Difference-in-Differences

Francis J. DiTraglia

University of Oxford

Core Empirical Research Methods

Example: Worker's Compensation and Injury Duration[1]

Background

- **Worker's compensation**: cash and medical care benefits for work-related injuries.
- Run by US States: coverage, amount, and type of benefits varies considerably.
- **Temporary Total Disability** (TTD): unable to work but full recovery expected.
- For TTD there is no fixed duration of benefits.

Research Question

Do more generous TTD benefits increase the duration of claims?

Kentucky (KY) Policy Experiment

Increase in maximum TTD benefit from \$131 to \$217/week on July 15th 1980

[1] Meyer, Viscusi & Durbin (1995)

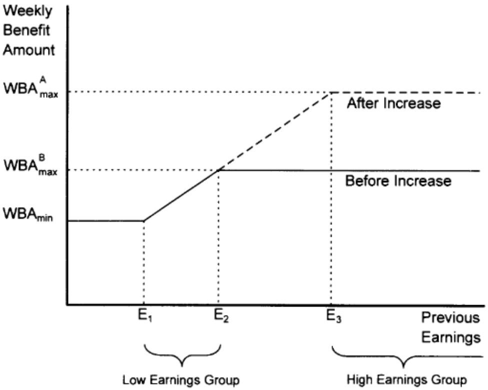
Example: Worker's Compensation and Injury Duration[2]

Treated / Untreated

- Low earners unaffected by change
- Earnings below old and new max
- Policy only affects high earners
- Weekly benefits increase

Data

- Repeated random samples
- Date of injury
- Earnings
- Duration of benefits



[2] Meyer, Viscusi & Durbin (1995)

Adding a Time Dimension

Old Notation

Y_0 and Y_1 are the potential outcomes at a single unspecified moment in time.

New Notation

$Y_t(0)$ and $Y_t(1)$ are the potential outcomes at a specified moment in time t.

Parentheses = **P**otential Outcomes

$Y_t(d)$ is the potential outcome at time t if we set your treatment to d.

New Idea

Use the time dimension to make *before-and-after treatment* comparisons.

Two-period Model

- Y_t is observed in two time periods: $t \in \{\text{Before}, \text{After}\}$.
- **Between** these periods some are treated $(D = 1)$ and the rest are not $(D = 0)$.
- **Before**: before *anyone* has been treated; **After**: after *some people* are treated.
- Each person has a pair of potential outcome **time series**:
  - $\{Y_{\text{Before}}(1), Y_{\text{After}}(1)\}$ if treated between the periods
  - $\{Y_{\text{Before}}(0), Y_{\text{After}}(0)\}$ otherwise
- **Observed Outcomes**: $\{Y_{\text{Before}}, Y_{\text{After}}\}$

$$
\begin{aligned}
Y_{\text{Before}} &= (1 - D)Y_{\text{Before}}(0) + DY_{\text{Before}}(1) \\
Y_{\text{After}} &= (1 - D)Y_{\text{After}}(0) + DY_{\text{After}}(1)
\end{aligned}
$$

- **2nd Period Effect**: $\Delta \equiv Y_{\text{After}}(1) - Y_{\text{After}}(0)$ is our causal effect of interest

Anticipation (aka Ashenfelter Dip)

What is $Y_{\text{Before}}(1)$?

- This is *not* the potential outcome if treated in the first period!
- It is the potential outcome in period "Before" if treated *after* this period.

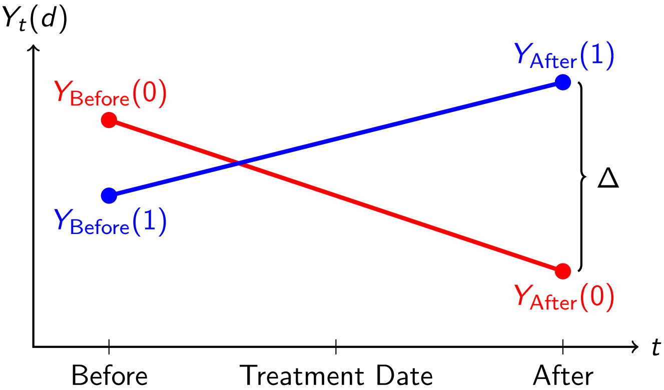
Why this distinction?

- In period "Before" everyone is untreated; why distinguish $Y_{\text{Before}}(0)$ and $Y_{\text{Before}}(1)$?
- How can a potential outcome depend on a **future** treatment?

Ashenfelter (1978)

- **Anticipation**: if I know that I will be treated tomorrow, I may change my behavior in ways that affect my outcomes today.
- "All of the trainee [treatment] groups suffered unpredicted earnings declines in the year prior to training" in a study of the effects of a government training program.

Anticipation (aka "Ashenfelter Dip")



- 2nd Period Treatment Effect: $\Delta \equiv Y_{\text{After}}(1) - Y_{\text{After}}(0)$
- Anticipation: $Y_{\text{Before}}(1) \neq Y_{\text{Before}}(0)$

More on our Target Causal Effect

$$\text{TOT}_{\text{After}} \equiv \mathbb{E}[Y_{\text{After}}(1) - Y_{\text{After}}(0)|D = 1] = \mathbb{E}[\Delta|D = 1]$$

- Natural to study the causal effect *after* treatment: $\Delta \equiv Y_{\text{After}}(1) - Y_{\text{After}}(0)$
- Want the causal effect on the *future* not on the past (anticipation)
- Focus on *the treated*: the people for whom we can make a before-and-after treatment comparison.
- I will write TOT for short, but remember: this is the *second period* effect.

Before-and-after Design

What is this?

- Comparison of observed outcomes for treated: after minus before ("within person")
- E.g. average TTD claim duration among high-income before and after 1980-07-15.
- Stepping stone to Difference-in-differences design.
- Two assumptions, one of which we'll relax later.

Assumption: No Anticipation

$$\mathbb{E}[Y_{\text{Before}}(1) - Y_{\text{Before}}(0) | D = 1] = 0$$

Assumption: No Trend

$$\mathbb{E}[Y_{\text{After}}(0) - Y_{\text{Before}}(0) | D = 1] = 0$$

Theorem

$$\text{TOT} \equiv \mathbb{E}[Y_{\text{After}}(1) - Y_{\text{After}}(0) | D = 1] = \mathbb{E}[Y_{\text{After}} - Y_{\text{Before}} | D = 1].$$

Derivation: Before-and-after Design

Let BA be the average difference of *observed* outcomes for the treated:

$$\text{BA} \equiv \mathbb{E}[Y_{\text{After}} - Y_{\text{Before}} | D = 1] = \mathbb{E}[Y_{\text{After}}(1) | D = 1] - \mathbb{E}[Y_{\text{Before}}(1) | D = 1].$$

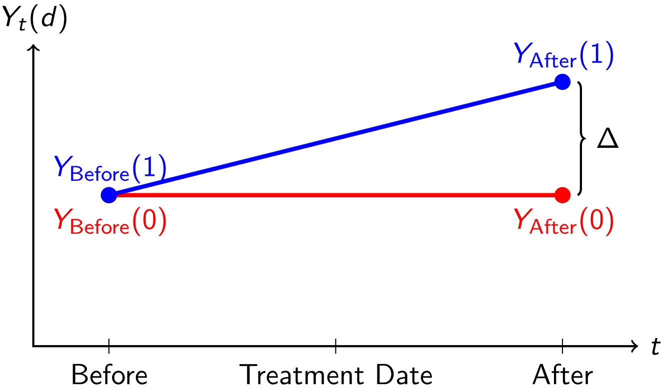
But since

$$\mathbb{E}[Y_{\text{Before}}(1) | D = 1] = \mathbb{E}[Y_{\text{Before}}(0) | D = 1] = \mathbb{E}[Y_{\text{After}}(0) | D = 1]$$

by **No Anticipation** and **No Trend**, we see that

$$\text{BA} = \mathbb{E}[Y_{\text{After}}(1) | D = 1] - \mathbb{E}[Y_{\text{After}}(0) | D = 1] = \text{TOT}.$$

Diagram: Before-and-after Design[3]



- No Anticipation: $Y_{\text{Before}} = Y_{\text{Before}}(1) = Y_{\text{After}}(0)$
- No Trend: $Y_{\text{After}}(0) = Y_{\text{Before}}(0) \implies Y_{\text{Before}} = Y_{\text{After}}(0)$

[3] In the derivation, potential outcomes merely need to be equal *on average for the treated*.

Difference-in-Differences Design

What's wrong with Before-and-after?

- **No Trend** assumption may be implausible: "nothing else changes"

Overview

- Rather than assuming no trend, *estimate* trend from untreated "control" group.
- Combine "between person" and "within person" comparisons.
- **Parallel Trends** assumption replaces **No Trend** assumption
- Retain **No Anticipation** assumption exactly as above.

Assumption: Parallel Trends

$$\mathbb{E}\left[Y_{\text{After}}(0) - Y_{\text{Before}}(0) \middle| D = 1\right] = \mathbb{E}\left[Y_{\text{After}}(0) - Y_{\text{Before}}(0) \middle| D = 0\right]$$

Parallel Trends Assumption

$$\mathbb{E}\left[\, Y_{\text{After}}(0) - Y_{\text{Before}}(0) \middle|\, D = 1\right] = \mathbb{E}\left[\, Y_{\text{After}}(0) - Y_{\text{Before}}(0) \middle|\, D = 0\right]$$

- Trend in untreated potential outcomes *same on average* for treated and untreated.
- Allows *time invariant* unobservables to drive selection into treatment.
- Rules out selection based on time-varying unobservables ("transitory shocks").
- **Fundamentally untestable**: never observe $Y_{\text{After}}(0)$ for the treated

Theorem

No Anticipation & **Parallel Trends** $\implies$

$$\text{TOT} = \mathbb{E}[Y_{\text{After}} - Y_{\text{Before}} | D = 1] - \mathbb{E}[Y_{\text{After}} - Y_{\text{Before}} | D = 0]$$

Derivation: Difference-in-Differences Design

Let DiD be the difference of differences: treated minus control and after minus before

$$\text{DiD} \equiv \mathbb{E}[Y_{\text{After}} - Y_{\text{Before}} | D = 1] - \mathbb{E}[Y_{\text{After}} - Y_{\text{Before}} | D = 0]$$

By the equations linking observe and potential outcomes:

$$\mathbb{E}[Y_{\text{After}} - Y_{\text{Before}} | D = 0] = \mathbb{E}[Y_{\text{After}}(0) - Y_{\text{Before}}(0) | D = 0]$$
$$\mathbb{E}[Y_{\text{After}} - Y_{\text{Before}} | D = 1] = \mathbb{E}[Y_{\text{After}}(1) - Y_{\text{Before}}(1) | D = 1].$$

By **No Anticipation** $\mathbb{E}[Y_{\text{Before}}(1) | D = 1] = \mathbb{E}[Y_{\text{Before}}(0) | D = 1]$ and hence

$$\mathbb{E}[Y_{\text{After}} - Y_{\text{Before}} | D = 1] = \mathbb{E}[Y_{\text{After}}(1) - Y_{\text{Before}}(0) | D = 1].$$

Thus,

$$\text{DiD} = \mathbb{E}[Y_{\text{After}}(1) - Y_{\text{Before}}(0) | D = 1] - \mathbb{E}[Y_{\text{After}}(0) - Y_{\text{Before}}(0) | D = 0]$$

Derivation Continued

Continuing from the previous slide:

$$\text{DiD} = \mathbb{E}[Y_{\text{After}}(1) - Y_{\text{Before}}(0)|D = 1] - \mathbb{E}[Y_{\text{After}}(0) - Y_{\text{Before}}(0)|D = 0].$$

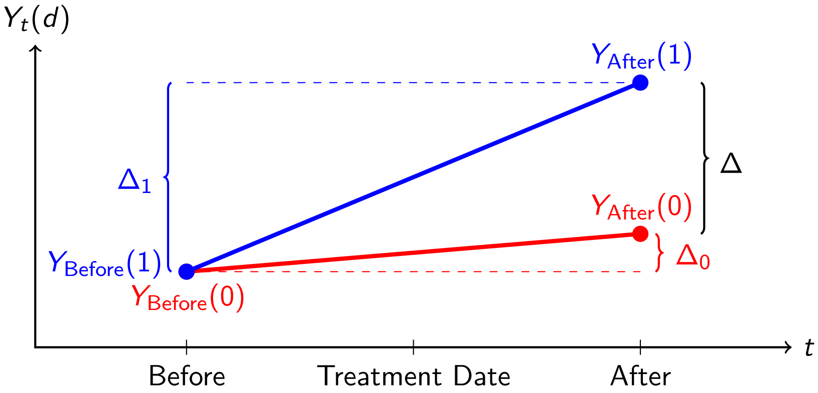
Now, by **Parallel Trends**:

$$\mathbb{E}\,[\,Y_{\text{After}}(0) - Y_{\text{Before}}(0)|\,D = 1] = \mathbb{E}\,[\,Y_{\text{After}}(0) - Y_{\text{Before}}(0)|\,D = 0]$$

Substituting this to replace the second term in the expression for DiD:

$$\begin{aligned}\text{DiD} &= \mathbb{E}[Y_{\text{After}}(1) - Y_{\text{Before}}(0)|D = 1] - \mathbb{E}[Y_{\text{After}}(0) - Y_{\text{Before}}(0)|D = 1] \\ &= \text{TOT} - 0.\end{aligned}$$

Diagram: Difference-in-Differences Design[4]



- No Anticipation: $Y_{\text{Before}}(1) = Y_{\text{Before}}(0) \implies \Delta = \Delta_1 - \Delta_0$
- Parallel Trends: Δ_0 same on average for treated and untreated

[4] For the Theorem

Two-period DiD Implementation

Comparison of Means

- Replace population expectations with sample means and take differences:

$$\widehat{\text{DiD}} = \left(\bar{Y}_{\text{After, Treated}} - \bar{Y}_{\text{Before, Treated}}\right) - \left(\bar{Y}_{\text{After, Untreated}} - \bar{Y}_{\text{Before, Untreated}}\right).$$

DiD Regression

- Equivalently: regress outcome on D_i, $\text{Post}_t = \mathbb{1}(t = \text{After})$ and interaction:

$$Y_{it} = \alpha + \beta D_i + \gamma \text{Post}_t + \delta \left(D_i \times \text{Post}_t\right) + U_{it}, \quad t \in \{\text{Before}, \text{After}\}.$$

- DiD estimand is the coefficient on the interaction: δ
- Makes it easier to compute SEs, add controls, etc.

Panel versus Repeated Cross-Section Data

Panel

Random sample of people observed in *multiple time periods*.

Repeated Cross-Section

Multiple random samples taken at different points in time.

Do we need panel data for DiD?

- Causal effects are *within person*, but use *between person* info to identify them.
- This works because expectation expectation is a linear operator.
- For the same reason, DiD works **just fine** with repeated cross-sections:

$$\widehat{\text{DiD}} = \left(\bar{Y}_{\text{After, Treated}} - \bar{Y}_{\text{Before, Treated}}\right) - \left(\bar{Y}_{\text{After, Untreated}} - \bar{Y}_{\text{Before, Untreated}}\right).$$

- Also works with the regression approach...

Example

```r
library(broom); library(estimatr); library(tidyverse)
library(modelsummary)
library(wooldridge) # injury dataset

KY <- injury |>
  filter(ky == 1) |> # dataset contains Kentucky and Michigan
  rename(treated = highearn, post = afchnge, duration = durat)

naive <- lm_robust(log(duration) ~ treated, KY, subset = (post == 1))
ba <- lm_robust(log(duration) ~ post, KY, subset = (treated == 1))
dd <- lm_robust(log(duration) ~ treated * post, KY)

results <- list(Naive = naive, BA = ba, DiD = dd)
```

```r
modelsummary(results, fmt = 2, gof_omit = 'R2 Adj.|AIC|BIC|RMSE',
             output = 'latex')
```

	Naive	BA	DiD
(Intercept)	1.13	1.38	1.13
	(0.03)	(0.04)	(0.03)
treated	0.45		0.26
	(0.05)		(0.05)
post		0.20	0.01
		(0.05)	(0.04)
treated × post			0.19
			(0.07)
Num.Obs.	2688	2394	5626
R2	0.029	0.006	0.021

Parallel Trends and Transformed Outcomes

- Following Meyer, Viscusi & Durbin (1995), we worked with log(Duration).
- This requires the assumption of parallel trends **in logs**.
- If parallel trends holds in logs, it likely fails in levels and vice-versa.
- See Athey & Imbens (2006), Kahn-Lang & Lang (2020), Roth & Sant'Anna (2023).

What about Anticipation?

- Assume **Parallel Trends** but *do not assume* **No Anticipation**
- **Anticipation Effect for Treated**: $\text{AET} \equiv \mathbb{E}[Y_{\text{Before}}(1) - Y_{\text{Before}}(0)|D = 1]$
- Slight modification of argument from above (lecture notes)

$$\text{DiD} = \text{TOT} - \text{AET}$$

- See Malani & Reif (2016) for a discussion of anticipation versus endogeneity.

Inference in DiD

- This is genuinely difficult; I chose an empirical example to avoid complications.
- Repeated cross-sections: multiple random samples before and after policy change.
- iid sampling in each cross-section so no need for clustering.
- In true "panel" settings, typical to cluster over i whether e.g. US States
- Inference based on the CLT requires **many clusters** but this is not the case in many DiD examples.
- Active area of research. See Roth et al (2023) Section 5.

Staggered Treatment Timing

- Above: two periods, treatment between the periods.
- Generally: $T > 2$; treatment in any period $1 < t < T$; once treated remain so.
- Potential outcomes for every possible **treatment start date**, including *never*
- E.g. $Y_{2016}(2014)$ is a US State's potential outcome in 2016 if it experienced a Medicaid expansion beginning in 2014.
- Homogeneous treatment effects over t and $i \implies$ extend regression from above:

$$Y_{it} = (\text{Fixed Effect})_i + (\text{Fixed Effect})_t + \delta D_{it} + U_{it}$$

- Heterogeneous treatment effects $\implies$ this regression approach can fail badly.
- See Roth et al (2023) Section 3 for a good summary of recent literature.

Relaxing / Evaluating Parallel Trends

Evaluating

- With more than two periods, researchers often **compare pre-trends** in treated versus untreated over periods before treatment occured.
- Additionally / Alternatively: **placebo tests**, i.e. DiD with "fake" treatment date.
- Strictly speaking, neither provides any direct evidence for or against the parallel trends assumption.

Relaxing

- Perhaps parallel trends only holds **conditional on covariates**
- If so, can combine DiD with selection-on-observables approaches.
- E.g. regression adjustment and propensity score weighting.
- Unsurprisingly, requires an overlap assumption.
- See Roth et al (2023) Section 4.2 for an overview.