Semester Review, Questions, Exam Practice

Econ 140, Section 12

Jonathan Old

Roadmap

1. Regression Discontinuity
2. Broad topics for the exam
3. IV Exam Exercise
4. Other exam practice

Course evaluation

Regression Discontinuity

The world is full of arbitrary rules

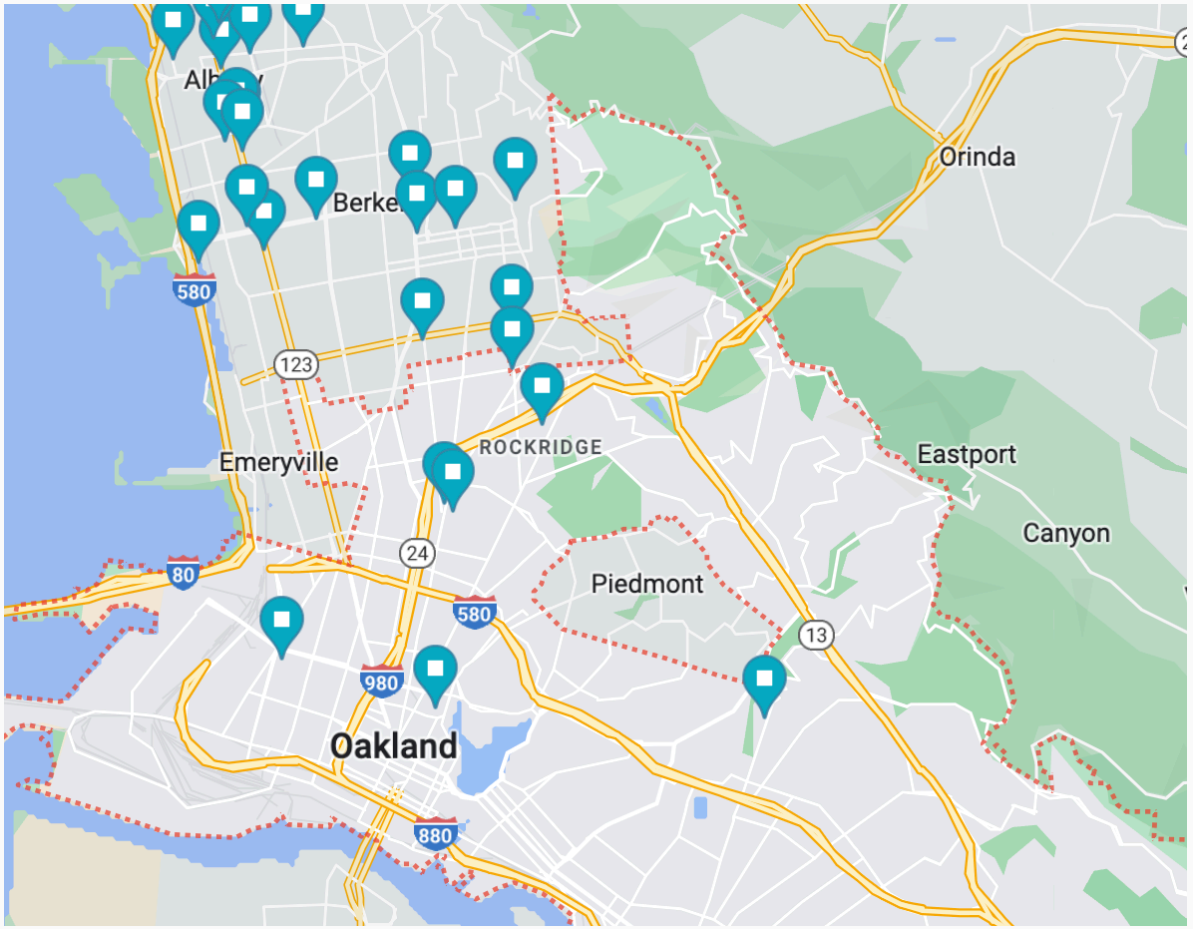
- Students receive a scholarship if their GPA is above 3.0
- Children are allowed to start school if they are five years old before January 31 of that year
- Individuals are eligible for a microfinance loan if they own less than 0.5 acres of land
- Legislators are elected if they receive over 50% of the vote

This creates wonderful discontinuities in the data that we can exploit for econometric analysis!

The most important rules of regression discontinuity

- The essence of RD is to compare people (units) just above the cut-off to others just below
- Crucial assumptions: There are no jumps but our jump at the jump // The effect of X (our running variable) and any other variable C on Y (our outcome) is smooth around the discontinuity
- Identifying assumption: In the absence of the ”treatment” (which was allocated by a discontinuous rule), the outcomes of the treated would have been essentially the same as the outcomes of the untreated
- Limitations:
  1. We only estimate a **local** effect!
  2. There may be strategic behavior at the cutoff (example: school district Oakland/Piedmont)

Oakland and Piedmont



Implementing regression discontinuity designs

With a running variable X, an outcome Y, and a treatment D allocated at the discontinuity, we estimate:

$$Y_i = \alpha + f(X_i) + \beta D_i + \varepsilon_i$$

- $f(X_i)$ is a smooth function of the running variable - e.g. just linear, or quadratic
- We could also estimate whether the **slope** changes at the discontinuity
- We are eventually interested in β: Does the outcome change at the discontinuity?

Broad topics for the exam

Introduction

- Interpreting regression tables
- p-values: How surprised we should be to observe the world as it is if your hypothesis about how it works were true?
- Omitted variable bias
- Interpretation of control variables and their role
- Interaction terms (dummy or continuous) – dummyXdummy: get means for every group

Instrumental variables

- Calculation
- Three assumptions: Relevance, Independence (Exogeneity), Exclusion restriction
- Think of three regressions: First stage, reduced form, ”target regression” (Second stage)
- Connection to RCTs (imperfect compliance)

Differences in Differences

- Parallel trends assumption
- Calculate with tables
- Calculate using regression tables

Panel Data and Fixed effects

- Fixed effects: Calculation
- Useful way to think about it: We add a lot of controls, but key problem remains (time-varying confounders: Things that change over time that affect different units differently, e.g.: climate change ...)
- Connection of two-way fixed effects with DiD

Regression Discontinuities

- ”No jump but our jump at the jump”
- Fuzzy DiD is just IV

General and overarching concepts

- Comparison of OLS and other estimates: What changed and why? (Hint: Always go back to OVB)
- Be creative but also clear: Most thinking questions can be solved with reasonable economic intuition, your knowledge of the world, and what you learned on the course.
- Keep in mind the big picture: We are just trying to find a valid counterfactual, to estimate causal effects!

IV Exam Exercise

IV Exam exercise i

The land of Econometrea consists of 34 islands. Each island has their own university and students attend uni on their home island. Students in Econometrea either have to attend lectures in person or watch them online. Priyanka has obtained data for the students from all the universities and would like to study the effect of watching lectures online on the students' exam scores. For each student she has the exam score (0 - 100, 70+ is a first, 40 and below a fail), which she uses as her left hand side variable, the fraction of lectures watched online, and how on how many days the student visited the library each week. Priyanka obtains the following regression results:

IV Exam exercise ii

	Estimation Method				
	OLS	IV	IV	IV	IV
Regressor	(1)	(2)	(3)	(4)	(5)
Fraction online	-4.91	-2.08	-0.56	-0.75	6.09
	(0.11)	(0.22)	(0.40)	(0.90)	(3.22)
Library visits			0.91		2.09
			(0.08)		(0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each but one of the universities. Robust standard errors are reported in parentheses.

IV Exam exercise iii

(a) What is the interpretation of the coefficient in column (1). If this were a causal effect, would it be big or small? Explain whether this estimate is likely to have a causal interpretation.
(b) Your friend realizes that you also have data from before Covid, when all students had to attend all lectures in person. She suggests to construct a panel dataset and to run a regression and to include time and individual fixed effects. What assumption would be required for this regression to deliver a causal effect?

IV Exam exercise iv

Solution

- Students who watch all lectures online, on average, have around 4.9 fewer points on the exam than students who went to all lecturs in person. This is a moderately strong effect, but likely not causal (think of omitted variables, such as student motivation)

IV Exam exercise v

Solution

- For this to work, we need the parallel assumption to hold:
  In the absence of Covid and the introduction of the online policy, individuals with more online attendance would have had the same change in exam scores as individuals with less online attendance.
  In this case, this might not be very plausible, because Covid probably affected individuals who are likely to stay at home differently than individuals who attend class in person.

IV Exam exercise vi

(c) Wei Min observes that some halls of residence are close to the lecture rooms while others are further away and suggests to use the distance of halls from lecture halls as an instrumental variable for the fraction of lectures watched online. Results are in column (2). Explain which assumptions need to be satisfied for the IV to deliver better causal estimates than OLS here. Discuss the validity of the assumptions in this case.

IV Exam exercise vii

	Estimation Method				
	OLS	IV	IV	IV	IV
Regressor	(1)	(2)	(3)	(4)	(5)
Fraction online	−4.91	−2.08	−0.56	−0.75	6.09
	(0.11)	(0.22)	(0.40)	(0.90)	(3.22)
Library visits			0.91		2.09
			(0.08)		(0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each but one of the universities. Robust standard errors are reported in parentheses.

IV Exam exercise viii

Solution

- **Relevance:** Proximity to lecture halls must correlate strongly with lecture attendance. This is plausible from our experience, and we could test this using an F-test.
- **Independence/Exogeneity:** Proximity must be as good as randomly allocated. This might not hold here, for example, if more motivated students live closer to campus.
- **Exclusion restruction:** Proximity to lecture halls only influences exam scores because it has an effect on lecture attendance. This can be violated if proximity also affects the probability to go and study in the library, or if residence halls closer to campus are also closer to bars

IV Exam exercise ix

(d) Alma points out to Wei Min that students can pick which hall they want to live in and that some halls are located in Study Village, close to lecture halls and the library while others are located further away in Party Town, surrounded by pubs. How does this information affect your assessment of the IV strategy?
(e) Manuel realises that the data also include a variable for the number of times a student has checked into the library per week. He suggests to rerun the instrumental variables regression adding this variable as a control. Results for this regression are displayed in column (3). Assess Manuel's strategy.

IV Exam exercise x

Solution

- This would be a violation of independence/exogeneity: The students who live closer to the campus are just inherently different than students who live far away
- Library visits are a bad control: They are another outcome of the instrument, and so controlling for them introduces additional bias.

IV Exam exercise xi

	Estimation Method				
	OLS	IV	IV	IV	IV
Regressor	(1)	(2)	(3)	(4)	(5)
Fraction online	−4.91	−2.08	−0.56	−0.75	6.09
	(0.11)	(0.22)	(0.40)	(0.90)	(3.22)
Library visits			0.91		2.09
			(0.08)		(0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each but one of the universities. Robust standard errors are reported in parentheses.

IV Exam exercise xii

(f) Jenny notices that there are five universities which assign students to their halls of residence by a lottery. She suggests to run the IV model from columns (2) and (3) for the subsample of students from these universities only. Results are displayed in columns (4) and (5). Assess Jenny's regressions.
(g) Drawing on the results in the table above, what have you learned from this exercise about the causal effect of watching lectures online on students' exam results?

IV Exam exercise xiii

Solution

- For these universities, the distance is as good as randomly assigned, and so we have no violations of the independence/exogeneity restriction. However, the exclusion restriction may still not hold.
- In a simply OLS regression, it may look as if watching lectures online causes worse grades. However, we can correct for this using a valid IV, and find that the effect is only very small and statistically insignificant.

Other exam practice

RDD question

Question 2

The local government wants to estimate the impact on future earnings of a job training program that it operated in 2016 and 2017. Access to the program is governed by an eligibility rule: only individuals whose income in the prior tax year was less than £12,000 can participate.

A. **(20 marks)** Explain how you could use this eligibility rule to estimate the causal effect of the program. Describe any data you would need, write down the regression equation(s) you would estimate, define all variables precisely, and explain how you would interpret the regression results. If you make any additional assumptions, state them clearly.

B. **(10 marks)** A colleague worries that because the eligibility rule was public, your estimates of the program's causal effect may be biased. Why might this pose a problem for identification? How could you use the data to assess the validity of this concern?

DiD question

Question 3

(10 marks) On 28 August 2017, a new state law forced the city of St. Louis, Missouri, United States to reduce its minimum wage from $10 per hour to $7.70 per hour. You have monthly employment data for at the municipal level for the entire state in the months of May 2017 and November 2017. How would you construct the difference-in-differences estimate for the effect of reducing the minimum wage? What assumption would have to hold for this to be a valid causal estimate of the effect? Precisely describe one concern that you may have with the validity of this assumption. Describe two things you could do to assess the validity of this assumption.