Gov 50: 5. Observational Studies

Matthew Blackwell

Harvard University

Fall 2018

1/ Today's agenda

What have you been up to?

1. Section

What have you been up to?

1. Section
   - Video posted if you missed it.

What have you been up to?

1. Section
   - Video posted if you missed it.
2. Reading

What have you been up to?

1. Section
   - Video posted if you missed it.
2. Reading
   - Read sections 2.5 of QSS.

Where are we going?

1. Quick review of randomized experiments

Where are we going?

1. Quick review of randomized experiments
2. Review sections 2.5 of QSS

Where are we going?

1. Quick review of randomized experiments
2. Review sections 2.5 of QSS
   - Observational studies

Where are we going?

1. Quick review of randomized experiments
2. Review sections 2.5 of QSS
    - Observational studies
    - Confounding bias

Where are we going?

1. Quick review of randomized experiments
2. Review sections 2.5 of QSS
    - Observational studies
    - Confounding bias
    - Cross-section, before-and-after, and differences-in-differences designs

Where are we going?

1. Quick review of randomized experiments
2. Review sections 2.5 of QSS
    - Observational studies
    - Confounding bias
    - Cross-section, before-and-after, and differences-in-differences designs
    - Newspaper endorsements in UK

Where are we going?

1. Quick review of randomized experiments
2. Review sections 2.5 of QSS
    - Observational studies
    - Confounding bias
    - Cross-section, before-and-after, and differences-in-differences designs
    - Newspaper endorsements in UK
3. Problem Set 1

2/ Review of randomized experiments

Reviewing experiments

- Fundamental problem of causal inference:

Reviewing experiments

- Fundamental problem of causal inference:
  - Comparison between factual and counterfactual.

Reviewing experiments

- Fundamental problem of causal inference:
  - Comparison between factual and counterfactual.
  - Counterfactuals not observed.

Reviewing experiments

- Fundamental problem of causal inference:
  - Comparison between factual and counterfactual.
  - Counterfactuals not observed.
- Solution: **randomized control trials** (RCTs)

Reviewing experiments

- Fundamental problem of causal inference:
  - Comparison between factual and counterfactual.
  - Counterfactuals not observed.
- Solution: **randomized control trials** (RCTs)
  - Treatment and control groups identical on average

Reviewing experiments

- Fundamental problem of causal inference:
  - Comparison between factual and counterfactual.
  - Counterfactuals not observed.
- Solution: **randomized control trials** (RCTs)
  - Treatment and control groups identical on average
  - Similar in all (observed and unobserved) characteristics

Reviewing experiments

- Fundamental problem of causal inference:
  - Comparison between factual and counterfactual.
  - Counterfactuals not observed.
- Solution: **randomized control trials** (RCTs)
  - Treatment and control groups identical on average
  - Similar in all (observed and unobserved) characteristics
- Difference in means as an estimate of the **Sample Average Treatment Effect** (SATE):

Reviewing experiments

- Fundamental problem of causal inference:
  - Comparison between factual and counterfactual.
  - Counterfactuals not observed.
- Solution: **randomized control trials** (RCTs)
  - Treatment and control groups identical on average
  - Similar in all (observed and unobserved) characteristics
- Difference in means as an estimate of the **Sample Average Treatment Effect** (SATE):

$$\text{difference-in-means estimator} = \overline{Y}_{\text{treated}} - \overline{Y}_{\text{control}}$$

Reviewing experiments

- Fundamental problem of causal inference:
  - Comparison between factual and counterfactual.
  - Counterfactuals not observed.
- Solution: **randomized control trials** (RCTs)
  - Treatment and control groups identical on average
  - Similar in all (observed and unobserved) characteristics
- Difference in means as an estimate of the **Sample Average Treatment Effect** (SATE):

$$\text{difference-in-means estimator} = \overline{Y}_{\text{treated}} - \overline{Y}_{\text{control}}$$

$$\text{SATE} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$$

Resume experiment

- Load the data:

Resume experiment

- Load the data:

```
## load data
resume <- read.csv("data/resume.csv")
head(resume)
```

Resume experiment

- Load the data:

```r
## load data
resume <- read.csv("data/resume.csv")
head(resume)
```

```
##   firstname    sex  race call
## 1   Allison female white    0
## 2   Kristen female white    0
## 3   Lakisha female black    0
## 4   Latonya female black    0
## 5    Carrie female white    0
## 6       Jay   male white    0
```

Resume experiment

- Load the data:

```r
## load data
resume <- read.csv("data/resume.csv")
head(resume)
```

```
##   firstname    sex  race call
## 1   Allison female white    0
## 2   Kristen female white    0
## 3   Lakisha female black    0
## 4   Latonya female black    0
## 5    Carrie female white    0
## 6       Jay   male white    0
```

- Estimate the SATE:

Resume experiment

- Load the data:

```
## load data
resume <- read.csv("data/resume.csv")
head(resume)
```

```
##   firstname    sex  race call
## 1   Allison female white    0
## 2   Kristen female white    0
## 3   Lakisha female black    0
## 4   Latonya female black    0
## 5    Carrie female white    0
## 6       Jay   male white    0
```

- Estimate the SATE:

```
mean(resume$call[resume$race == "black"]) -
  mean(resume$call[resume$race == "white"])
```

Resume experiment

- Load the data:

```r
## load data
resume <- read.csv("data/resume.csv")
head(resume)
```

```
##   firstname    sex  race call
## 1   Allison female white    0
## 2   Kristen female white    0
## 3   Lakisha female black    0
## 4   Latonya female black    0
## 5    Carrie female white    0
## 6       Jay   male white    0
```

- Estimate the SATE:

```r
mean(resume$call[resume$race == "black"]) -
  mean(resume$call[resume$race == "white"])
```

```
## [1] -0.032
```

3/ Observational Studies

Do newspaper endorsements matter?

- Can newspaper endorsements change voters’ minds?

Do newspaper endorsements matter?

- Can newspaper endorsements change voters' minds?
- Problem: people might read newspaper because of political leanings of paper

Do newspaper endorsements matter?

- Can newspaper endorsements change voters' minds?
- Problem: people might read newspaper because of political leanings of paper
  - Liberals read the New York Times, conservatives read the Wall Street Journal.

Do newspaper endorsements matter?

- Can newspaper endorsements change voters' minds?
- Problem: people might read newspaper because of political leanings of paper
  - Liberals read the New York Times, conservatives read the Wall Street Journal.
- Could do a lab experiment, but there are concerns over **external validity**

British newspaper readers

- Two political scientists tested this with British newspapers in the 1990s.

British newspaper readers

- Two political scientists tested this with British newspapers in the 1990s.

Name	Description
`tolabor`	Whether or not the respondent read a newspaper that switched endorsement to the Labour between 1992 and 1997
`vote_l_92`	Indicator for if the respondent voted for Labour in 1992 election
`vote_l_97`	Indicator for if the respondent voted for Labour in 1997 election
`parent_labor`	Did the respondent's parents vote for Labour?
`male`	Is the respondent male (1) or female (0)?

Loading the data

- First we load the data:

```
news <- read.csv("data/newspapers.csv")
dim(news)
```

```
## [1] 1593    7
```

Loading the data

- First we load the data:

```
news <- read.csv("data/newspapers.csv")
dim(news)
```

```
## [1] 1593    7
```

- Next we create subsets for readers of newspapers that switched to Labour (**treatment group**) and readers of those papers who didn't switch (**control group**):

Loading the data

- First we load the data:

```
news <- read.csv("data/newspapers.csv")
dim(news)
```

```
## [1] 1593    7
```

- Next we create subsets for readers of newspapers that switched to Labour (**treatment group**) and readers of those papers who didn't switch (**control group**):

```
switched <- subset(news, subset = tolabor == 1)
stayed <- subset(news, subset = tolabor == 0)
```

Observational studies

- Example of an **observational study**:

Observational studies

- Example of an **observational study**:
  - We as researchers observe a naturally assigned treatment

Observational studies

- Example of an **observational study**:
  - We as researchers observe a naturally assigned treatment
  - Very common: often can't randomize for ethical/logistical reasons.

Observational studies

- Example of an **observational study**:
  - We as researchers observe a naturally assigned treatment
  - Very common: often can't randomize for ethical/logistical reasons.
- **Internal validity**: are the causal assumption satisfied? Can we interpret this as a causal effect?

Observational studies

- Example of an **observational study**:
  - We as researchers observe a naturally assigned treatment
  - Very common: often can't randomize for ethical/logistical reasons.
- **Internal validity**: are the causal assumption satisfied? Can we interpret this as a causal effect?
  - RCTs usually have higher internal validity.

Observational studies

- Example of an **observational study**:
  - We as researchers observe a naturally assigned treatment
  - Very common: often can't randomize for ethical/logistical reasons.
- **Internal validity**: are the causal assumption satisfied? Can we interpret this as a causal effect?
  - RCTs usually have higher internal validity.
  - Observational studies less so, because pre-treatment variable may differ between treatment and control groups

Observational studies

- Example of an **observational study**:
  - We as researchers observe a naturally assigned treatment
  - Very common: often can't randomize for ethical/logistical reasons.
- **Internal validity**: are the causal assumption satisfied? Can we interpret this as a causal effect?
  - RCTs usually have higher internal validity.
  - Observational studies less so, because pre-treatment variable may differ between treatment and control groups
- **External validity**: can the conclusions/estimated effects be generalized beyond this study?

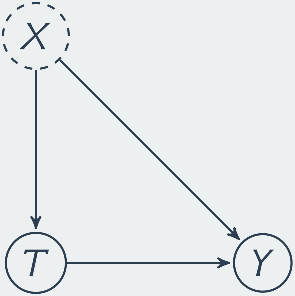
Observational studies

- Example of an **observational study**:
  - We as researchers observe a naturally assigned treatment
  - Very common: often can't randomize for ethical/logistical reasons.
- **Internal validity**: are the causal assumption satisfied? Can we interpret this as a causal effect?
  - RCTs usually have higher internal validity.
  - Observational studies less so, because pre-treatment variable may differ between treatment and control groups
- **External validity**: can the conclusions/estimated effects be generalized beyond this study?
  - RCTs weaker here because often very expensive to conduct on representative samples.

Observational studies

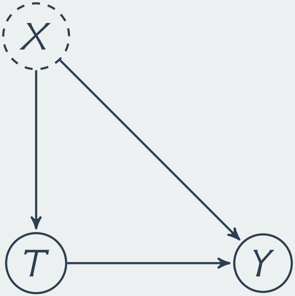
- Example of an **observational study**:
  - We as researchers observe a naturally assigned treatment
  - Very common: often can't randomize for ethical/logistical reasons.
- **Internal validity**: are the causal assumption satisfied? Can we interpret this as a causal effect?
  - RCTs usually have higher internal validity.
  - Observational studies less so, because pre-treatment variable may differ between treatment and control groups
- **External validity**: can the conclusions/estimated effects be generalized beyond this study?
  - RCTs weaker here because often very expensive to conduct on representative samples.
  - Observational studies often have larger/more representative samples that improve external validity.

Confounding



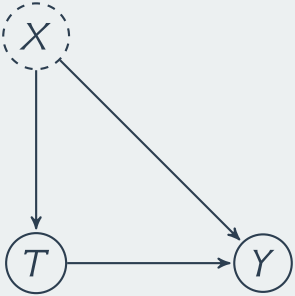
- **Confounder**: a pre-treatment variable affecting treatment and the outcome.

Confounding



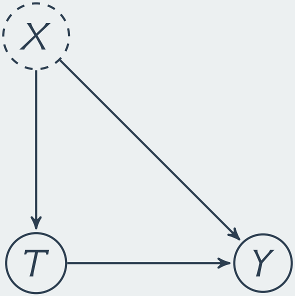
- **Confounder**: a pre-treatment variable affecting treatment and the outcome.
  - More liberal people (X) might read newspapers that switch to endorsing Labour (T).

Confounding



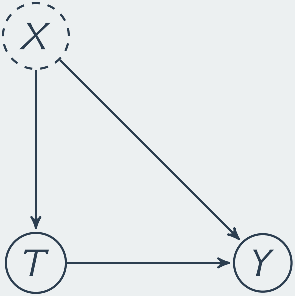
- **Confounder**: a pre-treatment variable affecting treatment and the outcome.
  - More liberal people (X) might read newspapers that switch to endorsing Labour (T).
  - More liberal people (X) also more likely to vote for Labour (Y).

Confounding



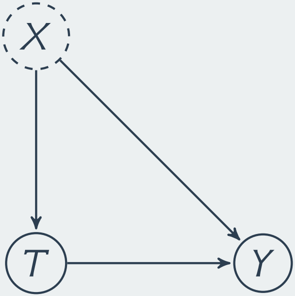
- **Confounder**: a pre-treatment variable affecting treatment and the outcome.
  - More liberal people (X) might read newspapers that switch to endorsing Labour (T).
  - More liberal people (X) also more likely to vote for Labour (Y).
- **Confounding bias** in the estimated SATE due to these differences

Confounding



- **Confounder**: a pre-treatment variable affecting treatment and the outcome.
  - More liberal people (X) might read newspapers that switch to endorsing Labour (T).
  - More liberal people (X) also more likely to vote for Labour (Y).
- **Confounding bias** in the estimated SATE due to these differences
  - $\overline{Y}_{\text{control}}$ not a good proxy for $Y_i(0)$ in treated group.

Confounding



- **Confounder**: a pre-treatment variable affecting treatment and the outcome.
    - More liberal people (X) might read newspapers that switch to endorsing Labour (T).
    - More liberal people (X) also more likely to vote for Labour (Y).
- **Confounding bias** in the estimated SATE due to these differences
    - $\overline{Y}_{\text{control}}$ not a good proxy for $Y_i(0)$ in treated group.
    - one type: **selection bias** from self-selection into treatment

Research designs

- How can we find a good comparison group?

Research designs

- How can we find a good comparison group?
- Depends on the data we have available.

Research designs

- How can we find a good comparison group?
- Depends on the data we have available.
- Three general types of observational study **reseach designs**:

Research designs

- How can we find a good comparison group?
- Depends on the data we have available.
- Three general types of observational study **reseach designs**:
    1. **Cross-sectional design**: compare outcomes treated and control units at one point in time.

Research designs

- How can we find a good comparison group?
- Depends on the data we have available.
- Three general types of observational study **reseach designs**:
    1. **Cross-sectional design**: compare outcomes treated and control units at one point in time.
    2. **Before-and-after design**: compare outcomes before and after a unit has been treated, but need over-time data on treated group.

Research designs

- How can we find a good comparison group?
- Depends on the data we have available.
- Three general types of observational study **reseach designs**:
    1. **Cross-sectional design**: compare outcomes treated and control units at one point in time.
    2. **Before-and-after design**: compare outcomes before and after a unit has been treated, but need over-time data on treated group.
    3. **Difference-in-differences design**: use before-and-after information for the treated and control group, but need over-time on treated and control group.

Cross-sectional design

- Compare treatment (readers of switching papers) to control after the switches.

Cross-sectional design

- Compare treatment (readers of switching papers) to control after the switches.
- Treatment and control groups assumed to be identical (on average) in terms of all confounders.

Cross-sectional design

- Compare treatment (readers of switching papers) to control after the switches.
- Treatment and control groups assumed to be identical (on average) in terms of all confounders.
  - Sometimes called **unconfoundedness**.

Cross-sectional design

- Compare treatment (readers of switching papers) to control after the switches.
- Treatment and control groups assumed to be identical (on average) in terms of all confounders.
  - Sometimes called **unconfoundedness**.
- SATE estimate:

Cross-sectional design

- Compare treatment (readers of switching papers) to control after the switches.
- Treatment and control groups assumed to be identical (on average) in terms of all confounders.
  - Sometimes called **unconfoundedness**.
- SATE estimate:

```
mean(switched$vote_l_97) - mean(stayed$vote_l_97)
```

Cross-sectional design

- Compare treatment (readers of switching papers) to control after the switches.
- Treatment and control groups assumed to be identical (on average) in terms of all confounders.
  - Sometimes called **unconfoundedness**.
- SATE estimate:

```
mean(switched$vote_l_97) - mean(stayed$vote_l_97)
```

```
## [1] 0.152
```

Cross-sectional design

- Compare treatment (readers of switching papers) to control after the switches.
- Treatment and control groups assumed to be identical (on average) in terms of all confounders.
    - Sometimes called **unconfoundedness**.
- SATE estimate:

```
mean(switched$vote_l_97) - mean(stayed$vote_l_97)
```

```
## [1] 0.152
```

- Could there be confounders?

Checking confounders

- Compare means of possible confounders in treated and control groups.

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

```
mean(switched$male)
```

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

```
mean(switched$male)
```

```
## [1] 0.455
```

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

```r
mean(switched$male)
```

```
## [1] 0.455
```

```r
mean(stayed$male)
```

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

```
mean(switched$male)
```

```
## [1] 0.455
```

```
mean(stayed$male)
```

```
## [1] 0.556
```

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

```r
mean(switched$male)
```

```
## [1] 0.455
```

```r
mean(stayed$male)
```

```
## [1] 0.556
```

- Proportion whose parents voted for Labour:

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

```
mean(switched$male)
```

```
## [1] 0.455
```

```
mean(stayed$male)
```

```
## [1] 0.556
```

- Proportion whose parents voted for Labour:

```
mean(switched$parent_labor)
```

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

```r
mean(switched$male)
```

```
## [1] 0.455
```

```r
mean(stayed$male)
```

```
## [1] 0.556
```

- Proportion whose parents voted for Labour:

```r
mean(switched$parent_labor)
```

```
## [1] 0.436
```

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

```
mean(switched$male)
```

```
## [1] 0.455
```

```
mean(stayed$male)
```

```
## [1] 0.556
```

- Proportion whose parents voted for Labour:

```
mean(switched$parent_labor)
```

```
## [1] 0.436
```

```
mean(stayed$parent_labor)
```

Checking confounders

- Compare means of possible confounders in treated and control groups.
- Proportion male:

```
mean(switched$male)
```

```
## [1] 0.455
```

```
mean(stayed$male)
```

```
## [1] 0.556
```

- Proportion whose parents voted for Labour:

```
mean(switched$parent_labor)
```

```
## [1] 0.436
```

```
mean(stayed$parent_labor)
```

```
## [1] 0.354
```

Statistical control

- **Statistical control**: adjust for confounders using statistical procedures.

Statistical control

- **Statistical control**: adjust for confounders using statistical procedures.
  - can help to reduce confounding bias.

Statistical control

- **Statistical control**: adjust for confounders using statistical procedures.
  - can help to reduce confounding bias.
- One type of statistical control: **subclassification**

Statistical control

- **Statistical control**: adjust for confounders using statistical procedures.
  - can help to reduce confounding bias.
- One type of statistical control: **subclassification**
  - Compare treated and control groups within levels of a confounding variable.

Statistical control

- **Statistical control**: adjust for confounders using statistical procedures.
    - can help to reduce confounding bias.
- One type of statistical control: **subclassification**
    - Compare treated and control groups within levels of a confounding variable.
    - Remaining effect can't be due to the confounder.

Statistical control

- **Statistical control**: adjust for confounders using statistical procedures.
  - can help to reduce confounding bias.
- One type of statistical control: **subclassification**
  - Compare treated and control groups within levels of a confounding variable.
  - Remaining effect can't be due to the confounder.
- Threat to inference: we can only control for observed variables $\rightsquigarrow$ threat of **unmeasured confounding**

Subclassifying on gender

- Estimate effect within levels of gender. First, for men:

Subclassifying on gender

- Estimate effect within levels of gender. First, for men:

```
switched.males <- switched[switched$male == 1,]
stayed.males <- stayed[stayed$male == 1,]

mean(switched.males$vote_l_97) - mean(stayed.males$vote_l_97)
```

Subclassifying on gender

- Estimate effect within levels of gender. First, for men:

```
switched.males <- switched[switched$male == 1,]
stayed.males <- stayed[stayed$male == 1,]

mean(switched.males$vote_l_97) - mean(stayed.males$vote_l_97)
```

```
## [1] 0.126
```

Subclassifying on gender

- Estimate effect within levels of gender. First, for men:

```
switched.males <- switched[switched$male == 1,]
stayed.males <- stayed[stayed$male == 1,]

mean(switched.males$vote_l_97) - mean(stayed.males$vote_l_97)
```

```
## [1] 0.126
```

- For women:

Subclassifying on gender

- Estimate effect within levels of gender. First, for men:

```
switched.males <- switched[switched$male == 1,]
stayed.males <- stayed[stayed$male == 1,]

mean(switched.males$vote_l_97) - mean(stayed.males$vote_l_97)
```

```
## [1] 0.126
```

- For women:

```
switched.females <- switched[switched$male == 0,]
stayed.females <- stayed[stayed$male == 0,]

mean(switched.females$vote_l_97) - mean(stayed.females$vote_l_97)
```

Subclassifying on gender

- Estimate effect within levels of gender. First, for men:

```
switched.males <- switched[switched$male == 1,]
stayed.males <- stayed[stayed$male == 1,]

mean(switched.males$vote_l_97) - mean(stayed.males$vote_l_97)
```

```
## [1] 0.126
```

- For women:

```
switched.females <- switched[switched$male == 0,]
stayed.females <- stayed[stayed$male == 0,]

mean(switched.females$vote_l_97) - mean(stayed.females$vote_l_97)
```

```
## [1] 0.172
```

Before-and-after comparison

- Compare readers of party-switching newspapers before and after switch.

Before-and-after comparison

- Compare readers of party-switching newspapers before and after switch.
- Advantage: all person-specific features held fixed

Before-and-after comparison

- Compare readers of party-switching newspapers before and after switch.
- Advantage: all person-specific features held fixed
  - comparing within a person over time.

Before-and-after comparison

- Compare readers of party-switching newspapers before and after switch.
- Advantage: all person-specific features held fixed
  - comparing within a person over time.
- Estimate:

Before-and-after comparison

- Compare readers of party-switching newspapers before and after switch.
- Advantage: all person-specific features held fixed
  - comparing within a person over time.
- Estimate:

```
switchedDiff <- mean(switched$vote_l_97) -
  mean(switched$vote_l_92)
switchedDiff
```

Before-and-after comparison

- Compare readers of party-switching newspapers before and after switch.
- Advantage: all person-specific features held fixed
  - comparing within a person over time.
- Estimate:

```
switchedDiff <- mean(switched$vote_l_97) -
  mean(switched$vote_l_92)
switchedDiff
```

```
## [1] 0.194
```

Before-and-after comparison

- Compare readers of party-switching newspapers before and after switch.
- Advantage: all person-specific features held fixed
  - comparing within a person over time.
- Estimate:

```
switchedDiff <- mean(switched$vote_l_97) -
  mean(switched$vote_l_92)
switchedDiff
```

```
## [1] 0.194
```

- Threat to inference: **time-varying confounders**

Before-and-after comparison

- Compare readers of party-switching newspapers before and after switch.
- Advantage: all person-specific features held fixed
  - comparing within a person over time.
- Estimate:

```
switchedDiff <- mean(switched$vote_l_97) -
  mean(switched$vote_l_92)
switchedDiff
```

```
## [1] 0.194
```

- Threat to inference: **time-varying confounders**
  - Time trend: Labour just did better overall in 1997 compared to 1992.

Differences in differences

- Key idea: use the before-and-after difference of **control group** to infer what would have happend to **treatment group** without treatment.

Differences in differences

- Key idea: use the before-and-after difference of **control group** to infer what would have happend to **treatment group** without treatment.
- **Parallel time trend assumption**

Differences in differences

- Key idea: use the before-and-after difference of **control group** to infer what would have happend to **treatment group** without treatment.
- **Parallel time trend assumption**
  - Changes in vote of readers of non-switching papers roughly the same as changes that readers of switching papers would have been if they read non-switching papers.

Differences in differences

- Key idea: use the before-and-after difference of **control group** to infer what would have happend to **treatment group** without treatment.
- **Parallel time trend assumption**
    - Changes in vote of readers of non-switching papers roughly the same as changes that readers of switching papers would have been if they read non-switching papers.
    - Threat to inference: non-parallel trends.

Differences in differences

- Key idea: use the before-and-after difference of **control group** to infer what would have happend to **treatment group** without treatment.
- **Parallel time trend assumption**
  - Changes in vote of readers of non-switching papers roughly the same as changes that readers of switching papers would have been if they read non-switching papers.
  - Threat to inference: non-parallel trends.
- Estimate:

Differences in differences

- Key idea: use the before-and-after difference of **control group** to infer what would have happend to **treatment group** without treatment.
- **Parallel time trend assumption**
    - Changes in vote of readers of non-switching papers roughly the same as changes that readers of switching papers would have been if they read non-switching papers.
    - Threat to inference: non-parallel trends.
- Estimate:

```
stayedDiff <- mean(stayed$vote_l_97) -
  mean(stayed$vote_l_92)
switchedDiff-stayedDiff
```

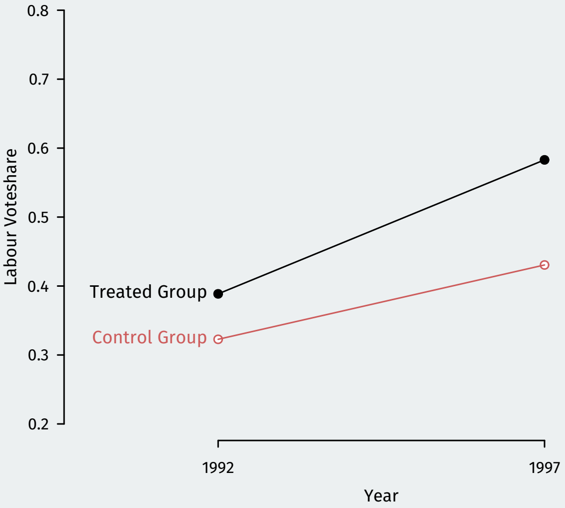
Differences in differences

- Key idea: use the before-and-after difference of **control group** to infer what would have happend to **treatment group** without treatment.
- **Parallel time trend assumption**
  - Changes in vote of readers of non-switching papers roughly the same as changes that readers of switching papers would have been if they read non-switching papers.
  - Threat to inference: non-parallel trends.
- Estimate:

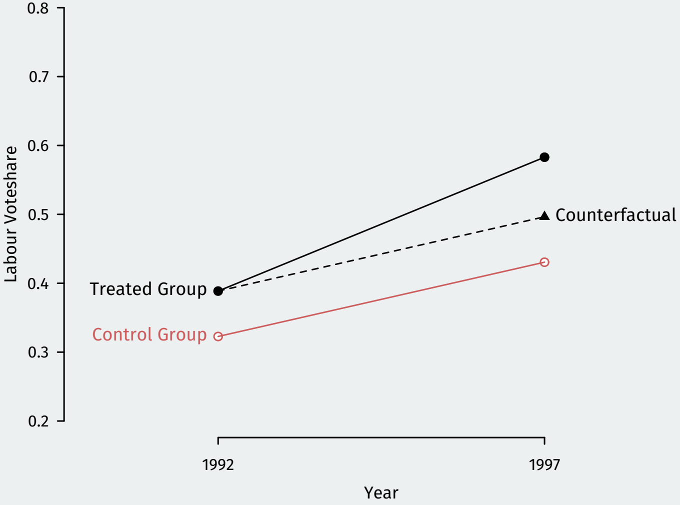
```
stayedDiff <- mean(stayed$vote_l_97) -
  mean(stayed$vote_l_92)
switchedDiff-stayedDiff
```

```
## [1] 0.0865
```

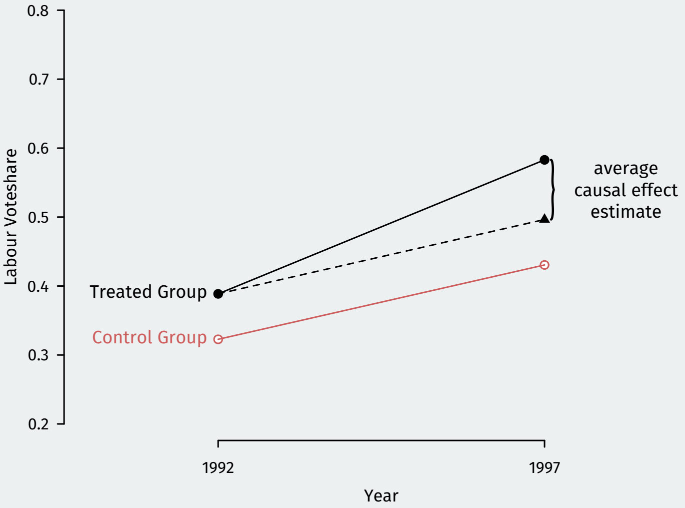
Visualizing DiD



Visualizing DiD



Visualizing DiD



Summarizing approaches

1. **Cross-sectional comparison**

Summarizing approaches

1. **Cross-sectional comparison**
   - Compare treated units with control units after treatment

Summarizing approaches

1. **Cross-sectional comparison**
   - Compare treated units with control units after treatment
   - Assumption: treated and controls units are comparable

Summarizing approaches

1. **Cross-sectional comparison**
   - Compare treated units with control units after treatment
   - Assumption: treated and controls units are comparable
   - Possible confounding

Summarizing approaches

1. **Cross-sectional comparison**
   - Compare treated units with control units after treatment
   - Assumption: treated and controls units are comparable
   - Possible confounding
2. **Before-and-after comparison**

Summarizing approaches

1. **Cross-sectional comparison**
   - Compare treated units with control units after treatment
   - Assumption: treated and controls units are comparable
   - Possible confounding
2. **Before-and-after comparison**
   - Compare the same units before and after treatment

Summarizing approaches

1. **Cross-sectional comparison**
   - Compare treated units with control units after treatment
   - Assumption: treated and controls units are comparable
   - Possible confounding
2. **Before-and-after comparison**
   - Compare the same units before and after treatment
   - Assumption: no time-varying confounding

Summarizing approaches

1. **Cross-sectional comparison**
   - Compare treated units with control units after treatment
   - Assumption: treated and controls units are comparable
   - Possible confounding
2. **Before-and-after comparison**
   - Compare the same units before and after treatment
   - Assumption: no time-varying confounding
3. **Differences-in-differences**

Summarizing approaches

1. **Cross-sectional comparison**
   - Compare treated units with control units after treatment
   - Assumption: treated and controls units are comparable
   - Possible confounding
2. **Before-and-after comparison**
   - Compare the same units before and after treatment
   - Assumption: no time-varying confounding
3. **Differences-in-differences**
   - Assumption: parallel trends assumptions

Summarizing approaches

1. **Cross-sectional comparison**
    - Compare treated units with control units after treatment
    - Assumption: treated and controls units are comparable
    - Possible confounding
2. **Before-and-after comparison**
    - Compare the same units before and after treatment
    - Assumption: no time-varying confounding
3. **Differences-in-differences**
    - Assumption: parallel trends assumptions
    - Under this assumption, it accounts for unit-specific and time-varying confounding.

Summarizing approaches

1. **Cross-sectional comparison**
    - Compare treated units with control units after treatment
    - Assumption: treated and controls units are comparable
    - Possible confounding
2. **Before-and-after comparison**
    - Compare the same units before and after treatment
    - Assumption: no time-varying confounding
3. **Differences-in-differences**
    - Assumption: parallel trends assumptions
    - Under this assumption, it accounts for unit-specific and time-varying confounding.

- All three rely on assumptions that can't be verified to handle confounding.

Summarizing approaches

1. **Cross-sectional comparison**
   - Compare treated units with control units after treatment
   - Assumption: treated and controls units are comparable
   - Possible confounding
2. **Before-and-after comparison**
   - Compare the same units before and after treatment
   - Assumption: no time-varying confounding
3. **Differences-in-differences**
   - Assumption: parallel trends assumptions
   - Under this assumption, it accounts for unit-specific and time-varying confounding.

- All three rely on assumptions that can't be verified to handle confounding.
- RCTs handle confounding by design.

4/ Wrapping up

Homework 1

- On Canvas/rstudio.cloud now—you should start.

Homework 1

- On Canvas/rstudio.cloud now—you should start.
- See Rmarkdown tutorials and section videos/notes for help with Rmarkdown.

Homework 1

- On Canvas/rstudio.cloud now—you should start.
- See Rmarkdown tutorials and section videos/notes for help with Rmarkdown.
- Post questions and answers to Canvas, but avoid posting solutions.

Homework 1

- On Canvas/rstudio.cloud now—you should start.
- See Rmarkdown tutorials and section videos/notes for help with Rmarkdown.
- Post questions and answers to Canvas, but avoid posting solutions.
- Submission:

Homework 1

- On Canvas/rstudio.cloud now—you should start.
- See Rmarkdown tutorials and section videos/notes for help with Rmarkdown.
- Post questions and answers to Canvas, but avoid posting solutions.
- Submission:
  - electronic copy of your Rmd file

Homework 1

- On Canvas/rstudio.cloud now—you should start.
- See Rmarkdown tutorials and section videos/notes for help with Rmarkdown.
- Post questions and answers to Canvas, but avoid posting solutions.
- Submission:
  - electronic copy of your Rmd file
  - electronic copy of your compiled PDF

Homework 1

- On Canvas/rstudio.cloud now—you should start.
- See Rmarkdown tutorials and section videos/notes for help with Rmarkdown.
- Post questions and answers to Canvas, but avoid posting solutions.
- Submission:
    - electronic copy of your Rmd file
    - electronic copy of your compiled PDF
- Harvard College students:

Homework 1

- On Canvas/rstudio.cloud now—you should start.
- See Rmarkdown tutorials and section videos/notes for help with Rmarkdown.
- Post questions and answers to Canvas, but avoid posting solutions.
- Submission:
  - electronic copy of your Rmd file
  - electronic copy of your compiled PDF
- Harvard College students:
  - Study Halls begin tonight 6-9pm in Lowell Dining Hall (location may change in future weeks).

Homework 1

- On Canvas/rstudio.cloud now—you should start.
- See Rmarkdown tutorials and section videos/notes for help with Rmarkdown.
- Post questions and answers to Canvas, but avoid posting solutions.
- Submission:
  - electronic copy of your Rmd file
  - electronic copy of your compiled PDF
- Harvard College students:
  - Study Halls begin tonight 6-9pm in Lowell Dining Hall (location may change in future weeks).
  - 3 CAs (T.J., Hana, and Kayla) will be there to help answer questions.

Next time

- Start to talk more about measurement and descriptive statistics.

Next time

- Start to talk more about measurement and descriptive statistics.
- Read: QSS 2.6, 3.1–3.2