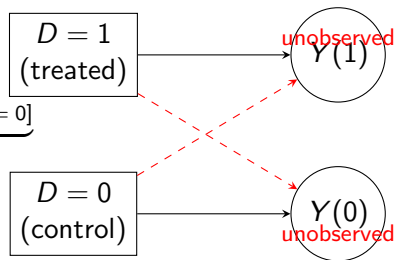


# Selection Bias: A Formal Illustration

## Difference in observed outcomes

$$\underbrace{\mathbb{E}[Y \mid D = 1] - \mathbb{E}[Y \mid D = 0]}_{\text{naïve difference}} = \underbrace{\mathbb{E}[Y(1) - Y(0)]}_{\text{Average Treatment Effect (ATE)}} + \underbrace{\mathbb{E}[Y(0) \mid D = 1] - \mathbb{E}[Y(0) \mid D = 0]}_{\text{Selection Bias}}$$

- $D \in \{0, 1\}$  is the treatment indicator.
- $Y(1)$ ,  $Y(0)$  are potential outcomes *with* and *without* treatment.
- The ATE is not directly observed; we only see  $Y = D Y(1) + (1 - D) Y(0)$ .
- If  $\mathbb{E}[Y(0) \mid D = 1] \neq \mathbb{E}[Y(0) \mid D = 0]$ , the naïve difference conflates the treatment effect with **pre-treatment differences**.



### Key takeaway:

Without random assignment ( $D \perp Y(0), Y(1)$ ), selection into treatment biases simple comparisons.

# DiD Trend

- Better computing resources, easy-to-use software.
- Budget constraint for experiments, unavailability of experimental data.
- With observational data, we have no choice but rely on assumptions to conduct causal inference.

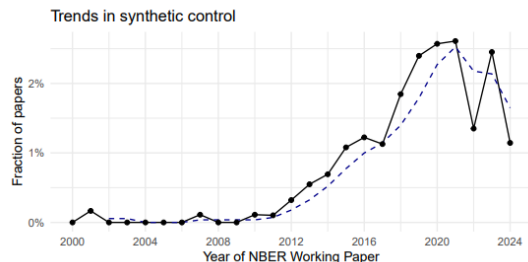
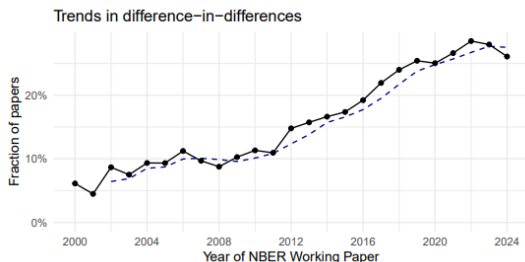


Figure: DiD and Synthetic Control Trend

Assuming **unconfoundedness (either unconditionally or conditional on observed covariates)**, we can estimate the treatment effect using:

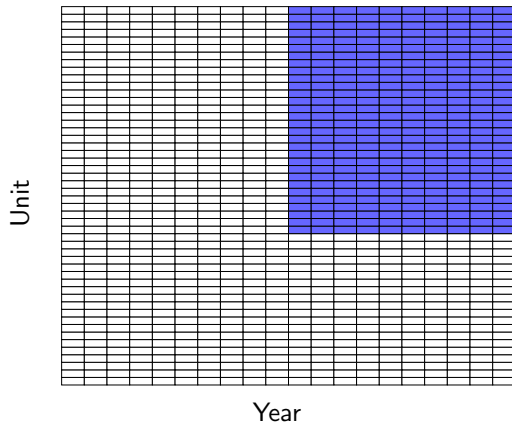
- regression
- matching
- reweighting
- double machine learning

**DiD Advantage:** Allow for selection on unobservables and time-trends.

# Parallel Trends

Absent the treatment and conditional on covariates (features), the outcome of interest would evolve similarly across treated and control groups.

## Panel View: Treatment Adoption (2000–2019)





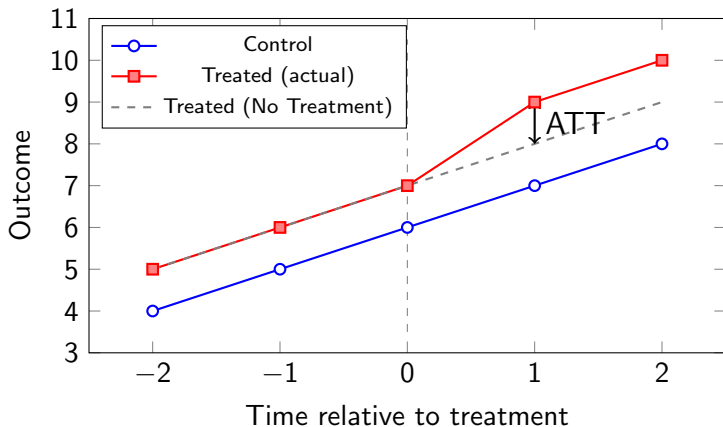
# The canonical $2 \times 2$ DiD estimator

$$\hat{\theta}^{\text{DiD}} = (\bar{Y}_{g=\text{treated}, t=\text{post}} - \bar{Y}_{g=\text{treated}, t=\text{pre}}) \\ - (\bar{Y}_{g=\text{untreated}, t=\text{post}} - \bar{Y}_{g=\text{untreated}, t=\text{pre}}),$$

## DiD Illustration: Table

	<b>Before</b>	<b>After</b>	<b>Diff (Aft-Bef)</b>
Control	1.5	4	2.5
Treat	3.5	12	8.5
Diff (T-C)	2	8	<b>6</b>

# Difference-in-Differences Illustration with Counterfactual



# Stable Unit Treatment Value Assumption (SUTVA)

## Assumption (SUTVA)

*Observed outcomes at time  $t$  are realized as*

$$Y_{i,t} = \sum_{g \in \mathcal{G}} \mathbf{1}\{G_i = g\} Y_{i,t}(g).$$

- Implicitly implies that potential outcomes for unit  $i$  are not affected by the treatment of unit  $j$ .
  - Rules out interference across units
  - Rules out spillover effects
  - Rules out general equilibrium effects

# No-Anticipation Assumption

## Assumption (No-Anticipation)

For all units  $i$ ,

$$Y_{i,t}(g) = Y_{i,t}(\infty) \quad \text{for all } t < g$$

*i.e. in every pre-treatment period.*

- Replace all “untreated” (or “not-yet-treated”) potential outcomes by  $Y_{i,t}(\infty)$ .
- Many times, this assumption is already “baked” into the potential-outcome notation (replace  $Y_{i,t}(\infty)$  with  $Y_{i,t}(0)$  in all pre-treatment periods).
- Potential outcome of control Group is equivalent to treatment group pre treatment.

## Parallel Trends Assumption – 2×2 setup

Since a simple comparison of means at time  $t = 2$  does not recover a parameter of interest (ATT), we can take a different route.

### Assumption (Parallel Trends Assumption)

$$\mathbb{E}[Y_{i,t=2}(\infty) \mid G_i = 2] - \mathbb{E}[Y_{i,t=1}(\infty) \mid G_i = 2] = \mathbb{E}[Y_{i,t=2}(\infty) \mid G_i = \infty] - \mathbb{E}[Y_{i,t=1}(\infty) \mid G_i = \infty]$$

The parallel trends (PT) assumption states that, in the absence of treatment, the evolution of the outcome among the treated units is, on average, the same as the evolution among the untreated units.

# Parallel Trends and the ATT

- We will start from the perspective that the *ATT* at time  $t = 2$  is the target parameter.
- From the definition of the ATT and SUTVA, we have

$$\begin{aligned} \text{ATT} &\equiv \mathbb{E}[Y_{i,t=2}(2) \mid G_i = 2] - \mathbb{E}[Y_{i,t=2}(\infty) \mid G_i = 2] \\ &= \underbrace{\mathbb{E}[Y_{i,t=2} \mid G_i = 2]}_{\text{by SUTVA}} - \mathbb{E}[Y_{i,t=2}(\infty) \mid G_i = 2] \end{aligned}$$

- Green object is estimable from data (under SUTVA).
- Red object still depends on potential outcomes, and we aim to find ways to “impute” it.
- This is where PT comes into play!

# Parallel Trends and the ATT

1. First, recall the PT assumption:

$$\mathbb{E}[Y_{i,t=2}(\infty) | G_i = 2] - \mathbb{E}[Y_{i,t=1}(\infty) | G_i = 2] = \mathbb{E}[Y_{i,t=2}(\infty) | G_i = \infty] - \mathbb{E}[Y_{i,t=1}(\infty) | G_i = \infty].$$

2. By simple manipulation, we can write it as

$$\mathbb{E}[Y_{i,t=2}(\infty) | G_i = 2] = \mathbb{E}[Y_{i,t=1}(\infty) | G_i = 2] + \left( \mathbb{E}[Y_{i,t=2}(\infty) | G_i = \infty] - \mathbb{E}[Y_{i,t=1}(\infty) | G_i = \infty] \right)$$

3. Now, exploiting No-Anticipation and SUTVA:

$$\mathbb{E}[Y_{i,t=2}(\infty) | G_i = 2] = \underbrace{\mathbb{E}[Y_{i,t=1}(2) | G_i = 2]}_{\text{by No-Anticipation}} + \left( \mathbb{E}[Y_{i,t=2}(\infty) | G_i = \infty] - \mathbb{E}[Y_{i,t=1}(\infty) | G_i = \infty] \right)$$

$$\mathbb{E}[Y_{i,t=2}(\infty) | G_i = 2] = \underbrace{\mathbb{E}[Y_{i,t=1} | G_i = 2]}_{\text{by SUTVA}} + \left( \mathbb{E}[Y_{i,t=2} | G_i = \infty] - \mathbb{E}[Y_{i,t=1} | G_i = \infty] \right).$$



# DiD Main Result

- Combining these results, we have that under SUTVA+No-Anticipation+PT assumptions:

$$\begin{aligned} \text{ATT} &= \mathbb{E}[Y_{i,t=2} \mid G_i = 2] - (\mathbb{E}[Y_{i,t=1} \mid G_i = 2] + (\mathbb{E}[Y_{i,t=2} \mid G_i = \infty] - \mathbb{E}[Y_{i,t=1} \mid G_i = \infty])) \\ &= (\mathbb{E}[Y_{i,t=2} \mid G_i = 2] - \mathbb{E}[Y_{i,t=1} \mid G_i = 2]) - (\mathbb{E}[Y_{i,t=2} \mid G_i = \infty] - \mathbb{E}[Y_{i,t=1} \mid G_i = \infty]) \end{aligned}$$

- This is “the birth” of the DiD estimand!

# Event Study Setup: Medicaid Expansion

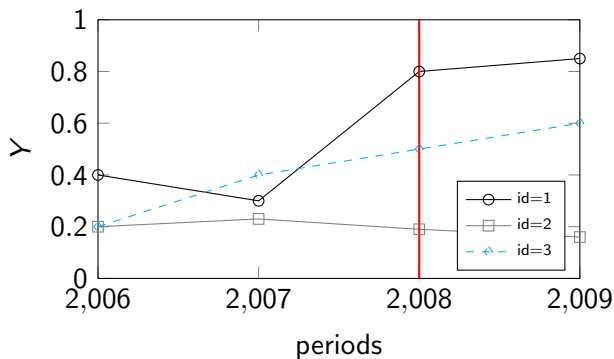
- Compare individuals where treatment took place vs control.
- For individual  $i$  at time  $t$ , and relative time with respect to treatment  $r = (-q, +R)$ :

$$Y_{it} = \gamma_s + \theta_t + \sum_{r=-q}^R \beta_r D_{st} + \epsilon_{it}$$

- We generate one dummy variable for each relative time period before and after treatment, and assign value 1 only when the unit is in that relative time (e.g. +1 in the year after the treatment (lags), -2 two years before the treatment (leads) ).

# EV: Example

id	time	treat	period	Y
1	2006	0	-2	0.4
1	2007	0	-1	0.3
1	2008	1	0	0.8
1	2009	1	1	0.85
2	2006	0	-2	0.2
2	2007	0	-1	0.23
2	2008	0	0	0.19
2	2009	0	1	0.16
3	2006	0	-2	0.2
3	2007	0	-1	0.4
3	2008	1	0	0.5
3	2009	1	1	0.6



- Combining these results, we have that under SUTVA+No-Anticipation+PT assumptions:

$$\begin{aligned} \text{ATT} &= \mathbb{E}[Y_{i,t=2} \mid G_i = 2] - (\mathbb{E}[Y_{i,t=1} \mid G_i = 2] + (\mathbb{E}[Y_{i,t=2} \mid G_i = \infty] - \mathbb{E}[Y_{i,t=1} \mid G_i = \infty])) \\ &= (\mathbb{E}[Y_{i,t=2} \mid G_i = 2] - \mathbb{E}[Y_{i,t=1} \mid G_i = 2]) - (\mathbb{E}[Y_{i,t=2} \mid G_i = \infty] - \mathbb{E}[Y_{i,t=1} \mid G_i = \infty]) \end{aligned}$$

- This is “the birth” of the DiD estimand!

# Two-Way Fixed Effects (TWFE)

- Standard Difference-in-Differences estimator when treatment timing is **uniform**:

$$Y_{it} = \alpha + \beta D_{it} + \delta_i + \gamma_t + \varepsilon_{it}$$

- $\delta_i$  absorbs **time-invariant heterogeneity** across units (states, firms, etc.).
- $\gamma_t$  absorbs shocks **common to all units** in each period.
- $\beta$  captures the *average treatment effect on the treated (ATT) under parallel trends*.

# Assumptions for 2WFE

$$Y_{it} = \tau D_{it} + X' \beta + \alpha_i + \xi_t + \varepsilon_{it}$$

in which  $D_{it}$  is dichotomous

## 1. Functional form

- Additive fixed effect: no interaction of omitted variables.
- *Constant* (heterogeneity) and *contemporaneous* (no anticipation) treatment effect
- Linearity in covariates

## 2. Strict exogeneity

$$\begin{aligned} \varepsilon_{it} &\perp\!\!\!\perp D_{js}, X_{js}, \alpha_j, \xi_s \quad \forall i, j, t, s \\ \Rightarrow \quad \{Y_{it}(0), Y_{it}(1)\} &\perp\!\!\!\perp D_{js} \mid X, \alpha, \xi \quad \forall i, j, t, s \end{aligned}$$

If only two groups, parallel trends implies:

$$\mathbb{E}[Y_{it}(0) - Y_{i't'}(0) \mid X] = \mathbb{E}[Y_{jt}(0) - Y_{j't'}(0) \mid X] \quad i \in \mathcal{T}, j \in \mathcal{C}, \forall t, t'$$

# Fixed Effects & Linear Trends

- **Fixed Effects only (FE):**

$$Y_{it} = \alpha + \beta D_{it} + \delta_i + \gamma_t + \varepsilon_{it}$$

- Controls for unobserved, time-invariant unit factors ( $\delta_i$ ).
- Controls for period shocks common to all units ( $\gamma_t$ ).

- **Adding Unit-Specific Linear Trends:**

$$Y_{it} = \alpha + \beta D_{it} + \delta_i + \gamma_t + \lambda_i t + \varepsilon_{it}$$

- $\lambda_i$  allows each unit to follow its own pre-treatment trajectory.
- Helps when parallel trends hold *after detrending*.
- Costs degrees of freedom and may absorb some treatment variation.

## When to Include Trends?

Use diagnostic plots / pre-trend tests; include trends only if parallel-trends plausibly fails without them.

# Event-Study (Leads/Lags) with Staggered Adoption in TWFE

## Step 1: Define Event-Time Dummies

$$E_{it}^{(k)} = \mathbf{1}\{\text{unit } i \text{ is } k \text{ periods from first treatment at } t\}, \quad k = -K, \dots, -1, 0, 1, \dots, L$$

- $k < 0 \Rightarrow$  *lead* (placebo / pre-trend test)
- $k = 0 \Rightarrow$  *contemporaneous treatment*
- $k > 0 \Rightarrow$  *lag* (dynamic effect)
- Omit one event time (e.g.  $k = -1$ ) as the reference period.

## Step 2: TWFE Event-Study Specification

$$Y_{it} = \alpha + \sum_{k \neq -1} \beta_k E_{it}^{(k)} + \delta_i + \gamma_t + \varepsilon_{it}$$

- $\delta_i$  = unit fixed effects,  $\gamma_t$  = time fixed effects.
- Coefficients  $\beta_k$  trace the dynamic path relative to treatment.



# The End