The "Credibility Revolution" Goes to Political Economics

Fernanda Brollo

(University of Warwick)

University of Warwick – January, 2015

The "Credibility Revolution"

Taking the "con" out of econometrics! Angrist and Pischke (2010)

> *"Design-based studies are distinguished by their prima facie credibility and by the attention investigators devote to making both an* **institutional** *and a* **data-driven** *case for causality"*

Taking the "econ" out of econometrics? Angrist and Pischke (2010)

> *"Critics of design-driven studies argue that in pursuit of clean and credible research designs, researchers seek good answers instead of* **good questions***"*

Identification strategies in political economics

1. Can we trust the CIA?
    - Studies based on some kind of conditional independence assumption
2. Time goes by (sometimes not)
    - Studies based on some kind of time-invariance assumption
3. Political discontinuities
    - Studies based on regression discontinuity designs
4. History as a lab
    - Studies drawing lessons from variation in historical data
5. Natural and field experiments
    - Studies based on random variation originated by policies (e.g., random auditing of corruption; random interruption of term length) or natural shocks (e.g., rain; natural disasters)
    - Studies based on random variation originated by researcher's intervention in randomized controlled trials

Research questions in political economics

- Institutions, culture, and public policy
- Do politicians respond to incentives? (salary & performance; rent-seeking & corruption; intrinsic motivations; term limit & accountability)
- Role of elections (political competition between and within party; incumbency advantage; voters' vs. politicians' preferences)
- Political selection (institutional, economic, and social determinants; policy effects)
- Role of information (media influence; voters' information set; political campaigns)

Research "waves" in empirical political economics

- Cross-country regressions à la Barro
- IV and matching estimators (Persson and Tabellini, 2003)
- IV in within-country setups (constant wave)
- Regression discontinuity (dying wave?)
- Randomized experiments (infant wave)

Some of the papers that we are going to discuss

- Incentives:
  - Ferraz-Finan (2011) – RD
  - Gagliarducci-Nannicini (2013) – RD
  - Ferraz-Finan (2011) – RD + CIA
- Selection:
  - Besley (2005) – Survey
  - Jones-Olken (2005) – IV
  - Galasso-Nannicini (2011) – CIA
- Information:
  - Ferraz-Finan (2008) – IV
  - Gentzkow-Shapiro-Sinkinson (2011) – DD
  - Kendall-Nannicini-Trebbi (2013) – RCT

A formal framework to think about causality

- The outcome of interest is denoted by $Y_i(D_i)$, the effect that we want to attribute to the treatment. The notation indicates that it may depends on D_i.

→ $Y_i(1)$ if $D_i = 1$

→ $Y_i(0)$ if $D_i = 0$

- The outcome for each individual i can be written as: $Y_i(D_i)$=$D_i Y_i(1)$-$(1 - D_i) Y_i(0)$ **eq.(1)**

- This is enough for correlation: $Cov(D_i, Y_i)/Var(D_i)$

- But does this imply causality?

- Example (1): disclosure of corruption and electoral outcomes
- Example (2): electoral rules and public policies
- Example (3): politicians' wage and their in-office performance

The fundamental problem of causal inference

For every individual i, the event $\{D_i = 1 \text{ instead of } D_i = 0\}$causes the effect Δ_i=$Y_i(1)$-$Y_i(0)$

Given this definition we would like to:

1. Establish whether the above causality link exists for an individual i
2. Measure the dimension of the effect of D_i on Y_i

The selection problem

The "Fundamental Problem of Causal Inference"

- It is impossible to observe for the same individual i the values $D_i = 1$ and $D_i = 0$ as well as the values $Y_i(1)$ and $Y_i(0)$ and, therefore, it is impossible to observe the effect of D on Y for unit i (Holland, 1986)
- Another way to express this problem is to say that we cannot infer the effect of a treatment because we do not have the counterfactual evidence

The statistical solution

Statistics proposes to approach the problem by focusing on the average causal effect for the entire population or for some interesting sub-groups

- The effect of treatment on a random individual (ATE):

  Suppose you pick a person at random in the population and you expose him/her to treatment

  What is the expected effect on the outcome for this person?
  $E\{\Delta_i\} = E\{Yi(1) - Yi(0)\} = E\{Yi(1)\} - E\{Yi(0)\}$ **eq.(2)**

  Apparently we are not making progress, because we cannot observe the outcome in both counterfactual situation for <u>all</u> individuals and therefore we cannot compute the expectations on the right-hand side
  Furthermore, the effect of treatment on a random person may not be an interesting treatment effect from the viewpoint of an economist

The statistical solution

- The effect of treatment on the treated (ATT):

  This second type of average effect is often more interesting for economists

  Let's consider a sub-population of those who are actually treated

  What is the average treatment effect for these persons?

  It is the difference between the average outcome in case of treatment (which we observe) minus the average outcome in the counterfactual situation of no-treatment (which we do not observe). Formally:

  $E\{\Delta_i|Di = 1\} = E\{Y_i(1) - Y_i(0)|D_i = 1\}$ **eq.(3)**
  $= E\{Yi(1)|Di = 1\} - E\{Yi(0)|Di = 1\}$

The statistical solution

A comparison of output by treatment status gives a biased estimate of the ATT:

$E\{Yi|Di = 1\} - E\{Y_i|D_i = 0\}$ **eq.(4)**

$= E\{Y_i(1)|D_i = 1\} - E\{Y_i(0)|D_i = 0\}$

$= E\{Y_i(1)|D_i = 1\} - \mathbf{E\{Y_i(0)|D_i = 1\}}$

$+\mathbf{E\{Y_i(0)|D_i = 1\}} - E\{Y_i(0)|D_i = 0\} =$

$\tau + \mathbf{E\{Y_i(0)|D_i = 1\}} - \mathbf{E\{Y_i(0)|D_i = 0\}}$

where $\tau = E\{\Delta_i|D_i = 1\}$ is the ATT.

The statistical solution

- The observed difference in treatment status adds to this causal effect a term called **selection bias**
- This selection bias term is the difference in average $Y_i(0)$ between those who where and those who were not treated
- Let's go back to our examples!

Randomize experiments

Randomization solves this problem

Random assignment makes D_i independent of potential outcomes: $Y(1), Y(0) \perp D$

Ex: The release of the audit reports

Consider two random samples C and T from the population. Since by construction these samples are statistically identical to the entire population we can write:

$$E\{Y_i(0)|i \in C\} = E\{Y_i(0)|i \in T\} = E\{Y_i(0)\} \textbf{ eq.(5)}$$

and

$$E\{Y_i(1)|i \in C\} = E\{Y_i(1) \in T\} = E\{Y_i(1)\}. \textbf{ eq.(6)}$$

Randomize experiments

Substituting 5 and 6 in 2 it is immediate to obtain:

$E\{\Delta_i \equiv E\{Y_i(1)\}$ - $E\{Y_i(0)\}$ **eq**.(**7**)

$= E\{Y_i(1)|i \in T\}$ - $E\{Y_i(0)|i \in C\}$.

Randomization allows us to use the control units C as an image of what would happen to the treated units T in the counterfactual situation of no treatment, and vice-versa.

Randomize experiments

However randomization is rarely a feasible solution for economists:

- ethical concerns
- technical implementation

But: always useful benchmark. And increasingly feasible in some setting (also because of increasing awarness among policy makers and researchers)

Regression analysis of experiments

Regression is useful tool for the study of causal questions

Suppose the treatment effect is the same for everyone:

$Y_i(1) - Y_i(0) = \mu$, a constant

Recall $Y_i(D_i)$=$D_i Y_i(1)$-$(1 - D_i) Y_i(0)$
$= Y_i(0)$+$[Y_i(1)$-$(Y_i(0)]D_i$

With constant treatment effects we can write:
$Y_i = \alpha + \mu D_i + \epsilon_i$

$\alpha = E(Y_i(0))$
$\mu = E(Y_i(1) - Y_i(0))$
$\epsilon =$ regression error term: idiosyncratic gain from treatment

Regression analysis of experiments

Evaluating the conditional expectation of this equation:

$E(Y_i|D_i = 1) = \alpha + \mu D_i + E(\epsilon_i|D_i = 1)$

$E(Y_i|D_i = 0) = \alpha + E(\epsilon_i|D_i = 0)$

So that,

$E(Y_i|D_i = 1) - E(Y_i|D_i = 0)) = \mu + E(\epsilon_i|D_i = 1) - E(\epsilon_i|D_i = 0)$

Where:

$E(\epsilon_i|D_i = 1) - E(\epsilon_i|D_i = 0)$ is the selection bias;

$E(\epsilon_i|D_i = 1)$ unobservable outcome of the treated in case of treatment;

$E(\epsilon_i|D_i = 0)$ unobservable outcome of the control in case of no treatment.

Assumptions for the IV estimation of the effect of treatment on the treated person

We want to estimate the effects on of D on Y

We assume that there exist a variable Z such that:

$COV\{Z, D\} \neq 0$
$COV\{Z, \eta\} = 0$

Then $E\{\Delta_i | D_i = 1\} = \frac{COV\{Y,Z\}}{COV\{D,Z\}}$

Notation

Consider the following framework:

- N individuals denoted by i;
- They are subject to two possible levels of treatment: D_i=0 and D_i=1;
- Y_i is a measure of the outcome;

Notation

Consider the following framework:

- Z_i is a binary indicator that denotes the assignment; it is crucial to observe that:

  $\Rightarrow$ assignment to treat may or may not be random;
  $\Rightarrow$ the correspondence between assignment and treatment may not be perfect.
  Angrist 1990 uses draft-lotteries number as an instrument to identify earning effect of the military service: the average gain of those who go is independent of the draft, that is: the average gain of those who are not drafted and go and the average gain of those who are draft and go must both be equal to the average gain of all those who go.

Definition of potential outcomes

The participation into treatment for individual i is a function of **Z**

$\Rightarrow$ $\mathbf{D_i}$=$D_i(Z)$

The outcome of individual i is a function of **Z** and **D**.

Note that in this framework we can define three (main) causal effects:

$\Rightarrow$ the effect of assignment Z_i on treatment D_i;
$\Rightarrow$ the effect of assignment Z_i on outcome Y_i;
$\Rightarrow$ the effect of assignment D_i on outcome Y_i;

The first two of this effect is called **intention-to-treat** effects.

Assumptions

Our goal is to establish which of these effects can be identified and estimated.

To do so we need to begin with a set of assumptions and definitions.

Assumption 1: SUTVA

Stable Unit Treatment Value Assumption (SUTVA).

The potential outcomes and treatments of individual i are independent of the potential assignments, treatments and outcomes of individual j ($j \neq i$):

$\Rightarrow \mathbf{D_i} = D_i(Z) \Rightarrow Y_i(\mathbf{Z}, \mathbf{D}) = Y_i(Z_i, D_i)$

where **Z** and **D** are the N dimensional vectors of assignments and treatments.

Given this assumption we can define the **intention-to-treat** effects:

$\Rightarrow$ **The causal effect of Z on D for individual i is**
$D_i(1)$-$D_i(0)$

$\Rightarrow$ **The causal effect of Z on Y for individual i is**

$Y_i(1, D_i(1))$-$Y_i(0, D_i(0))$

Potential outcomes definition

It is crucial to imagine that for each individual the full sets of

- possible outcomes $[Y_i(0,0), Y_i(1,0), Y_i(0,1), Y_i(1,1)]$
- possible treatments $[D_i(0) = 0, D_i(0) = 1, D_i(1) = 0, D_i(1) = 1]$
- possible assignments $[Z_i = 0, Z_i = 1]$

Definition of potential outcomes

Table : Classification of individuals according to assignment and treatment

		$Z_i = 0$ $D_i(0) = 0$	$Z_i = 0$ $D_i(0) = 1$
$Z_i = 1$	$D_i(1) = 0$	Never-taker	Defier
$Z_i = 1$	$D_i(1) = 1$	Complier	Always-taker

Note that each individual i effectively falls in one and only one of these four cells, even if all the full sets of assignments, treatments and outcomes are conceivable.

LATE is not informative about never-takers - by definition treatment status by these two groups is unchanged by the instrument.

LATE is the effect on the population of compliers.

IV solves the problem of causal inference in a randomized trail with partial compliance

Assumption 2: Random Assignment

All individuals have the same probability to be assigned to the treatment:

$$Pr\{Z_i = 1\} = Pr\{Z_j = 1\}$$

Given these first two assumptions we can consistently estimate the two intention to treat average effects by substituting sample statistics on the RHS of the following population equations:

$$E\{D_i|Z_i = 1\}\text{-}E\{D_i|Z_i = 0\}=\frac{COV\{D_i Z_i\}}{VAR\{Z_i\}}$$

$$E\{Y_i|Z_i = 1\}\text{-}E\{Y_i|Z_i = 0\}=\frac{COV\{Y_i Z_i\}}{VAR\{Z_i\}}$$

Note that the ratio between the causal effect of Z_i on Y_i and the causal effect of Z_i on D_i gives the conventional IV estimator

$$\frac{COV\{Y_i Z_i\}}{COV\{D_i Z_i\}}$$

Assumption 2: Random Assignment

All individuals have the same probability to be assigned to the treatment:

The questions we have to answer are:

- Under which assumptions this IV estimator gives an estimate of the average causal effect of D_i on Y_i and for which (sub-)group in the population?
- Does the estimate depend on the instrument we use?

Assumption 3: Non-zero causal effect of Z on D

The probability of treatment must be different in the two assignment groups:

$$Pr\{D_i(1) = 1\} \neq Pr\{D_i(0) = 1\}$$

This assumption requires that the instrument is correlated with the endogenous regressor.

Assumption 4: Exclusion Restriction

The assignment affects the outcome only through the treatment and we can write $Y_i(0, D_i) = Y_i(1, D_i) = Y_i(D_i)$

This assumption says that given the treatment, assignment does not affect the outcome. So we can define the causal effect of on with the following simpler notation:

The causal effect of D on Y for individual i is $Y_i(1) - Y_i(0)$

Assumption 4: Exclusion Restriction

We know that we cannot compute the causal effect because there is no individual for which we observed both its components.

We can, nevertheless, compare sample averages of the two components for individuals who are in the two treatment groups only because of different assignments.

Provided that assignment affects outcomes only through treatment, the difference between these two sample averages seems to allow us to make inference on the causal effect of D on Y. But...

Are the first four assumptions enough?

The four assumptions that we made so far allow us to establish the relation at the individual level between the intention to treat causal effects of Z on D

$$Y_i(1, D_i(1)) - Y_i(0, D_i(0)) = Y_i(D_i(1)) - Y_i(D_i(0))$$

$$= [Y_i(1)(D_i(1)) + Y_i(0)(1 - D_i(1))]$$

$$-[Y_i(1)(D_i(0)) + Y_i(0)(1 - D_i(0))]$$

$$= (D_i(1) - D_i(0))(Y_i(1) - Y_i(0))$$

Are the first four assumptions enough?

By taking expectations on both sides...

$E\{Y_i(1, D_i(1)) - Y_i(0, D_i(0))\}$

$= E\{D_i(1) - D_i(0))(Y_i(1) - Y_i(0)\}$

$= E\{Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = 1)\}Pr\{D_i(1) - D_i(0) = 1\}$

$-E\{Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = -1)\}Pr\{D_i(1) - D_i(0) = -1\}$

This equation clearly shows that even with the four assumptions that were made so far we still have an identification problem: the average treatment effect for compliers may cancel with the average effect for defiers.

To solve this problem we need a further and last assumption

Assumption 5: Monotonicity

This assumption amounts to excluding the possibility of **defiers**.

The probability that you are treated given that you are assigned to treatment is higher than the probability that you are not treated if you are assigned to treatment (and vice-versa)

Definition and relationship with IV

Given the monotonicity assumption:

$$E\{Y_i(1, D_i(1)) - Y_i(0, D_i(0))\}$$

$$= E\{Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = 1)\}Pr\{D_i(1) - D_i(0) = 1\}$$

Rearranging this equation we get the equation that defines the Local Average Treatment Effect:

$$= E\{Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = 1)\} = \frac{E\{Y_i(1,D_i(1)) - Y_i(0,D_i(0))\}}{Pr\{D_i(1) - D_i(0) = 1\}}$$

The local average treatment effect is the average effect of treatment for those who change treatment status because of a change of the instrument; i.e. the average effect of treatment for compliers.

Under the assumption made above, IV estimates are estimates of LATE

Introduction

Regression Discontinuity Design (RDD),

- first introduced by Thislethwaite and Campbell (1960):

→ a way of estimating treatment effects in a non-experimental setting where treatment is determined by whether an observed "assignment" variable exceeds a known cut-off point;

→ impact of merit awards on future academic outcomes

→ allocation of these awards were based on observed test scores;

→ individuals just below the cut-off were good comparison to those just above the cut-off.

Introduction

Regression Discontinuity Design (RDD),

- Disregarded for many years, then:
- Since 1990 a growing number of studies have relied on RDD to estimate program effects in a wide variety of economic contexts.

→ Van der Klaauw (2002), financial aid on students enrollment decisions (assignment rule based on a continuous measure of academic ability with a given cutoff).

→ Angrist and Lavy (1999), class size on education outcomes (Maimonides' rule).

→ Black (1999), discontinuities at the geographical level (school districts boundary) to estimate the wiliness to pay for good schools.

- In the last 8 years a range of questions have been answered using RDD (political, labor, development economics)→ Provide a highly credible and transparent way of estimating program effects

Introduction

- Hahn, J., P. Todd and W. Van der Klaauw (2001), "Identification and estimation of treatment effects with a regression discontinuity design", Econometrica, 69, 201-209
- Imbens, G. and T. Lemieux (2007), "Regression Discontinuity Designs: A guide to practice", Journal of Econometrics, (special issue devoted to RDD).
- Lee, D.S. and T. Lemieux (2010), "Regression Discontinuity Designs in economics", Journal of Economic Literature.

Introduction

- RDD exploit precise knowledge of the rule determining treatment.
- The idea is that some rules are arbitrary and therefore provide good experiments.
- The treatment is assigned to individuals with a value of X greater than or equal to a cut-off value c.
- We refer to the assignment variable as X.

Introduction

- RDD comes in two styles:

→ deterministically: SHARP regression discontinuity (SRD); and can be seen as a selection on observables story;

→ probabilistically: FUZZY regression discontinuity (FRD); leads to IV type of setup.

Sharp RDD

- Let the recipient of the treatment be denoted by a dummy variable $D \in \{0, 1\}$, so that we have:

→ $D = 1$ if $X \geq c$, and

→ $D = 0$ if $X < c$

- Treatment is a deterministic function of the forcing variable

→ mandatory for treated, not manipulable by control

→ e.g., Democratic government if share votes $>$ .5; not otherwise; students with score above the cutoff receive scholarship,...

- Fig. 1: Probability of Treatment Conditional on X. [Source: Imbens-Lemieux 2007]

Figure 1: Assignment probabilities (SRD)

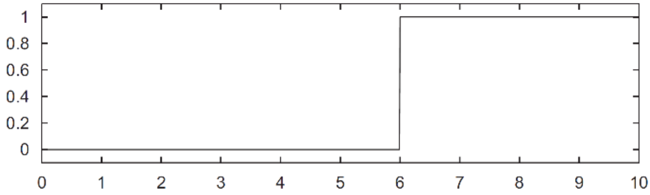


Fig. 1. Assignment probabilities (SRD).

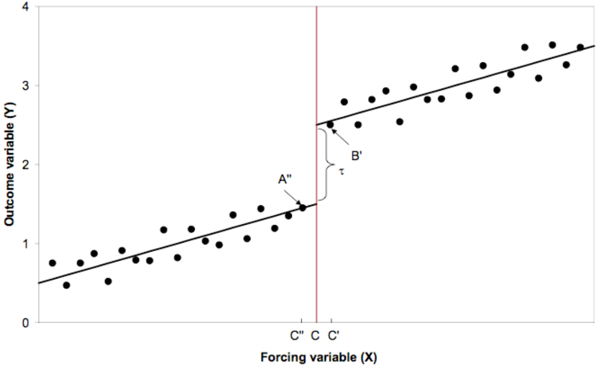
Sharp RDD

- This simple reasoning suggests attributing the discontinuous jump in Y at c to the causal effect of the merit award.
- Assuming that the relationship between Y and X is linear, we can estimate the treatment effect τ is by fitting the linear regression.

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

- Fig 2: Simple Linear RDD Setup [Source: Lee-Lemieux 2009]

Figure 2: Simple Linear RDD Setup



Functional form

- This figure illustrates two important features of the RDD design.

→ "all other factors" determining Y must be evolving "smoothly" with respect to X. If the other variables also jump at c, then the gap τ will be potentially biased for the treatment effect of interest.

→ Since RDD estimate requires data away from the cutoff; the estimate will be dependent on the chosen functional form.

RDD and the Potential Outcomes Framework

- For each individual i there is a pair of "potential" outcomes:

→ $Y_i(1)$ if the unit were exposed to the treatment and;

→ $Y_i(0)$ if not exposed

→ the causal effect of the treatment is represented by the difference $Y_i(1) - Y_i(0)$.

But...

RDD and the Potential Outcomes Framework

- → fundamental problem of causal inference.
- → focus on average effects of the treatment.
- → in RDD there are two underline relationships between X and $E[Y_i(1)|X]$ and $E[Y_i(0)|X]$

Potential Outcome Framework

- With what is observable, we could try to estimate the quantity: This is the average treatment effect at the cutoff c.

  $$B - A = \lim_{\epsilon \downarrow c} E[Y_i | X_i = c + \epsilon] - \lim_{\epsilon \uparrow c} E[Y_i | X_i = c + \epsilon],$$

  which would equal

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

Potential Outcome Framework

- Note that this inference is possible because of the continuity assumption of the underline functions $E[Y_i(1)|X]$ and $E[Y_i(0)|X]$

$\rightarrow$ In order to estimate the causal effect $\rightarrow$ ass. no omitted variables are correlated with treatment dummy.

$\rightarrow$ In the RDD this ass. is trivially satisfied $\rightarrow$ Conditioning on X there is no variation left in D, so it cannot be correlated with any other factor.

- Fig 3: RDD as a Local Randomized Experiment [Source: Lee-Lemieux 2009]
- Fig 4: Non Linear RDD [Source: Lee-Lemieux 2009]

RDD as a Local Randomized Experiment

- Advantages of RDD over most other existing methods becomes clear when we compare RDD to randomized experiments.
  in a Randomized Experiment:

→ Units are typically divided into treatment and control groups on the basis of a randomly generated number, v

→ Suppose v follows an uniform distribution over the range [0,4]

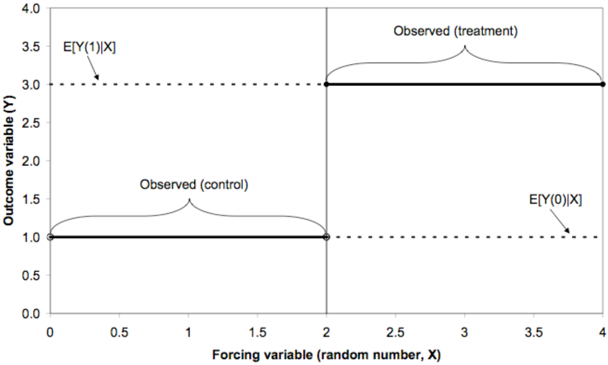
→ For units with v greater than 2 are given the treatment, units v<2 are denied treatment.

→ RDD can be seen as RCT where X=v and c=2

→ The difference is that because X is now completely random it is independent of potential outcomes and the curves are flat.

→ Since curves are flat they are also continuous at the cutoff point c → continuity is a direct consequence of randomization

Figure 4: RDD as a Local Randomized Experiment



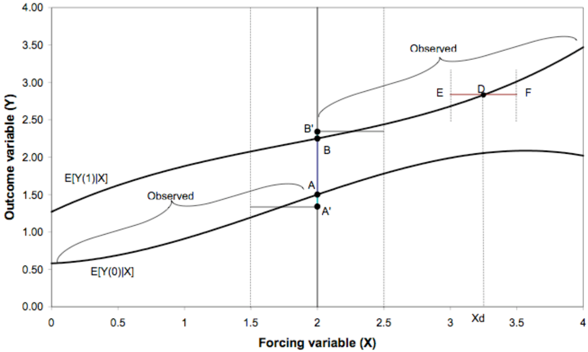
RDD as a Local Randomized Experiment

- Now suppose that people are compensate for having received a "bad draw" by getting a monetary compensation inversely proportional to the random number v (X)

  Example:

- Treatment $\rightarrow$ job search assistance for the unemployed
- Outcome $\rightarrow$ whether one find a job within one month of treatment
- Potential outcome curve will no longer be flat if monetary compensation change the incentives of people in finding a job as soon as possible.

Figure 3: Non Linear RDD



RDD as a Local Randomized Experiment

- A simple comparison of means no longer yields a consistent estimate of the treatment effect :-(

  But...

- By focusing right around the threshold, RDD would still yield a consistent estimate of the treatment effect associated with job search assistance :-)

- How come?

RDD as a Local Randomized Experiment

- Since people just above and below the cutoff receive (essentially) the same monetary compensation, we still have locally a randomized experiment around the cutoff point.
- It is also possible to test whether randomization "worked" by comparing the local values of baseline covariates on the two sides of the cutoff value.

Identification and Interpretation

- How I know whether an RDD is appropriate for my context? When are identification assumptions plausible or implausible?
- Is there any way I can test those assumptions?
- To what extent are results from RDD generalizable?

Identification and Interpretation

- Continuity of conditional regression functions

$\rightarrow$ $E[Y_i(0)|X_i = x]$ and $E[Y_i(1)X_i = x]$ are continuous in x.

- This means that all other unobserved determinants of Y are continuously related to the running variable X.
- This assumption allows us to use the average outcome of units right below the cutoff as a valid counterfactual for units righ above the cutoff.
- Can it be tested? Not directly, but the distribution of observed baseline covariates should not change discontinuously at the threshold c.

Valid or Invalid RDD

- Are individuals able to influence the assignment variable, and if so, what is the nature of this control?
- RDD can be invalid if individuals can precisely manipulate the "assignment variable".
- 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.
- Precise sorting around the threshold is self-selection.

Valid or Invalid RDD

- RDD can be analyzed and tested like randomized experiments.
- Graphical representation is helpful and informative.

Valid or Invalid RDD

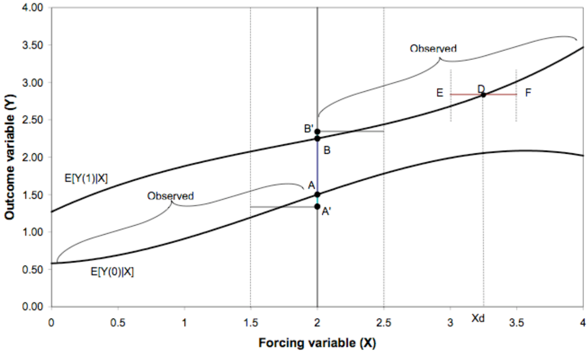
- Non parametric estimation does not represent a solution to functional form issues raised by RDD. It is helpful to view as a complement to - rather than substitute for - a parametric estimation (boundary problem).
- It is essential to explore how RDD estimates are robust to the inclusion of high order polynomial terms and to changes in the window width around the cutoff point.

Valid or Invalid RDD

- There is no particular reason to believe that the true model is linear
- Misspecification of the functional form typically generates bias in the treatment effect
- But we need to estimate regressions at the cutoff point $\rightarrow$ boundary problem

  Solutions:
- relaxing linearity assumption including polynomial functions of X $\rightarrow$ use data far from cutoff to predict Y at the cutoff point.
- Kernel regressions.

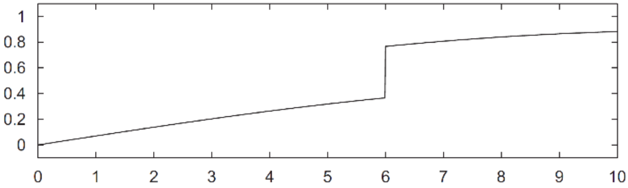
Figure 3: Non Linear RDD



Fuzzy RDD (Imperfect Compliance)

- Treatment is determined partly by whether the assignment variable crosses a cutoff point.
- The probability of receiving treatment changes discontinuously at the threshold c; but need not go from 0 to 1:
- e.g., incentives to participate in program change discontinuously at the threshold, but are not powerful enough to move everyone from non-participation to participation (e.g., encouragement designs)
- Fig 5. Assignment probabilities [Source: Imbens-Lemieux 2007]

Figure 6: Assignment Probabilities (FRD)



Fuzzy RDD (Imperfect Compliance)

Under continuity of potential outcomes at $X = c$ (Hahn et al. 2001), we can identify the first-stage/reduced-form effects:

$$E[D(h_c) - D(\ell_c)|X = c] = \lim_{x \downarrow c} E[D|X = x] - \lim_{x \uparrow c} E[D|X = x]$$
$$E[Y(h_c) - Y(\ell_{c)})|X = c] = \lim_{x \downarrow c} E[Y|X = x] - \lim_{x \uparrow c} E[Y|X = x]$$

Fuzzy RDD (Imperfect Compliance)

The next step is to use these reduced-form effects to identify the causal effect of D on Y:

$\tau_{FRD} = \frac{E[D(h_c) - D(\ell_c) | X=c]}{E[D(h_c) - D(\ell_c) | X=c]}$,

- $\rightarrow$ Numerator: jump in the regression of the outcome on the covariate
- $\rightarrow$ Denom: Jump in the regression of the treatment indicator on the covariate

Fuzzy RDD (Imperfect Compliance)

Which can be implemented by estimating:

$$Y_i = g(X_i) + \tau D_i + \epsilon_i \tag{1}$$

where T is used as instrument for D

in this setting, τ can be interpreted as an "intent to treat" effect.

Fuzzy RDD (Imperfect Compliance)

- The interpretation of this ratio as a causal effect requires the same assumptions as in Imbens and Angrist (1994):

$\rightarrow$ monotonicity: crossing the cutoff cannot simultaneously cause units to take up and others to reject the treatment.

$\rightarrow$ excludability: crossing the cutoff cannot impact Y except through impaction receipt of treatment

- When these assumptions are made, it follows that:

$\rightarrow$ $\tau_{FRD} = E[Y_i(1) - Y_i(0)|$ unit is complier,$X{=}c$,
where compliers are unites that received the treatment when they satisfy the cutoff rule $(X_i \geq c)$

Summary

- If there is a local random assignment (do to the plausibility of individuals' imprecise control over X), we can apply what we know about the assumptions and interpretability of instrumental variables.
- The difference between sharp and fuzzy RDD is exactly parallel to the difference between the randomized experiment with perfect compliance and the case of imperfect compliance, when only the intent to treat is randomized.

Summary

- In the case of imperfect compliance, even if a proposed binary instrument Z is randomized, it is necessary to rule out the possibility that Z affects the outcome, outside of its influence through treatment receipt, D.
- Only then will the instrumental variable estimand - the ratio of the reduced form effects of Z on Y and of Z on D - be properly interpreted as a causal effect of D on Y.

Summary

- It is important to verify a strong first-stage relationship in an IV design.
- It is important to verify that a discontinuity exists in the relationship between D and X in a FRDD
- The ratio of the FRDD gaps in Y and D can be interpreted as a weighted LATE, where the weights reflect the ex ante likelihood the individual's X is near the threshold.
- Exclusion restriction and monotonicity should hold

Test of RDD validity

Challenges to RDD

- Treatment is not as good as randomly assigned around the cutoff when agents are able to manipulate their scores. This happens when (i) the assignment rule is known in advance, (ii) agents are interested in adjusting, and (iii) agents have time to adjust.

  Examples: re-take exam, self-reported income, etc.

- Some other unobservable characteristic changes at the threshold, and this has a direct effect of outcome.

  Examples: population thresholds or age thresholds used for several policies.

  We can formally assess the extent of these problems

Test of RDD validity

Test 1: Manipulation of the running variable

What to look for.

- Consider a desirable treatment where the assignment rule is $X \geq c$. If there is manipulation (or sorting on the running variable) you would expect "bunching" of obs just to the right of c; and relatively few obs just to the left of c.
- It turns out that if the running variable is not entirely under agent's control (partial manipulation), identification can still be achieved (Lee 2007). Complete manipulation instead undermines identification.

Test of RDD validity

Test 1: Manipulation of the running variable

How to detect it

- Graphically, plot density of the running variable (see above)
- McCrary (2008) proposes a formal test, known as the McCrary Density Test. This has now become a *must* for every analysis using RDD.

  The test employs a local linear density estimator and develops a test statistic for the null $f^{+} - f^{-} = 0$.

  Bandwidth is optimally selected.

  .ado file available at www.econ.berkeley.edu/ jmccrary/DCdensity

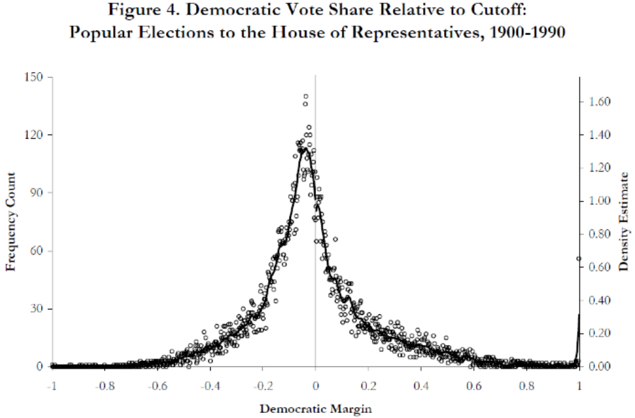
Test of RDD validity

Test 1: Manipulation of the running variable

Example: Incumbency effect, (Lee 2001)

- Pol. sci literature on incumbency advantage. Having won election once helps win subsequent ones.
- Empirical challenge: selection effects
- Lee (2001): in 2-party system with majority rule, incumbency is assigned discontinuously at 50%
- Complete manipulation unlikely due to difficulty of coordination among voters in a popular election (in the absence of fraud!)
- Data on votes in elections to US House of Representatives, 1900-1990 Fig. 4 in McCrary (2008): no evidence of manipulation

Test of RDD validity



2. Test of RDD validity

Contrast with other types of election

- Roll call voting in US House of Representatives
- Coordination is expected because: (i) repeated game, (ii) vote record is public, (iii) side payments possible in the form of future votes
- Data on roll call votes in the House, 1857-2004
- Bills around the cutoff are more likely to be passed than not — cannot use RDD to generate quasi-random assignment of policy decisions!

  Fig. 5 in McCrary (2008): evidence of manipulation

2. Test of RDD validity

Figure 5. Percent Voting Yeay:
Roll Call Votes, U.S. House of Representatives, 1857-2004