Bias versus Precision in Regression Discontinuity Design Estimators

Nirav Mehta[1], Rebecca Maynard, and Nianbo Dong

March 4, 2010

[1]The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C050041-05 to the University of Pennsylvania. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

Why use RD Designs

- Experimentation can be expensive.
- Policy makers often have priors belief about the effective targeting.
  - They may assign "treatment" accordingly.
  - E.g. those scoring above "x" get the treatment, those below don't.
- RD designs may yield unbiased impact estimates
  - Near *the "cut-off"*.
  - Local Average Treatment Effects (LATE)

Limitations of RD Designs

Bias-Precision Trade-off

- Discard many observations to mitigate bias.
- Include observations further from cut-off to increase precision.

Our Question:

- How good are RD estimators?
  - How far do they diverge from the true impacts?
  - How is this divergence affected by sample inclusion criteria (caliper size)?
    - Bias
    - Precision

Our Strategy

- Two data sets for which we know the true impacts (LATE) *throughout the sample*.
    1. Observational data set with hypothetical treatment impacts (ECLS-K)
    2. Real experimental data set (New Chance)

Simulated Treatment Impacts: ECLS-K

Focus: Bias in RD estimates of the LATE under 3 treatment impact scenarios?

Simulated Treatment Impacts: ECLS-K

Focus: Bias in RD estimates of the LATE under 3 treatment impact scenarios?

- Each scenario specifies the treatment impact as a different function of the criterion measure used to determine the cut-off (z_i)
  1. constant (i.e. independent of z_i)
  2. linear function of z_i
  3. quadratic function of z_i

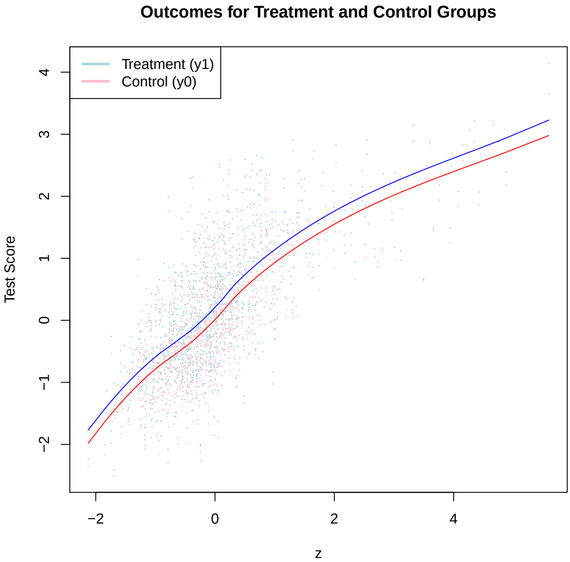
Simulated Treatment Impacts: ECLS-K

Focus: Bias in RD estimates of the LATE under 3 treatment impact scenarios?

- Each scenario specifies the treatment impact as a different function of the criterion measure used to determine the cut-off (z_i)
  1. constant (i.e. independent of z_i)
  2. linear function of z_i
  3. quadratic function of z_i
- z_i is child i's kindergarten reading achievement level (sd units)
- Outcome: Reading Achievement scores (standard deviation units)
  - $y0_i$ is her first grade reading achievement level (control group) (sd units)
  - $y1_i$ is the *simulated* outcome for i if she is in the treatment group (sd units)

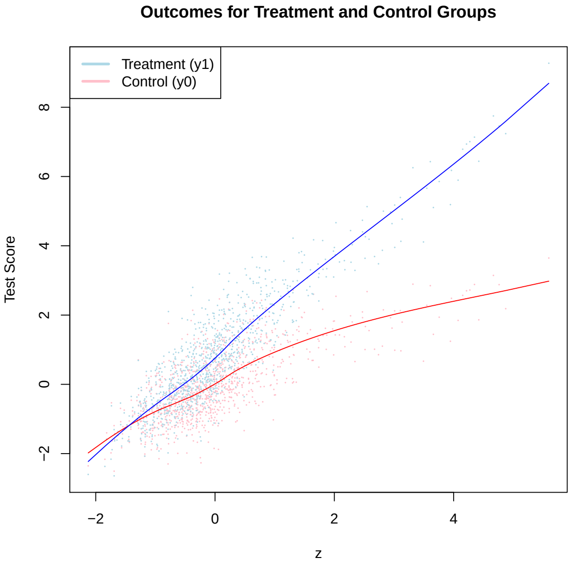
Simulated Treatment Impacts: ECLS-K

Scenario 1: Constant Treatment Impact



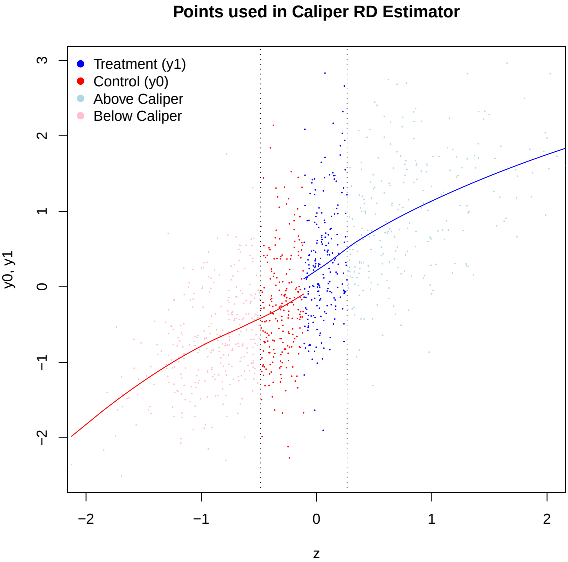
Simulated Treatment Impacts: ECLS-K

Scenario 2: Quadratic Treatment Impact



Simulated Treatment Impacts: ECLS-K

Non-parametric RD Estimator: Simple to Compute



Non-parametric vc Regression RD Estimator

- Non-parametric RD Estimator can't accomodate simple trend in data
- Local Linear Regression RD Estimator *can*.
  - Estimate a regression below and above the cut-off
  - Difference in intercepts is treatment effect.

Simulation Algorithm for Impact Scenarios

1. Pick $p \in P$, where $P = \{\text{constant}, \text{linear}, \text{quadratic}\}$.

Simulation Algorithm for Impact Scenarios

1. Pick $p \in P$, where $P = \{\text{constant}, \text{linear}, \text{quadratic}\}$.
2. Given the impact scenario p, for simulation $s \in S$, where $S = \{1, ..., 10,000\}$:

Simulation Algorithm for Impact Scenarios

1. Pick $p \in P$, where $P = \{\text{constant}, \text{linear}, \text{quadratic}\}$.
2. Given the impact scenario p, for simulation $s \in S$, where $S = \{1, ..., 10,000\}$:
   2.1 Create a synthetic dataset by:
      - Sampling original $(z_i, y0_i)$ pairs with replacement N times ($N = 1,000$). Store as $(z_i^s, y0_i^s)$
      - For each pair $(z_i^s, y0_i^s)$, simulate treatment impact according to $\Delta(z_i)$
      - shock ϵ_i is drawn i.i.d.

Simulation Algorithm for Impact Scenarios

1. Pick $p \in P$, where $P = \{\text{constant}, \text{linear}, \text{quadratic}\}$.
2. Given the impact scenario p, for simulation $s \in S$, where $S = \{1, ..., 10,000\}$:
   2.1 Create a synthetic dataset by:
   - Sampling original $(z_i, y0_i)$ pairs with replacement N times ($N = 1,000$). Store as $(z_i^s, y0_i^s)$
   - For each pair $(z_i^s, y0_i^s)$, simulate treatment impact according to $\Delta(z_i)$
   - shock ϵ_i is drawn i.i.d.

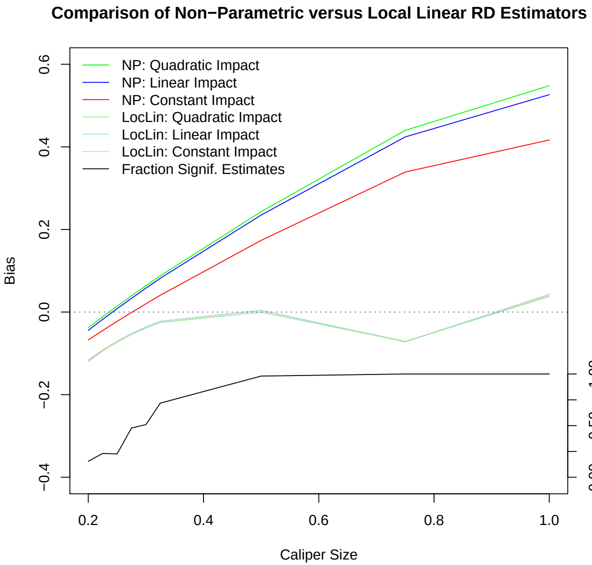
   2.2 For each caliper size, calculate non-parametric and regression RD estimates of treatment effect, store in a table X_s.

Simulation Algorithm for Impact Scenarios

1. Pick $p \in P$, where $P = \{\text{constant}, \text{linear}, \text{quadratic}\}$.
2. Given the impact scenario p, for simulation $s \in S$, where $S = \{1, ..., 10,000\}$:
   2.1 Create a synthetic dataset by:
      - Sampling original $(z_i, y0_i)$ pairs with replacement N times ($N = 1,000$). Store as $(z_i^s, y0_i^s)$
      - For each pair $(z_i^s, y0_i^s)$, simulate treatment impact according to $\Delta(z_i)$
      - shock ϵ_i is drawn i.i.d.

   2.2 For each caliper size, calculate non-parametric and regression RD estimates of treatment effect, store in a table X_s.
3. Average the RD estimates and significance
   - for each RD method and caliper size over all the results, X_s, for $s \in S$.

Simulated Treatment Impacts: ECLS-K



What we learned from the ECLS-K simulations

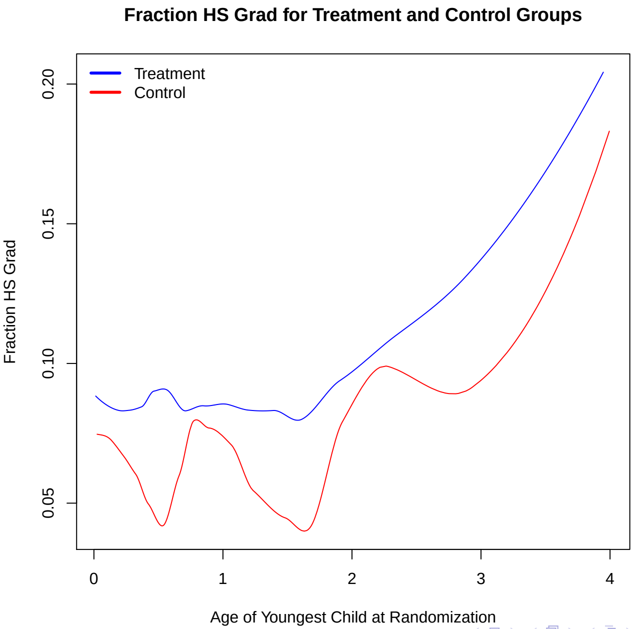
- Non-parametric RD estimators can't really handle a simple trend in the data, let alone a treatment effect which is a function of the index.
  - Problem is exacerbated by increased caliper size.
- The Local Linear Regression RD estimator *can* deal with smoothly varying, monotonic treatment impacts.

New Chance Experimental Data

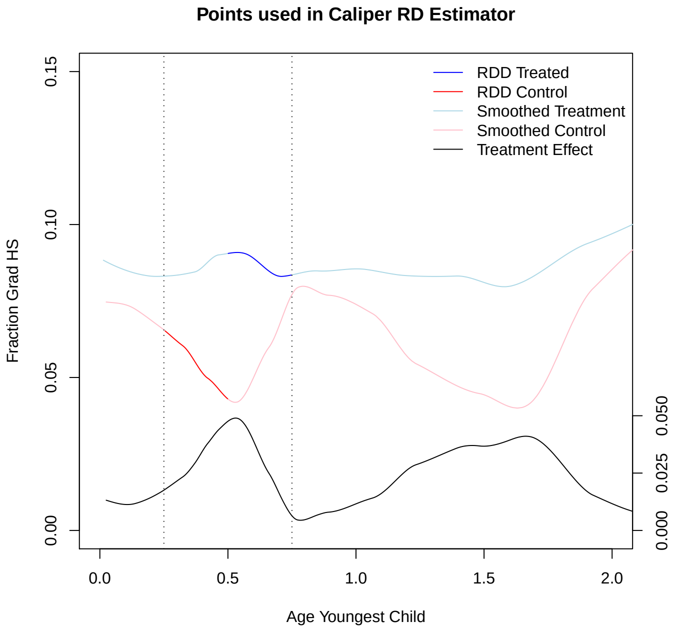
New Chance Demonstration: experiment in social services provision for teenage mothers.

- "True" (i.e. unbiased) impact is a function of criterion variable (z_i).
- z_i is the age of mother i's youngest child
- Outcome: High school graduation
  - $y0_i$ is whether she graduated from high school or obtained her GED 18 mo. after start of program if she was in the control group
  - $y1_i$ is her outcome if she is in the treatment group

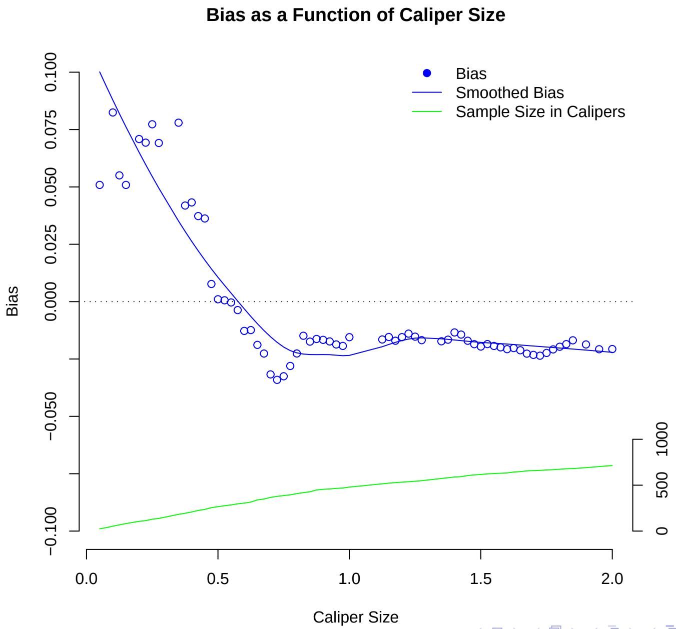
Treatment effects in New Chance, Function of Youngest Child Age



RD plot for New Chance, assuming treatment cut-off is 6 months



RD Estimates, by Caliper Size



What we learned from New Chance

- RD estimator is *very sensitive* to changes in caliper size.
- RD estimator is also *very sensitive* to individual data points when using small calipers.

Conclusions

- RD estimates often diverge from the "true" treatment effects.
- Precision considerations can exacerbate this divergence.
- LATE often isn't the same as ATE! (e.g. New Chance)
  - unless there is a uniform treatment effect