

Regression Discontinuity

Jeremy Magruder

UC-Berkeley

May 2009

Evaluating Programs with Rules

- Many Development programs have strict eligibility criteris
 - In South Africa, Gov't Pensions given to women who are at least 60, but not 59
 - Grameen bank requires less than 1/2 acre of land to be eligible
 - In China, counties with less than a certain amount of income are eligible for rural development assistance
- This means two things:
 - Program recipients are systematically different from non-recipients
 - However, for people/counties who are near the threshold, differences are slight.

Case Study: RD in China

- The 8-7 plan is an effort of the Chinese government to reduce rural poverty
- Eligible counties receive three things:
 - subsidized loans for rural enterprise
 - budgetary funds
 - public works projects
- counties were designated eligible if, in 1992, had <400 yuan per capita income
- Program accounts for about 6% of govt spending – is it effective?

How to evaluate the 8-7 plan?

- Clearly, program is not randomly placed.
- Could, in principle, do a difference-in-differences:

$$\bar{Y}_{2004}^{8-7} - \bar{Y}_{1994}^{8-7} - \left(\bar{Y}_{2004}^{No8-7} - \bar{Y}_{1994}^{No8-7} \right)$$

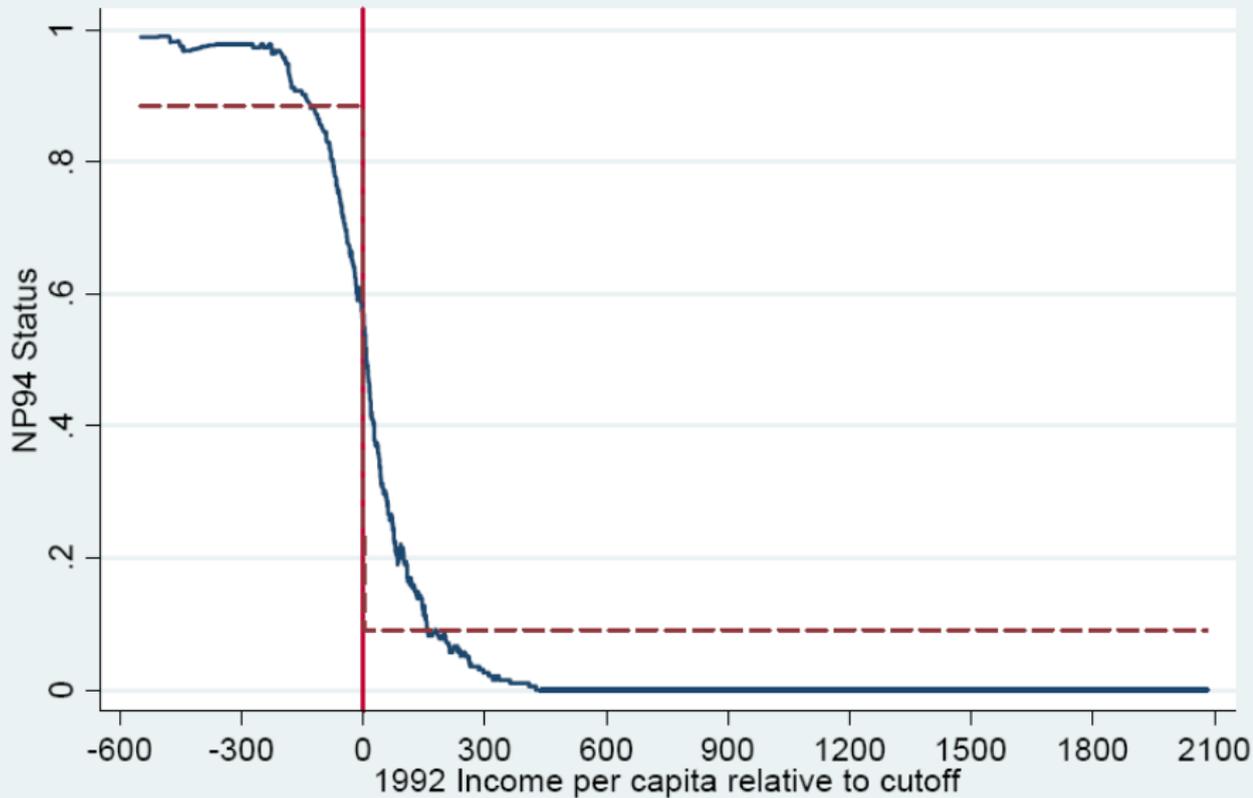
- But, worry about common trends assumption
- Eligible counties are very different from ineligible ones (intentionally so, in fact)
- Some eligible counties are very poor, some ineligible very rich. Different geographies, local government, labor market trends...

Difference in differences results

Post*Treatment	-114.57** (45.77)	9.4 (55.26)	146.45*** (52.15)
County Level Controls	No	Yes	Yes
Provincial dummies	No	No	Yes
R-square	0.004	0.09	0.38
Sample Size	1305	1268	1268

Regression Discontinuity

- Can use the fact that counties which are just barely eligible are probably similar to those which are just barely ineligible, though
- A county with a GDP of 395 Yuan/capita would be eligible; one with 405 Yuan/capita would not
- However, their total income is pretty similar – similar to a propensity-score match, we might think that in general these counties would have similar baseline patterns were it not for this difference in program eligibility.
- We might worry about whether the criteria was really enforced or not. If this will be a successful RD, should be able to see difference in eligibility graphically



— Nonparametric Prediction - - - OLS Prediction

Regression Discontinuity (2)

Formally, suppose we regress

$$Y_{it} = \beta_1 T_i + \beta_2 Post_t + \beta_3 T_i * Post_t + \varepsilon_{it}$$

This is the difference-in-difference specification.

The concern is that $\text{corr}(P_i * Post_t, \varepsilon_{it}) \neq 0$ because different trends are happening in treatment areas than in control areas

If this is true, then we can write

$$Y_{it} = \beta_1 T_i + \beta_2 Post_t + \beta_3 T_i * Post_t + u_{it} + \omega_{it}$$

where u_{it} is correlated with $P_i * Post_t$ and ω_{it} is not

Regression Discontinuity (3)

The basic idea in regression discontinuity is that the endogenous component of the error, u_{it} , is similar in expectation as we approach the eligibility threshold.

That is, if the threshold income (I_{it}) to receive the 8-7 Program is 400 Yuan/yr, then RD would say

$$E[u_{it} | 400 - B < I_{it} < 400] \cong E[u_{it} | 400 < I_{it} < 400 + B]$$

For some small bandwidth B (say $B = 50$).

Regression Discontinuity (4)

- The conditional expectation above is a statement for small movements in income.
- In other words, income is surely related to income growth. That means that, when a program is targeted on income, program treatment is correlated with the error terms, and estimates are biased.
- However, if there is a strict cutoff, then you have some observations which are just below the cutoff and some which are just above it. These observations have (about) the same income level, but very different program status
- Since we know why program status is endogenous – it is purposefully related to income – and we know the allocation rule, we can control for this. In other words, we know which observations will be similarly endogenous, but very different in program status.

Regression Discontinuity in Practice

- One approach would keep only observations within a very small bandwidth of the eligibility threshold
- However, in the real world, we have finite data. This means two things
 - Data may not be as fine as we would like. Bandwidths may be larger than we really want them to be
 - We often want to control for other variables also. We may be able to learn about these variables from observations which are not immediately at the discontinuity
 - For example, time trends, effects of geography on income, etc.
- Also in the real world, discontinuity thresholds are often not enforced perfectly ("fuzzy" discontinuity versus "sharp" discontinuity)
 - Procedure works so long as probability of treatment changes sharply at the eligibility threshold

Fuzzy Discontinuity and Intent to Treat

- Because enforcement is not perfect, that means that some counties which should be ineligible (according to the eligibility cutoff) actually receive the program, and some counties which are eligible would not
- There is a direct link to the Intent-to-Treat that we discussed earlier. If we were to examine counties which actually receive the program and try to draw comparisons, we would be stuck with this problem that counties which are eligible but do not receive the program may be very different from eligible counties which do
- So, the solution is to either
 - 1 Ignore program take-up, and compare differences of eligible counties vs. ineligible ones
 - 2 Exactly as before, use program eligibility as an instrument for program take-up. This scales up the point estimates to create the "treatment on the treated" effect

Regression Discontinuity in Practice (2)

- The usual practice is:

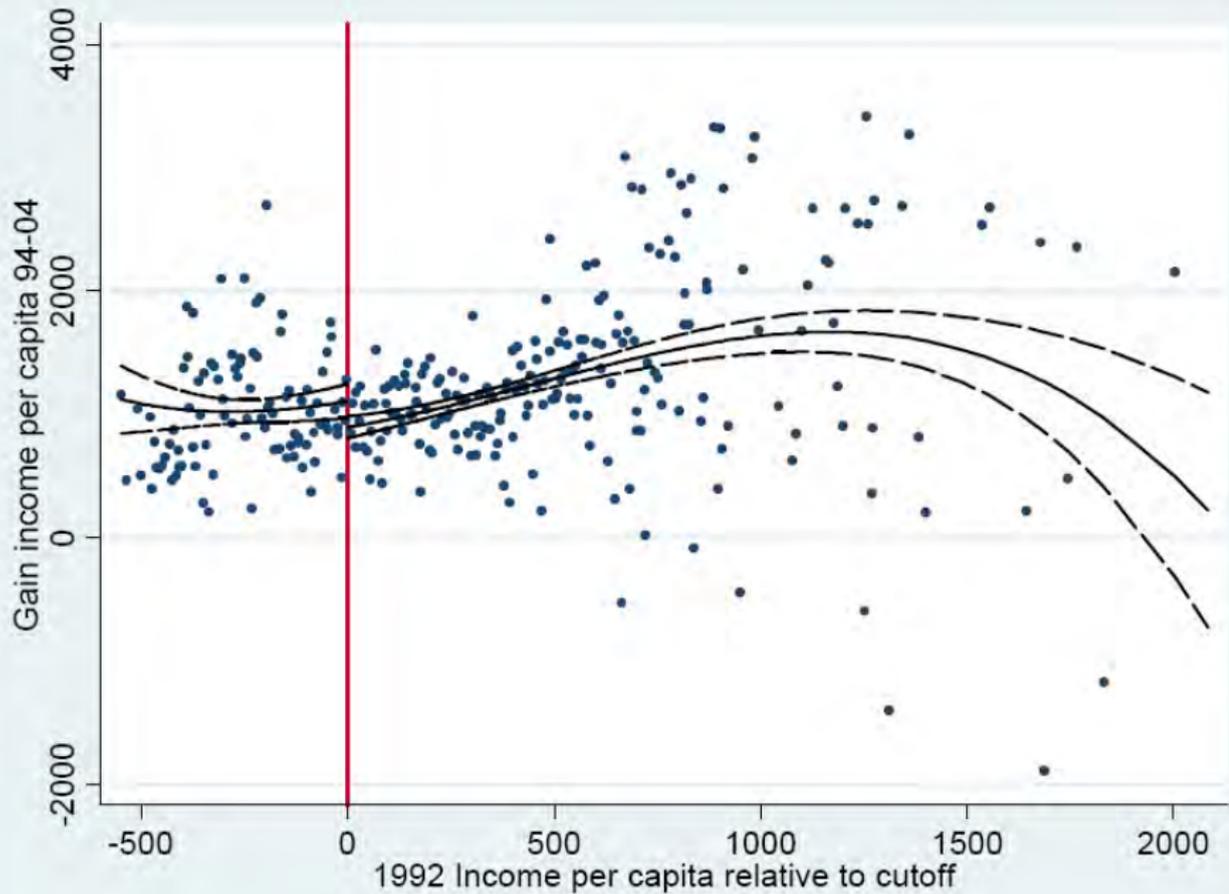
- ① Keep observations which are within a fairly broad bandwidth of the discontinuity
- ② Use, as a control variable a polynomial in the variable which determines eligibility "running variable"
- ③ Allow this polynomial to be different before and after the cutoff
- ④ Instrument program receipt with program eligibility, that is, regress

$$P_{it} = \Gamma X_{it} + \delta_1 f(l_{it}|l_{it} > 400) + \delta_2 f(l_i|l_{it} < 400) + \delta_3 (l_{it} > 400) + \varepsilon_{it}$$

- ⑤ and use the predicted values to estimate our effect, regressing

$$Y_{it} = \Gamma X_{it} + \delta_1 f(l_{it}|l_{it} > 400) + \delta_2 f(l_i|l_{it} < 400) + \beta \hat{P}_{it} + \omega_{it}$$

- ⑥ Graphical analysis is often the most convincing here



Analysis (2)

- Picture seemed to show some change at the eligibility threshold
- Not super clear, though
- Can explore this by statistical analysis
- 2sls using eligible as an instrument for program receipt
- Want to be sure that this is not just an artifact of functional form/bandwidth assmps
 - Examine analysis using a variety of higher level polynomials in running variable
 - Examine analysis using data only within several bandwidths of eligibility

Analysis (3)

First stage

Y < 400	0.70*** (0.04)	0.61*** (0.04)	0.42*** (0.05)	0.43*** (0.05)
---------	-------------------	-------------------	-------------------	-------------------

Second Stage

Post*treatment	415.66*** (86.47)	268.94** (64.53)	362.36** (74.16)	339.58** (75.39)
----------------	----------------------	---------------------	---------------------	---------------------

Controls:

Y in 92	No	Yes	Yes	Yes
Cubic in Y	No	No	Yes	Yes
Other county Controls	No	No	No	Yes

Bandwidths	Full Sample		± 1000 Yuan		± 500 Yuan
Post*treatment	339.58*	365.04**	366.30**	363.49*	398.90**
	(183.65)	(185.52)	(187.33)	(193.46)	(203.71)

Polynomial in Y in 92 Sample	Cubic 1268	Quartic	Cubic 1268	Quartic	Cubic 1268

What do we learn?

- Looks like counties that were just beneath the threshold in 1992 income outperformed counties just over the threshold
- This suggests that the program was effective. RD designs, when they work, can be very convincing
- One Caveat: analysis was based on the assumption that endogeneity did not change discontinuously at the threshold
 - Could be violated if other laws/programs used the same or similar targetting
 - Or if there aren't many observations near the threshold
 - Or if the threshold wasn't enforced

What didn't we learn?

- RD design only identifies local differences at the threshold
 - Tells us nothing about interior observations so:
 - don't learn if program was effective in the poorest counties
 - don't learn if program would be effective in all counties
 - Means we only identify effect locally at these marginally poor counties, not sure that that's the most interesting group
- However, for this group we identify the effect quite well
 - Didn't require an intervention
 - Didn't require strong assumptions like selection on observables