Review Session

Jonathan Roth

Mathematical Econometrics I
Brown University
Fall 2023

Overview

1. Structure of final and logistics
2. Key course concepts
3. Your questions

Structure of final

- The structure of the final will be similar to the final from last year
- There will be 1-2 analytical questions that will ask you to use mathematical tools we've learned in this course
- There will be questions about a new empirical application. You will be required to state and evaluate assumptions under which we can learn about causal effects, and comment on empirical results.

Logistics for final

- The final exam will be in-person and 2 hours long
  - Please email me and your TAs if you have a SAS accommodation for extra time
- The final exam will be December 15 at 2pm in MacMillan Hall 117
- You may bring 2 8.5x11" pieces of paper with notes
  - E.g. your "cheatsheet" for the midterm plus a new one

Overview of course

- In this class, we focused primarily on how we can answer causal economic questions with data
  - What is the effect of going to Brown versus URI on earnings?
  - What is the effect of gaining health insurance on depression?
  - What is the effect of the minimum wage on employment?
- We formalized the idea of a causal effect with *potential outcomes*
  - Each unit has a potential outcome under both treatment and control, $Y_i(1)$ and $Y_i(0)$
  - E.g. $Y_i(1) =$ earnings if go to Brown; $Y_i(0) =$ earnings if go to URI
  - We observe only $Y_i(1)$ for treated units and $Y_i(0)$ for control units
  - We are interested in the causal effect $Y_i(1) - Y_i(0)$ (or averages of this over the population)

Two key challenges

There are two key challenges in answering causal questions with data:

- We never observe the counterfactual outcome for each unit
  - E.g., we observe earnings for Brown students ($Y_i(1)$), but not their earnings if they'd gone to URI ($Y_i(0)$)
- We typically only observe data for a sample rather than for the full population that we care about
  - E.g., we only have data from a survey of a small fraction of recent graduates

Identification versus Statistical Inference

We typically tackle these two problems separately:

- **Identification:** what could we learn about the causal effect if we had the observable data from the full population?
  - Typically start with some assumptions about how treatment is assigned
  - Under these assumptions, show that this causal effect is a function of observable population means
- **Statistical Inference:** what can we learn about the observable features of the population given the sample?
  - Typically estimate population means by plugging in sample means
  - When we need to estimate conditional means, we use regression to approximation the CEF
  - We can use statistical tools to test hypotheses and construct confidence intervals

5 Different Types of Identification Arguments

- **Experiments:** if treatment is randomized, can compare outcomes for treated pop to control pop
- **Conditional unconfoundedness:** assume treatment assignment is like an experiment conditional on observable characteristics — compare treated/control populations with the same covariates
- **Difference-in-differences:** allow for there to be selection into treatment, but assume selection bias is constant over time — compare differences after treatment occurs to differences before treatment
- **Instrumental variables:** assume that assignment of an instrument is random and affects outcome only through treatment — compare populations with different values of the instrument
- **Regression discontinuity:** assume that confounding factors evolve continuously around the cutoff — compare population with scores just below the cutoff to just above

Experiments

Sample means of depression

	Control Group	Treated Group
Mean	0.329	0.306
SD	0.470	0.461
N	10426	13315

- Identification:
  $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0)) \implies ATE = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$

  Make sure you know how to derive this!

- Estimation: replace population means with sample means!

- Estimation 2: can also estimate with OLS

$$Y_i = \alpha + \beta D_i + \varepsilon_i$$

Conditional unconfoundedness

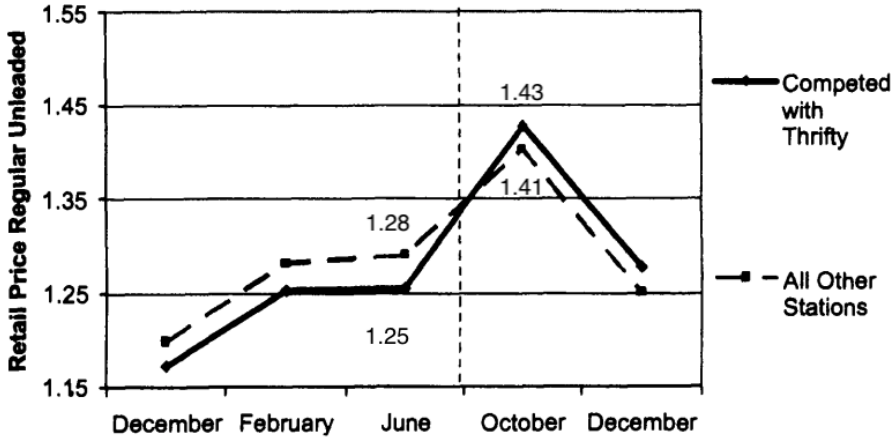
- Example: maybe where you go to college is as-good-as-random conditional on where you get in (Dale & Krueger)
- Identification: $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0))|X_i \implies CATE(x) = E[Y_i|D_i = 1, X_i = x] - E[Y_i|D_i = 0, X_i = x]$
- Estimation: typically, we need to approximate the CEF $\rightarrow$ use OLS!
- Common to approximate ATE with OLS estimates of

$$Y_i = \alpha + \beta D_i + \boldsymbol{\gamma}' \boldsymbol{X_i} + \varepsilon_i$$

- This works well if (i) conditional unconfoundedness holds, and (ii) the CEF is approximately linear: $E[Y_i|D_i, \boldsymbol{X}_i] \approx \alpha + \beta D_i + \boldsymbol{\gamma}' \boldsymbol{X_i}$

	Basic model: no selection controls		Matched-applicant model
	Full sample	Restricted sample	Similar school-SAT matches*
Variable	1	2	3
School-average SAT score/100	0.076	0.082	−0.016
	(0.016)	(0.014)	(0.022)
Predicted log(parental income)	0.187	0.190	0.163
	(0.024)	(0.033)	(0.033)
Own SAT score/100	0.018	0.006	−0.011
	(0.006)	(0.007)	(0.007)
Female	−0.403	−0.410	−0.395
	(0.015)	(0.018)	(0.024)
Black	−0.023	−0.026	−0.057
	(0.035)	(0.053)	(0.053)
Hispanic	0.015	0.070	0.020
	(0.052)	(0.076)	(0.099)
Asian	0.173	0.245	0.241
	(0.036)	(0.054)	(0.064)
Other/missing race	−0.188	−0.048	0.060
	(0.119)	(0.143)	(0.180)
High school top 10 percent	0.061	0.091	0.079
	(0.018)	(0.022)	(0.026)
High school rank missing	0.001	0.040	0.016
	(0.024)	(0.026)	(0.038)
Athlete	0.102	0.088	0.104
	(0.025)	(0.030)	(0.039)

Difference-in-differences I



(a) Los Angeles

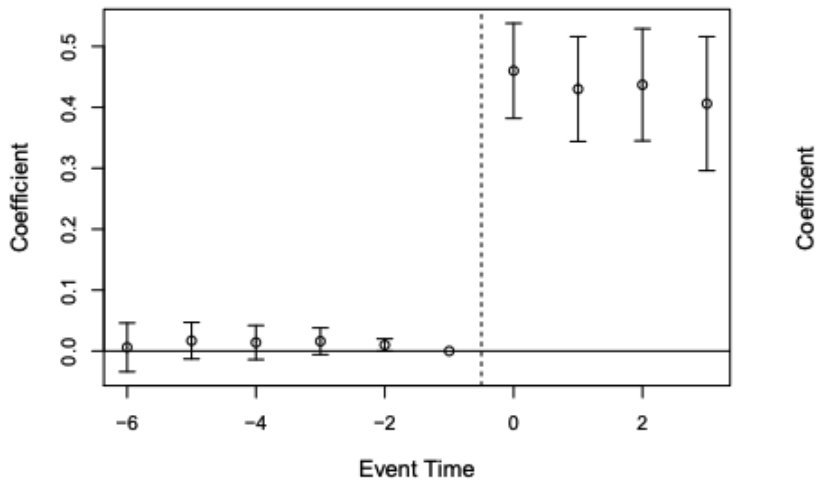
- Key identification assumption: parallel trends — selection bias is constant over time (make sure you know the formal definition)
- Identification: under parallel trends (and no anticipation),

$$\tau_{ATT} = \underbrace{(\mu_{12} - \mu_{11})}_{\text{Change for treated pop}} - \underbrace{(\mu_{02} - \mu_{01})}_{\text{Change for control pop}}$$

- Estimation: plug in sample means instead of population means!
- Estimation 2: can also estimate with OLS (make sure you know how to do this!)

Difference-in-differences II

- It is common to assess plausibility of DiD assumption by looking at pre-treatment trends with an "event-study"



(a) Medicaid Eligibility

- We gain confidence in the research design if (i) pre-trends close to 0, and (ii) can't draw a straight line through all the CIs (there is a break from trend!)

- Estimation of the event-study can be done via OLS (make sure you know how!)

IV

PANEL A: WALD ESTIMATES FOR 1970 CENSUS—MEN BORN 192

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year
ln (wkly. wage)	5.1484	5.1574
Education	11.3996	11.5252

- In IV, we have an instrument that is as-good-as-randomly assigned and affects the outcome only through its effect on the treatment
- Four key identifying assumptions (make sure you understand them!):
  - Independence: instrument is as good as randomly assigned
  - Exclusion: instrument affects outcome only through treatment
  - Monotonicity: no defiers
  - Relevance: instrument affects treatment status
- Under the four key assumptions, $\beta_{IV} = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$ identifies a LATE (average treatment effect for compliers)

IV - Estimation

PANEL A: WALD ESTIMATES FOR 1970 CENSUS—MEN BORN 192

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year
ln (wkly. wage)	5.1484	5.1574
Education	11.3996	11.5252

- Under the four key assumptions, $\beta_{IV} = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$ identifies a LATE (average treatment effect for compliers)
- Estimation is done by plugging in sample analogs for reduced form and first-stage, then taking the ratio
- Estimation 2: can also be done by "two-stage least squares", which allows us to incorporate multiple instruments (make sure you know how)
  - Regress D_i on instrument Z_i (and controls)
  - Regress Y_i on predictions $\hat{D}_i$ (and controls)

RDD

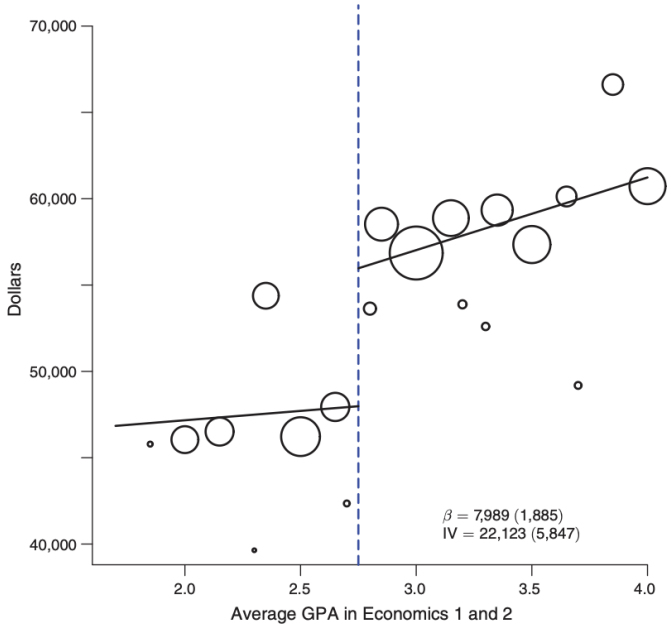


FIGURE 2. THE EFFECT OF THE UCSC ECONOMICS GPA THRESHOLD ON ANNUAL WAGES

- In RDD, we compare people just above/below a threshold that (partially) determines treatment status
- Identification: potential outcomes are continuous at the cutoff
  - Make sure you understand when this might fail!
- Estimation: estimate CEF at the cutoff using OLS or local linear regression
  - Note: I will not test you on the details of local linear regression

Concluding thoughts I

We've done a lot in one semester!

- Learned about the challenges of estimating causal effects
- Developed statistical language to help us think about when/how we can learn about causal effects
- Learned several applicable "identification strategies" for learning about causal effects
- Developed tools for estimating and testing hypotheses about causal effects in finite samples (often using regression!)

There's still a lot more to learn, if you're interested in taking more classes!

- Non-parametrics, machine learning approaches, Bayesian econometrics, time series and forecasting, etc.

Concluding thoughts II

There are many different ways that I hope you can apply these tools going forward:

- Do research that helps to improve policies or firm decisions, thereby increasing social welfare (or corporate profits - your choice!)
  - If you're interested in academic research, I encourage you to work as a research assistant for profs at Brown and/or write an honors thesis
- Better understand empirical evidence as you read articles in the newspaper, online, etc.
- Become an econometrician and help to develop tools for getter policy analysis in the future :-)

Concluding thoughts III

- I encourage you to fill out the course evaluation/feedback form: https://brown.evaluationkit.com
- Course feedback will be used both for evaluation purposes and for trying to improve the course!
- This is my third time teaching this course, so any comments on what worked or could have been done better are much appreaciated. Thanks!


Good luck on the final!