Regression Discontinuity

Chungsang Tom Lam[1]

[1]Department of Economics
Clemson University

April 9, 2020

Idea

- Regression Discontinuity design achieves identification by some arbitrary cutoff point or threshold.
- Many economic policies endogenously affect people's behaviour which makes it difficult to identify the causal effect.
- However, in some cases, while people's actions are endogenous, they can be assigned to either side of the cutoff point arbitrarily and hence create the randomness needed for identifying the treatment effect.

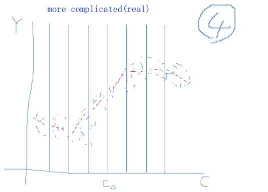
Regression Model

$$D_i = 1 \text{ if } c_i \geq c_0$$
$$D_i = 0 \text{ if } c_i < c_0$$

C0 is the border, you live in Clemson or Central
only one area is affected by the policy

- D_i indicates whether i is affected by the policy and c_0 is the arbitrary cutoff point. D = 1 means individual i is affected by policy
- This case is called the Sharp Regression Discontinuity.

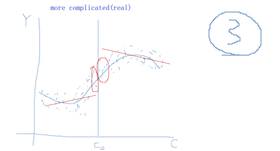
Regression Model



normally, IO professor will ask you to chop many times
and regress each group. you can do this by adding C as dummy

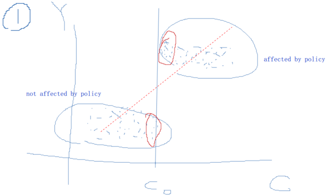
el:

$$y_i = \beta_0 + \beta_1 x_{1i} + \gamma c_i + \rho D_i + \epsilon$$

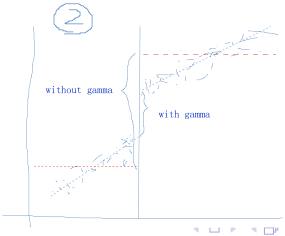


Does the policy have an effect?
No! if you run regression without gamma C you should probably
see the policy have effect, as the expect value are different(red line)
however, after add gamma C(only focus on border(red circle), those ones'
expected values are very close. thus, we know the policy does not have
an effect. you can also solve this problem by add c^2 in your regression.
your regression line is the blue curve. it also tell us the policy doesn't
have an effect.

- x_1 is just a control variable, the effect of the policy identified by an estimator of ρ. the effect of the policy
- Since whether the observation falls into which side of c_0 is arbitrary, we can just estimate it by the OLS estimator $\hat{\rho}$.
- However, what about γc_i?



we cannot run regression like red dashed line.
we should focus on those who are at the border (in red circle).
if these guys have very different expected value, we say, policy have a significant effect
the more they close to the border, the more random they are. that is what we want (arbitrary)



without gamma*C(i), the expected value(run
regression) is really big.
after you include gamma*C, you focus of the
ones around the border, the blue line is your
regression line. much more accurate.

Regression Model

$$y_i = \beta_0 + \beta_1 x_{1i} + f(c_i) + \rho D_i + \epsilon \qquad (2)$$

- We can consider a even more general form and make γc_i a special case of $f(c_i)$
- While ρ captures the "jump" in expected y at the arbitrary cutoff point c_0, interpretation of ρ (and hence the estimator $\hat{\rho}$ is affected by the functional form of $f(c_i)$.
- (figures shown in class)

if you do not have sharp cutoff point, you can recognize D as endogenous variable, and use c as instrument variable.
you can also get a consistent estimator(but biased, as IV is biased)

Regression Model

- How about the situation when the policy cutoff point is not 'sharp'?
- In other words, instead of a jump from $D_i = 0$ to $D_i = 1$, there is a jump in the probability.

$$P(D_i = 1) = g_1(c_i) \text{ if } c_i \geq c_0$$
$$P(D_i = 0) = g_0(c_i) \text{ if } c_i < c_0$$

- Instead of OLS, we use a 2SLS estimation
- This is called "Fuzzy Regression Discontinuity"