PSC 504: Regression Discontinuity Designs

Matthew Blackwell

4/4/2013

Sharp RD

Setup

- The basic idea behind regression discontinuity designs is that we have a variable, X_i, that we call the **forcing variable**, which determines (partly or wholly) the treatment assignment on either side of a fixed threshold.
- This variable may or may not be related to the potential outcomes, but we assume that relationship is smooth, so that changes in the outcome around the threshold can be interpretted as a causal effect.
- The classic example of this in political science is the Lee study of the incumbency effect. We want to know if a party holding a House seat gives that party an advantage in the next election. But candidates who win (the incumbent) tend to better than challengers from the same party. To overcome this, Lee used an RDD with the Democratic share of the two-party vote in the last election as the forcing variable for Democratic incumbency in the current election. The key idea is that, in close elections, seats where a Democratic candidate won will have similar characteristics to districts where a Democratic candidate lost.

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

$$A_i = 1\{X_i \geq c\}$$

- Thus, all units with the forcing variable above c receive treatment and those below c receive control. In the incumbency example, we know that a district is only a "Democratic incumbent" district if the Democratic share of the two-party vote is greater than 0.5.
- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment. But note that overlap here is explicitly violated for the forcing variable. At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see control values. Thus, for a given value of the forcing variables, we only observe treated or control units. Thus, we want to investigate the behavior of the outcome around the threshold:

$$\lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]$$

- Under certain assumptions, this quantity identifies the ATE at the threshold:

$$\tau_{SRD} = E[Y_i(1) - Y_i(0)|X_i = c]$$

- The basic idea behind sharp RD can be summarized in this plot

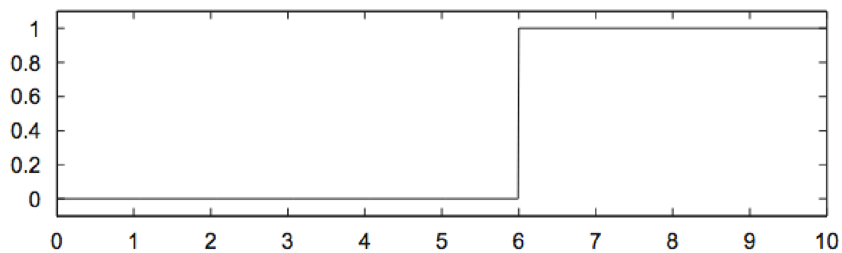

Fig. 1. Assignment probabilities (SRD).

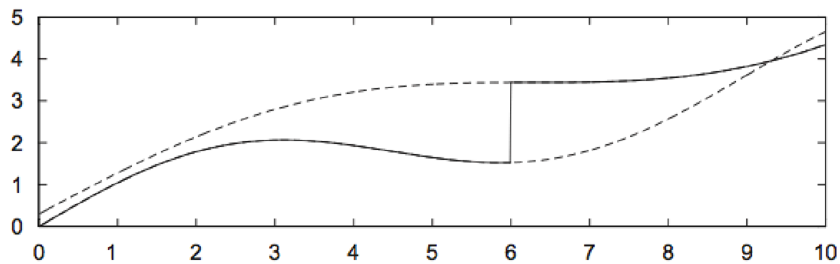

Fig. 2. Potential and observed outcome regression functions.

- Here, we have the forcing variable on each x axis, with the propensity score on the upper y axis and the CEF of the potential outcomes distribution on the lower y-axis. The threshold here is 6. Each of the dotted lines represents the CEF of the potential outcomes: $\mu_a(x) = E[Y_i(a)|X_i = x]$ for $a = 0, 1$. The solid line in the lower graph is the conditional expectation of the observed outcomes. Clearly the discontinuity in the treatment assignment creates a strong discontinuity in the observed outcome.

Assumptions

- Note that ignorability here hold by design, because condition on the forcing variable, the treatment is deteministic.

$$Y_i(1), Y_i(0) \perp\!\!\!\perp A_i|X_i$$

- Again, we can't directly use this because we know that the usual posivity assumption is violated. Remember that positivity is an overlap condition:

$$0 < \Pr[A_i = 1|X_i = x] < 1$$

- Here, obviously, the propensity score is only 0 or 1, depending on the value of the forcing variable. Thus, we need to extrapolate from the treated to the control group and vice versa.

- Extrapolation, even at short distances, requires a certain smoothness in the functions we are extraplating. Thus, we will make a continuity assumption: that $E[Y_i(1)|X_i = x]$ and $E[Y_i(0)|X_i = x]$ are continuous in x. This continuity implies the following:

$$E[Y_i(0)|X_i = c] = \lim_{x \uparrow c} E[Y_i(0)|X_i = x] = \lim_{x \uparrow c} E[Y_i(0)|A_i = 0, X_i = x] = \lim_{x \uparrow c} E[Y_i|X_i = x]$$

- The first equality here comes from continuity, the second from ignorability and the third from the sharp design of the treatment assignment. Note that this is the same for the treated group:

$$E[Y_i(1)|X_i = c] = \lim_{x \downarrow c} E[Y_i|X_i = x]$$

- Thus, under the ignorability assumption, the sharp RD assumption, and the continuity assumption, we have:

$$\tau_{SRD} = \lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]$$

Fuzzy RD

Setup

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

$$\lim_{x \downarrow c} \Pr[A_i = 1|X_i = x] \neq \lim_{x \uparrow c} \Pr[A_i = 1|X_i = x]$$

- In the sharp RD, this is also true, but it further requried the jump in probability to be from 0 to 1. Here, we allow for small jumps. This design is often useful when the a threshold encourages participation in program, but does not actually force units to participate.

- Note that now the story looks slightly different:

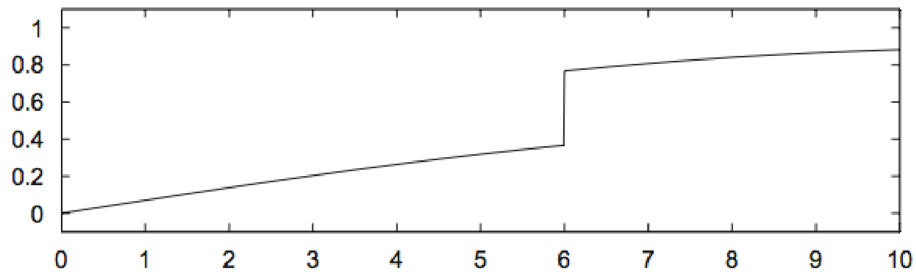

Fig. 3. Assignment probabilities (FRD).

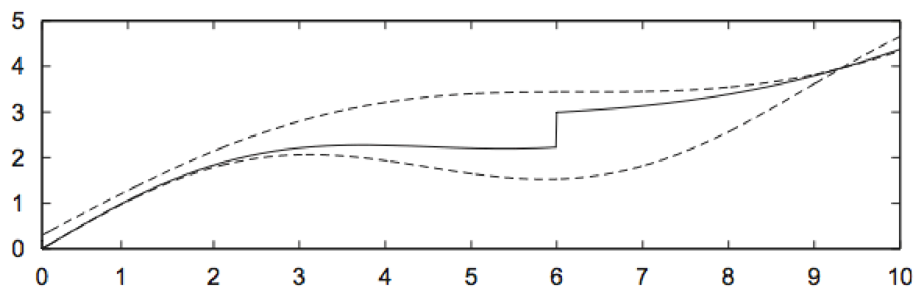

Fig. 4. Potential and observed outcome regression (FRD).

- One way to think about this set up is that the forcing variable is an instrument—it affects the distribution of the treatment, but only affects the outcome through its effect on the treatment (at least in neighborhood around the threshold). Thus, in this case, we have a potential outcome for the treatment again: $A_i(x)$, where this is defined as the potential treatment given a value of the forcing variable.

- We have to think a little bit about how to define "compliers" and so on in this case. A complier would be someone who takes the treatment when encouraged to do so (above the treshold) and takes control when not encouraged (below the threshold). Thus, they would have $A_i(c+e) = 1$ and $A_i(c-e) = 0$.

- We have to make a monotonicity assumption here, just like with IV. Here is a version for our needs: $A_i(x)$ is non-decreasing in x at $x = c$. What does this say? It says that increasing the cutoff point never makes someone more likely to be treated.

- In this setup we also have always-takers:

$$\lim_{x \downarrow X_i} A_i(x) = 1 \quad \text{and} \quad \lim_{x \uparrow X_i} A_i(x) = 1$$

- Of course we have never-takers:

$$\lim_{x \downarrow X_i} A_i(x) = 0 \quad \text{and} \quad \lim_{x \uparrow X_i} A_i(x) = 0$$

- We can define an estimator that is in the spirit of IV:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x]}{\lim_{x \downarrow c} E[A_i | X_i = x] - \lim_{x \uparrow c} E[A_i | X_i = x]}$$

- We assume that $\tau_i = Y_i(1) - Y_i(0)$ and $A_i(x)$ are independent of X_i near c, then we can write that the estimator is equal to the effect at the threshold for compliers.

$$\tau_{FRD} = \lim_{e \downarrow 0} E[\tau_i | A_i(c+e) > A_i(c-e), X_i = c]$$

- This follows from the same argument as last week, which has notes that:

$$E[Y_i|X_i = c+e] = E[Y_i(0) + \tau_i A_i | X_i = c+e] = E[Y_i(0) + \tau_i A_i(c+e)]$$

$$\begin{aligned}
E[Y_i|X_i = c+e] - E[Y_i|X_i = c-e] =& E[\tau_i(A_i(c+e) - A_i(c-e))] \\
=& E[\tau_i(A_i(c+e) - A_i(c-e))|A_i(c+e) > A_i(c-e)] \Pr[A_i(c+e) > A_i(c-e)] \\
&+ E[\tau_i(A_i(c+e) - A_i(c-e))|A_i(c+e) < A_i(c-e)] \Pr[A_i(c+e) < A_i(c-e \\
=& E[\tau_i|A_i(c+e) > A_i(c-e)] \Pr[A_i(c+e) > A_i(c-e)] \\
=& E[\tau_i|A_i(c+e) > A_i(c-e)] E[A_i(c+e)|Z_i = c+e] - E[A_i(c-e)|X_i = c-e] \\
=& E[\tau_i|A_i(c+e) > A_i(c-e)] E[A_i|Z_i = c+e] - E[A_i|X_i = c-e]
\end{aligned}$$

- Note that the FRD estimator emcompasses the SRD estimator because with a sharp design:

$$\lim_{x \downarrow c} E[A_i|X_i = x] - \lim_{x \uparrow c} E[A_i|X_i = x] = 1$$

- A note on external validity: obsviously, FRD puts even more restrictions on the external validity of our estimates because not only are we discussing a LATE, but also the effect is at the threshold. That might give us pause about generalizing other populations for the both the SRD and FRD.

Estimation

Graphical approaches

- First, it is a good idea to simply investigate a plot of the outcome as a function of the forcing variable to see if there is a visually obvious discontinuity in the outcome at the threshold. It is probably not enough to simply look at one outcome, but many. This is to assess the plausibility of the design. Are there other disconuitities that we can't explain?

- Next, it's a good idea to plot covariates by the forcing variable to see if these covariates also jump at the discontinuity. If these are fixed before the assignment of the forcing variable, we might be worried that there could be sorting around the discontinuity which could be related to the outcome.

- Lee paper for examples.

Local linear regression

- The main goal in RD is to estimate the limit of the functions in each of the estimands above. It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point. As a result, the usual kinds of nonparametric estimators perform poorly.

- In general, we are going to have to choose some way of estimating the regression functions around the cutpoint. Using the entire sample on either side will obviously lead to bias because those values that are far from the cutpoint are clearly different than those nearer to the cutpoint. We might think, then about restricting our estimation to units close to the threshold.

- Let's define $\mu_1(x) = \lim_{z \downarrow x} E[Y_(1)|X_i = z]$ and $\mu_0(x) = \lim_{z \uparrow x} E[Y_(0)|X_i = z]$. For the SRD, we have $\tau_{SRD} = \mu_1(x) - \mu_0(x)$.

- One nonparametric approach is to estimate nonparametrically $\mu_1(x)$ with values such that $X_i \in [c, c+h]$ and $\mu_0(x)$ with $X_i \in [c-h, c)$. That is, we just compute means in those bins. This turns out to have very large bias as the we increase the bandwidth.

- A useful semiparametric approach is to run a linear regression of Y_i on $X_i - c$ in the group $X_i \in [c, c+h]$ to estimate $\mu_1(x)$ and the same regression for group with $X_i \in [c-h, c)$.

- Obviously, we can estimate this with an interaction term between the treatment status and the forcing variable.

- The choice of bandwidth is fairly important here and we want it to be smaller as N grows. In general, we can use cross-validation techniques to choose the optimal bandwidth. Of course, we probably also want to show that this choice of bandwidth is not crucial for results.