Regression Discontinuity Design

POST 8000 - Foundations of Social Science Research for Public Policy

Steven V. Miller

Department of Political Science



Goal for Today

Discuss regression discontinuity design (RDD) as a means to causal inference.

Introduction

RDD allows us to estimate causal treatment effects in non-experimental settings.

- It exploits precise knowledge of the rules determining treatment.
- Identification is based on the idea that some rules are arbitrary and provide good quasi-experiments.

Recent flurry of applied economics research using RDD:

- Seemingly mild assumptions (Hahn, Todd, and van der Klaauw 2001)
- More credible than other non-experimental identification strategies (Lee 2008)

Origins: Thistlethwaite and Campbell (1960)

Question: what is the effect of merit awards on future academic outcomes?

• However, awards were allocated on the basis of test scores.

There's a cutoff point c.

• Below which: no award. Above which: award.

Origins: Thistlethwaite and Campbell (1960)

Couldn't you just compare those with/without award?

• No: factors that influence the test score are also related to future academic outcomes.

But you could compare individuals just above and below the cutoff point c.

• That gives you the estimated causal effect of a particular treatment.

An Example of a Regression Discontinuity Design from Thistlethwaite and Campbell (1960)

Notice the different effect for both groups, but more importantly notice the discontinuity.



Aptitude Score Interval (1:20)

Data: Thistlethwaite and Campbell (1960)

The Intuition Behind RDD (In This Case)

- Assignment mechanism is known.
- The probability of treatment jumps to 1 if test score $\geq c$:
 - We call this a **sharp** discontinuity.
- Individuals cannot manipulate their assignment variable.
- Individuals near cutoff are comparable/similar, save for that one distinction.

Thus, the discontinuous jump in outcome variables at \boldsymbol{c} amounts to the causal effect of merit award.

• i.e. the local average treatment effect (LATE)

Identification in RDD

First:

- All other factors determining the outcome variable should be evolving smoothly with respect to the assignment variable.
- If other factors (variables) also jump at the cutoff point, then the estimates of treatment effect will be biased.

Second:

• Since RD estimate requires data away from *c*, the estimate will be dependent on a chosen functional form.

Thistlethwaite and Campbell (1960) is a clear case of a "sharp" RDD.

- Discontinuity precisely determines treatment (i.e. p(treatment) jumps to 1 at c).
- Equivalent to random assignment.

Social Security is a nice example of a sharp design in the wild.

- You can elect to take it earlier, but after a certain threshold, you have to take it.
- (Or at least there's no benefit from delaying payment)

"Fuzzy" RDDs are when treatments are just highly correlated with treatments.

• Probability of treatment jumps by less than one when *x* crosses the threshold *c*.

"Fuzzy" situations make the assignment rule-as-IV an appropriate solution.

RDD has a lot of promise for researchers.

- Intuition for eliminating selection bias is appealing.
- Potential applications *everywhere*.

The Pitfalls of RDD

- It's greedy, often information-poor.
- Breaks in the real world are never super obvious and clear.
- Potentially prone to manipulation.

Conclusion

A regression discontinuity design is a way of undertaking causal inference, usually of some policy intervention.

• It can provide robust, convincing estimates of causal impacts under fairly weak conditions or minimal assumptions.

The nature of the intervention will determine whether an RDD is appropriate.

• Caveat: even when it is, data demands are often great

Table of Contents

Regression Discontinuity Design

Introduction Intuition and Identification Sharp and Fuzzy The Promise/Pitfalls of RDD Conclusion