Chapter 6: Panel Data and Difference-in-Differences

Jonathan Roth

Mathematical Econometrics I
Brown University

Motivation

- We've seen how to estimate causal effects when a treatment is as good as randomly assigned conditional on observable characteristics
- But often we're worried that there are unobservable characteristics we haven't properly accounted for (i.e. confounding variables)

Motivation

- We've seen how to estimate causal effects when a treatment is as good as randomly assigned conditional on observable characteristics
- But often we're worried that there are unobservable characteristics we haven't properly accounted for (i.e. confounding variables)
- Next we'll think about how we can deal with certain types of unobserved confounding variables when we have **panel data**

What is Panel Data?

- Panel data refers to a situation where we observe observations for each unit i (say a person or state) across multiple periods t

What is Panel Data?

- Panel data refers to a situation where we observe observations for each unit i (say a person or state) across multiple periods t

- Why is this useful? It allows us to look at differences in outcomes between treated/untreated units *before* the treatment occurred

What is Panel Data?

- Panel data refers to a situation where we observe observations for each unit i (say a person or state) across multiple periods t
- Why is this useful? It allows us to look at differences in outcomes between treated/untreated units *before* the treatment occurred
- If treated/control outcomes are different before the treatment, this must be the result of confounding factors.

What is Panel Data?

- Panel data refers to a situation where we observe observations for each unit i (say a person or state) across multiple periods t
- Why is this useful? It allows us to look at differences in outcomes between treated/untreated units *before* the treatment occurred
- If treated/control outcomes are different before the treatment, this must be the result of confounding factors.
- So we can potentially use pre-treatment differences to learn about the confounds and adjust for them

What is Panel Data?

- Panel data refers to a situation where we observe observations for each unit i (say a person or state) across multiple periods t
- Why is this useful? It allows us to look at differences in outcomes between treated/untreated units *before* the treatment occurred
- If treated/control outcomes are different before the treatment, this must be the result of confounding factors.
- So we can potentially use pre-treatment differences to learn about the confounds and adjust for them
- Let's see how this works in an example of **difference-in-differences**, which is the most common panel data method used in applied microeconomic research

Hastings (2004)

- In 2004, Justine Hastings (a former Brown prof!) wrote a study analyzing how mergers in the gas industry affect gas prices

Hastings (2004)

- In 2004, Justine Hastings (a former Brown prof!) wrote a study analyzing how mergers in the gas industry affect gas prices
- In particular, she studied an episode in California where a refinery, ARCO, bought one of the largest gas stations, Thrifty

Hastings (2004)

- In 2004, Justine Hastings (a former Brown prof!) wrote a study analyzing how mergers in the gas industry affect gas prices
- In particular, she studied an episode in California where a refinery, ARCO, bought one of the largest gas stations, Thrifty
- How do you think such a merger might affect prices?

Hastings (2004)

- In 2004, Justine Hastings (a former Brown prof!) wrote a study analyzing how mergers in the gas industry affect gas prices
- In particular, she studied an episode in California where a refinery, ARCO, bought one of the largest gas stations, Thrifty
- How do you think such a merger might affect prices?
  - On the one hand, it could reduce competition and increase prices
  - On the other, a merger could reduce costs of providing gas and decrease prices (synergies)

Hastings (2004)

- In 2004, Justine Hastings (a former Brown prof!) wrote a study analyzing how mergers in the gas industry affect gas prices
- In particular, she studied an episode in California where a refinery, ARCO, bought one of the largest gas stations, Thrifty
- How do you think such a merger might affect prices?
  - On the one hand, it could reduce competition and increase prices
  - On the other, a merger could reduce costs of providing gas and decrease prices (synergies)
- Hastings attempted to answer this question empirically using data on gas prices by neighborhood in CA
  - Data contains info on neighborhoods both with/without Thrifty stations

- Suppose first that we only had data on gas prices from after the merger occurred.
- We could compare prices in areas that had a Thrifty beforehand ($D_i = 1$) and places that didn't have a Thrifty beforehand ($D_i = 0$) to estimate the causal effect of a Thrifty conversion

- Suppose first that we only had data on gas prices from after the merger occurred.
- We could compare prices in areas that had a Thrifty beforehand ($D_i = 1$) and places that didn't have a Thrifty beforehand ($D_i = 0$) to estimate the causal effect of a Thrifty conversion
- Why might this not give us the causal effect of converting Thrifties?

- Suppose first that we only had data on gas prices from after the merger occurred.
- We could compare prices in areas that had a Thrifty beforehand ($D_i = 1$) and places that didn't have a Thrifty beforehand ($D_i = 0$) to estimate the causal effect of a Thrifty conversion
- Why might this not give us the causal effect of converting Thrifties? Omitted variables!

- Suppose first that we only had data on gas prices from after the merger occurred.
- We could compare prices in areas that had a Thrifty beforehand ($D_i = 1$) and places that didn't have a Thrifty beforehand ($D_i = 0$) to estimate the causal effect of a Thrifty conversion
- Why might this not give us the causal effect of converting Thrifties? Omitted variables!
- In particular, places that already had a Thrifty beforehand likely had more competition than places without a Thrifty. We thus might expect them to have lower prices.

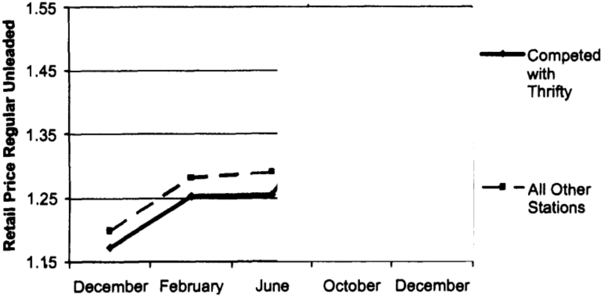
- Suppose first that we only had data on gas prices from after the merger occurred.

- We could compare prices in areas that had a Thrifty beforehand ($D_i = 1$) and places that didn't have a Thrifty beforehand ($D_i = 0$) to estimate the causal effect of a Thrifty conversion

- Why might this not give us the causal effect of converting Thrifties? Omitted variables!

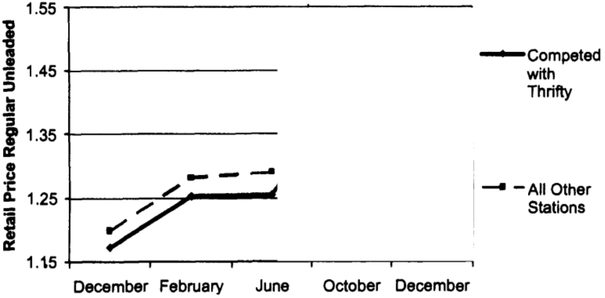
- In particular, places that already had a Thrifty beforehand likely had more competition than places without a Thrifty. We thus might expect them to have lower prices.

- With panel data, we can test this empirically by looking at prices before the merger!



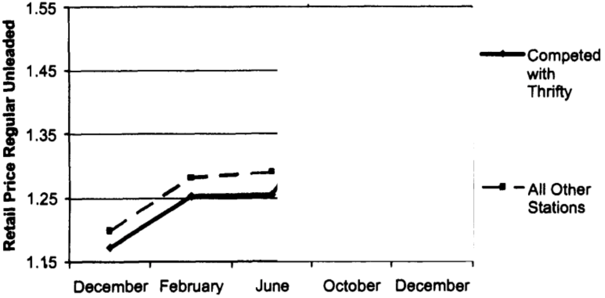
(a) Los Angeles

- Before the merger, stations in markets competing with Thrifty had gas prices about 3 cents lower in every period



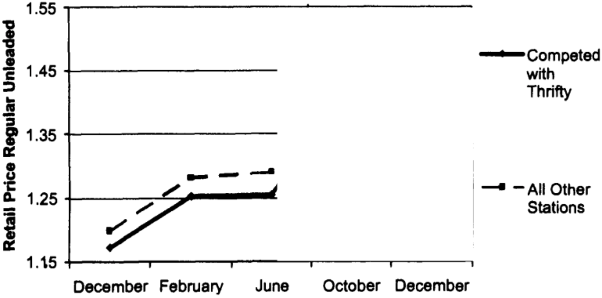
(a) Los Angeles

- Before the merger, stations in markets competing with Thrifty had gas prices about 3 cents lower in every period
- Is it reasonable to assume unconfoundedness after the merger?



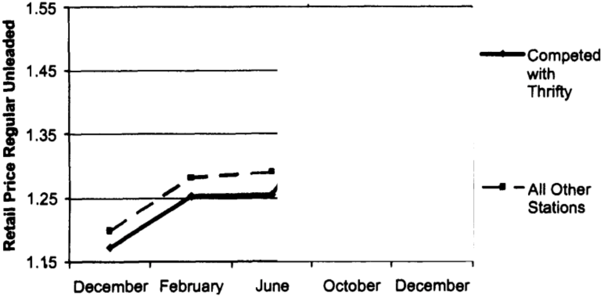
(a) Los Angeles

- Before the merger, stations in markets competing with Thrifty had gas prices about 3 cents lower in every period
- Is it reasonable to assume unconfoundedness after the merger? No!



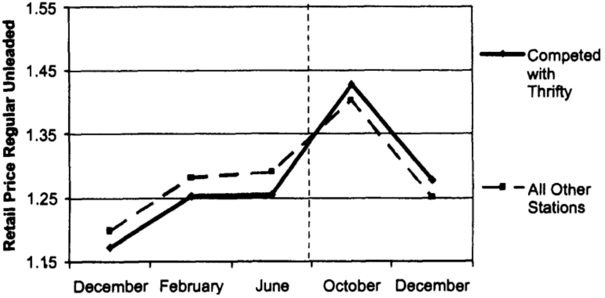
(a) Los Angeles

- Before the merger, stations in markets competing with Thrifty had gas prices about 3 cents lower in every period
- Is it reasonable to assume unconfoundedness after the merger? No!
- A better assumption might be that the gap would haved remained 3c if not for the merger!



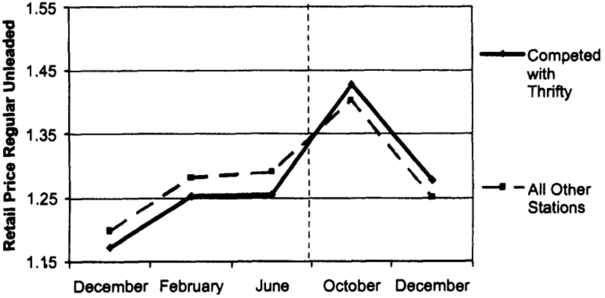
(a) LOS ANGELES

- Before the merger, stations in markets competing with Thrifty had gas prices about 3 cents lower in every period
- Is it reasonable to assume unconfoundedness after the merger? No!
- A better assumption might be that the gap would haved remained 3c if not for the merger! This is the idea of **difference-in-differences**



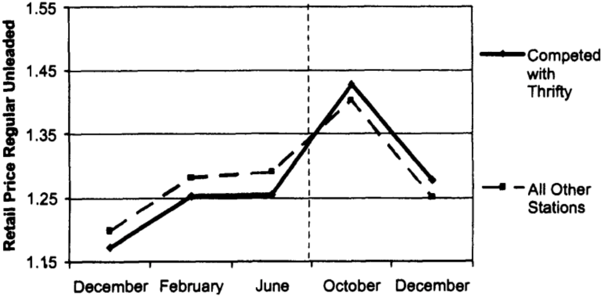
(a) LOS ANGELES

- After the merger, stations in areas with a Thrifty had *higher prices* by about 2c



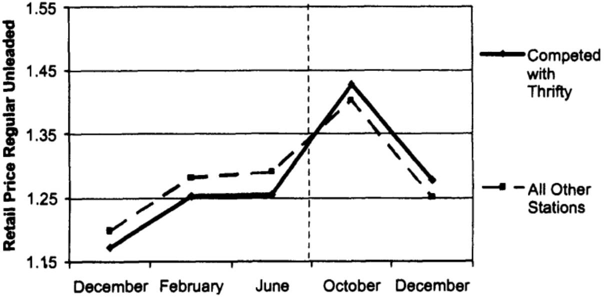
(a) LOS ANGELES

- After the merger, stations in areas with a Thrifty had *higher prices* by about 2c
- If we assume that they would have had *lower* prices by 3c (as before the merger), then this implies a treatment effect of



(a) Los Angeles

- After the merger, stations in areas with a Thrifty had *higher prices* by about 2c
- If we assume that they would have had *lower* prices by 3c (as before the merger), then this implies a treatment effect of $2-(-3)=5$



(a) Los Angeles

- After the merger, stations in areas with a Thrifty had *higher prices* by about 2c
- If we assume that they would have had *lower* prices by 3c (as before the merger), then this implies a treatment effect of $2-(-3)=5$
- This is the post-treatment difference (2) between treatment & control minus the pre-treatment difference (-3), i.e. a *difference-in-differences*

Formalizing the Assumptions of DiD

- Assume there are 2 periods, $t = 1, 2$. Treated units ($D_i = 1$) are treated in period 2; control units never-treated.

- Let Y_{it} be the observed outcome for unit i in period t.
  Assume $Y_{it} = D_i Y_{it}(1) + (1 - D_i) Y_{it}(0)$

Formalizing the Assumptions of DiD

- Assume there are 2 periods, $t = 1, 2$. Treated units ($D_i = 1$) are treated in period 2; control units never-treated.

- Let Y_{it} be the observed outcome for unit i in period t.
  Assume $Y_{it} = D_i Y_{it}(1) + (1 - D_i) Y_{it}(0)$

- **No anticipation assumption**: $Y_{i1}(0) = Y_{i1}(1)$
  - Your treatment in period 2 doesn't affect your outcome in period 1

- **Parallel trends assumption**:

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

Formalizing the Assumptions of DiD

- Assume there are 2 periods, $t = 1,2$. Treated units ($D_i = 1$) are treated in period 2; control units never-treated.

- Let Y_{it} be the observed outcome for unit i in period t.
  Assume $Y_{it} = D_i Y_{it}(1) + (1 - D_i) Y_{it}(0)$

- **No anticipation assumption**: $Y_{i1}(0) = Y_{i1}(1)$
  - Your treatment in period 2 doesn't affect your outcome in period 1

- **Parallel trends assumption**:

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

Equivalently,

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

- Under these assumptions, we have

$$\underbrace{E[Y_{i2} - Y_{i1}|D_i = 1]}_{\text{Observed change for treated}} - \underbrace{E[Y_{i2} - Y_{i1}|D_i = 0]}_{\text{Observed change for control}} =$$

$$= E[Y_{i2}(1) - Y_{i1}(1)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0] \text{ (Observed data rule)}$$

- Under these assumptions, we have

$$\underbrace{E[Y_{i2} - Y_{i1}|D_i = 1]}_{\text{Observed change for treated}} - \underbrace{E[Y_{i2} - Y_{i1}|D_i = 0]}_{\text{Observed change for control}} =$$

$$= E[Y_{i2}(1) - Y_{i1}(1)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0] \text{ (Observed data rule)}$$

$$= E[Y_{i2}(1) - Y_{i1}(0)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0] \text{ (No anticipation)}$$

- Under these assumptions, we have

$$\underbrace{E[Y_{i2} - Y_{i1}|D_i = 1]}_{\text{Observed change for treated}} - \underbrace{E[Y_{i2} - Y_{i1}|D_i = 0]}_{\text{Observed change for control}} =$$

$$= E[Y_{i2}(1) - Y_{i1}(1)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0] \text{ (Observed data rule)}$$

$$= E[Y_{i2}(1) - Y_{i1}(0)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0] \text{ (No anticipation)}$$

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

$$E[Y_{i2}(0) - Y_{i1}(0)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0] \text{(Adding and subtracting)}$$

- Under these assumptions, we have

$$\underbrace{E[Y_{i2} - Y_{i1} | D_i = 1]}_{\text{Observed change for treated}} - \underbrace{E[Y_{i2} - Y_{i1} | D_i = 0]}_{\text{Observed change for control}} =$$

$$= E[Y_{i2}(1) - Y_{i1}(1) | D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0) | D_i = 0] \text{ (Observed data rule)}$$

$$= E[Y_{i2}(1) - Y_{i1}(0) | D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0) | D_i = 0] \text{ (No anticipation)}$$

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

$$E[Y_{i2}(0) - Y_{i1}(0) | D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0) | D_i = 0] \text{(Adding and subtracting)}$$

$$= E[Y_{i2}(1) - Y_{i2}(0) | D_i = 1] \text{ (Parallel trends)}$$

- Under these assumptions, we have

$$\underbrace{E[Y_{i2} - Y_{i1}|D_i = 1]}_{\text{Observed change for treated}} - \underbrace{E[Y_{i2} - Y_{i1}|D_i = 0]}_{\text{Observed change for control}} =$$

$$= E[Y_{i2}(1) - Y_{i1}(1)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0] \text{ (Observed data rule)}$$
$$= E[Y_{i2}(1) - Y_{i1}(0)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0] \text{ (No anticipation)}$$
$$= E[Y_{i2}(1) - Y_{i2}(0)|D_i = 1] +$$
$$E[Y_{i2}(0) - Y_{i1}(0)|D_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0] \text{(Adding and subtracting)}$$
$$= E[Y_{i2}(1) - Y_{i2}(0)|D_i = 1] \text{ (Parallel trends)}$$

- Thus, the difference-in-difference of sample means identifies $\tau_{ATT} = E[Y_{i2}(1) - Y_{i2}(0)|D_i = 1]$.

- This is called the **average treatment effect on the treated** (ATT). It is the average effect in period 2 for treated units.

Estimating the ATT

- We've shown that under the DiD assumptions (parallel trends and no anticipation), the ATT is identified as

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

- How can we estimate this?

Estimating the ATT

- We've shown that under the DiD assumptions (parallel trends and no anticipation), the ATT is identified as

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

- How can we estimate this? Plug in sample means!

Estimating the ATT

- We've shown that under the DiD assumptions (parallel trends and no anticipation), the ATT is identified as

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

- How can we estimate this? Plug in sample means!

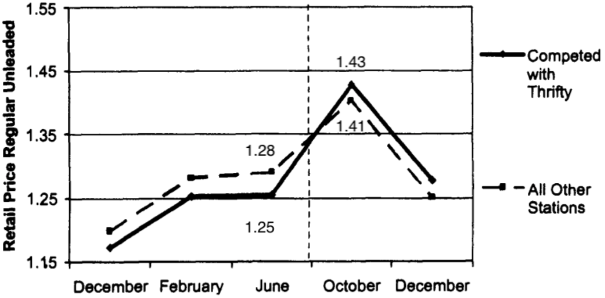
- Our estimate is:

$$\hat{\tau}_{ATT} = \underbrace{\bar{Y}_{12} - \bar{Y}_{11}}_{\text{Change in sample mean for treated}} - \underbrace{\bar{Y}_{02} - \bar{Y}_{01}}_{\text{Change in sample mean for control}},$$

where $\bar{Y}_{dt}$ is the sample mean for units with $D_i = d$ in period t.

Example

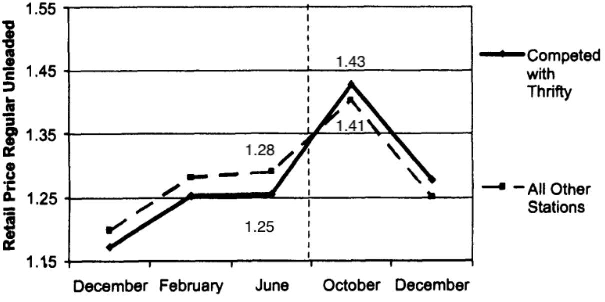
- Consider Hasting's example, comparing June (period 1) to October



(a) Los Angeles

Example

- Consider Hasting's example, comparing June (period 1) to October

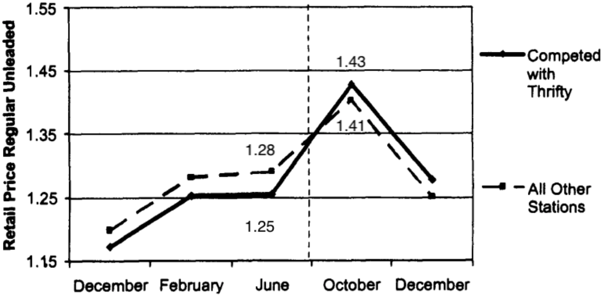


(a) LOS ANGELES

$$\hat{\tau}_{ATT} = \underbrace{\bar{Y}_{12} - \bar{Y}_{11}}_{\text{Change in sample mean for treated}} - \underbrace{\bar{Y}_{02} - \bar{Y}_{01}}_{\text{Change in pop mean for control}} =$$

Example

- Consider Hasting's example, comparing June (period 1) to October



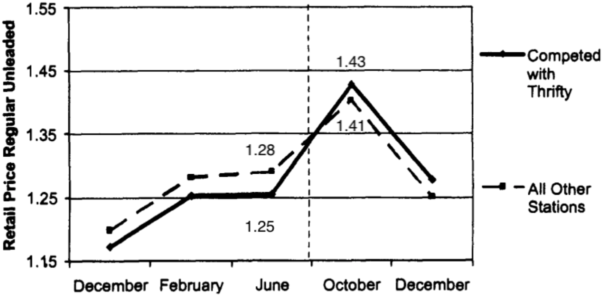
(a) Los Angeles

$$\hat{\tau}_{ATT} = \underbrace{\bar{Y}_{12} - \bar{Y}_{11}}_{\text{Change in sample mean for treated}} - \underbrace{\bar{Y}_{02} - \bar{Y}_{01}}_{\text{Change in pop mean for control}} =$$

$$(1.43 - 1.25) - (1.41 - 1.28) =$$

Example

- Consider Hasting's example, comparing June (period 1) to October



(a) Los Angeles

$$\hat{\tau}_{ATT} = \underbrace{\bar{Y}_{12} - \bar{Y}_{11}}_{\text{Change in sample mean for treated}} - \underbrace{\bar{Y}_{02} - \bar{Y}_{01}}_{\text{Change in pop mean for control}} =$$

$$(1.43 - 1.25) - (1.41 - 1.28) = 0.05$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

where $Post_t = 1[t = 2]$.

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

  where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

$$E[Y_{it}|D_i = 0, Post_t = 0] =$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

  where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

$$E[Y_{it} | D_i = 0, Post_t = 0] = \beta_0$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

  where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

$$E[Y_{it}|D_i = 0, Post_t = 0] = \beta_0$$
$$E[Y_{it}|D_i = 0, Post_t = 1] =$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

$$E[Y_{it}|D_i = 0, Post_t = 0] = \beta_0$$
$$E[Y_{it}|D_i = 0, Post_t = 1] = \beta_0 + \beta_1$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

  where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

$$E[Y_{it}|D_i = 0, Post_t = 0] = \beta_0$$
$$E[Y_{it}|D_i = 0, Post_t = 1] = \beta_0 + \beta_1$$
$$E[Y_{it}|D_i = 1, Post_t = 0] =$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

  where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

$$E[Y_{it}|D_i = 0, Post_t = 0] = \beta_0$$
$$E[Y_{it}|D_i = 0, Post_t = 1] = \beta_0 + \beta_1$$
$$E[Y_{it}|D_i = 1, Post_t = 0] = \beta_0 + \beta_2$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

  where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

$$E[Y_{it}|D_i = 0, Post_t = 0] = \beta_0$$
$$E[Y_{it}|D_i = 0, Post_t = 1] = \beta_0 + \beta_1$$
$$E[Y_{it}|D_i = 1, Post_t = 0] = \beta_0 + \beta_2$$
$$E[Y_{it}|D_i = 1, Post_t = 1] =$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

  where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

$$E[Y_{it}|D_i = 0, Post_t = 0] = \beta_0$$
$$E[Y_{it}|D_i = 0, Post_t = 1] = \beta_0 + \beta_1$$
$$E[Y_{it}|D_i = 1, Post_t = 0] = \beta_0 + \beta_2$$
$$E[Y_{it}|D_i = 1, Post_t = 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

where $Post_t = 1[t = 2]$.

- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

$$\begin{aligned} E[Y_{it}|D_i = 0, Post_t = 0] &= \beta_0 \\ E[Y_{it}|D_i = 0, Post_t = 1] &= \beta_0 + \beta_1 \\ E[Y_{it}|D_i = 1, Post_t = 0] &= \beta_0 + \beta_2 \\ E[Y_{it}|D_i = 1, Post_t = 1] &= \beta_0 + \beta_1 + \beta_2 + \beta_3 \end{aligned}$$

- Thus,

$$\begin{aligned} \beta_3 =& (E[Y_{it}|D_i = 1, Post_t = 1] - E[Y_{it}|D_i = 1, Post_t = 0]) - \\ & (E[Y_{it}|D_i = 0, Post_t = 1] - E[Y_{it}|D_i = 0, Post_t = 0]) = \tau_{ATT} \end{aligned}$$

DiD as Regression

- Consider the regression

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

  where $Post_t = 1[t = 2]$.
- Claim: the population regression coefficient β_3 is equal to τ_{ATT} under the DiD assumptions.
- Why? The regression above models the CEF as:

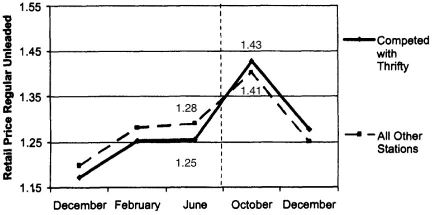
$$\begin{aligned} E[Y_{it}|D_i = 0, Post_t = 0] &= \beta_0 \\ E[Y_{it}|D_i = 0, Post_t = 1] &= \beta_0 + \beta_1 \\ E[Y_{it}|D_i = 1, Post_t = 0] &= \beta_0 + \beta_2 \\ E[Y_{it}|D_i = 1, Post_t = 1] &= \beta_0 + \beta_1 + \beta_2 + \beta_3 \end{aligned}$$

- Thus,

$$\begin{aligned} \beta_3 =& (E[Y_{it}|D_i = 1, Post_t = 1] - E[Y_{it}|D_i = 1, Post_t = 0]) - \\ & (E[Y_{it}|D_i = 0, Post_t = 1] - E[Y_{it}|D_i = 0, Post_t = 0]) = \tau_{ATT} \end{aligned}$$

- Analogously, $\hat{\beta}_3 = (\bar{Y}_{12} - \bar{Y}_{11}) - (\bar{Y}_{02} - \bar{Y}_{01}) = \hat{\tau}_{ATT}$

Example



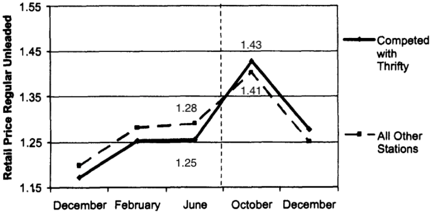
(a) LOS ANGELES

- Suppose we take the Hastings data from June/October and estimate

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

via OLS, where $Post_t$ is 1 for October and 0 for June.

Example



(a) Los Angeles

- Suppose we take the Hastings data from June/October and estimate

$$Y_{it} = \beta_0 + \beta_1 \times Post_t + \beta_2 D_i + \beta_3 D_i \times Post_t + \varepsilon_{it},$$

via OLS, where $Post_t$ is 1 for October and 0 for June.

- We get the regression coefficients:

Constant ($\hat{\beta}_0$)	1.28
Post ($\hat{\beta}_1$)	0.13
Treated ($\hat{\beta}_2$)	-0.03
Treated $\times$ Post ($\hat{\beta}_3$)	0.05

DiD with Multiple Periods

- Often we have more that 2 periods for a DiD analysis

DiD with Multiple Periods

- Often we have more that 2 periods for a DiD analysis
- This is useful for two reasons:

DiD with Multiple Periods

- Often we have more that 2 periods for a DiD analysis
- This is useful for two reasons:
  1. We can test whether parallel trends appears to hold *prior* to treatment

DiD with Multiple Periods

- Often we have more that 2 periods for a DiD analysis
- This is useful for two reasons:
  1. We can test whether parallel trends appears to hold *prior* to treatment

DiD with Multiple Periods

- Often we have more that 2 periods for a DiD analysis
- This is useful for two reasons:
  1. We can test whether parallel trends appears to hold *prior* to treatment
  2. We can analyze how the ATT changes over time

DiD with Multiple Periods

- Often we have more that 2 periods for a DiD analysis
- This is useful for two reasons:
  1. We can test whether parallel trends appears to hold *prior* to treatment
  2. We can analyze how the ATT changes over time
- How do we do this?

DiD with Multiple periods

- Suppose that we have periods $t = -\underline{T}, ..., \bar{T}$. Treated units begin getting treatment at period 1.

DiD with Multiple periods

- Suppose that we have periods $t = -\underline{T}, ..., \bar{T}$. Treated units begin getting treatment at period 1.

- For each period $s \neq 0$, we can estimate a 2-period DiD between period s and period 0:

$$\hat{\beta}_s = \underbrace{(\bar{Y}_{1s} - \bar{Y}_{0s})}_{\text{Diff in period } s} - \underbrace{(\bar{Y}_{10} - \bar{Y}_{00})}_{\text{Diff in period } 0}$$

where $\bar{Y}_{dt}$ is the average for treatment group d in period t.

DiD with Multiple periods

- Suppose that we have periods $t = -\underline{T}, ..., \bar{T}$. Treated units begin getting treatment at period 1.
- For each period $s \neq 0$, we can estimate a 2-period DiD between period s and period 0:

$$\hat{\beta}_s = \underbrace{(\bar{Y}_{1s} - \bar{Y}_{0s})}_{\text{Diff in period } s} - \underbrace{(\bar{Y}_{10} - \bar{Y}_{00})}_{\text{Diff in period 0}}$$

where $\bar{Y}_{dt}$ is the average for treatment group d in period t.

- Conveniently, the $\hat{\beta}_s$ are equal to the OLS estimates of the regression

$$Y_{it} = \phi_t + D_i\gamma + \sum_{s\neq 0} D_i \times 1[t=s] \times \beta_s + \varepsilon_{it}$$

DiD with Multiple periods

- Suppose that we have periods $t = -\underline{T}, ..., \bar{T}$. Treated units begin getting treatment at period 1.

- For each period $s \neq 0$, we can estimate a 2-period DiD between period s and period 0:

$$\hat{\beta}_s = \underbrace{(\bar{Y}_{1s} - \bar{Y}_{0s})}_{\text{Diff in period } s} - \underbrace{(\bar{Y}_{10} - \bar{Y}_{00})}_{\text{Diff in period } 0}$$

  where $\bar{Y}_{dt}$ is the average for treatment group d in period t.

- Conveniently, the $\hat{\beta}_s$ are equal to the OLS estimates of the regression

$$Y_{it} = \phi_t + D_i\gamma + \sum_{s \neq 0} D_i \times 1[t = s] \times \beta_s + \varepsilon_{it}$$

- You can also replace $D_i\gamma$ with a unit fixed effect λ_i and you get the exact same $\hat{\beta}_s$.

Example - Medicaid Expansion

- The Affordable Care Act (ACA, aka Obamacare) expanded Medicaid coverage to people with income up to 138% of the federal poverty line
- Medicaid expansion went into effect in 2014. However, some Republican-leaning states opted out of expanded coverage.
- By 2015, 24 states had expanded Medicaid (more have done so since)

Example - Medicaid Expansion

- The Affordable Care Act (ACA, aka Obamacare) expanded Medicaid coverage to people with income up to 138% of the federal poverty line
- Medicaid expansion went into effect in 2014. However, some Republican-leaning states opted out of expanded coverage.
- By 2015, 24 states had expanded Medicaid (more have done so since)
- Carey, Miller, and Wherry (2020) study the impacts of Medicaid expansion using a DiD design comparing early-adopting states to non-adopters.

Example - Medicaid Expansion

- A slightly simplified version of their regression specification is

$$Y_{its} = \phi_t + \lambda_s + \sum_{r \neq -1} D_i \times 1[t = 2014 + r] \times \beta_r + \varepsilon_{it}$$

where Y_{its} is outcome for person i in year t in state s, and $D_i = 1$ if in an expansion state.

Example - Medicaid Expansion

- A slightly simplified version of their regression specification is

$$Y_{its} = \phi_t + \lambda_s + \sum_{r \neq -1} D_i \times 1[t = 2014 + r] \times \beta_r + \varepsilon_{it}$$

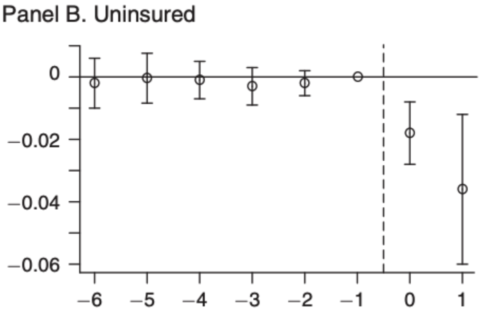
where Y_{its} is outcome for person i in year t in state s, and $D_i = 1$ if in an expansion state. Lets plot the β_s estimates and 95% CIs:

Example - Medicaid Expansion

- A slightly simplified version of their regression specification is

$$Y_{its} = \phi_t + \lambda_s + \sum_{r \neq -1} D_i \times 1[t = 2014 + r] \times \beta_r + \varepsilon_{it}$$

where Y_{its} is outcome for person i in year t in state s, and $D_i = 1$ if in an expansion state. Lets plot the β_s estimates and 95% CIs:

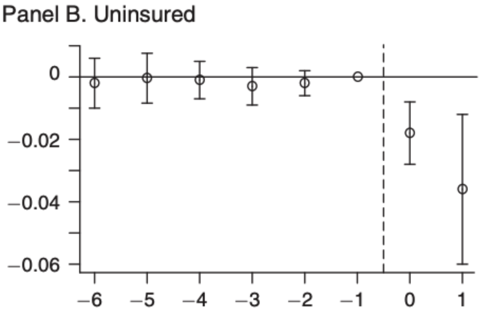


Example - Medicaid Expansion

- A slightly simplified version of their regression specification is

$$Y_{its} = \phi_t + \lambda_s + \sum_{r \neq -1} D_i \times 1[t = 2014 + r] \times \beta_r + \varepsilon_{it}$$

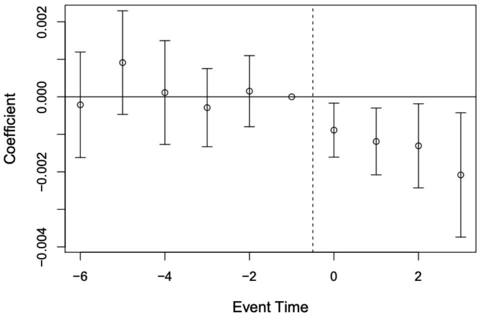
where Y_{its} is outcome for person i in year t in state s, and $D_i = 1$ if in an expansion state. Lets plot the β_s estimates and 95% CIs:



- Results show similar "pre-trends" but negative effects after treatment

In a related paper, some of the same authors used a similar research design to estimate the impacts on mortality

Figure 2: Effect of the ACA Medicaid Expansions on Annual Mortality



Some Caution about Parallel Trends

- DiD relies on the parallel trends assumption, which allows for selection bias but requires it to be stable over time.

Some Caution about Parallel Trends

- DiD relies on the parallel trends assumption, which allows for selection bias but requires it to be stable over time. This rules out time-varying confounding factors.

Some Caution about Parallel Trends

- DiD relies on the parallel trends assumption, which allows for selection bias but requires it to be stable over time. This rules out time-varying confounding factors.

- Often we will be worried about time-varying confounds — e.g., macro-economic factors might differentially affect Democratic versus Republican states

Some Caution about Parallel Trends

- DiD relies on the parallel trends assumption, which allows for selection bias but requires it to be stable over time. This rules out time-varying confounding factors.

- Often we will be worried about time-varying confounds — e.g., macro-economic factors might differentially affect Democratic versus Republican states

- Testing for pre-treatment differences ("pre-trends") can help increase our confidence in the research design.

Some Caution about Parallel Trends

- DiD relies on the parallel trends assumption, which allows for selection bias but requires it to be stable over time. This rules out time-varying confounding factors.
- Often we will be worried about time-varying confounds — e.g., macro-economic factors might differentially affect Democratic versus Republican states
- Testing for pre-treatment differences ("pre-trends") can help increase our confidence in the research design. But they're not perfect. Why?
    1. Just because trends were parallel beforehand doesn't mean that they would continue to be afterwards

Some Caution about Parallel Trends

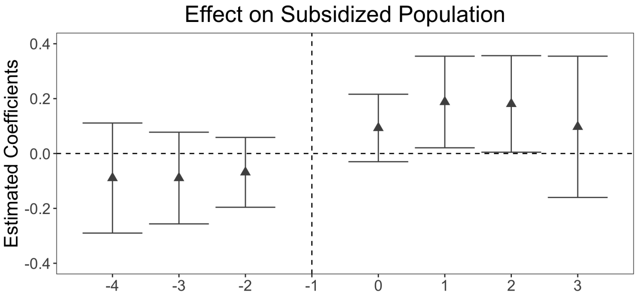
- DiD relies on the parallel trends assumption, which allows for selection bias but requires it to be stable over time. This rules out time-varying confounding factors.
- Often we will be worried about time-varying confounds — e.g., macro-economic factors might differentially affect Democratic versus Republican states
- Testing for pre-treatment differences ("pre-trends") can help increase our confidence in the research design. But they're not perfect. Why?
  1. Just because trends were parallel beforehand doesn't mean that they would continue to be afterwards
  2. Often our estimates of pre-trends are noisy so we're not sure whether they're actually zero or not.

- In addition to looking at the point estimates of pre-trends, it's important to consider what the CIs rule out

- In addition to looking at the point estimates of pre-trends, it's important to consider what the CIs rule out

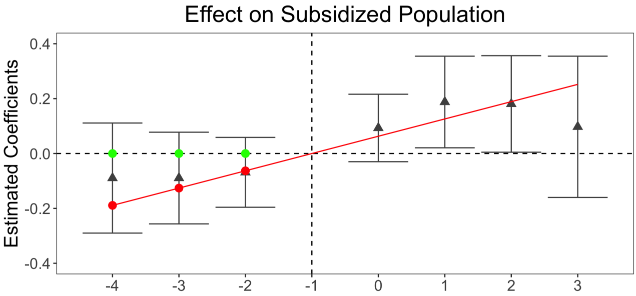
- A good rule of thumb for whether a plot is convincing is whether you can draw a smooth line through all the confidence intervals

- In addition to looking at the point estimates of pre-trends, it's important to consider what the CIs rule out
- A good rule of thumb for whether a plot is convincing is whether you can draw a smooth line through all the confidence intervals



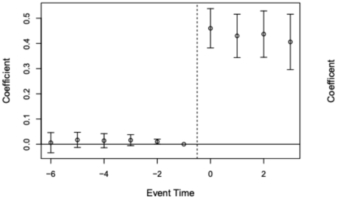
- Are you convinced there's an effect here?

- In addition to looking at the point estimates of pre-trends, it's important to consider what the CIs rule out
- A good rule of thumb for whether a plot is convincing is whether you can draw a smooth line through all the confidence intervals



- Are you convinced there's an effect here? Maybe not!

- In addition to looking at the point estimates of pre-trends, it's important to consider what the CIs rule out
- A good rule of thumb for whether a plot is convincing is whether you can draw a smooth line through all the confidence intervals



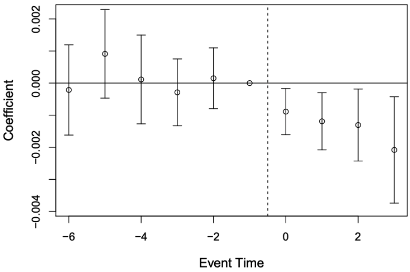
(a) Medicaid Eligibility

- What about here?

- In addition to looking at the point estimates of pre-trends, it's important to consider what the CIs rule out

- A good rule of thumb for whether a plot is convincing is whether you can draw a smooth line through all the confidence intervals

Figure 2: Effect of the ACA Medicaid Expansions on Annual Mortality



- And here?

Standard Errors for Panel Regressions

- We know how to get standard errors for OLS estimates of

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i$$

when $(Y_i, \boldsymbol{X}_i)$ are drawn *iid*.

Standard Errors for Panel Regressions

- We know how to get standard errors for OLS estimates of

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i$$

when $(Y_i, \boldsymbol{X}_i)$ are drawn *iid*.

- Now, we have

$$Y_{it} = \boldsymbol{X}_{it}'\boldsymbol{\beta} + e_{it}$$

Standard Errors for Panel Regressions

- We know how to get standard errors for OLS estimates of

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i$$

when $(Y_i, \boldsymbol{X}_i)$ are drawn *iid*.

- Now, we have

$$Y_{it} = \boldsymbol{X}_{it}'\boldsymbol{\beta} + e_{it}$$

- Is it reasonable to assume that $(Y_{it}, \boldsymbol{X}_{it})$ are *iid* across i and t?

Standard Errors for Panel Regressions

- We know how to get standard errors for OLS estimates of

$$Y_i = \boldsymbol{X}_i' \boldsymbol{\beta} + e_i$$

  when $(Y_i, \boldsymbol{X}_i)$ are drawn *iid*.

- Now, we have

$$Y_{it} = \boldsymbol{X}_{it}' \boldsymbol{\beta} + e_{it}$$

- Is it reasonable to assume that $(Y_{it}, \boldsymbol{X}_{it})$ are *iid* across i and t? No

Standard Errors for Panel Regressions

- We know how to get standard errors for OLS estimates of

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i$$

  when $(Y_i, \boldsymbol{X}_i)$ are drawn *iid*.

- Now, we have

$$Y_{it} = \boldsymbol{X}_{it}'\boldsymbol{\beta} + e_{it}$$

- Is it reasonable to assume that $(Y_{it}, \boldsymbol{X}_{it})$ are *iid* across i and t? No
  1. We expect Y_{i1} to be correlated with Y_{i2}, e.g., people with high earnings in 2010 also tend to have higher earnings in 2011. This is called *serial autocorrelation*

Standard Errors for Panel Regressions

- We know how to get standard errors for OLS estimates of

$$Y_i = \boldsymbol{X}_i' \boldsymbol{\beta} + e_i$$

  when $(Y_i, \boldsymbol{X}_i)$ are drawn *iid*.

- Now, we have

$$Y_{it} = \boldsymbol{X}_{it}' \boldsymbol{\beta} + e_{it}$$

- Is it reasonable to assume that $(Y_{it}, \boldsymbol{X}_{it})$ are *iid* across i and t? No
  1. We expect Y_{i1} to be correlated with Y_{i2}, e.g., people with high earnings in 2010 also tend to have higher earnings in 2011. This is called *serial autocorrelation*
  2. More subtlely, if treatment is assigned at the state level, all people in a given state will have the same value of D_{it} (which is included in $\boldsymbol{X}_{it}$)

Clustered Standard Errors

- **Clustered standard errors** extend the OLS variance formula to allow $(Y_{it}, \boldsymbol{X}_{it})$ to be correlated across observations in the same "cluster"

Clustered Standard Errors

- **Clustered standard errors** extend the OLS variance formula to allow $(Y_{it}, \boldsymbol{X}_{it})$ to be correlated across observations in the same "cluster"

- The assumption is that each cluster is sampled independently.

Clustered Standard Errors

- **Clustered standard errors** extend the OLS variance formula to allow $(Y_{it}, \boldsymbol{X}_{it})$ to be correlated across observations in the same "cluster"
- The assumption is that each cluster is sampled independently.
- For example, if we cluster at the individual level (i), then we allow for Y_{i1} and Y_{i2} to be dependent, but assume (Y_{i1}, Y_{i2}) is independent of (Y_{j1}, Y_{j2}) for $j \neq i$

Clustered Standard Errors

- **Clustered standard errors** extend the OLS variance formula to allow $(Y_{it}, \boldsymbol{X}_{it})$ to be correlated across observations in the same "cluster"

- The assumption is that each cluster is sampled independently.

- For example, if we cluster at the individual level (i), then we allow for Y_{i1} and Y_{i2} to be dependent, but assume (Y_{i1}, Y_{i2}) is independent of (Y_{j1}, Y_{j2}) for $j \neq i$

- In panel analyses, you should at minimum cluster at the individual level to allow for autocorrelation.

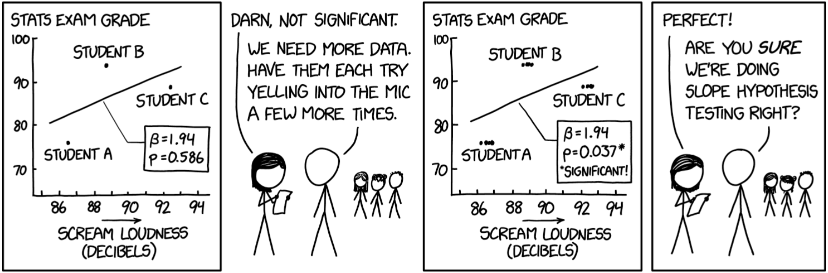
Clustered Standard Errors

- **Clustered standard errors** extend the OLS variance formula to allow $(Y_{it}, \boldsymbol{X}_{it})$ to be correlated across observations in the same "cluster"
- The assumption is that each cluster is sampled independently.
- For example, if we cluster at the individual level (i), then we allow for Y_{i1} and Y_{i2} to be dependent, but assume (Y_{i1}, Y_{i2}) is independent of (Y_{j1}, Y_{j2}) for $j \neq i$
- In panel analyses, you should at minimum cluster at the individual level to allow for autocorrelation.
- If treatment is assigned at a more aggregate level, it is best to cluster at the level where treatment is assigned.

Clustered Standard Errors

- **Clustered standard errors** extend the OLS variance formula to allow $(Y_{it}, \boldsymbol{X}_{it})$ to be correlated across observations in the same "cluster"
- The assumption is that each cluster is sampled independently.
- For example, if we cluster at the individual level (i), then we allow for Y_{i1} and Y_{i2} to be dependent, but assume (Y_{i1}, Y_{i2}) is independent of (Y_{j1}, Y_{j2}) for $j \neq i$
- In panel analyses, you should at minimum cluster at the individual level to allow for autocorrelation.
- If treatment is assigned at a more aggregate level, it is best to cluster at the level where treatment is assigned.
- Keep in mind: the number of "effective observations" (used for CLT) is the number of clusters
  - Clustered SEs will not be reliable when the number of clusters is very small (e.g. < 20)

XKCD



STATS EXAM GRADE
100
90
80
70
STUDENT B
STUDENT C
STUDENT A
β=1.94
p=0.586
86 88 90 92 94
SCREAM LOUDNESS (DECIBELS)
DARN, NOT SIGNIFICANT.
WE NEED MORE DATA. HAVE THEM EACH TRY YELLING INTO THE MIC A FEW MORE TIMES.
STATS EXAM GRADE
100
90
80
70
STUDENT B
STUDENT C
STUDENT A
β=1.94
p=0.037*
*SIGNIFICANT!
86 88 90 92 94
SCREAM LOUDNESS (DECIBELS)
PERFECT!
ARE YOU SURE WE'RE DOING SLOPE HYPOTHESIS TESTING RIGHT?

Implementing Clustered SEs

- Implementing clustered SEs in Stata is very easy
- Just replace
  reg y x, robust

  with
  reg y x, cluster(clustervar)

```
. reg vehicle_fatality_rate beertax i.state i.year, r

Linear regression                               Number of obs     =        336
                                                F(54, 281)        =     128.89
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9089
                                                Root MSE          =     1.9e-05

------------------------------------------------------------------------------
             |               Robust
vehicle_fa~e | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     beertax |   -.000064   .0000255    -2.51   0.013    -.0001141   -.0000139
             |
       state |
         AZ  |  -.0000547   .0000357    -1.53   0.127    -.0001249    .0000156
         AR  |  -.0000639   .0000293    -2.18   0.030    -.0001214   -6.26e-06
         CA  |  -.0001485   .0000412    -3.60   0.000    -.0002296   -.0000674
```

```
. reg vehicle_fatality_rate beertax i.state i.year, cluster(state)

Linear regression                               Number of obs     =        336
                                                F(6, 47)          =          .
                                                Prob > F          =          .
                                                R-squared         =     0.9089
                                                Root MSE          =     1.9e-05

                                 (Std. err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
vehicle_fa~e | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     beertax |   -.000064   .0000386    -1.66   0.104    -.0001416    .0000136
             |
       state |
         AZ  |  -.0000547   .0000506    -1.08   0.286    -.0001566    .0000472
         AR  |  -.0000639   .0000399    -1.60   0.116     -.000144    .0000163
         CA  |  -.0001485   .0000589    -2.52   0.015    -.0002671     -.00003
```

The Quarterly Journal of Economics, February 2004

HOW MUCH SHOULD WE TRUST DIFFERENCES-IN-DIFFERENCES ESTIMATES?*

MARIANNE BERTRAND
ESTHER DUFLO
SENDHIL MULLAINATHAN

Most papers that employ Differences-in-Differences estimation (DD) use many years of data and focus on serially correlated outcomes but ignore that the resulting standard errors are inconsistent. To illustrate the severity of this issue, we randomly generate placebo laws in state-level data on female wages from the Current Population Survey. For each law, we use OLS to compute the DD estimate of its "effect" as well as the standard error of this estimate. These conventional DD standard errors severely understate the standard deviation of the estimators: we find an "effect" significant at the 5 percent level for up to 45 percent of the placebo interventions. We use Monte Carlo simulations to investi-

A Very Famous DiD

- Card and Krueger (1994) ask: how does the minimum wage affect employment?

A Very Famous DiD

- Card and Krueger (1994) ask: how does the minimum wage affect employment?
- How would you expect the MW to affect employment, based on what you learned in micro-economic theory?

A Very Famous DiD

- Card and Krueger (1994) ask: how does the minimum wage affect employment?
- How would you expect the MW to affect employment, based on what you learned in micro-economic theory?
  - In a competitive market, a floor on wages (i.e. the price of labor), should induce a decrease in demand

A Very Famous DiD

- Card and Krueger (1994) ask: how does the minimum wage affect employment?
- How would you expect the MW to affect employment, based on what you learned in micro-economic theory?
  - In a competitive market, a floor on wages (i.e. the price of labor), should induce a decrease in demand
- To study this, CK study an episode in 1992 where NJ raised its minimum wage from $4.25 to $5.05

A Very Famous DiD

- Card and Krueger (1994) ask: how does the minimum wage affect employment?
- How would you expect the MW to affect employment, based on what you learned in micro-economic theory?
  - In a competitive market, a floor on wages (i.e. the price of labor), should induce a decrease in demand
- To study this, CK study an episode in 1992 where NJ raised its minimum wage from \$4.25 to \$5.05
- They use a DiD comparing change in employment in fast food restaurants in NJ to that in neighboring PA , where the MW was flat at \$4.25

Variable		PA (i)	NJ (ii)	Difference, NJ-PA (iii)
1.	FTE employment before,	23.33	20.44	-2.89
	all available observations	(1.35)	(0.51)	(1.44)
2.	FTE employment after,	21.17	21.03	-0.14
	all available observations	(0.94)	(0.52)	(1.07)
3.	Change in mean FTE	-2.16	0.59	2.76
	employment	(1.25)	(0.54)	(1.36)

Notes: Adapted from Card and Krueger (1994), Table 3. The table reports average full-time equivalent (FTE) employment at restaurants in Pennsylvania and New Jersey before and after a minimum wage increase in New Jersey. The sample consists of all stores with data on employment. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing. Standard errors are reported in parentheses

Variable	PA (i)	NJ (ii)	Difference, NJ-PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	-2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	-0.14 (1.07)
3. Change in mean FTE employment	-2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

Notes: Adapted from Card and Krueger (1994), Table 3. The table reports average full-time equivalent (FTE) employment at restaurants in Pennsylvania and New Jersey before and after a minimum wage increase in New Jersey. The sample consists of all stores with data on employment. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing. Standard errors are reported in parentheses

- Point estimates suggest an *increase* in employment of 2.76 FTEs, but not statisticially signfiicant.

Why?!

- The result that an increase in the MW does not seem to decrease employment was very surprising (and controversial) at the time

Why?!

- The result that an increase in the MW does not seem to decrease employment was very surprising (and controversial) at the time

- One explanation for this finding is that labor markets are *not perfectly competitive*. Rather, firms are *monopsonistic*

Why?!

- The result that an increase in the MW does not seem to decrease employment was very surprising (and controversial) at the time
- One explanation for this finding is that labor markets are *not perfectly competitive*. Rather, firms are *monopsonistic*
  - Consider a firm the employs 100 workers at $7/hour.
  - Suppose hiring another worker would produce an extra $10 of profit, but would require raising the wage to $8/hour.

Why?!

- The result that an increase in the MW does not seem to decrease employment was very surprising (and controversial) at the time
- One explanation for this finding is that labor markets are *not perfectly competitive*. Rather, firms are *monopsonistic*
  - Consider a firm the employs 100 workers at $7/hour.
  - Suppose hiring another worker would produce an extra $10 of profit, but would require raising the wage to $8/hour.
  - Should the firm raise the wage to $8/hour?

Why?!

- The result that an increase in the MW does not seem to decrease employment was very surprising (and controversial) at the time
- One explanation for this finding is that labor markets are *not perfectly competitive*. Rather, firms are *monopsonistic*
  - Consider a firm the employs 100 workers at $7/hour.
  - Suppose hiring another worker would produce an extra $10 of profit, but would require raising the wage to $8/hour.
  - Should the firm raise the wage to $8/hour? Not if it means they have to pay all 100 workers an extra $1!

Why?!

- The result that an increase in the MW does not seem to decrease employment was very surprising (and controversial) at the time
- One explanation for this finding is that labor markets are *not perfectly competitive*. Rather, firms are *monopsonistic*
    - Consider a firm the employs 100 workers at \$7/hour.
    - Suppose hiring another worker would produce an extra \$10 of profit, but would require raising the wage to \$8/hour.
    - Should the firm raise the wage to \$8/hour? Not if it means they have to pay all 100 workers an extra \$1!
    - However, if the MW is raised to \$8/hour, then the firm has to pay the first 100 workers \$8 anyway, and would gladly hire the 101st worker at \$8/hour since this brings \$10 of profit.

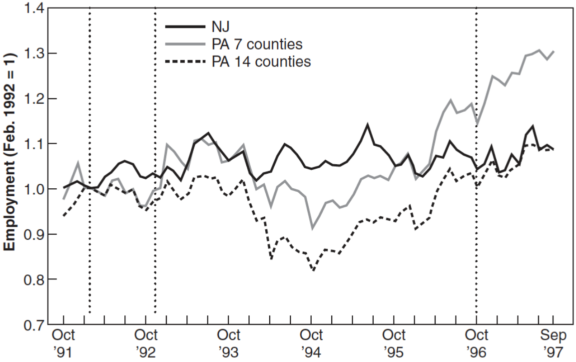


Figure 5.2.2 Employment in New Jersey and Pennsylvania fast food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum wage increase.

- By modern standards, the CK analysis is perhaps not the most convincing

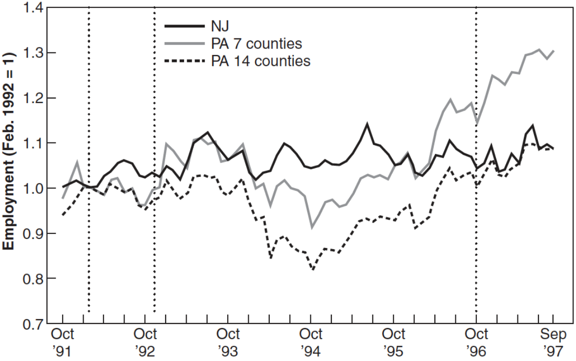


Figure 5.2.2 Employment in New Jersey and Pennsylvania fast food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum wage increase.

- By modern standards, the CK analysis is perhaps not the most convincing
- The two states do not move exactly in parallel even before the policy change in April 1992.

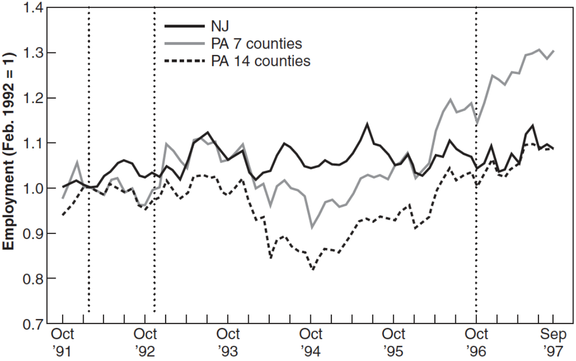


Figure 5.2.2 Employment in New Jersey and Pennsylvania fast food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum wage increase.

- By modern standards, the CK analysis is perhaps not the most convincing
- The two states do not move exactly in parallel even before the policy change in April 1992. We also only have 2 states!

Staggered Timing

- Next I'll show you some more modern evidence on the MW.
- But first we need to discuss DiD when treatment timing is staggered – e.g., states pass minimum wages in different years

Staggered Timing

- Next I'll show you some more modern evidence on the MW.
- But first we need to discuss DiD when treatment timing is staggered – e.g., states pass minimum wages in different years
- Until about 5 years ago, people extended DiD to the staggered setting by running OLS regressions like:

$$Y_{it} = \phi_i + \lambda_t + D_{it}\beta + e_{it}$$

where $D_{it} = 1$ if unit i is treated in period t.

Staggered Timing

- Next I'll show you some more modern evidence on the MW.

- But first we need to discuss DiD when treatment timing is staggered – e.g., states pass minimum wages in different years

- Until about 5 years ago, people extended DiD to the staggered setting by running OLS regressions like:

$$Y_{it} = \phi_i + \lambda_t + D_{it}\beta + e_{it}$$

  where $D_{it} = 1$ if unit i is treated in period t.

- In the two-period model, this corresponds to the diff-in-diff in sample means between treatment and control

Staggered Timing

- Next I'll show you some more modern evidence on the MW.
- But first we need to discuss DiD when treatment timing is staggered – e.g., states pass minimum wages in different years
- Until about 5 years ago, people extended DiD to the staggered setting by running OLS regressions like:

  $$Y_{it} = \phi_i + \lambda_t + D_{it}\beta + e_{it}$$

  where $D_{it} = 1$ if unit i is treated in period t.
- In the two-period model, this corresponds to the diff-in-diff in sample means between treatment and control
- Unfortunately, it turns out that this estimator is *not* an average of DiDs between treated and untreated units in the staggered case.
  - See Borusyak and Jaravel (2016), de Chaisemartin and D'Haultfoeuille (2020), Goodman-Bacon (2021)

- Over the last few years, there has been a lot of research about “fixing” the issues with these regressions
- The solutions typically involve making “clean comparisons” by hand

- Over the last few years, there has been a lot of research about “fixing” the issues with these regressions
- The solutions typically involve making “clean comparisons” by hand
    1. For units first treated in year g, compare outcome change between $g-1$ and $g+k$ to that of units who weren't treated over that period

- Over the last few years, there has been a lot of research about "fixing" the issues with these regressions

- The solutions typically involve making "clean comparisons" by hand
    1. For units first treated in year g, compare outcome change between $g-1$ and $g+k$ to that of units who weren't treated over that period
    2. This is an estimate of the effect k years after treatment for cohort g

- Over the last few years, there has been a lot of research about "fixing" the issues with these regressions
- The solutions typically involve making "clean comparisons" by hand
  1. For units first treated in year g, compare outcome change between $g-1$ and $g+k$ to that of units who weren't treated over that period
  2. This is an estimate of the effect k years after treatment for cohort g
  3. Do this for every g, and then aggregate them to get an average effect

- Over the last few years, there has been a lot of research about "fixing" the issues with these regressions
- The solutions typically involve making "clean comparisons" by hand
  1. For units first treated in year g, compare outcome change between $g-1$ and $g+k$ to that of units who weren't treated over that period
  2. This is an estimate of the effect k years after treatment for cohort g
  3. Do this for every g, and then aggregate them to get an average effect
- There are many implementations of this and related approaches, including Callaway and Sant'Anna (2020), Sun and Abraham (2020), Borusyak, Jaravel & Spiess (2021)

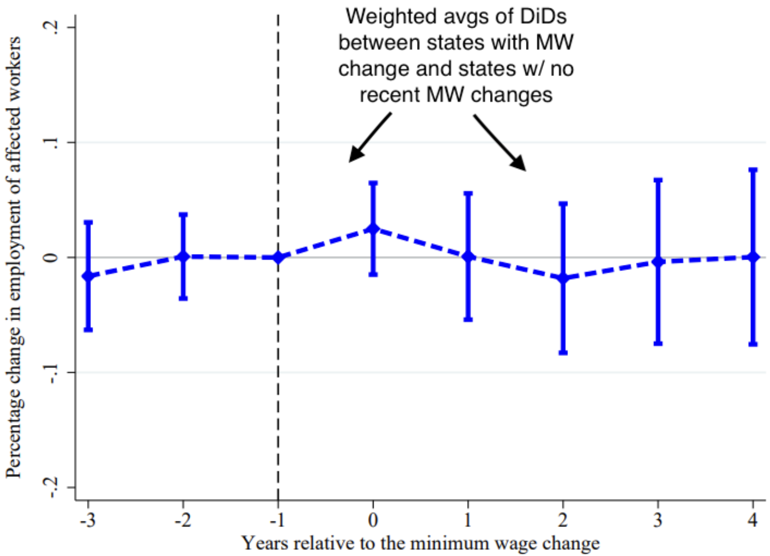
- Cengiz et al (2019) do a modern version of C&K using 138 MW changes between 1976 and 2016

- Cengiz et al (2019) do a modern version of C&K using 138 MW changes between 1976 and 2016
- For each state that changes its MW, they take a "control group" of states that didn't change their MW in the 4 years before/after

- Cengiz et al (2019) do a modern version of C&K using 138 MW changes between 1976 and 2016
- For each state that changes its MW, they take a "control group" of states that didn't change their MW in the 4 years before/after
- They compute a DiD between the treated state and the matched control states

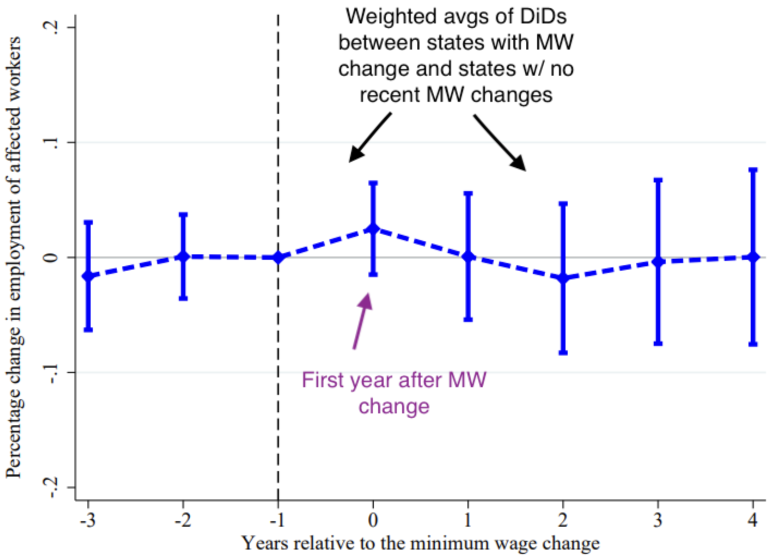
- Cengiz et al (2019) do a modern version of C&K using 138 MW changes between 1976 and 2016
- For each state that changes its MW, they take a “control group” of states that didn’t change their MW in the 4 years before/after
- They compute a DiD between the treated state and the matched control states
- They then take a weighted average of these DiDs to get an overall average effect

(a) Evolution of the missing and excess jobs



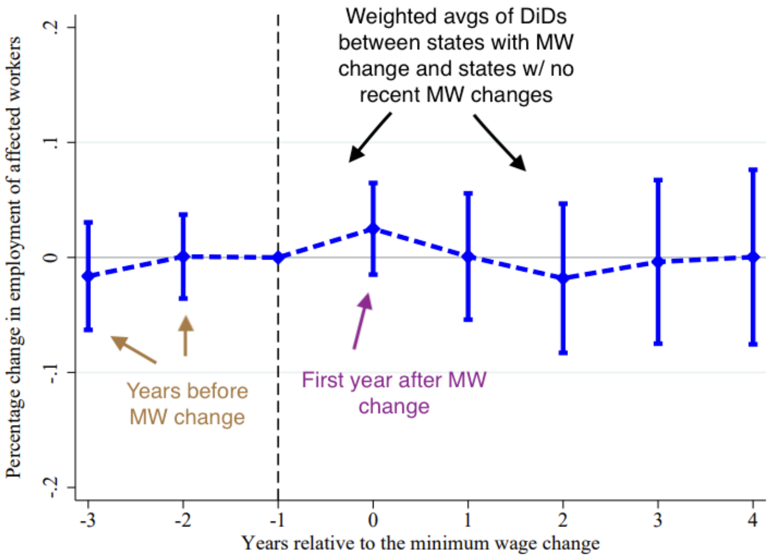
(b) Evolution of the employment of the affected workers

(a) Evolution of the missing and excess jobs



(b) Evolution of the employment of the affected workers

(a) Evolution of the missing and excess jobs



(b) Evolution of the employment of the affected workers

Important Considerations/Caveats

- Historical MW changes in the MW have been fairly modest
  - Not clear that changes in MW from \$4.25 to \$5.05 are informative abour raises from \$7.25 to \$15!

Important Considerations/Caveats

- Historical MW changes in the MW have been fairly modest
  - Not clear that changes in MW from \$4.25 to \$5.05 are informative abour raises from \$7.25 to \$15!
- Historical analyses of MW are typically relatively short-run
  - Over long-run, MW increases may induce shifts in technology that replace workers

Important Considerations/Caveats

- Historical MW changes in the MW have been fairly modest
  - Not clear that changes in MW from $4.25 to $5.05 are informative abour raises from $7.25 to $15!
- Historical analyses of MW are typically relatively short-run
  - Over long-run, MW increases may induce shifts in technology that replace workers
- There is still some debate among economists over whether MWs reduce employment!

Other Panel Data Methods

- We've focused on DiD, which is the most commonly-used panel data method in applied micro-economics
- But there are many others:
    - Controls for lagged dependent variables
    - Synthetic control
    - Matrix completion
- We won't have time to cover these, but if you're interested, I suggest taking more econometrics classes :)