Modern Regression Discontinuity Analysis
For Education Research

Howard Bloom
MDRC

Overview of Today’s Workshop

**Part #1**: Getting Started

**Part #2**: Identifying intervention effects

**Part #3**: Estimating intervention effects

**Part #4**: Generalizing RD Findings
Part 1
Getting Started
Introductions

**Who** is in this workshop?

**What** is our prior experience with RD analysis?
- What do these experiences have in common?
- How do they differ?

**Where** did RD analysis come from?
- *Genesis*: 1960s
- *Dark Ages*: 1970s and 80s
- *Renaissance*: 1990s
- *New Age*: 21st century
What is RD Analysis?

Characterizations
– Discontinuity at a point
– Local randomization
– Known selection process

Timeframes
– Ex ante
– Ex post

Components
– Candidates
– Ratings
– Cut-point
– Continuity
Two Ways to Characterize Regression Discontinuity Analysis

1. Discontinuity at a point
2. Local randomization
What Conditions are Necessary for the Internal Validity of an RD Analysis?

**Independence of Ratings and Cut-Point**
- Ratings must be set independently of the cut-point.
- The cut-point must be set independently of ratings.

**Functional Form**
- The functional form of the relationship between the outcome and rating must be properly specified throughout the region of data used.
- The smaller this region is, the more likely the relationship is to be linear.
How Might These Conditions Be Violated?

Non-independence of Ratings and Cut-Point

– Decision makers might move the cut-point to include or exclude specific candidates if their ratings are known.
– Decision makers might assign ratings to include or exclude specific candidates if the cut-point is known.
– What are examples of these threats to internal validity?

An Incorrect Functional Form

– If the slope of the outcome/rating relationship is changing at the cut-point in a way that is not properly accounted for, the estimated discontinuity will conflate the changing slope with the true discontinuity.
– This can produce an over-estimate or an under-estimate of the intervention effect.
Part 2
Identifying Intervention Effects
RD Identification

**A Sharp RD Analysis** (with full compliance)
- can identify an average treatment effect at the cut-point \((ATE_c)\)

**A Type I Fuzzy RD Analysis** (with no-shows)
- can identify an average effect of treatment on the treated at the cut-point \((TOT_c)\)

**A Type II Fuzzy RD Analysis** (with no-shows and crossovers)
- can identify a local average treatment effect at the cut-point \((LATE_c)\)
Illustrative Regression Discontinuity Analyses

**Sharp Regression Discontinuity (Full Compliance)**

Expected Outcome ($\overline{Y}_{(i)}$) (student scores)

- Control
  - $\overline{Y}_1 = 470$
- Treatment
  - $\overline{Y}_1^* = 500$

Cut-point ($r^*$) Rating ($r$)

**Type I Fuzzy Regression Discontinuity (No-Shows)**

Expected Outcome ($\overline{Y}_{(i)}$) (student scores)

- Control
  - $\overline{Y}_2 = 470$
- Treatment
  - $\overline{Y}_2^* = 495$

Cut-point ($r^*$) Rating ($r$)

**Type II Fuzzy Regression Discontinuity (No-Shows and Crossovers)**

Expected Outcome ($\overline{Y}_{(i)}$) (student scores)

- Control
  - $\overline{Y}_3 = 475$
- Treatment
  - $\overline{Y}_3^* = 495$

Cut-point ($r^*$) Rating ($r$)
The Probability of Receiving Treatment As a Function of the Rating

Sharp Regression Discontinuity (Full Compliance)

Type I Fuzzy Regression Discontinuity (No-Shows)

Type II Fuzzy Regression Discontinuity (No-Shows and Crossovers)
How a Sharp RD (with full compliance) Identifies An Average Treatment Effect at the Cut-Point ($ATE_c$)

$$ATE_c = \frac{\bar{Y}^+ - \bar{Y}^-}{\bar{T}^+ - \bar{T}^-} = \frac{\bar{Y}^+ - \bar{Y}^-}{1 - 0}$$

$$= \bar{Y}^+ - \bar{Y}^-$$

$$= 500 - 470 = 30 \text{ points}$$
How a Type I Fuzzy RD Identifies the Average Effect of Treatment on the Treated

\[ ITT_c = \overline{T}^+ \cdot TOT_c + (1 - \overline{T}^+) \cdot 0 \]
\[ = \overline{T}^+ \cdot TOT_c \]
\[ TOT_c = \frac{ITT_c}{\overline{T}^+} = \frac{\overline{Y}^+ - \overline{Y}^-}{\overline{T}^+} \]
How a Type I Fuzzy RD (with no-shows) Identifies An Average Effect of Treatment on the Treated at the Cut-Point (TOTc)

\[ TOT_c = \frac{\bar{Y}^+ - \bar{Y}^-}{\bar{T}^+ - \bar{T}^-} = \frac{\bar{Y}^+ - \bar{Y}^-}{\bar{T}^+ - 0} \]

\[ = \frac{\bar{Y}^+ - \bar{Y}^-}{\bar{T}^+} \]

\[ = \frac{495 - 470}{0.8} = 31.25 \text{ points} \]
The Angrist, Imbens, and Rubin Causal Categories (Absent Defiers)

Treatment Group

T = 1

Compliers $P_{co}^{(a)}$

Control Group

T = 0

Always-Takers $P_{at}^{(a)}$

T = 1 (Crossovers)

Never-Takers $P_{nt}^{(a)}$

T = 0

(T = 1 (No-Shows))
How a Type II Fuzzy RD (with no-shows and crossovers) Identifies a Local Average Treatment Effect at the Cut-Point (LATE$_c$)
Part 3

Estimating Intervention Effects
Estimation Approaches

Graphical analysis
  – What you see is what you get.
  – But much is in the eye of the beholder.

Parametric estimation
  – is where RD began.

Non- or Semi-Parametric Estimation
  – is all about flexibility.
Graphical Analysis

**Step 1:** Plot the probability of receiving treatment vs. the rating
  – To confirm that a treatment contrast exists

**Step 2:** Plot the outcome vs. the rating for all data points
  – to look directly at the data

**Step 3:** Plot bin means for the outcome vs. the rating
  – to smooth the data

**Step 4:** Choose an optimal bin width
  – to tradeoff noise against specificity
A Graphical Regression Discontinuity Analysis

Probability of Receiving Treatment

Cut-point Rating (r)

Original Data Points

Cut-point Rating (r)

“Narrow Bin” Means

Cut-point Rating (r)

“Wide Bin” Means
Parametric Estimation #1: The Simple Linear Model

\[ Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \eta_i \]

\[ \hat{\beta}_0 = \text{estimated average effect of intent to treat} \]

\[
\text{at the cutpoint } (ITT_C)
\]

\[ \hat{\beta}_1 = \text{estimated regression slope} \]
Simple Linear Regression Discontinuity Analysis

\[ \hat{\beta}_0 = \text{estimated treatment effect} \]

\[ \hat{\beta}_1 = \text{estimated slope of the outcome / rating relationship} \]

Probability of Receiving Treatment

Cut-point Rating

Outcome

Cut-point Rating

\( \bar{T}^+ \)

\( \bar{T}^- \)
Parametric Estimation #2:
Translating Estimates of ITT<sub>c</sub> into Estimates of LATE<sub>c</sub>

\[
\text{LATE}_C = \frac{\text{ITT}_C}{\bar{T}^+ - \bar{T}^-} = \frac{\hat{\beta}_0}{\bar{T}^+ - \bar{T}^-}
\]

\[
= \frac{495 - 475}{0.80 - 0.15} = 30.77 \text{ points}
\]

\[
\text{se}(\text{LATE}_C) = \frac{\text{se}(\hat{\beta}_0)}{\bar{T}^+ - \bar{T}^-}
\]

\[
t_{\text{LATE}_C} = \frac{\text{LATE}_C}{\text{se}(\text{LATE}_C)} = \frac{\hat{\beta}_0}{\text{se}(\hat{\beta}_0)} = t_{\hat{\beta}_0}
\]
Regression Discontinuity Estimation with an Incorrect Functional Form
Parametric Estimation #3: Common Functional Forms

Separate linear model for Ts and Cs

\[ Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i \cdot T_i + \eta_i \]

A quadratic, cubic or other polynomial

\[ Y_i = \alpha + \beta_0 \cdot T_i + \sum_{m=1}^{M} \beta_m \cdot r_i^m + \eta_i \]

N.B. Center values of the rating on the cut-point (i.e. set \( r = 0 \) at the cut-point)
Nonparametric Estimation

It is a complement to parametric estimation, not a substitute.

Kernel regression
  – What is it?
  – Why does it have poor boundary properties?

Local linear (or polynomial) regression
  – What is it?
  – How does it improve boundary properties?

Limitations of the approach
  – It requires very large samples.
  – It can be sensitive to selection of a bandwidth.
Boundary Bias from Kernel Regression Versus Local Linear Regression (Given Zero Treatment Effect)
Validation and Robustness Tests

Test the assumptions of the design

- Learn about the selection process! (Were ratings and the cut-point in fact determined independently?)
- Look for (estimate) a discontinuity in the density of observations at the cut-point to test for “gaming”.

Test the validity of the estimator

- Look for (estimate) discontinuities in baseline characteristics (especially pretests) at the cut-point.

Test the sensitivity of the estimator

- Try alternative functional forms for parametric estimators.
- Try alternative bandwidths for non-parametric estimators.
- Try regions of varying distance from the cut-point for all estimators (e.g. trim outliers).
Statistical Precision #1:
Parametric Estimators of the Average Effect of Intent to Treat

Minimum Detectable Effect Size (MDES)

\[ MDES \approx 2.8 \sqrt{\frac{1 - R_Y^2}{N \cdot P(1 - P)(1 - R_T^2)}} \]

MDES Relative to an Otherwise Comparable Randomized Trial

\[ \frac{MDES_{RD}}{MDES_{randomized}} = \frac{1}{\sqrt{1 - R_T^2}} \]

Sample Size Relative to an Otherwise Comparable Randomized Trial

\[ \frac{N_{RD}}{N_{randomized}} = \frac{1}{1 - R_T^2} \]
Collinearity Coefficient and Sample Size Multiple for a Regression Discontinuity Design Relative to an Otherwise Comparable Randomized Trial

<table>
<thead>
<tr>
<th>Regression Discontinuity Model</th>
<th>Balanced Design (P = 0.5)</th>
<th>Unbalanced Design (P = 0.33 or 0.67)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Uniform Rating Distribution</td>
<td>Normal Rating Distribution</td>
</tr>
<tr>
<td>Simple linear</td>
<td>0.750</td>
<td>0.637</td>
</tr>
<tr>
<td>Simple quadratic</td>
<td>0.750</td>
<td>0.637</td>
</tr>
<tr>
<td>Simple cubic</td>
<td>0.859</td>
<td>0.744</td>
</tr>
<tr>
<td>Separate treatment/control linear</td>
<td>0.750</td>
<td>0.637</td>
</tr>
<tr>
<td>Separate treatment/control quadratic</td>
<td>0.889</td>
<td>0.802</td>
</tr>
</tbody>
</table>

| Regression Discontinuity Model                  |   | Collinearity Coefficient ($R^2$) |   |
|------------------------------------------------|   |---------------------------------|---|
| Simple linear                                   |   |                                 | 0.637 |
| Simple quadratic                                |   |                                 | 0.637 |
| Simple cubic                                    |   |                                 | 0.744 |
| Separate treatment/control linear               |   |                                 | 0.637 |
| Separate treatment/control quadratic            |   |                                 | 0.802 |

<table>
<thead>
<tr>
<th>Regression Discontinuity Model</th>
<th>Sample Size Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple linear</td>
<td>4.00</td>
</tr>
<tr>
<td>Simple quadratic</td>
<td>4.00</td>
</tr>
<tr>
<td>Simple cubic</td>
<td>7.11</td>
</tr>
<tr>
<td>Separate treatment/control linear</td>
<td>4.00</td>
</tr>
<tr>
<td>Separate treatment/control quadratic</td>
<td>9.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression Discontinuity Model</th>
<th>Sample Size Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple linear</td>
<td>2.75</td>
</tr>
<tr>
<td>Simple quadratic</td>
<td>2.75</td>
</tr>
<tr>
<td>Simple cubic</td>
<td>3.91</td>
</tr>
<tr>
<td>Separate treatment/control linear</td>
<td>2.75</td>
</tr>
<tr>
<td>Separate treatment/control quadratic</td>
<td>5.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression Discontinuity Model</th>
<th>Sample Size Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple linear</td>
<td>2.97</td>
</tr>
<tr>
<td>Simple quadratic</td>
<td>4.79</td>
</tr>
<tr>
<td>Simple cubic</td>
<td>5.22</td>
</tr>
<tr>
<td>Separate treatment/control linear</td>
<td>4.00</td>
</tr>
<tr>
<td>Separate treatment/control quadratic</td>
<td>5.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression Discontinuity Model</th>
<th>Sample Size Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple linear</td>
<td>2.46</td>
</tr>
<tr>
<td>Simple quadratic</td>
<td>2.86</td>
</tr>
<tr>
<td>Simple cubic</td>
<td>3.52</td>
</tr>
<tr>
<td>Separate treatment/control linear</td>
<td>2.72</td>
</tr>
<tr>
<td>Separate treatment/control quadratic</td>
<td>3.89</td>
</tr>
</tbody>
</table>
Statistical Precision #2: Non-parametric Estimators

Precision for non-parametric estimators is typically much less than that for parametric estimators because the former only use observations that are close to the cut-point.

It is difficult to know in advance what the precision of an analysis will be because its bandwidth will not yet have been determined.

Not much is generally known about the precision of non-parametric estimators because of their extreme flexibility and multiple possibilities for the sample and approach to be used.
Part 4

Generalizing RD Findings
A Strict Constructionist View

RD findings only generalize to candidates at or very near the cut-point.

Thus RD findings are only relevant to decisions about marginal increases or decreases in a program.

They are not relevant to decisions about whether to open or close a program.
A More Expansive View

Imprecise control over ratings creates heterogeneity at the cut-point.

The less precise control over ratings is the more heterogeneity there is at the cut-point.

In the extreme, random ratings produce a sample at the cut-point that reflects the full eligible population.
How Imprecise Control Over Ratings Affects the Distribution of Counterfactual Outcomes at the Cut-Point of a Regression Discontinuity Design
An Even More Expansive “Old School” View: Extrapolation Beyond the Cut-point

A simple linear model

\[ \bar{Y} = \alpha + \beta_0 \cdot T + \beta_1 \cdot r \]

\[ \frac{d\bar{Y}}{dt} = \beta_0 \]

Separate T/C Linear Models

\[ \bar{Y} = \alpha + \beta_0 \cdot T + \beta_1 \cdot r + \beta_2 \cdot T \cdot r \]

\[ \frac{d\bar{Y}}{dT} = \beta_0 + \beta_2 \cdot r \]
Extrapolating Regression Discontinuity Findings Beyond the Cut-Point

Using a Simple Linear Model

Using Separate Treatment/Control Linear Models
For Further Reference and a Bibliography


http://www.mdrc.org/publications/539/abstract.html