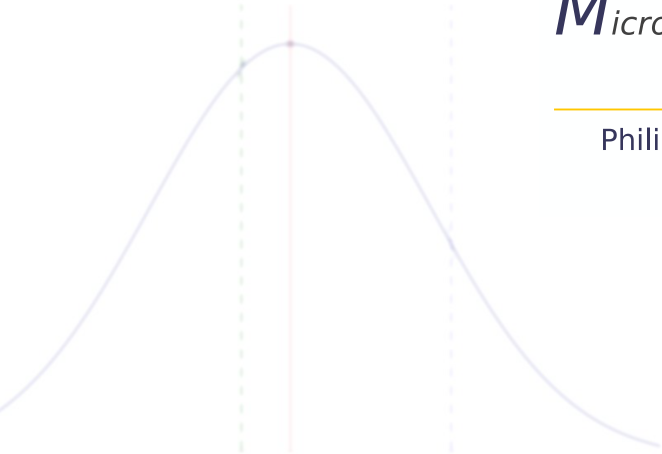
Microeconometrics

Philipp Eisenhauer



Material available on

Visit us!

UNIVERSITÄT BONN

Regression-discontinuity design

Philipp Eisenhauer

I heavily draw on the material presented in:

- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, *48*(2), 281–355.

Issues

- intuition
- identification
- interpretation
- estimation

Key points

- RD designs can be invalid if individuals can precisely manipulate the assignment variable.
  - discontinuity rules might generate incentives
- If individuals - even while having some influence - are unable to **precisely** manipulate the assignment variable, a **consequence** of this is that the variation in treatment near the threshold is randomized as though from a randomized experiment.
  - contrast to IV assumption

Key points

- RD designs can be analyzed - and tested - like randomized experiments.
- Graphical representation of an RD design is helpful and informative, but the visual presentation should not be tilted toward either finding an effect or finding no effect.
- Nonparametric estimation does not represent a "solution" to functional form issues raised by RD designs. It is therefore helpful to view it as a complement to - rather than a substitute for - parametric estimation.

Key points

- Goodness-of-fit and other statistical tests can help rule out overly restrictive specifications.

Baseline

A simple way to estimating the treatment effect τ is to run the following linear regression.

$$Y = \alpha + D\tau + X\beta + \epsilon,$$

where $D \in [0, 1]$ and we have $D = 1$ if $X \geq c$ and $D = 0$ otherwise.

Baseline setup

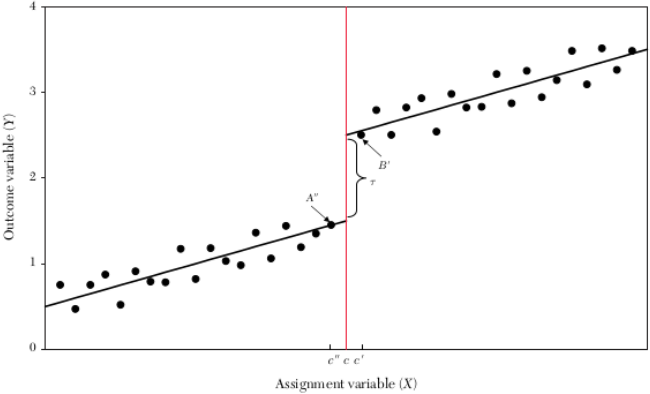


Figure 1. Simple Linear RD Setup

Potential outcome framework

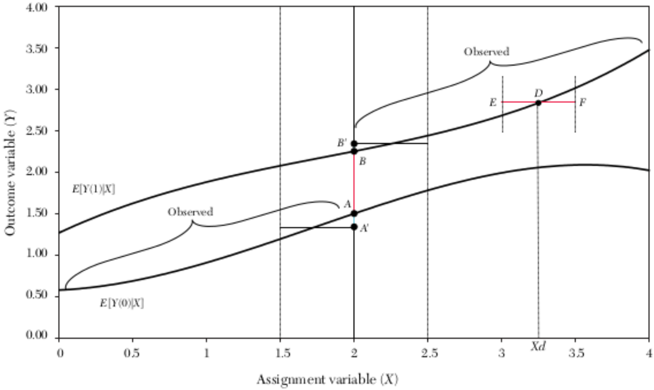


Figure 2. Nonlinear RD

Potential outcome framework

$$E[Y_i(1) - Y_i(0) \mid X = c]$$

$\Rightarrow$ average treatment effect at the cutoff

Alternatives

Consider the standard assumptions for matching:

- ignorability
  - trivially satisfied by research design
- common support
  - cannot be satisfied and replaced by continuity

Alternatives

Lee and Lemieux (2010) emphasize the close connection of RDD to randomized experiments.

- How does the graph in the potential outcome framework change?

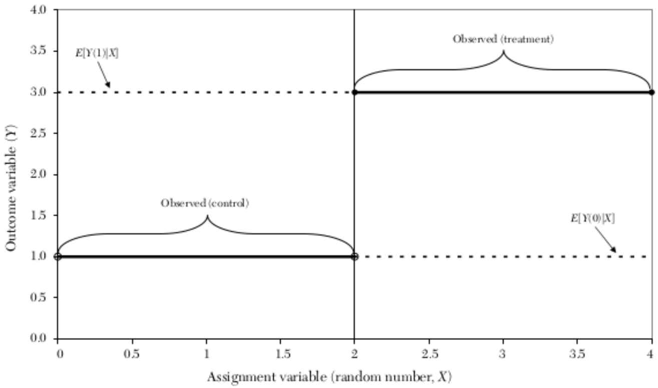


Figure 3. Randomized Experiment as a RD Design

- *Continuity*, the key assumption of RDD, is a **consequence** of the research design and not simply imposed.

Identification

Question

- How do I know whether an RD design is appropriate for my context? When are the identification assumptions plausable or implausable?

Answers

- × An RD design will be appropriate if it is plausible that all other unobservable factors are "continuously" related to the assignment variable.

- ✓ When there is a continuously distributed stochastic error component to the assignment variable - which can occur when optimizing agents do not have *precise* control over the assignment variable - then the variation in the treatment will be as good as randomized in a neighborhood around the discontinuity threshold.

Question

- Is there any way I can test those assumptions?

Answers

- × No, the continuity assumption is necessary so there are no tests for the validity of the design.
- ✓ Yes. As in randomized experiment, the distribution of observed baseline covariates should not change discontinuously around the threshold.

Simplified setup

$$Y = D\tau + W\delta_1 + U$$
$$D = \mathrm{I}[X \geq c]$$
$$X = W\delta_2 + V$$

- W is the vector of all predetermined and observable characteristics.

What are the source of heterogeneity in the outcome and assignment variable?

Simplified setup

The setup for an RD design is more flexible than other estimation strategies.

- We allow for W to be endogenously determined as long as it is determined prior to V.
- We take no stance as to whether some elements δ_1 and δ_2 are zero (exclusion restrictions)
- We make no assumptions about the correlations between W, U, and V.

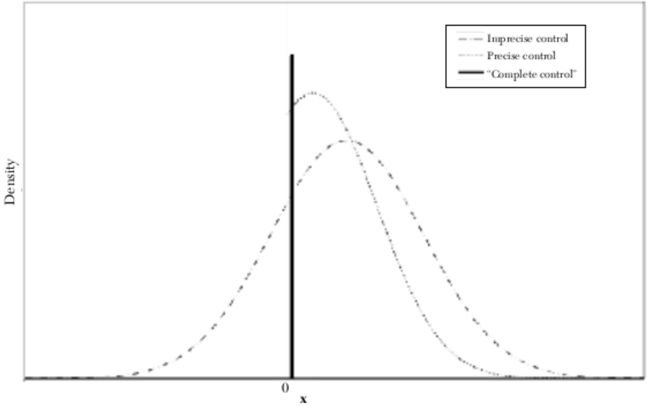


Figure 4. Density of Assignment Variable Conditional on $W = w$, $U = u$

Local randomization

Definition We say individuals have imprecise control over X when conditional on $W = w$ and $U = u$ the density of V (and hence X) is continuous.

Applying Baye’s rule

$$\Pr[W = w, U = u \mid X = x]$$
$$= f(x \mid W = w, U = u) \quad \frac{\Pr[W = w, U = u]}{f(x)}$$

Local randomization If individuals have imprecise control over X as defined above, then $\Pr[W = w, U = u \mid X = x]$ is continuous in x: the treatment is "as good as" randomly assigned around the cutoff.

$\Rightarrow$ the **behavioral** assumption of imprecise control of X around the threshold has the **prediction** that treatment is locally randmized.

Consequences

- testing prediction that $\Pr[W = w, U = u \mid X = x]$ is continuous in x
- irrelevance of including baseline covariates

Interpretation

Questions

- To what extent are results from RD designs generalizable?

Answers

- × The Rd estimate of the treatment effect is only applicable to the subpopulation of individuals at the discontinuity threshold and uninformative about the effect everywhere else.
- ✓ The RD estimand can be interpreted as a weighted average treatment effect, where the weights are relative ex ante probability that the value of an individual's assignment variable will be in the neighborhood of the threshold.

Accounting for treatment effect heterogeneity

$$Y = D\tau(W, U) + W\delta_1 + U$$

What is creating treatment effect heterogeneity?

Accounting for treatment effect heterogeneity

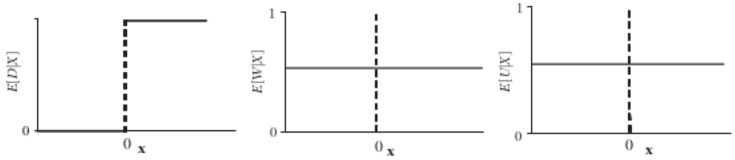
$$\lim_{\epsilon \downarrow 0} E(Y \mid X = c + \epsilon) - \lim_{\epsilon \uparrow 0} E(Y \mid X = c + \epsilon) = ?$$

Alternative evaluation strategies

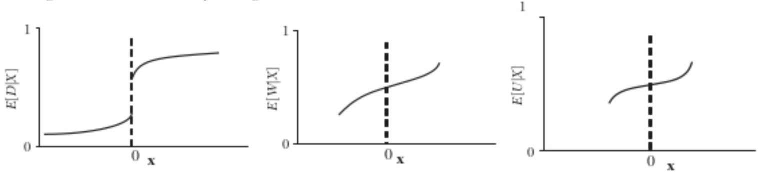
- randomized experiment
- regression discontinuity design
- matching on observables
- instrumental variables

How do the (assumed) relationships between the observables and unobservable differ?

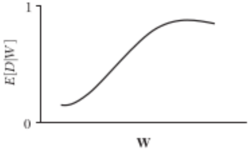
A. Randomized Experiment

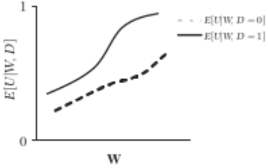


B. Regression Discontinuity Design



C. Matching on Observables





D. Instrumental Variables

Estimation

We will explore issues in estimation using a Python note-book.

http://bit.ly/2WGjWNI

Checklist

Recommendations

- To assess the possibility of manipulations of the assignment variable, show its distribution.
- Present the main RD graph using binned local averages.
- Graph a benchmark polynomial specification.

Recommendations

- Explore the sensitivity of teh results to a range of bandwidth, and a range of orders to the polynomial.
- conduct a parallel RD analysis on the baseline co-variates.
- Explore the sensitivity of the results to the inclusion of baseline covariates.

Resources

Technical

- Hahn, J., Todd, P. E., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209.

Applications

-
- Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology*, *51*(6), 309–317.

Appendix

References

Hahn, J., Todd, P. E., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, *48*(2), 281–355.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology*, *51*(6), 309–317.