Difference-in-Differences (D-i-D) Methods

Moshi Alam

Evaluation of policies

- *What is the causal impact of raising minimum wage on employment?*
- What does economic theory say?
  - Under perfect competition?

Evaluation of policies

- *What is the causal impact of raising minimum wage on employment?*

- What does economic theory say?
  - Under perfect competition?

  - However, empirical evidence shows markets are not perfectly competitive

Evaluation of policies

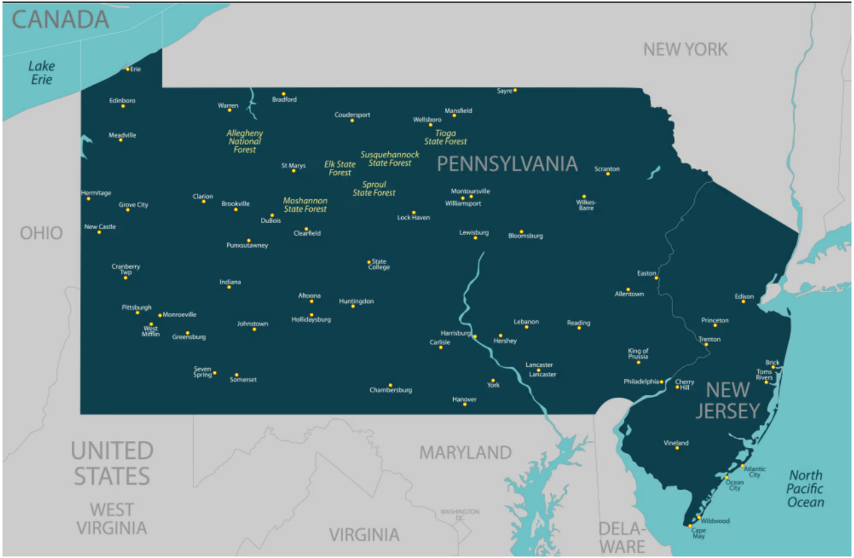
- *What is the causal impact of raising minimum wage on employment?*
- What does economic theory say?
  - Under perfect competition?
  - However, empirical evidence shows markets are not perfectly competitive
  - Then how about under monopsony?

Evaluation of policies

- *What is the causal impact of raising minimum wage on employment?*
- What does economic theory say?
  - Under perfect competition?
  - However, empirical evidence shows markets are not perfectly competitive
  - Then how about under monopsony?
- NJ raised minimum wage in 1992, from 4.25 to 5.05 per hour

Evaluation of policies

- *What is the causal impact of raising minimum wage on employment?*
- What does economic theory say?
  - Under perfect competition?
  - However, empirical evidence shows markets are not perfectly competitive
  - Then how about under monopsony?
- NJ raised minimum wage in 1992, from 4.25 to 5.05 per hour
- Observed change in employment was 0.59% in NJ
- Is this the causal effect of the minimum wage increase? Why or why not?


CANADA
Lake Erie
NEW YORK
PENNSYLVANIA
OHIO
UNITED STATES
WEST VIRGINIA
VIRGINIA
MARYLAND
DELA-WARE
NEW JERSEY
North Pacific Ocean
Erie
Edinboro
Meadville
Warren
Bradford
Coudersport
Wellsboro
Mansfield
Sayre
Allegheny National Forest
Tioga State Forest
Susquehannock State Forest
Elk State Forest
St Marys
Sproul State Forest
Moshannon State Forest
Scranton
Wilkes-Barre
Montoursville
Williamsport
Lock Haven
Hermitage
Grove City
Clarion
Brookville
DuBois
Clearfield
New Castle
Punxsutawney
Lewisburg
Bloomsburg
State College
Cranberry Twp
Indiana
Altoona
Huntingdon
Easton
Allentown
Edison
Pittsburgh
Monroeville
West Mifflin
Greensburg
Johnstown
Hollidaysburg
Lebanon
Reading
Princeton
Harrisburg
Carlisle
Hershey
Trenton
King of Prussia
Lancaster
Brick
Toms Rivers
Seven Spring
Somerset
Chambersburg
York
Hanover
Philadelphia
Cherry Hill
Vineland
Atlantic City
Ocean City
Wildwood
Cape May
WASHINGTON DC

Difference-in-Differences: Introduction

- Key issue in selection bias is that we do not observe the counterfactual
- D-i-D provides an often-plausible method for estimating the counterfactual (untreated) potential outcome of a treated group
- Difference-in-differences is a combination of two research designs:
    - Cross-section (treated versus untreated/control) comparison
    - Across time (pre-treatment versus post-treatment) comparison
- D-i-D is most commonly used to study the effect of some policy change that is applied to a subset of the population.

Difference in Differences: Introduction

- We begin by comparing difference-in-differences with each of these designs.
- Define two groups:
  - $D_i = 1$ denoting the treatment group
  - $D_i = 0$ denoting the control group
- Each group is observed in two periods:
  - $T_t = 1$ denoting the post-treatment period
  - $T_t = 0$ denoting the pre-treatment period

D1: The Cross-sectional Design

- A cross-section design compares outcomes for the treatment and control groups **in the post-treatment period**:

$$\Delta^{CC} = E[Y_{it} \mid D_i = 1, T_t = 1] - E[Y_{it} \mid D_i = 0, T_t = 1]$$

- Δ^{CC} can be interpreted as a causal effect of the treatment in the population if and only if individuals' average unobserved characteristics are equal across the two groups

- This is the standard conditional mean independence assumption you have seen before in the class:

$$E[\varepsilon_{it} \mid D_i = 1, T_t = 1] = E[\varepsilon_{it} \mid D_i = 0, T_t = 1]$$

- Probably will not hold.

D2: The Across Time Comparison Design

- An Across Time Comparison design compares outcomes for the treatment group in the pre- and post- treatment periods:

$$\Delta^{ATC} = E[Y_{it} \mid D_i = 1, T_t = 1] - E[Y_{it} \mid D_i = 1, T_t = 0]$$

- Δ^{ATC} can be interpreted as a causal effect of the treatment if and only if individuals' average unobserved characteristics do not change through time.

- This is a variant on the standard conditional mean independence assumption.

- This condition may fail for a range of reasons
  - Year-specific macro shocks
  - Changes in group composition
  - Changes in the institutional environment facing the group
  - Secular time trends

D-i-D: The Difference in Differences Design

- A difference-in-differences (DiD) design compares the pre-/post-treatment time period change in outcomes for the treatment group to that of the control group:

$$\Delta^{DiD} = \overbrace{\Big(E[Y_{it} \mid D_i = 1, T_t = 1] - E[Y_{it} \mid D_i = 1, T_t = 0]\Big)}^{\text{Time change in treatment group}} - \underbrace{\Big(E[Y_{it} \mid D_i = 0, T_t = 1] - E[Y_{it} \mid D_i = 0, T_t = 0]\Big)}_{\text{Time change in control group}} \tag{1}$$

Difference in Differences: Identifying Assumption

- Δ^{DiD} can be interpreted as a causal effect of the treatment only if:

$$E[\varepsilon_{it} \mid D_i = 1, T_t = 1] - E[\varepsilon_{it} \mid D_i = 1, T_t = 0] = E[\varepsilon_{it} \mid D_i = 0, T_t = 1] - E[\varepsilon_{it} \mid D_i = 0, T_t = 0] \quad (2)$$

- In words: individuals' average unobserved characteristics would have changed through time in the same way for the control and treatment groups in the absence of the treatment

- This is commonly referred to as the **parallel trends assumption** or **parallel trends in the absence of treatment**

Difference in Differences: Identifying Assumption

- The plausibility of "parallel trends in the absence of treatment" identifying assumption is context-specific
- Typically more plausible than the **CC** and **ATC** identifying assumptions above
- DiD **allows** for strategic/non-random/confounded selection into the treatment group. Give examples
- **The key is that these forms of selection must not be time-varying**

Implementing a D-i-D

Steps in implementing

- Data
- Estimation/model(s)
- Choosing the CG
- Violations of identifying assumptions: testing
- Generalization to multiple time periods and/or groups
- Inference: standard errors

Data

Data

- Must have data for units belonging tothe TG and the CG, both in the pre- and post-treatment periods

- DiD can be implemented using either panel data or repeated cross-sections

- Interpretation and methods for DiD are almost identical in both types of data.

Data - impositions on the identifying assumptions

- But the identifying assumption imposes some restrictions on changes through time in the data
    - Attrition for panel datasets
    - Consistency of the sampling frame for repeated cross-sections
- Need to defend the plausibility

Estimation

Estimation

- Assumption (2) is an identification assumption, describing which data generating processes identify ATT in D-i-D models
- In practice, almost all of the difference-in-differences models discussed in this lecture are estimated by ordinary least squares
- But the idea is to formulate the specification in such a way that will allow us to use the identification assumption
- At times we may and will need some additional assumptions, but for now lets deal with the 2x2 case

Canonical 2x2 DiD

- 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 2x2 DiD

- Panel data on Y_{it} for $t = 1, 2$ and $i = 1, \ldots, N$
- **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

Parallel Trends (PT) Assumption

- 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

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

Parallel Trends (PT) Assumption

- 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

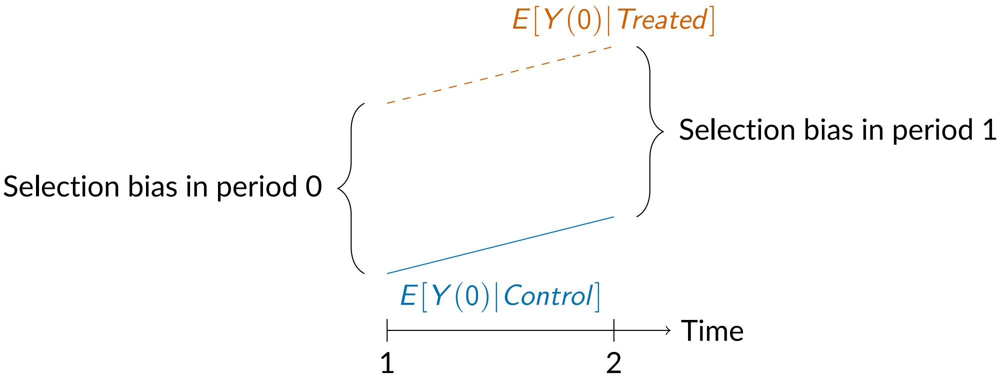
$$\underbrace{E[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1]}_{\text{Counterfactual change for treated group}} = \underbrace{E[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0]}_{\text{Change for untreated group}}$$

- 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



main identifying assumptions

- **Parallel trends:**

$$\mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 1] = \mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid D_i = 0]. \tag{3}$$

- **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

- Average treatment effect on the treated (ATT) in period 2 defined:

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

Identification

- Average treatment effect on the treated (ATT) in period 2 defined:

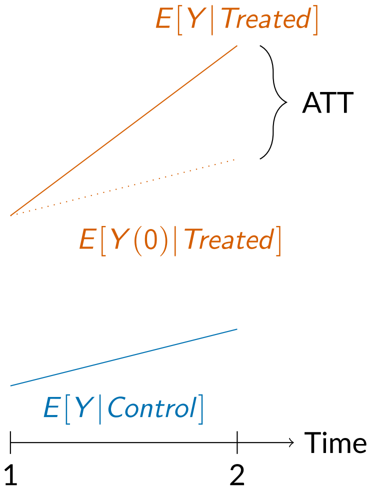
$$\tau_{ATT} = E[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1]$$

- 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

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

Proof of Identification

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

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

Proof of Identification

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

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

  Use No Anticipation

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

Proof of Identification

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

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

Use No Anticipation
$$E[Y_{i2}(1) - Y_{i1}(0) | D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0) | D_i = 0]$$

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

Proof of Identification

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

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

  Use No Anticipation
$$E[Y_{i2}(1) - Y_{i1}(0) | D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0) | D_i = 0]$$

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

- Cancel the last terms using PT to get $E[Y_{i2}(1) - Y_{i2}(0) | D_i = 1] = \tau_{ATT}$

2x2 DiD can be estimated in several ways:

1. Compute avg. outcome ($\overline{Y}$) in each group/period. Then compute,

$$\widehat{\Delta^{DiD}} = \left(\overline{Y}_{D=1,T=1} - \overline{Y}_{D=1,T=0}\right) - \left(\overline{Y}_{D=0,T=1} + \overline{Y}_{D=0,T=0}\right)$$

2. Estimate the regression model ,

$$Y_{it} = \alpha + \beta D_i + \gamma T_t + \Delta^{DiD} D_i T_t + \varepsilon_{it}$$

Equivalence of methods 1 and 2 and pictorial intuition

$$Y_{it} = \alpha + \beta D_i + \gamma T_t + \Delta^{DiD} D_i T_t + \varepsilon_{it}$$

$$\mathbb{E}[Y_{it}|D_i = 1, t = 0] =$$

$$\mathbb{E}[Y_{it}|D_i = 0, t = 0] =$$

$$\mathbb{E}[Y_{it}|D_i = 1, t = 1] =$$

$$\mathbb{E}[Y_{it}|D_i = 0, t = 1] =$$

Card and Kreuger (1994)

The problem

- Impact of increase in minimum wage on labor market unemployment
- What does Economic theory tell us the answer will be?
- Different theoretical assumptions imply different answers - hence it is an empirical question
- But this was not credibly answered empirically before
- There are some caveats to this study, but those are beyond the scope of this class
- Also their results have not been qualitatively changed with better data and better methods

Whats the Ideal Way to solve the problem

- We do a RCT
- We sample a large number of stores/ local labor markets
- Collect data at base-line on wages and employment
- Randomly assign whether there is an increase in minimum wage or not
- Then collect end-line data on employment and wages

In absence of a RCT

- Card and Kreuger found the second best way to answer this question
- An exogenous change in the wage structure by means of a policy change on minimum wages
- Before and after the policy went in place they collected data from the treated state and the comparison control state and implemented a D-i-D

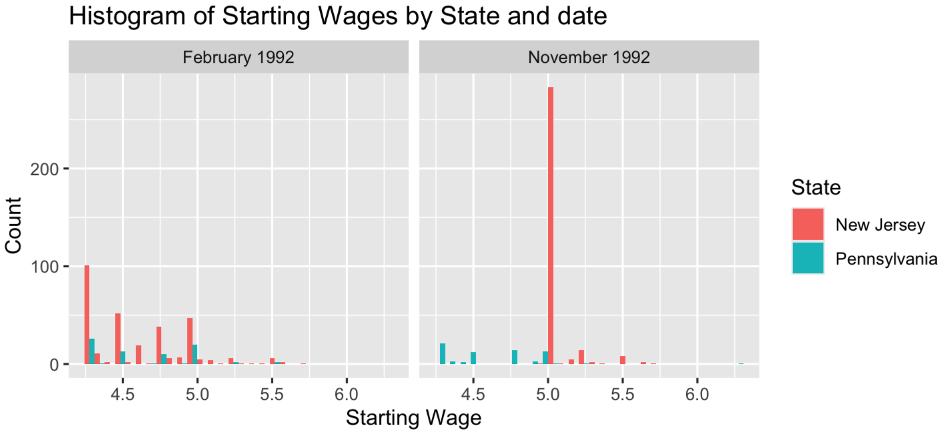
Instituitional Details

- New Jersey announced an increase in minimum wages from \$ 4.25 to \$ 5.05 in Nov 1992
- Neighboring state Pennsylvania did not have any change and its minimum wage stayed at \$ 4.25
- Created a perfect opportunity for a before and after comparison with NJ as the treated state and PA as the control state
- Classic 2 x 2 D-i-D

Data

- Suvey of around 400 fast food restaurants
- In both NJ and PA
- Before policy (Feb 1992) and after policy (Nov 1992)

Lets look at the raw data on wages in both states in R!



Histogram of Starting Wages by State and date
February 1992
November 1992
Count
200
100
0
4.5
5.0
5.5
6.0
Starting Wage
State
New Jersey
Pennsylvania

Averages in the data

	Stores by State		
Dependent Variable	PA	NJ	NJ – PA
FTW before	23.3	20.44	-2.89
	(1.35)	(0.51)	(1.44)
FTE after	21.147	21.03	-0.14
	(0.94)	(0.52)	(1.07)
Change in mean FTE	-2.16	0.59	2.76
	(1.25)	(0.54)	(1.36)

Estimating equation

$$Y_{ist} = \alpha + \beta NJ_s + \gamma Post_t + \delta NJ_s \times Post_t + \varepsilon_{ist}$$

- PA pre: $E[Y \mid NJ_s = 0, Post_t = 0] = \alpha + E[\varepsilon_{ist} \mid NJ_s = 0, Post_t = 0]$
- PA post: $E[Y \mid NJ_s = 0, Post_t = 1] = \alpha + \gamma + E[\varepsilon_{ist} \mid NJ_s = 0, Post_t = 1]$
- NJ pre: $E[Y \mid NJ_s = 1, Post_t = 0] = \alpha + \beta + E[\varepsilon_{ist} \mid NJ_s = 1, Post_t = 0]$
- NJ post: $E[Y \mid NJ_s = 1, Post_t = 1] = \alpha + \beta + \gamma + \delta + E[\varepsilon_{ist} \mid NJ_s = 1, Post_t = 1]$
- D-i-D: $\delta = \Big(E[Y \mid NJ_s = 1, Post_t = 1] - E[Y \mid NJ_s = 1, Post_t = 0]\Big) - \Big(E[Y \mid NJ_s = 0, Post_t = 1] - E[Y \mid NJ_s = 0, Post_t = 0]\Big)$

	Pennsylvania (PA)	New Jersey (NJ)	**Δ (NJ - PA)**
Pre mean	$\alpha + E[\varepsilon_{ist} \mid PA, pre]$	$\alpha + \beta + E[\varepsilon_{ist} \mid NJ, pre]$	β
Post mean	$\alpha + \gamma + E[\varepsilon_{ist} \mid PA, post]$	$\alpha + \beta + \gamma + \delta + E[\varepsilon_{ist} \mid NJ, post]$	$\beta + \delta$
Δ (post - pre)	$\gamma + E[\varepsilon_{ist} \mid PA, post] - E[\varepsilon_{ist} \mid PA, pre]$	$\gamma + \delta + E[\varepsilon_{ist} \mid NJ, post] - E[\varepsilon_{ist} \mid NJ, pre]$	δ

- Assuming parallel trends, the conditional expectations of the errors cancel out
- δ identifies the causal effect of the minimum wage increase on employment
- The equations in the paper are slightly different. Can add x_{ist} to control for other factors not affected by the treatment.

Estimating equation

$$Y_{ist} = \alpha + \beta NJ_s + \gamma Post_t + \delta NJ_s \times Post_t + \varepsilon_{ist}$$

- PA pre: $E[Y \mid NJ_s = 0, Post_t = 0] = \alpha + E[\varepsilon_{ist} \mid NJ_s = 0, Post_t = 0]$
- PA post: $E[Y \mid NJ_s = 0, Post_t = 1] = \alpha + \gamma + E[\varepsilon_{ist} \mid NJ_s = 0, Post_t = 1]$
- NJ pre: $E[Y \mid NJ_s = 1, Post_t = 0] = \alpha + \beta + E[\varepsilon_{ist} \mid NJ_s = 1, Post_t = 0]$
- NJ post: $E[Y \mid NJ_s = 1, Post_t = 1] = \alpha + \beta + \gamma + \delta + E[\varepsilon_{ist} \mid NJ_s = 1, Post_t = 1]$
- D-i-D: $\delta = \Big(E[Y \mid NJ_s = 1, Post_t = 1] - E[Y \mid NJ_s = 1, Post_t = 0]\Big) - \Big(E[Y \mid NJ_s = 0, Post_t = 1] - E[Y \mid NJ_s = 0, Post_t = 0]\Big)$

	Pennsylvania (PA)	New Jersey (NJ)	**Δ (NJ - PA)**
Pre mean	$\alpha + E[\varepsilon_{ist} \mid PA, pre]$	$\alpha + \beta + E[\varepsilon_{ist} \mid NJ, pre]$	β
Post mean	$\alpha + \gamma + E[\varepsilon_{ist} \mid PA, post]$	$\alpha + \beta + \gamma + \delta + E[\varepsilon_{ist} \mid NJ, post]$	$\beta + \delta$
Δ (post - pre)	$\gamma + E[\varepsilon_{ist} \mid PA, post] - E[\varepsilon_{ist} \mid PA, pre]$	$\gamma + \delta + E[\varepsilon_{ist} \mid NJ, post] - E[\varepsilon_{ist} \mid NJ, pre]$	δ

- Assuming parallel trends, the conditional expectations of the errors cancel out
- δ identifies the causal effect of the minimum wage increase on employment
- The equations in the paper are slightly different. Can add x_{ist} to control for other factors not affected by the treatment.

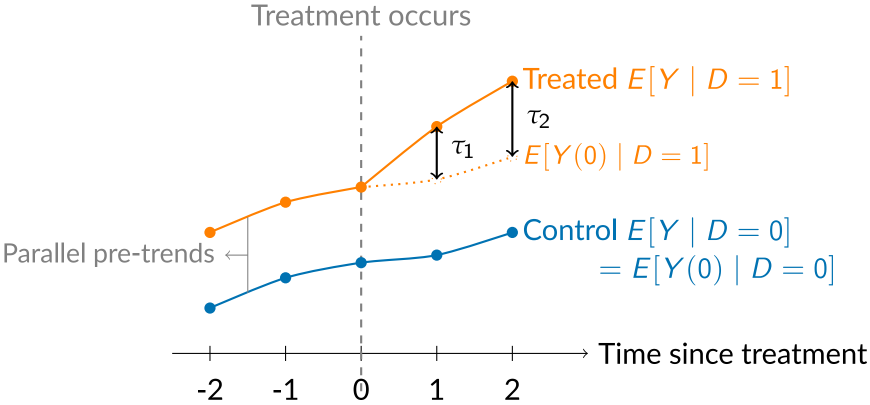
- Underlying assumptions is:
  1. In absence of the minimum wage increase, average employment in fast food restaurants in NJ would have followed the same trend as in PA

$$E[Y_{i2}(0) - Y_{i1}(0) \mid NJ_i = 1] = E[Y_{i2}(0) - Y_{i1}(0) \mid NJ_i = 0]$$

  2. No anticipation - firms did not change employment in Feb 1992 in anticipation of the Nov 1992 increase
- PT can also be defended by exploring -
  - Number of stores
  - Alternative specifications
  - Other sub samples
  - Other explanations of off setting effects - prices
- This will be part of your assignments
- But we can do much more if we have multiple periods

Multiple periods and Event Studies

Visually: Multiple Periods 2 groups assuming PT



PT: $E[Y_{it}(0) \mid D_i = 1] - E[Y_{it}(0) \mid D_i = 0]$ is constant over time t

Estimating equation in the 2xT case

- If you had multiple time periods but still only two groups (treatment and comparison [NJ and PA in Card and Kreuger]), you could estimate:

$$Y_{ist} = \alpha + \beta NJ_s + \gamma Post_t + \delta NJ_s \times Post_t + \varepsilon_{ist}$$

- $Post_t$ is an indicator equal to one for all periods after treatment
- What does δ measure here, based on the previous graph, assuming parallel trends and no anticipation?
- Let us work with an example in R!

We can do much more!

- With more than two time periods, we can estimate more flexible models
- we can estimate dynamic treatment effects (event studies)
- we can test the parallel trends assumption more directly

Placebo Tests: pre-trends

- With longer panels we can perform a “falsification test” or “placebo test” of the DID identifying assumptions
- Consider a case with two pre-treatment time periods $T = -1$ and $T = -2$
- If PT holds between periods -1 and -2, then changes between those periods should be identical in both groups

$$\begin{aligned}\tau_{-2} \quad = \quad & \overbrace{E[Y|D=1, T=-2] - E[Y|D=1, T=-1]}^{\text{Pre-treatment change among treated}} \\ & - \Big(\underbrace{E[Y|D=0, T=-2] - E[Y|D=0, T=-1]}_{\text{Pre-treatment change among controls}} \Big)\end{aligned}$$

Placebo Tests: pre-trends

- If τ_{-2} is non-zero, we reject the “equal time changes” assumption between periods -1 and -2
- This is not a direct test of the “parallel trends” identifying assumption
- It is possible that the treatment and control group time changes between periods 0 and 1 are equal even if they are not equal between periods -1 and -2
- However, this argument is difficult to sustain.
- We can do all this by including leads and lags of the treatment indicator in the simple model

Placebo Tests: graphing pre-trends

- More generally, DD designs with more than two periods typically show a graph of the treatment and control group means through time
- The graph is most convincing when the time series are approximately parallel in all pre-treatment periods and move apart after the treatment

Event study

we go from estimating a single δ to estimating a series of τ_k's:

$$Y_{ist} = \alpha + \beta D_s + \gamma_t + \sum_{k \neq -1} \tau_k \left(D_s \times 1\{t = k\} \right) + \epsilon_{ist} \qquad \text{(Event Study)} \qquad \text{where:}$$

- where $1\{t = k\}$ is an indicator equal to one if time period t is period k relative to treatment
- γ_t is a vector of time fixed effects equivalent to having indicators for each time period (except the reference period -1, to avoid collinearity)
- We omit one period (usually the period just before treatment, $k = -1$) to avoid perfect collinearity with D_s.

Let us expand it!

- What does τ_1, τ_2 measure here?

Event study

we go from estimating a single δ to estimating a series of τ_k's:

$$Y_{ist} = \alpha + \beta D_s + \gamma_t + \sum_{k \neq -1} \tau_k \left(D_s \times 1\{t = k\} \right) + \epsilon_{ist} \qquad \text{(Event Study)} \qquad \text{where:}$$

- where $1\{t = k\}$ is an indicator equal to one if time period t is period k relative to treatment
- γ_t is a vector of time fixed effects equivalent to having indicators for each time period (except the reference period -1, to avoid collinearity)
- We omit one period (usually the period just before treatment, $k = -1$) to avoid perfect collinearity with D_s.

Let us expand it!

- What does τ_1, τ_2 measure here? What does τ_{-2} measure here?

Event study

we go from estimating a single δ to estimating a series of τ_k's:

$$Y_{ist} = \alpha + \beta D_s + \gamma_t + \sum_{k \neq -1} \tau_k \big(D_s \times 1\{t = k\}\big) + \epsilon_{ist} \qquad \text{(Event Study)} \qquad \text{where:}$$

- where $1\{t = k\}$ is an indicator equal to one if time period t is period k relative to treatment
- γ_t is a vector of time fixed effects equivalent to having indicators for each time period (except the reference period -1, to avoid collinearity)
- We omit one period (usually the period just before treatment, $k = -1$) to avoid perfect collinearity with D_s.

Let us expand it!

- What does τ_1, τ_2 measure here? What does τ_{-2} measure here?
- τ_k: ATT parameter in period k relative to pre-treatment period -1
- Let us jump to R again!

Generalizing D-i-D to multiple groups and time periods: TWFE models