Lecture 08 - Logic of Regression Discontinuity

Samuel DeWitt

General Outline of Lecture

1) Review of important concepts: selection on...
2) Logic of regression discontinuity-design (RDD)
3) Advantages and Disadvantages of RDD
4) Examples of RDD

Review of Important Concepts - Selection on. . .

Two important terms to remember for this lecture on the regression discontinuity-design (RDD from now on) are **selection on observables** and **selection on unobservables**.

For these terms, **selection** is used to reference the process that results in an individual being assigned to the **treated** group (loosely defined).

As applied to the social sciences, assignment to **treatment** could be the result of an individual's own decision or due to a researcher assigning them to treatment.

Review of Important Concepts - Selection on. . .

Selection on observables refers to our ability to *predict* who is in treatment and who is not in treatment based upon observable (and **measured**) characteristics of the unit that influence which of these statuses they fall into.

This is the goal of **propensity score techniques** - to predict treatment using **observable** characteristics of units such that assignment to treatment becomes **as good as random**.

We can test whether we have satisfied this criterion by assessing **covariate balance** before and after matching pairs of units that were and were not treated.

Review of Important Concepts - Selection on. . .

Selection on unobservables instead refers to our ability to account for the influence of **unobserved** variables on a unit's probability to have been treated.

We can only assume we have done so in the context of a **true experiment** since units are assigned randomly to treatment and control groups.

This assumption is trickier using other methods - your **observable** measures must have very strong correlations with **unobserved** measures we think matter. This is inherently **untestable**.

Review of Important Concepts - Selection on. . .

Random assignment allows for us to assume groups are randomly different on things we **can** (observables) and **cannot** (unobservables) measure.

This is because units cannot self-select into treatment - a random process determines who gets treated and who does not, so the expectation is that **treated** and **untreated** units are **randomly** different on both **observables** and **unobservables**.

Review of Important Concepts - Selection on. . .

The problem for propensity score matching (PSM), and RDD for that matter, is that we cannot actually test to see if we have accounted for **selection on unobservables**.

This makes sense, because we of course cannot assess whether treatment and control groups are similar on variables we have not (and/or cannot) measure.

This is why social scientists value **true experiments** so much and why we try to emulate them when **random assignment** is not possible (or ethical!).

Some Background on the Regression Discontinuity-Design (RDD)

The origins of RDD date back to 1960 with a study of student merit awards by Thistlewaite and Campbell.

They argued that they could better **identify** the causal effect of a student merit award on future academic outcomes if they compared students just **under** the threshold to receive the award to students just **over** it.

That is, they recognized that it was clearly inappropriate to compare all students receiving the award to all students not receiving one - why do you think that is?

Some Background on the Regression Discontinuity-Design (RDD)

If you said the students receiving the awards probably had significantly different characteristics than the students who did not, you are correct.

One of the more important characteristics is **academic ability** which is notoriously difficult to measure (GPAs, GREs, SATs, etc... all are best predictive in the tails of their distributions, mid-range scores are hard to predict from).

Some Background on the Regression Discontinuity-Design (RDD)

Thus, Thistlewaite and Campbell argued that students just **over** and just **under** the threshold for the award are likely fairly similar with respect to academic ability as well as other important characteristics we can and cannot measure.

This was the birth of the RDD, but it went dormant for some time before returning to popularity in economics in the mid-1990s.

It's only now becoming more common in other social sciences (as such, there are some good examples of how **not** to employ RDD in Criminology out there).

Underlying Logic of the RDD

We know that measured and unmeasured characteristics of units influence their **selection** into different types of treatment (broadly defined) and we cannot obtain causal inference without holding these constant somehow.

RDD capitalizes on a particular circumstance where we do have some concrete information on why some units are treated and others are not - that they are over or under some **threshold** (or **cutoff**) along a **running variable** that determines their eligibility for treatment.

Underlying Logic of the RDD

The term **threshold** or **cutoff** is meant to indicate that treatment is made available to individuals over or under this value and is unavailable to those on the other side of it.

A **running variable** is a continuous measure which individuals can be placed along to determine whether they are above or below the threshold value for treatment.

A **continuous** variable is simply another term for an interval/ratio level variable. It has a theoretically infinite number of values between each whole integer (e.g., 2.1, 2.21111, 2.3323, etc. . .).

Underlying Logic of the RDD

Examples of running variables in criminal justice may include: age of eligibility, risk scores, time since last arrest, and the number of prior arrests or convictions.

We make various **treatment** decisions based upon specific values on these variables, which makes treatment status more predictable than with PSM (all else equal).

Underlying Logic of the RDD

If we take advantage of situations like these, where individuals cannot self-select into treatment (only their scores on the running variable) then we can estimate a valid treatment effect around the cutoff.

This treatment effect is sometimes referred to as a **local average treatment effect** or **LATE**, since the effect is only applicable to individuals within a certain range of the cutoff score.

The area around the cutoff we consider for *just under* and *just above* is called a **bandwidth**. All else equal, larger bandwidths have larger amounts of bias.

Advantages of the RDD

1) More ethical than randomizing access to beneficial treatments

2) Unbiased treatment effect around cutoff (assuming scores are not manipulated)

3) Can be used in conjunction with experimental methods around the cutoff.

  – E.g., assigning individuals above the cutoff to control, treatment, or random assignment groups

Disadvantages of the RDD

1) Relies on a smaller number of cases around the cutoff as opposed to a pure randomized experiment with more cases.

2) Treatment is generally *probabilistic* around the cutoff.

   – Cases above or below not **guaranteed** to (not) receive treatment.

3) Noisy distributions of the running variable

   – Often called **kinks** in the RDD literature

RDD - An Important Caveat

There's an important element to the running variable that needs to be true for the RDD to be valid.

Specifically, the running variable must not be **manipulable**. This means that the entity doing the scoring must not be able to adjust scores such that they can *make* units go above or under the threshold.

Alternatively, this also means that units cannot manipulate their *own* scores to be over or under the threshold.

Manipulation of the Running Variable - McCrary, 2008

The following slide shows the anticipated and actual distribution of a running variable around an income-based cutoff for eligibility in a job training program.

Those below the cutoff are eligible, while those above are not.

A and C are two different ways to demonstrate the frequency of observations along the cutoff. B and D are a separate pair of graphs with the same function.

Can you tell which is the simulated vs. actual data?

Manipulation of the Running Variable - McCrary, 2008

Figure 2. Hypothetical Example:
Gaming the System with an Income-Tested Job Training Program
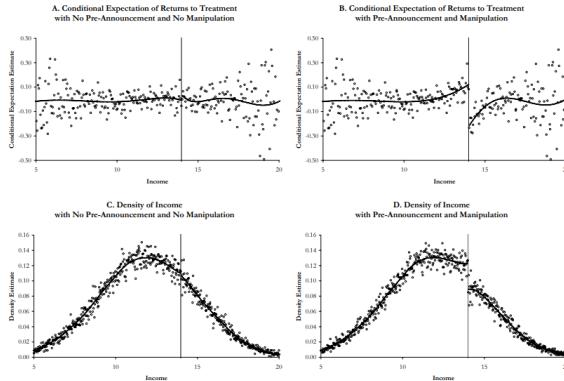
Manipulation of the Running Variable - McCrary, 2008

1) A and C are simulated data assuming **no** manipulation, whereas B and D are the actual data of what happens in practice.

2) There is no evidence of manipulation in A and C - the distribution is continuous about the cutoff.

3) There **is** evidence of manipulation in B and D - individuals are **manipulating** their incomes to fall just under the threshold and, therefore, to be eligible for treatment.

Manipulation of the Running Variable - Summary

Any manipulation of the running variable calls into question whether units are **randomly** distributed around the cutoff. This casts doubt over the **unbiasedness** of the LATE we estimate from RDD.

Manipulation can come in two forms:

1) Individuals manipulate their own scores to fall under or over a cutoff score.
2) Individuals scoring the cutoff variable manipulate scores so that a unit being scores gets treatment (or not).

In either case, it is no longer safe to assume that individuals are only **randomly** different about the cutoff.

Manipulation of the Running Variable - Summary

Either issue can be mitigated if cutoff scores are not published or made known in advance of scoring the running variable. This is not always possible in practice, however.

Manipulation of the Running Variable - Summary

More recent problems with RDD have been highlighted with the use of risk scores as running variables (for sentencing, probation, parole, etc. . .).

The scorers get to see something the designers of the instrument do not and may be inclined to manipulate scores based upon these other (unobserved/unmeasured) variables.

To sum it up - if the density of cases around the cutoff is abnormal, there is doubt about whether the LATE is unbiased.

Examples of RDD - Thistlewaite & Campbell (1960)

► National Merit Scholarship Program
 – Effect of certificate of merit vs. letter of commendation

► Test scores determine eligibility for certificate of merit
 – Cutoff is **fuzzy** and not **sharp** b/c treatment probability is discontinuous, but
 not definite.

Examples of RDD - Thistlewaite & Campbell (1960)
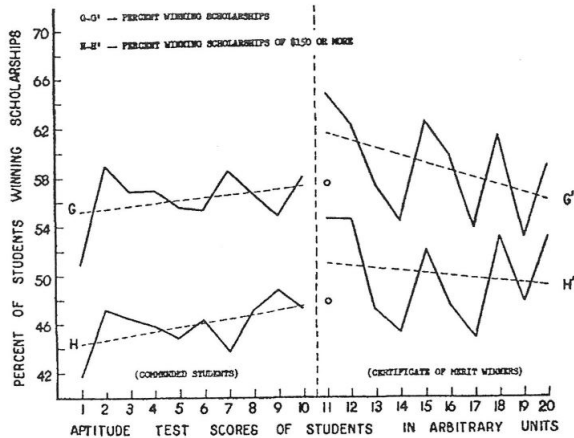

FIG. 2. Regression of success in winning scholarships on exposure determiner.

Examples of RDD - Thistlewaite & Campbell (1960)

A few things to notice:

1) Clear discontinuity in the percent of students winning scholarships just over the threshold for the certificate of merit.
   – This suggests that being *just over* the threshold has a positive treatment effect on receiving a scholarship of some kind.
2) Effect appears to diminish as the value of the running variable increases.
   – Difficult to say for sure what is going on there, but should be careful since comparisons farther away from the cutoff are more biased.
3) What about other characteristics around the cutoff?
   – Does RDD have a **covariate balance** equivalent? (Hint: Yes it does)
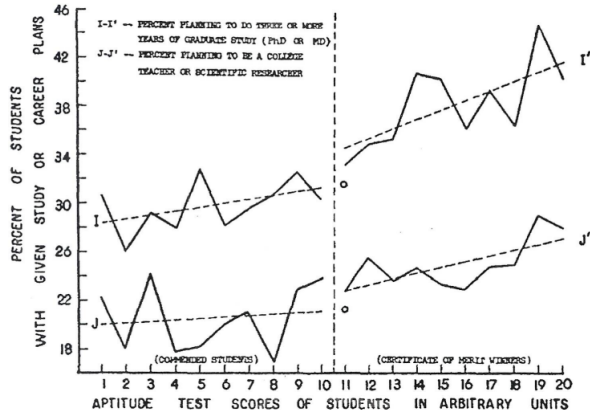
Examples of RDD - Thistlewaite & Campbell (1960)

But, what about other characteristics of these youth?

Are youth above/below the cutoff significantly different with respect to characteristics that affect scholarships?

If they are, this means that we have **covariate imbalance** around the cutoff - calls into question our estimate of the LATE.

Examples of RDD - Thistlewaite & Campbell (1960)

How about....intentions for graduate study?

Examples of RDD - Thistlewaite & Campbell (1960)

1) A slight jump over the discontinuity, but not enough to be significant.

2) Looks to be a fairly continuous distribution across the threshold - which is precisely what we want to see if we want an unbiased LATE.

Examples of RDD - Thistlewaite & Campbell (1960)

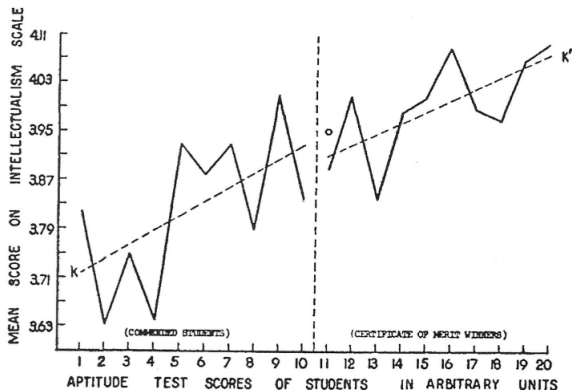How about….attitudes toward intellectualism? (Weird choice, but okay…)

FIG. 4. Regression of attitudes toward intellectualism on exposure determiner.

Examples of RDD - Thistlewaite & Campbell (1960)

1) Odd choice for covariate balance, but you work with what you have.

2) Much more continuous distribution across the cutoff.

3) Further demonstration that our LATE is unbiased (at least with respect to grad school intentions and intellectualism).

Examples of RDD - Tahamont (2019)

Now for a more interesting example for you all - what is the impact of inmate security classification on rules violations?

As background, we assign security classification to inmates for various purposes, but the overall intent is to increase safety for staff and other inmates.
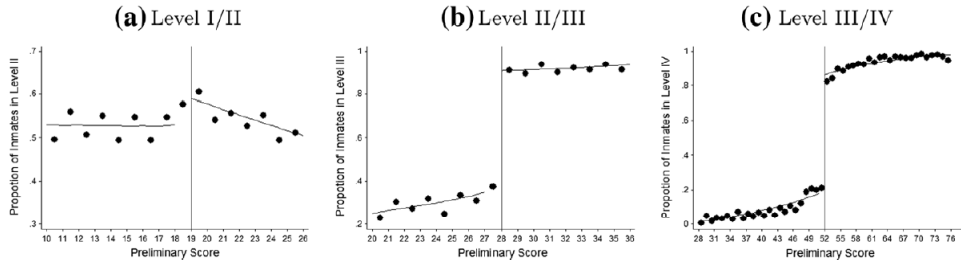
However, we do not know the causal effect of security classification: e.g., the impact of Level II (medium) versus Level III (close).

Examples of RDD - Tahamont (2019)

First, we want to know that scores are correlated with treatment (are scorers using the cutoffs to determine treatment?):
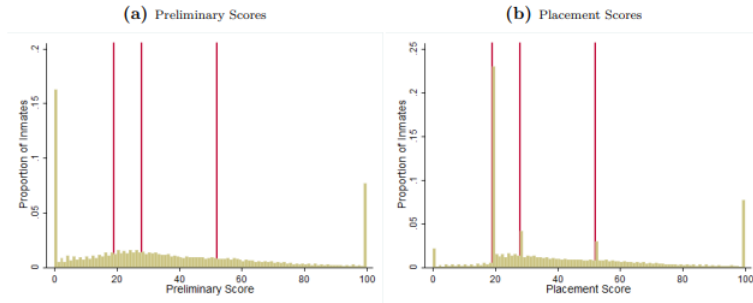
Examples of RDD - Tahamont (2019)

1) Looks as if they are, perhaps with the exception of the Level I/Level II cutoff.

2) To be clear, what we want to see here is a jump in the probability of being assigned to the next level if the inmate's score places them above the threshold. That looks to be mostly true.

Examples of RDD - Tahamont (2019)

Now, we want to know if there are any abnormalities in the distribution of the running variable about the cutoffs:

Figure 3: Distribution of Inmates by Preliminary and Placement Scores

Examples of RDD - Tahamont (2019)

1) The graph is separated into preliminary v. placement scores - just two different running variables representing two different risk scores.

2) It's slightly hard to see given the size of the graphs, but you can see slight increases in the height of the bars above each cutoff (the red vertical lines).

3) These aren't significant enough to cause large issues with the analysis, but there is slight evidence of manipulation here - slightly more common for individuals to be just over the cutoff as it is for them to be just under it.

Examples of RDD - Tahamont (2019)

Now, we want to assess covariate balance.

For this, the tables are simply too large to display them on a slide, so we will examine Table 3 on page 18 of the pdf file.

Examples of RDD - Tahamont (2019)

1) Level II and Level III groups are well balanced across the threshold - most covariates are not significantly different across the cutoff.

2) Level III and Level IV groups are less balanced - a number of covariates are significantly different across these groups.

3) Overall, though, these results aren't too troubling.

Examples of RDD - Tahamont (2019)

Finally, we want to see if security classification affects rule violation frequencies about the thresholds.

What we *want* to see, if security classification is successful in making prisons safer, is that rule violations are less likely to occur among inmates just over the security classification thresholds.

Again, the graphs are too large to fit on a slide, so we will go directly to the paper (Figure 3 on page 22)

Examples of RDD - Tahamont (2019)

1) Effects of security classification appear to be most clearly present for the Level II/Level III distinction.

2) Further, effects most pronounced for Division E and F rule violations (these are less serious).

3) Overall, security classifications do seem to have some effect on rule violations, but the effects are not that strong and only influence a limited range of behaviors.

RDD - A Summary

A conclusion, of sorts: RDD is helpful in situations where treatment assignment has some non-randomoness to it. Namely, that we know the assignment mechanism and it's outside the control of individual units whether or not they get treated (i.e., they do not create the cutoffs!).

This doesn't preclude units or scorers from manipulating values on the running variable, though!

We still need to check for covariate balance as well as for manipulation.

RDD - A Summary

So long as the above checks out, we have an unbiased LATE (only applicable to just under/over the cutoff, though!).

That is, our treatment effect **may not** be generalized to more distant parts of the distribution, only to those who are marginally close to getting treatment.

The End

It's over. Go home.