Causal Inference
Day 1

J.J. Naddeo

Massive Data Institute

April 18, 2023

Introduction

- Graduated May 2022 with PhD in Economics

Introduction

- Graduated May 2022 with PhD in Economics
- Focused on applied problems (specifically in CLS and US Politics)

Introduction

- Graduated May 2022 with PhD in Economics
- Focused on applied problems (specifically in CLS and US Politics)
- I am not a theorist/econometrician

Introduction

- Graduated May 2022 with PhD in Economics
- Focused on applied problems (specifically in CLS and US Politics)
- I am not a theorist/econometrician
- This class will (hopefully) give beginners some foundation to skeptically look at empirical results...

Introduction

- Graduated May 2022 with PhD in Economics
- Focused on applied problems (specifically in CLS and US Politics)
- I am not a theorist/econometrician
- This class will (hopefully) give beginners some foundation to skeptically look at empirical results...
- For more in depth dives I strongly recommend attending Mixtape workshops

Correlation $\neq$ Causation

- "Thing X is linked to Y, coming up at 11pm"

Correlation $\neq$ Causation

- “Thing X is linked to Y, coming up at 11pm”
- Always be super skeptical!!! Correlations appear everywhere!

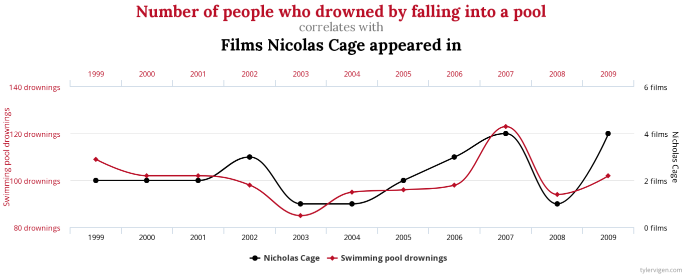
Correlation $\neq$ Causation

- "Thing X is linked to Y, coming up at 11pm"
- Always be super skeptical!!! Correlations appear everywhere!
- In fact Tyler Vigen wrote a Book and has a Webiste with over 30,000 examples of "spurious correlations"

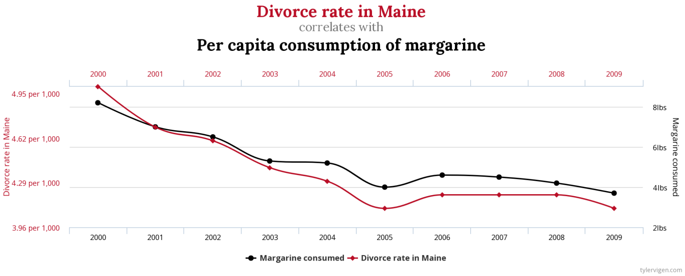
Correlation ≠ Causation

- "Thing X is linked to Y, coming up at 11pm"
- Always be super skeptical!!! Correlations appear everywhere!
- In fact Tyler Vigen wrote a Book and has a Webiste with over 30,000 examples of "spurious correlations"
- Some examples...

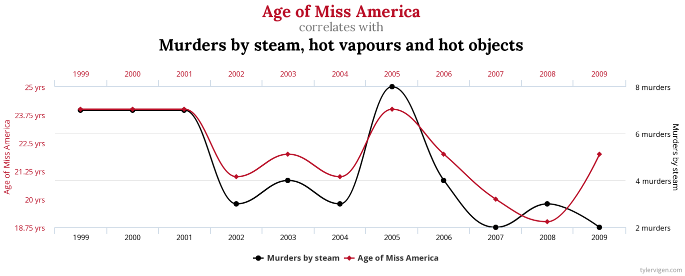
Correlation ≠ Causation



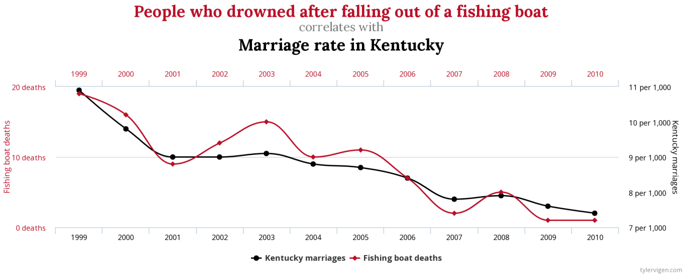
Correlation ≠ Causation



Correlation ≠ Causation



Correlation ≠ Causation



Not so crazy examples

- My guess is no one has a good theory why any of these things are correlated

Not so crazy examples

- My guess is no one has a good theory why any of these things are correlated
- We find that successful people make their beds

Not so crazy examples

- My guess is no one has a good theory why any of these things are correlated
- We find that successful people make their beds
- Is bed making causing success?

Not so crazy examples

- My guess is no one has a good theory why any of these things are correlated
- We find that successful people make their beds
- Is bed making causing success?
- Red wine consumption linked to higher income?

Not so crazy examples

- My guess is no one has a good theory why any of these things are correlated

- We find that successful people make their beds
- Is bed making causing success?

- Red wine consumption linked to higher income?
- Is red wine helping people earn more money?

Not so crazy examples

- My guess is no one has a good theory why any of these things are correlated
- We find that successful people make their beds
- Is bed making causing success?
- Red wine consumption linked to higher income?
- Is red wine helping people earn more money?
- Student who attend office hours do better on tests

Not so crazy examples

- My guess is no one has a good theory why any of these things are correlated
- We find that successful people make their beds
- Is bed making causing success?
- Red wine consumption linked to higher income?
- Is red wine helping people earn more money?
- Student who attend office hours do better on tests
- Do office hours actually increase test scores?

Not so crazy examples

- My guess is no one has a good theory why any of these things are correlated

- We find that successful people make their beds
- Is bed making causing success?

- Red wine consumption linked to higher income?
- Is red wine helping people earn more money?
- Student who attend office hours do better on tests
- Do office hours actually increase test scores?

- Always look for theory that explain mechanisms why, this gets you closer to causality

Not so crazy examples

- My guess is no one has a good theory why any of these things are correlated
- We find that successful people make their beds
- Is bed making causing success?
- Red wine consumption linked to higher income?
- Is red wine helping people earn more money?
- Student who attend office hours do better on tests
- Do office hours actually increase test scores?
- Always look for theory that explain mechanisms why, this gets you closer to causality
- Even when there is a theory, it is not clear empirical tests recover the things we want (more on this later)

Also No Correlation $\neq$ No Causation

Figure: Figure 1.1 from Causal Inference: The Mixtape by Scott Cunningham

Roadmap

- I will pull from Causal Inference: The Mixtape by Scott Cunningham as well as some material from "Mixtape" Workshops

Roadmap

- I will pull from Causal Inference: The Mixtape by Scott Cunningham as well as some material from "Mixtape" Workshops
- **Day 1:**
  1. Introduce some notation/theory
  2. Review OLS
  3. Code some simulations to show bias that can happen
  4. Introduce IV

Roadmap

- I will pull from Causal Inference: The Mixtape by Scott Cunningham as well as some material from "Mixtape" Workshops
- **Day 1:**
  1. Introduce some notation/theory
  2. Review OLS
  3. Code some simulations to show bias that can happen
  4. Introduce IV
- **Day 2:**
  1. Finish IV
  2. Example of IV
  3. Explain Difference in Differences
  4. Code simple DiD example

Bi-variate

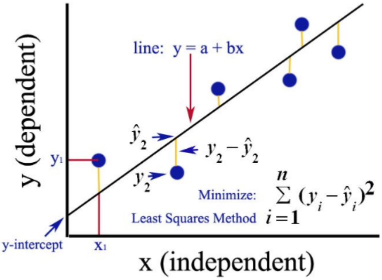
$$Y = \alpha + \beta X + \epsilon$$

- $\alpha \rightarrow$ constant (Y intercept)
- $\beta \rightarrow$ slope of linear line, measure of correlation between X and Y (after removing the constant)
- $\epsilon \rightarrow$ error term

Bi-variate

$$Y = \alpha + \beta X + \epsilon$$

- $\alpha \rightarrow$ constant (Y intercept)
- $\beta \rightarrow$ slope of linear line, measure of correlation between X and Y (after removing the constant)
- $\epsilon \rightarrow$ error term



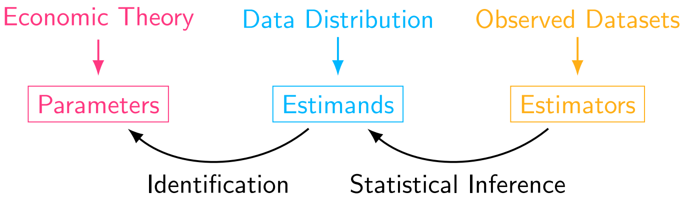
OLS

- OLS aims to find the (α, β) that minimize the sum squared ϵ's
- Luckily this isn't something you need to do guess and check with
- Given data (Y, X) there is a formula for $(\alpha^{OLS}, \beta^{OLS})$

$$\hat{\beta} = \frac{n \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n \sum_i x_i^2 - (\sum_i x_i)^2}$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

What does OLS get us?



- From Peter Hull's Mixtape Session
- The OLS *estimator* $\widehat{\beta}^{OLS}$ consistently estimates the regression *estimand* β^{OLS} under relatively weak conditions
- It *doesn't* directly tell us about the relationship between β^{OLS} and β (e.g. the causal effect of some treatment)

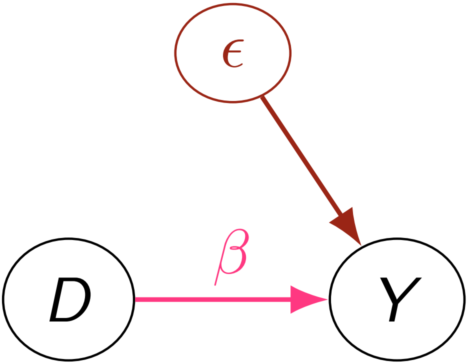
Bias

- The model parameter in $Y_i = \alpha + \beta D_i + \varepsilon_i$ need not coincide with the regression coefficient in $Y_i = \alpha^{OLS} + \beta^{OLS} D_i + U_i$
  - I.e. we may not have $Cov(D_i, \varepsilon_i) = 0$

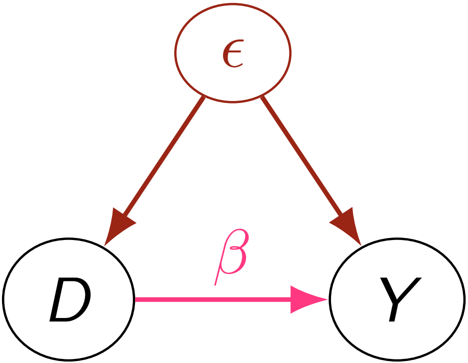
Bias

- The model parameter in $Y_i = \alpha + \beta D_i + \varepsilon_i$ need not coincide with the regression coefficient in $Y_i = \alpha^{OLS} + \beta^{OLS} D_i + U_i$
  - I.e. we may not have $Cov(D_i, \varepsilon_i) = 0$
- Selection bias (a.k.a. omitted variables bias): students with higher latent skill ε_i are more likely to receive treatment D_i
  - $Cov(D_i, \varepsilon_i) > 0$ means $\beta^{OLS} > \beta$: overstate the returns-to-mixtape

Regression $Cov(D_i, \varepsilon_i) = 0$



Regression $Cov(D_i, \varepsilon_i) \neq 0$



Controls

- What if I just add controls?
- Usually it is very hard to make a convincing argument that you controlled for *everything*

Controls

- What if I just add controls?
- Usually it is very hard to make a convincing argument that you controlled for *everything*
- Caveat: This does not mean that running regressions with controls and inspecting heterogenity in $\widehat{\beta}^{OLS}$ isn't interesting
- Just don't claim that $\widehat{\beta}^{OLS}$ represents the causal effect of X on Y (without more work on theory)

Controls

- What if I just add controls?
- Usually it is very hard to make a convincing argument that you controlled for *everything*
- Caveat: This does not mean that running regressions with controls and inspecting heterogenity in $\widehat{\beta}^{OLS}$ isn't interesting
- Just don't claim that $\widehat{\beta}^{OLS}$ represents the causal effect of X on Y (without more work on theory)
- Fixed effects
  - Add intercepts for groups
  - β^{OLS} is then identified by variation within-group
  - Can get you closer to causality (reduces confounding factors that may occur across groups)

Let's Get Our Hands Dirty

- We will use Python in this workshop (sorry? or you're welcome?)
- To make this as painless as possible I have made a ".yml" file that should get us all on the same page quickly by providing a "recipe" that should allow us all to use the same environment (i.e. same installed packages)

Let's Get Our Hands Dirty

- We will use Python in this workshop (sorry? or you're welcome?)
- To make this as painless as possible I have made a ".yml" file that should get us all on the same page quickly by providing a "recipe" that should allow us all to use the same environment (i.e. same installed packages)

Road Map:

1. Pull files down onto your computer using GitHub
2. Use Anaconda to build a virtual environment
3. Open the folder we pulled from GitHub using VS Code
4. Run Jupyter notebooks and have fun!

GitHub

- To easily share code and data I made a GitHub repo
- First, make a GitHub account
- To get the files onto your computer "Clone": `https://github.com/jnaddeo/mdi-workshop-spring-2023`
  1. Download GitHub desktop
  2. Follow link and open with GitHub Desktop Screenshot
  3. Make sure that you are in the main branch and click fetch to make sure you are up to date
  4. Note where the folder is located!

Build Virtual Environment

1. Download Anaconda
2. Install and **add Anaconda to PATH** Screenshot
3. Open terminal (with admin privileges)
4. Type: **conda --version** to test if anaconda is properly installed
5. Navigate to the repo folder and enter:

   conda env create --name mdi_workshop_env_SP23 --file environments/environment_unpinned.yml
6. Download VS code (if ananconda didn't automatically), download python and git extensions in VS Code
7. Use VS code to open the repo folder
8. Open **" mdi-workshop-spring-2023\code\OLS.ipynb"**
9. Make sure to select correct env and run code! Screenshot

Coding Exercise

- Code will help you learn python basics
  - Loading packages
  - Setting up directories that are independent of OS (important!!!)
  - Write functions

Coding Exercise

- Code will help you learn python basics
  - Loading packages
  - Setting up directories that are independent of OS (important!!!)
  - Write functions

Questions:

- What happens to $\widehat{\beta}^{OLS}$ when we introduce a simple omitted variable?

Coding Exercise

- Code will help you learn python basics
  - Loading packages
  - Setting up directories that are independent of OS (important!!!)
  - Write functions

Questions:

- What happens to $\widehat{\beta}^{OLS}$ when we introduce a simple omitted variable?
- Does controlling for this OV help?
  - What if we only have a proxy for this variable?

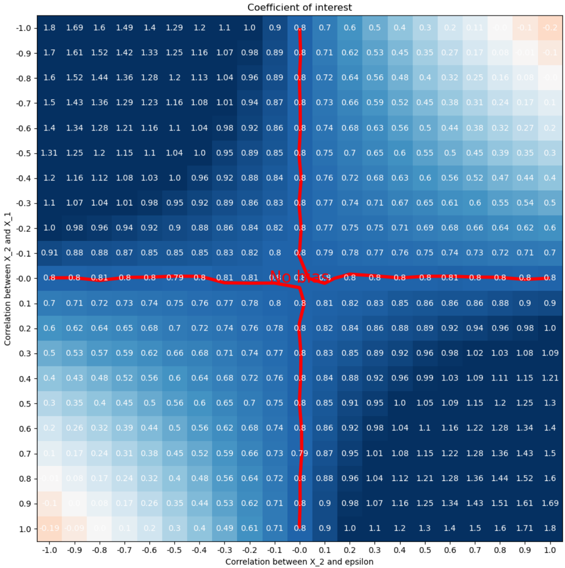
Coding Exercise

- Code will help you learn python basics
  - Loading packages
  - Setting up directories that are independent of OS (important!!!)
  - Write functions

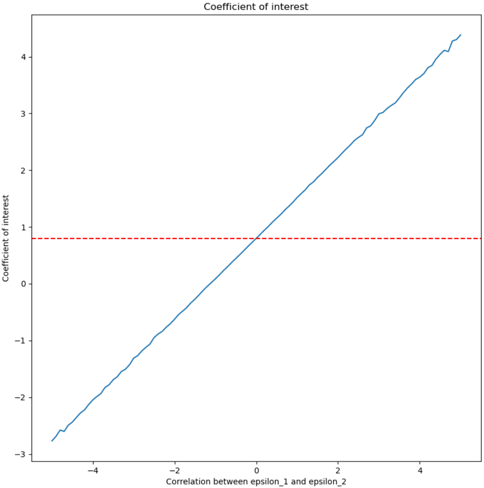
Questions:

- What happens to $\widehat{\beta}^{OLS}$ when we introduce a simple omitted variable?
- Does controlling for this OV help?
  - What if we only have a proxy for this variable?
- Simulate selection bias (a more "real-life" example of OVB)

Results: Simple OVB



Results: Selection Bias



Summary

- So are we just doomed to pray for data from an RCT?

Summary

- So are we just doomed to pray for data from an RCT?
- No! Smart people (nobel prize winners) have developed tools that allow us to get at β of interest

Summary

- So are we just doomed to pray for data from an RCT?
- No! Smart people (nobel prize winners) have developed tools that allow us to get at β of interest
- Use random (or exogenous) variation in another variable to purge the bias from our estimates

Summary

- So are we just doomed to pray for data from an RCT?
- No! Smart people (nobel prize winners) have developed tools that allow us to get at β of interest
- Use random (or exogenous) variation in another variable to purge the bias from our estimates
- Operationally IV estimator using Two-Stage Least Squares

Summary

- So are we just doomed to pray for data from an RCT?
- No! Smart people (nobel prize winners) have developed tools that allow us to get at β of interest
- Use random (or exogenous) variation in another variable to purge the bias from our estimates
- Operationally IV estimator using Two-Stage Least Squares
- 2SLS
  1. $D_i = \mu^{OLS} + \pi^{OLS} Z_i + W_i$ "first stage"
  2. $Y_i = \kappa^{OLS} + \rho^{OLS} Z_i + V_i$ "reduced form"
  3. $\beta^{IV} = \frac{\rho^{OLS}}{\pi^{OLS}}$

Alternative

- $D_i = \mu^{OLS} + \pi^{OLS} Z_i + W_i$ "first stage"
- Use first stage to calculate $\hat{D}_i$

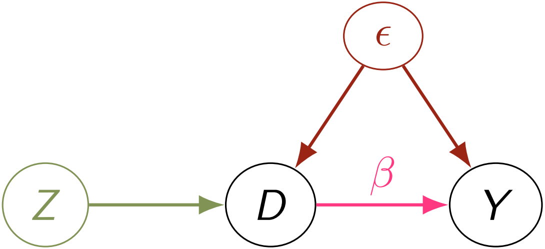
$$\hat{D}_i = \hat{\mu}^{OLS} + \hat{\pi}^{OLS} Z_i$$

- Run second stage to directly get

$$Y_i = \kappa^{IV} + \beta^{IV} \hat{D}_i + V_i$$

- Small note: The standard errors will be wrong if you do this because procedure does not take into account that $\hat{D}_i$ has error in it

The IV Solution



Theory

- Imbens Angrist (1994) set up
  - Binary treatment D
  - Discrete instruments Z
  - No controls
  - Allow for heterogeneous treatment effects (β_i)

Theory

- Imbens Angrist (1994) set up
  - Binary treatment D
  - Discrete instruments Z
  - No controls
  - Allow for heterogeneous treatment effects (β_i)
- Let $Y_i(0)$ and $Y_i(1)$ denote individual i's potential outcomes given a binary treatment $D_i \in \{0, 1\}$
  - Observed outcomes:
    $Y_i = (1 - D_i)Y_i(0) + D_i Y_i(1)$

Theory

- Imbens Angrist (1994) set up
  - Binary treatment D
  - Discrete instruments Z
  - No controls
  - Allow for heterogeneous treatment effects (β_i)
- Let $Y_i(0)$ and $Y_i(1)$ denote individual i's potential outcomes given a binary treatment $D_i \in \{0, 1\}$
  - Observed outcomes:
    $Y_i = (1 - D_i)Y_i(0) + D_i Y_i(1) = \alpha_i + \beta_i D_i$

Theory

- Imbens Angrist (1994) set up
  - Binary treatment D
  - Discrete instruments Z
  - No controls
  - Allow for heterogeneous treatment effects (β_i)
- Let $Y_i(0)$ and $Y_i(1)$ denote individual i's potential outcomes given a binary treatment $D_i \in \{0, 1\}$
  - Observed outcomes:
    $Y_i = (1 - D_i)Y_i(0) + D_i Y_i(1) = \alpha_i + \beta_i D_i$
- Recall that we want to find $\beta_i = Y_i(1) - Y_i(0)$

Theory

- Imbens Angrist (1994) set up
  - Binary treatment D
  - Discrete instruments Z
  - No controls
  - Allow for heterogeneous treatment effects (β_i)
- Let $Y_i(0)$ and $Y_i(1)$ denote individual i's potential outcomes given a binary treatment $D_i \in \{0, 1\}$
  - Observed outcomes: $Y_i = (1 - D_i)Y_i(0) + D_i Y_i(1) = \alpha_i + \beta_i D_i$
- Recall that we want to find $\beta_i = Y_i(1) - Y_i(0)$
- However for each i one is always unobserved

Imbens and Angrist (1994) Assumptions

1. *As-good-as-random assignment*:
   $Z_i \perp (Y_i(0), Y_i(1), D_i(0), D_i(1))$

Imbens and Angrist (1994) Assumptions

1. *As-good-as-random assignment*: $Z_i \perp (Y_i(0), Y_i(1), D_i(0), D_i(1))$
2. *Exclusion*: Z_i only affects Y_i through its effect on D_i

Imbens and Angrist (1994) Assumptions

1. *As-good-as-random assignment*: $Z_i \perp (Y_i(0), Y_i(1), D_i(0), D_i(1))$
2. *Exclusion*: Z_i only affects Y_i through its effect on D_i
3. *Relevance*: Z_i is correlated with D_i

Imbens and Angrist (1994) Assumptions

1. *As-good-as-random assignment*: $Z_i \perp (Y_i(0), Y_i(1), D_i(0), D_i(1))$
2. *Exclusion*: Z_i only affects Y_i through its effect on D_i
3. *Relevance*: Z_i is correlated with D_i
4. *Monotonicity*: $D_i(1) \geq D_i(0)$ for all i (i.e., almost-surely)
   - The instrument can only shift the treatment in one direction
   - Can be relaxed to on average (see de Chaisemartin (2012) and Frandsen et al. (2022))

Theory

- Using those assumptions, Imbens and Angrist show that:

$$\beta^{IV} = \mathbb{E}\big[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)\big]$$

Theory

- Using those assumptions, Imbens and Angrist show that:

$$\beta^{IV} = \mathbb{E}\big[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)\big]$$

- In English: β^{IV} represents the average treatment effect for the subsample of people that were impacted by the instrument

Theory

- Using those assumptions, Imbens and Angrist show that:

$$\beta^{IV} = \mathbb{E}\big[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)\big]$$

- In English: β^{IV} represents the average treatment effect for the subsample of people that were impacted by the instrument
- **LATE: L**ocal **A**verage **T**reatment **E**ffect
- **CACE: C**omplier **A**verage **C**ausal **E**ffect

Theory

- Using those assumptions, Imbens and Angrist show that:

$$\beta^{IV} = \mathbb{E}\big[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)\big]$$

- In English: β^{IV} represents the average treatment effect for the subsample of people that were impacted by the instrument
- **LATE: L**ocal **A**verage **T**reatment **E**ffect
- **CACE: C**omplier **A**verage **C**ausal **E**ffect
- Just remember that different instruments will get you different LATE/CACE if they induce changes in treatment in different subpopulations!

Lab Assignment

- Now that we have gotten far enough into the weeds lets back out and code
- Use VS code to open the repo folder
- Open " **mdi-workshop-spring-2023\code\IV.ipynb**"
- Make sure to select correct env and run code! Screenshot

Lab Assignment

- Now that we have gotten far enough into the weeds lets back out and code
- Use VS code to open the repo folder
- Open " **mdi-workshop-spring-2023\code\IV.ipynb**"
- Make sure to select correct env and run code! Screenshot

Questions:

- Simulate data same way as before when looking at selection bias
- Add a randomized instrument that imapacts if someone receives treatment
- Run IV
  1. By hand (doing both stages seperate)
  2. Using canned package

Empirical Estimation

- Quasi-random assignment of cases$\rightarrow$ prosecutors
  - Within prosecutorial team **X** time **X** broad crime type cells
- Use propensity of non-prosecution (dismissal/diversion) by race by prosecutor as an instrument for non-prosecution

Empirical Estimation

- Quasi-random assignment of cases$\rightarrow$ prosecutors
  - Within prosecutorial team **X** time **X** broad crime type cells
- Use propensity of non-prosecution (dismissal/diversion) by race by prosecutor as an instrument for non-prosecution

Assumptions for IV:

1. Conditional random assignment of instrument Evidence
2. Exclusion: instrument only impacts outcome through effect on decision
3. Relevance: instrument impacts decision (in a few slides)
4. Monotonicity: instrument only shifts treatment in one direction Evidence

Estimating Causal Effect: 2SLS

Instrument (by race):

$$L_{cp} = \left(\frac{1}{n_p - n_p^i}\right)\left(\sum_{c \in n_p} D_{cp} - \sum_{c \in n_p^i} D_{cp}\right) \tag{1}$$

First Stage:

$$D_i = \zeta L_i + \gamma \boldsymbol{X_i} + \tau + \nu_i \tag{2}$$

Reduced Form:

$$F_i = \beta \hat{D}_i + \boldsymbol{\eta} \boldsymbol{X_i} + \tau + \epsilon_i \tag{3}$$

Heterogeneity in Non-prosecution

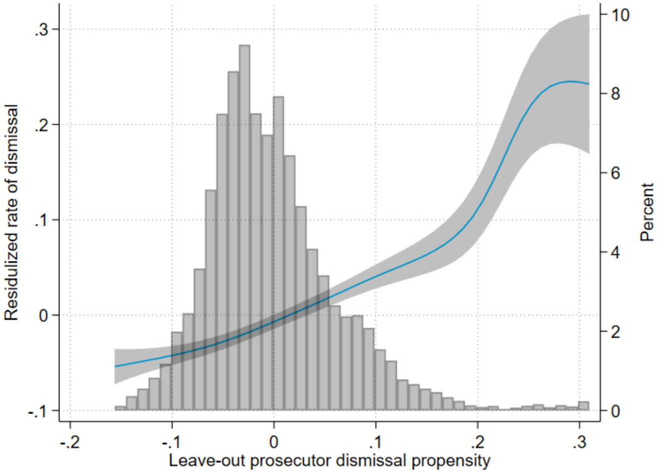
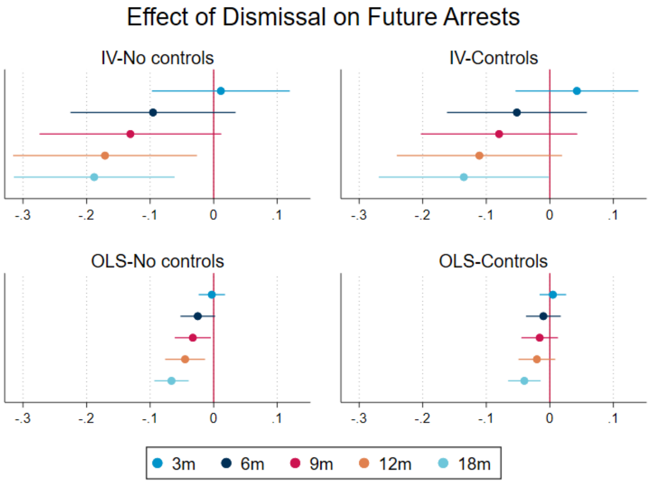
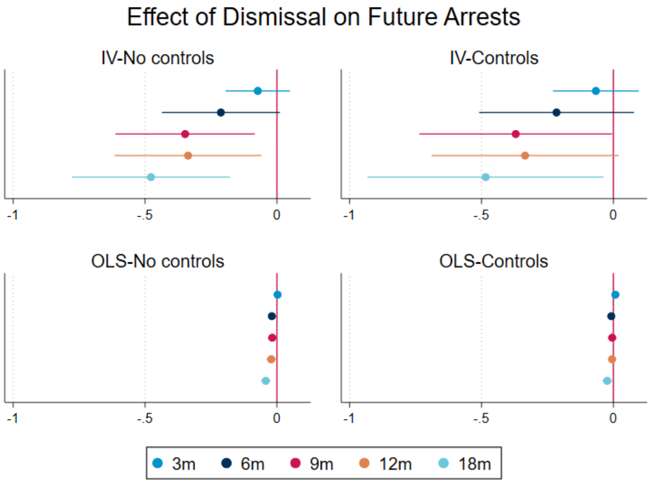


Figure: Representation of first stage in 2SLS. Residualized with respect to team by arrest crime type by time of arrest fixed effects.

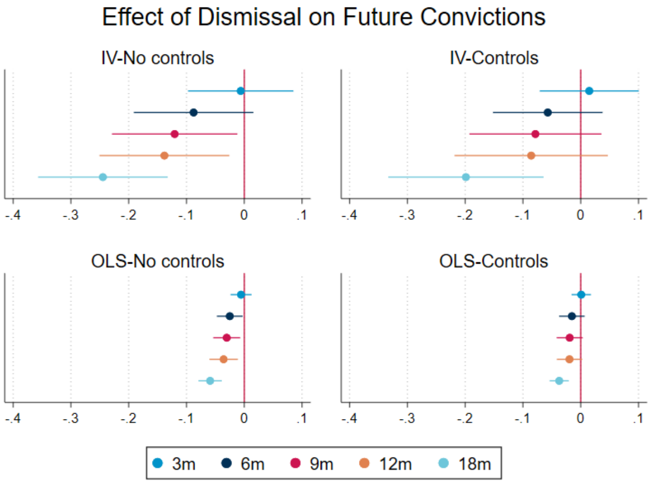
Arrests-White (Δ_A^W)



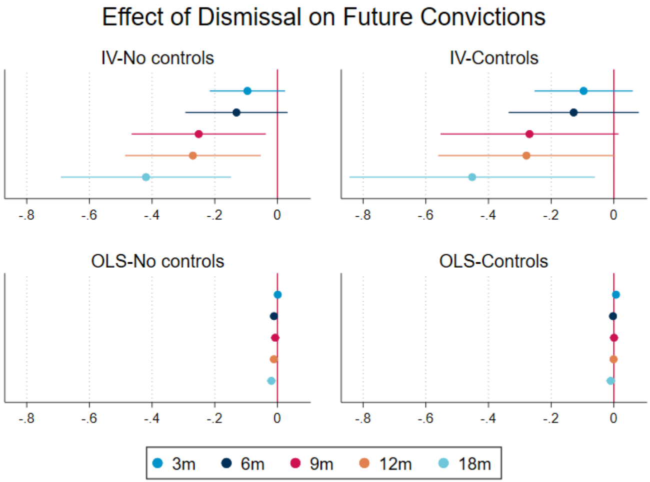
Arrests-Black (Δ_A^B)



Convictions-White (Δ_C^W)

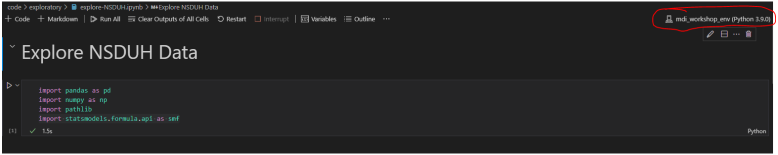


Convictions-Black (Δ_C^B)



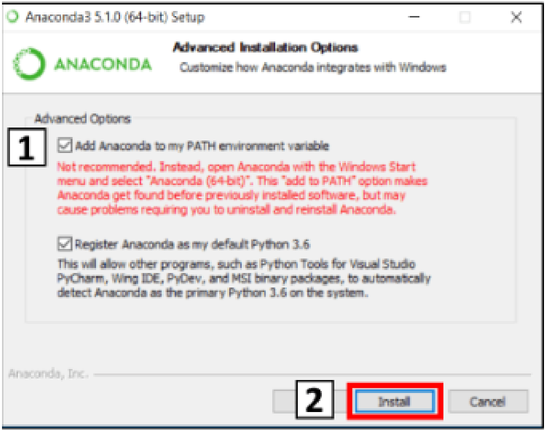
START OF APPENDIX SLIDES

Select Virtual Environment Go Back

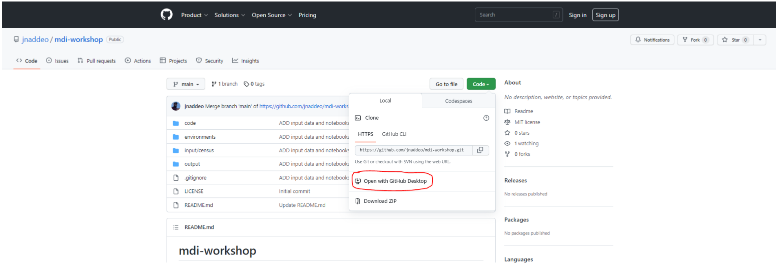


Add conda to PATH

Go Back



Open Repo with Desktop Go Back



Random Assignment Black

Go back

figures/randomization_check_dismissal_black.png

Random Assignment White

Go back

figures/randomization_check_dismissal_white.png

Monotonicity

Go back

- **?** provides joint test for null hypothesis that the exclusion and monotonicity assumptions hold, I fail to reject the null
- Also show assumption can be relaxed to average monotonicity $\implies$ first stage should hold in all subsamples

figures/first_stage_subsamples.png