ECON 3510: Poverty and Economic Development
Lecture 7: Regression Discontinuity

Instructor: Weizheng Lai

Bowdoin College

Fall 2025

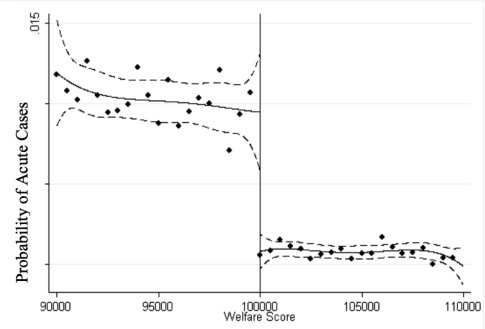
Motivation

- The key challenge in causal inference is finding a control group that is comparable to treated units in all ways except for treatment status.
- Today we will see a new strategy: **regression discontinuity (RD)**.
- The idea behind RD is that sometimes treatment status is (partially) determined by whether a score is above/below a threshold.
  - For example, whether you are admitted to a school or university may depend on whether your test score exceeds a certain cutoff score.
- In such cases, RD compares outcomes for people with scores just above the threshold to people with scores just below the threshold.

Example - Hou and Chao (2008)

- Hou and Chao (2008) are interested in the question: does lack of health insurance prevent poor people from getting important medical care?
- Why don't they just compare medical care utilization between people with and without health insurance?
  - Confounding variables! People with health insurance may be richer, which also may affect health directly
- Context: the Republic of Georgia (not the state!) created a health insurance program in 2006. Each household received a "poverty score" derived from 80 household variables, and households with a score $\leq$100,000 received health insurance
- Key idea: people with scores just above 100,000 may be very similar to people with scores just below 100,000
  - Except those below 100,000 had the health insurance treatment!

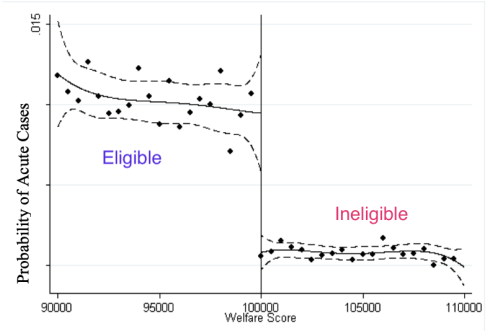
Figure 3: The Effect of MAP on Utilization of Acute Surgeries/In-Patient Services



Note: This figure plots probability of utilization of acute surgeries/inpatient services against welfare scores. Each dot is the average probability within 500 intervals of welfare scores. Solid lines are fitted values from 4^{th} order polynomial regressions on either side of the discontinuity. Dotted lines are 95% confidence intervals.

- This plot shows outcomes (acute surgeries) as a function of the score

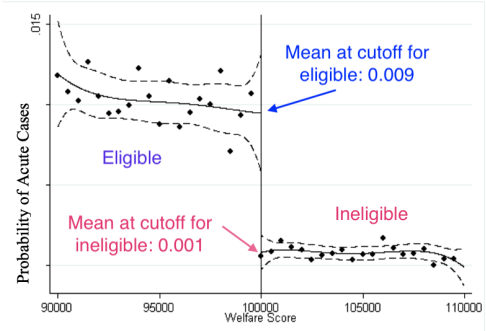
Figure 3: The Effect of MAP on Utilization of Acute Surgeries/In-Patient Services



Note: This figure plots probability of utilization of acute surgeries/inpatient services against welfare scores. Each dot is the average probability within 500 intervals of welfare scores. Solid lines are fitted values from 4th order polynomial regressions on either side of the discontinuity. Dotted lines are 95% confidence intervals.

- People to the left of the vertical line (at 100,000) receive health insurance from the government
- We might expect people with scores just below 100,000 to be very similar to people just above 100,000 on all factors other than healthcare eligibility

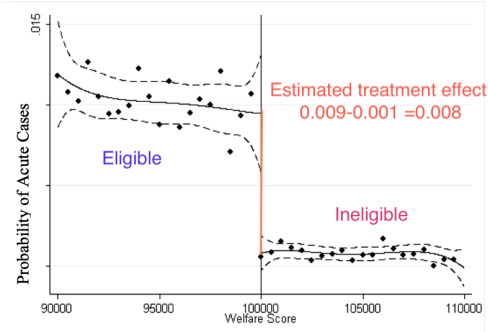
Figure 3: The Effect of MAP on Utilization of Acute Surgeries/In-Patient Services



Note: This figure plots probability of utilization of acute surgeries/inpatient services against welfare scores. Each dot is the average probability within 500 intervals of welfare scores. Solid lines are fitted values from 4th order polynomial regressions on either side of the discontinuity. Dotted lines are 95% confidence intervals.

- We might expect people with scores just below 100,000 to be very similar to people just above 100,000 on all factors other than healthcare eligibility
- But people just below the threshold seem to get a lot more surgeries!

Figure 3: The Effect of MAP on Utilization of Acute Surgeries/In-Patient Services



Note: This figure plots probability of utilization of acute surgeries/inpatient services against welfare scores. Each dot is the average probability within 500 intervals of welfare scores. Solid lines are fitted values from 4^{th} order polynomial regressions on either side of the discontinuity. Dotted lines are 95% confidence intervals.

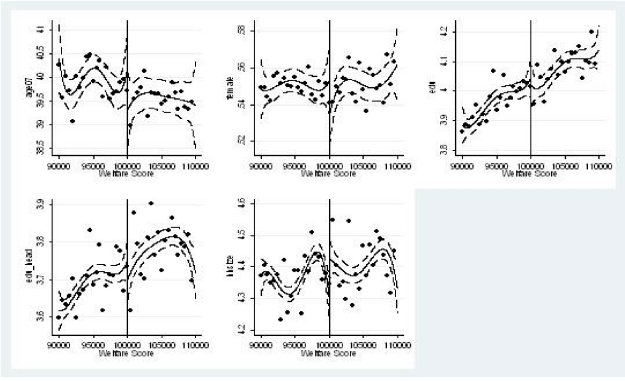
- But people just below the threshold seem to get a lot more surgeries!
- If everything else is continuous at the threshold, the difference is the causal effect (for people at the threshold)!

Evaluating Continuity

- Why might the continuity assumption be violated?
- #1: confounding factors change discontinuously at the cutoff
- #2: people can manipulate scores to get just above/below the cutoff

- To partially address these types of concerns, it is common to show that observable features don't vary discontinuously at the cutoff

- The figure below shows that there don't appear to be any discontinuities in age, sex, education, and household size in the Hou and Chao paper



Age, sex, education, education of head and household size

- But as usual, there may still be concern about other **unobserved confounding factors** varying at the cutoff

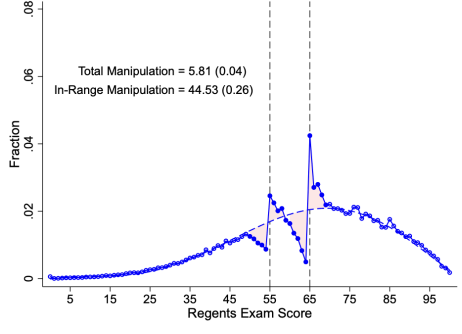
Testing for Manipulation

- To test for such RD manipulation, it is common to check whether there are a similar number of units on both sides of the cutoff.
- If there is bunching on one side of the cutoff, this is typically interpreted as evidence of manipulation.
- The continuity assumption will usually be much more questionable if there is bunching.
- Plot the density/histogram to judge!

Example - Manipulation

- In NYC, students must take the Regents exams and get a score of at least 55 to get a diploma (and 65 to get a more prestigious diploma)
- Might be tempted to use this to study diploma effects... But it turns out there are way more students with scores just above the thresholds

Figure 1
Test Score Distributions for Core Regents Exams, 2004-2010



- Likely reason: teachers cheat to bump students over the threshold

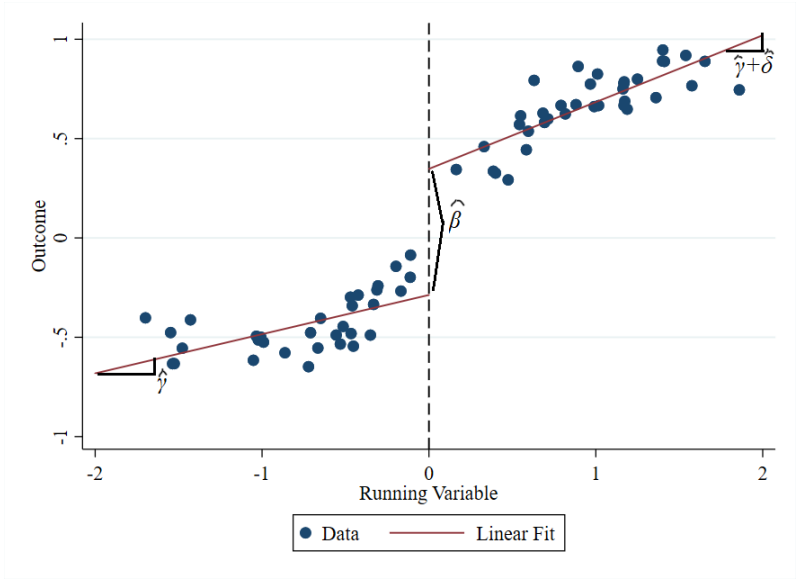
Doing RD in One Regression

- Suppose a cutoff $c = 0$ (without loss of generality).
- Consider the regression:

$$Y_i = \alpha + \beta 1[R_i \geq 0] + \gamma R_i + \delta R_i 1[R_i \geq 0] + U_i$$

- Then

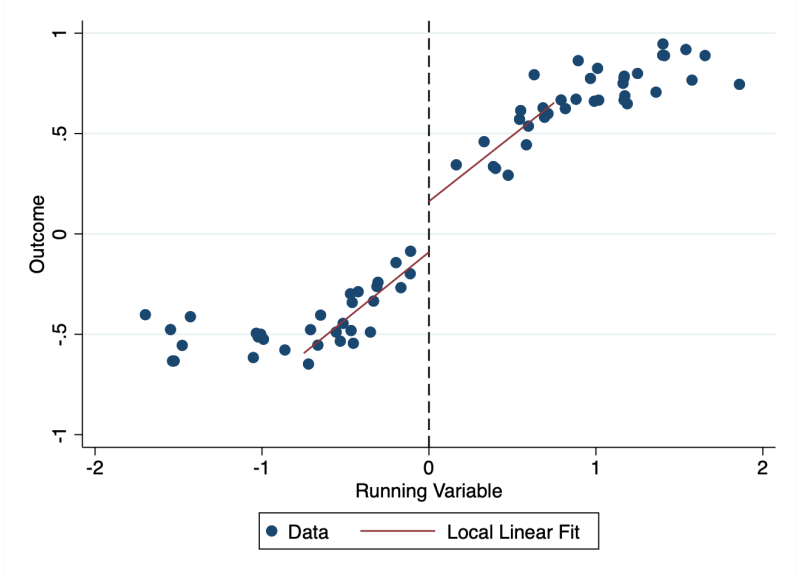
$$\begin{aligned} &\text{if } R_i < 0, && Y_i = \alpha + \gamma R_i + U_i \\ &\text{if } R_i \geq 0, && Y_i = \alpha + \beta + \gamma R_i + \delta R_i + U_i \end{aligned}$$


Outcome
1
.5
0
-.5
-1
-2
-1
0
1
2
Running Variable
$\hat{\gamma}$
$\hat{\beta}$
$\hat{\gamma}+\hat{\delta}$
Data
Linear Fit

What to Do in Practice?

- In practice, the usual approach is to use **local linear regression**
- The basic idea is to fit a linear regression but to only use points that are “close” to the boundary

Local Linear Regression



RD Estimation is Tricky!

- Good to trust your eyes – does it look like there's a discontinuity on the plot?
- The most convincing RDs are obvious from the plot and don't need any fancy econometrics.

Fuzzy RD

- Sometimes crossing a threshold doesn't completely determine treatment status, but discontinuously increases **treatment probability**
- This is called a *fuzzy* RD
- The basic idea is similar to IV — we estimate the effect of being above the threshold on the outcome, then divide this by the effect on the treatment (i.e. the change treatment probability)

Example - Bleemer and Mehta (2022)

- Bleemer and Mehta ask a very important question: does majoring in economics make you rich?
- They study this Q in the context of UC Santa Cruz (UCSC), where the econ department only allowed people with GPA below 2.8 in intro classes to major in econ “at the discretion of the department”

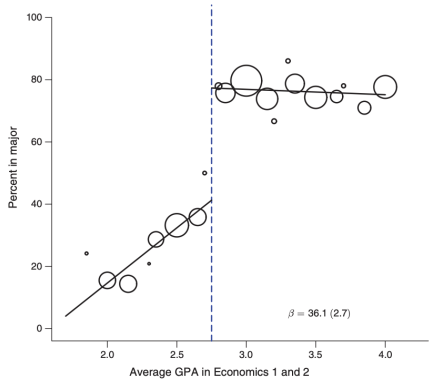


FIGURE 1. THE EFFECT OF THE UCSC ECONOMICS GPA THRESHOLD ON MAJORING IN ECONOMICS

Notes: Each circle represents the percent of economics majors (y-axis) among 2008–2012 UCSC students who earned a given *EGPA* in Economics 1 and 2 (x-axis). The size of each circle corresponds to the proportion of students who earned that *EGPA*. *EGPA*s below 1.8 are omitted, leaving 2,839 students in the sample. Fit lines and beta estimate (at the 2.8 GPA threshold) from linear RD specification; standard error (clustered by *EGPA*) in parentheses.

Students above the threshold about 36 pp more likely to major in econ

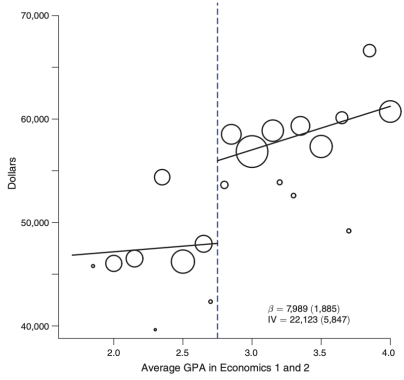


FIGURE 2. THE EFFECT OF THE UCSC ECONOMICS GPA THRESHOLD ON ANNUAL WAGES

- Students about the threshold earn about \$8K more
- The fuzzy RD estimate of the effect of majoring in econ is then $\$8K/0.36 \approx 22K$, or 40% of mean earnings

References I

Bleemer, Zachary and Aashish Mehta (2022). "Will studying economics make you rich? A regression discontinuity analysis of the returns to college major". *American Economic Journal: Applied Economics* 14.2, pp. 1–22.

Hou, Xiaohui and Shiyan Chao (2008). "An evaluation of the initial impact of the medical assistance program for the poor in Georgia". *World Bank Policy Research Working Paper* 4588.