Regression Discontinuity Designs

Classic Theories and Recent Development

Ye Wang

Outline

- We start from considering the ideal experiment behind sharp RDD.

Outline

- We start from considering the ideal experiment behind sharp RDD.
- We then proceed to discuss the the asymptotic performance of the classic RDD estimators.

Outline

- We start from considering the ideal experiment behind sharp RDD.
- We then proceed to discuss the the asymptotic performance of the classic RDD estimators.
- Next, we discuss a special case in RDD where the running variable is discrete.

Outline

- We start from considering the ideal experiment behind sharp RDD.
- We then proceed to discuss the the asymptotic performance of the classic RDD estimators.
- Next, we discuss a special case in RDD where the running variable is discrete.
- We finally introduce two modern perspectives to understand RDD: local randomization and noise-induced randomization.

Roadmap

- Thistlethwaite and Campbell (1960) first develop this idea in psychology.

Roadmap

- Thistlethwaite and Campbell (1960) first develop this idea in psychology.
- Hahn, Todd, and Van der Klaauw (2001) formally discuss the identification assumptions in RDD.

Roadmap

- Thistlethwaite and Campbell (1960) first develop this idea in psychology.
- Hahn, Todd, and Van der Klaauw (2001) formally discuss the identification assumptions in RDD.
- Porter (2003) proposes the local regression estimator based on the idea in Fan and Gijbels (1996).

Roadmap

- Thistlethwaite and Campbell (1960) first develop this idea in psychology.
- Hahn, Todd, and Van der Klaauw (2001) formally discuss the identification assumptions in RDD.
- Porter (2003) proposes the local regression estimator based on the idea in Fan and Gijbels (1996).
- G. Imbens and Kalyanaraman (2012) introduce the first data-driven bandwidth selector.

Roadmap

- Thistlethwaite and Campbell (1960) first develop this idea in psychology.
- Hahn, Todd, and Van der Klaauw (2001) formally discuss the identification assumptions in RDD.
- Porter (2003) proposes the local regression estimator based on the idea in Fan and Gijbels (1996).
- G. Imbens and Kalyanaraman (2012) introduce the first data-driven bandwidth selector.
- Calonico, Cattaneo, and Titiunik (2014) address the insufficiency in G. Imbens and Kalyanaraman (2012) and derive the asymptotic distributions for a group of RDD estimators.

Roadmap

- Thistlethwaite and Campbell (1960) first develop this idea in psychology.
- Hahn, Todd, and Van der Klaauw (2001) formally discuss the identification assumptions in RDD.
- Porter (2003) proposes the local regression estimator based on the idea in Fan and Gijbels (1996).
- G. Imbens and Kalyanaraman (2012) introduce the first data-driven bandwidth selector.
- Calonico, Cattaneo, and Titiunik (2014) address the insufficiency in G. Imbens and Kalyanaraman (2012) and derive the asymptotic distributions for a group of RDD estimators.
- Calonico, Cattaneo, and Farrell (2020) suggest a bandwidth selector that is optimal for inference.

Roadmap

- Thistlethwaite and Campbell (1960) first develop this idea in psychology.
- Hahn, Todd, and Van der Klaauw (2001) formally discuss the identification assumptions in RDD.
- Porter (2003) proposes the local regression estimator based on the idea in Fan and Gijbels (1996).
- G. Imbens and Kalyanaraman (2012) introduce the first data-driven bandwidth selector.
- Calonico, Cattaneo, and Titiunik (2014) address the insufficiency in G. Imbens and Kalyanaraman (2012) and derive the asymptotic distributions for a group of RDD estimators.
- Calonico, Cattaneo, and Farrell (2020) suggest a bandwidth selector that is optimal for inference.
- McCrary (2008) develops a test for self-selection in RDD.

Roadmap

- Thistlethwaite and Campbell (1960) first develop this idea in psychology.
- Hahn, Todd, and Van der Klaauw (2001) formally discuss the identification assumptions in RDD.
- Porter (2003) proposes the local regression estimator based on the idea in Fan and Gijbels (1996).
- G. Imbens and Kalyanaraman (2012) introduce the first data-driven bandwidth selector.
- Calonico, Cattaneo, and Titiunik (2014) address the insufficiency in G. Imbens and Kalyanaraman (2012) and derive the asymptotic distributions for a group of RDD estimators.
- Calonico, Cattaneo, and Farrell (2020) suggest a bandwidth selector that is optimal for inference.
- McCrary (2008) develops a test for self-selection in RDD.
- Lee (2008) first applies RDD to analyze congressional elections in the US.

Roadmap

- Theoretical results in Calonico, Cattaneo, and Titiunik (2014) can be directly applied to analyze fuzzy RDD and kink

Roadmap

- Theoretical results in Calonico, Cattaneo, and Titiunik (2014) can be directly applied to analyze fuzzy RDD and kink
- Card et al. (2015) first introduce kink and fuzzy kink designs into econometrics.

Roadmap

- Theoretical results in Calonico, Cattaneo, and Titiunik (2014) can be directly applied to analyze fuzzy RDD and kink
- Card et al. (2015) first introduce kink and fuzzy kink designs into econometrics.
- Cattaneo et al. (2016) and Cattaneo et al. (2020) discuss RDD with multiple cutoffs and their extrapolation.

Roadmap

- Theoretical results in Calonico, Cattaneo, and Titiunik (2014) can be directly applied to analyze fuzzy RDD and kink
- Card et al. (2015) first introduce kink and fuzzy kink designs into econometrics.
- Cattaneo et al. (2016) and Cattaneo et al. (2020) discuss RDD with multiple cutoffs and their extrapolation.
- Kolesár and Rothe (2018) solve the problem of inference in RDD with a discrete running variable.

Roadmap

- Theoretical results in Calonico, Cattaneo, and Titiunik (2014) can be directly applied to analyze fuzzy RDD and kink
- Card et al. (2015) first introduce kink and fuzzy kink designs into econometrics.
- Cattaneo et al. (2016) and Cattaneo et al. (2020) discuss RDD with multiple cutoffs and their extrapolation.
- Kolesár and Rothe (2018) solve the problem of inference in RDD with a discrete running variable.
- Cattaneo, Frandsen, and Titiunik (2015) suggest the perspective of local randomization for RDD.

Roadmap

- Theoretical results in Calonico, Cattaneo, and Titiunik (2014) can be directly applied to analyze fuzzy RDD and kink
- Card et al. (2015) first introduce kink and fuzzy kink designs into econometrics.
- Cattaneo et al. (2016) and Cattaneo et al. (2020) discuss RDD with multiple cutoffs and their extrapolation.
- Kolesár and Rothe (2018) solve the problem of inference in RDD with a discrete running variable.
- Cattaneo, Frandsen, and Titiunik (2015) suggest the perspective of local randomization for RDD.
- G. Imbens and Wager (2019) and Eckles et al. (2020) develop the perspective of noise-induced randomization.

Set up

- We possess a dataset with N units.

Set up

- We possess a dataset with N units.
- We observe (Y_i, D_i, Z_i) for each unit i, $i \in \{1, 2, \ldots, N\}$.
- Z_i is called the running variable with density $f(z)$, which decides the value of the binary treatment D_i.

Set up

- We possess a dataset with N units.
- We observe (Y_i, D_i, Z_i) for each unit i, $i \in \{1, 2, \ldots, N\}$.
- Z_i is called the running variable with density $f(z)$, which decides the value of the binary treatment D_i.
- There exists a cutoff c (often set to be 0) such that $D_i = \mathbf{1}\{Z_i \geq c\}$.

Set up

- We possess a dataset with N units.
- We observe (Y_i, D_i, Z_i) for each unit i, $i \in \{1, 2, \ldots, N\}$.
- Z_i is called the running variable with density $f(z)$, which decides the value of the binary treatment D_i.
- There exists a cutoff c (often set to be 0) such that $D_i = \mathbf{1}\{Z_i \geq c\}$.
- The causal parameter of interest, τ_{SRD}, is defined as

$$\mathbb{E}[Y_i(1) - Y_i(0) | Z_i = c].$$

Set up

- We possess a dataset with N units.
- We observe (Y_i, D_i, Z_i) for each unit i, $i \in \{1, 2, \ldots, N\}$.
- Z_i is called the running variable with density $f(z)$, which decides the value of the binary treatment D_i.
- There exists a cutoff c (often set to be 0) such that $D_i = \mathbf{1}\{Z_i \geq c\}$.
- The causal parameter of interest, τ_{SRD}, is defined as

$$\mathbb{E}[Y_i(1) - Y_i(0) | Z_i = c].$$

- Notice that the causal parameter is a local one by definition.

Identification

- Identification in the classic RDD relies the assumption of continuity (Hahn, Todd, and Van der Klaauw 2001).

Identification

- Identification in the classic RDD relies the assumption of continuity (Hahn, Todd, and Van der Klaauw 2001).
- Define $\mu(z) = \mathbb{E}[Y_i|Z_i = z]$, $\mu_+ = \lim_{z \to c^+} \mu(z)$, and $\mu_- = \lim_{z \to c^-} \mu(z)$.

Identification

- Identification in the classic RDD relies the assumption of continuity (Hahn, Todd, and Van der Klaauw 2001).
- Define $\mu(z) = \mathbb{E}[Y_i | Z_i = z]$, $\mu_+ = \lim_{z \to c^+} \mu(z)$, and $\mu_- = \lim_{z \to c^-} \mu(z)$.
- Obviously, if $\mu_+ = \mathbb{E}[Y_i(1) | Z_i = c]$ and $\mu_- = \mathbb{E}[Y_i(0) | Z_i = c]$, then

$$\tau_{SRD} = \mu_+ - \mu_-.$$

Identification

- Identification in the classic RDD relies the assumption of continuity (Hahn, Todd, and Van der Klaauw 2001).
- Define $\mu(z) = \mathbb{E}[Y_i|Z_i = z]$, $\mu_+ = \lim_{z \to c^+} \mu(z)$, and $\mu_- = \lim_{z \to c^-} \mu(z)$.
- Obviously, if $\mu_+ = \mathbb{E}[Y_i(1)|Z_i = c]$ and $\mu_- = \mathbb{E}[Y_i(0)|Z_i = c]$, then

$$\tau_{SRD} = \mu_+ - \mu_-.$$

- All we need to do is to estimate μ_+ and μ_-.

Ideal experiment behind RDD

- Intuitively, the assumption of continuity requires that all the confounders (observable or not) vary smoothly around the cutoff c.

Ideal experiment behind RDD

- Intuitively, the assumption of continuity requires that all the confounders (observable or not) vary smoothly around the cutoff c.
- We can treat RDD as a random assignment in the block defined by $Z_i = c$.

Ideal experiment behind RDD

- Intuitively, the assumption of continuity requires that all the confounders (observable or not) vary smoothly around the cutoff c.
- We can treat RDD as a random assignment in the block defined by $Z_i = c$.
- Suppose there are 1,000 congressional elections, in all of which both candidates win 50% of all votes.

Ideal experiment behind RDD

- Intuitively, the assumption of continuity requires that all the confounders (observable or not) vary smoothly around the cutoff c.
- We can treat RDD as a random assignment in the block defined by $Z_i = c$.
- Suppose there are 1,000 congressional elections, in all of which both candidates win 50% of all votes.
- We then randomly select a winner for each election with a coin flip, which implies that $D_i \perp \{Y_i(1), Y_i(0)\} | Z_i = c$.

Ideal experiment behind RDD

- Intuitively, the assumption of continuity requires that all the confounders (observable or not) vary smoothly around the cutoff c.
- We can treat RDD as a random assignment in the block defined by $Z_i = c$.
- Suppose there are 1,000 congressional elections, in all of which both candidates win 50% of all votes.
- We then randomly select a winner for each election with a coin flip, which implies that $D_i \perp \{Y_i(1), Y_i(0)\} | Z_i = c$.
- The difference in the outcome reflects the impact of the winner's attributes (e.g party affiliation).

Ideal experiment behind RDD

- Unfortunately, we do not have that many elections with a 50-50 split in practice.

Ideal experiment behind RDD

- Unfortunately, we do not have that many elections with a 50-50 split in practice.
- All we can do is to approximate the two limits using data at hand.

Ideal experiment behind RDD

- Unfortunately, we do not have that many elections with a 50-50 split in practice.
- All we can do is to approximate the two limits using data at hand.
- The RDD estimate is thus biased by definition.

Ideal experiment behind RDD

- Unfortunately, we do not have that many elections with a 50-50 split in practice.
- All we can do is to approximate the two limits using data at hand.
- The RDD estimate is thus biased by definition.
- Yet the bias diminishes to zero as we have a larger sample, which usually leads to a smaller bandwidth.

Ideal experiment behind RDD

- Unfortunately, we do not have that many elections with a 50-50 split in practice.
- All we can do is to approximate the two limits using data at hand.
- The RDD estimate is thus biased by definition.
- Yet the bias diminishes to zero as we have a larger sample, which usually leads to a smaller bandwidth.
- In other words, we approach the ideal experiment as sample size increases.

Ideal experiment behind RDD

- Unfortunately, we do not have that many elections with a 50-50 split in practice.
- All we can do is to approximate the two limits using data at hand.
- The RDD estimate is thus biased by definition.
- Yet the bias diminishes to zero as we have a larger sample, which usually leads to a smaller bandwidth.
- In other words, we approach the ideal experiment as sample size increases.
- Remember that we need extra structural assumptions (continuity) for the trick to work.

Estimation

- To estimate μ_+ and μ_-, the most common choice is local regression.

Estimation

- To estimate μ_+ and μ_-, the most common choice is local regression.
- Suppose we have selected a bandwidth h, we just need to fit kernelized regression within the windows $[c-h, c]$ and $[c, c+h]$:

$$
\begin{aligned}
&(\hat{\mu}_+, \hat{\beta}_+) \\
&= \arg\min_{\alpha,\beta} \sum_{i=1}^{N} \mathbf{1}\{Z_i \geq c\} \left(Y_i - \mu - \beta(Z_i - c)\right)^2 \mathbf{K}\left(\frac{Z_i - c}{h_N}\right)
\end{aligned}
$$

- $(\hat{\mu}_-, \hat{\beta}_-)$ are similarly estimated.

Estimation

- To estimate μ_+ and μ_-, the most common choice is local regression.
- Suppose we have selected a bandwidth h, we just need to fit kernelized regression within the windows $[c-h, c]$ and $[c, c+h]$:

$$
\begin{aligned}
&(\hat{\mu}_+, \hat{\beta}_+) \\
&= \arg\min_{\alpha,\beta} \sum_{i=1}^{N} \mathbf{1}\{Z_i \geq c\} \left(Y_i - \mu - \beta(Z_i - c)\right)^2 \mathbf{K}\left(\frac{Z_i - c}{h_N}\right)
\end{aligned}
$$

- $(\hat{\mu}_-, \hat{\beta}_-)$ are similarly estimated.
- Then,

$$
\hat{\tau}_{SRD} = \hat{\mu}_+ - \hat{\mu}_-.
$$

Estimation

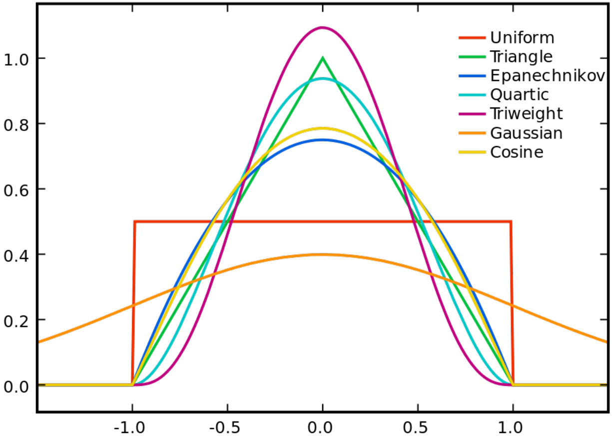
- The two regressions are local since we apply the kernel weights.

Estimation

- The two regressions are local since we apply the kernel weights.
- Intuitively, observations that are closer to the cutoff should receive a larger weight.

Estimation

- The two regressions are local since we apply the kernel weights.
- Intuitively, observations that are closer to the cutoff should receive a larger weight.
- There are multiple choices for the kernel.



Estimation in practice

- In practice, some scholars select the bandwidth using an existent algorithm and then manually fit two regressions within the selected window.

Estimation in practice

- In practice, some scholars select the bandwidth using an existent algorithm and then manually fit two regressions within the selected window.
- What is the problem of this approach?

Estimation in practice

- In practice, some scholars select the bandwidth using an existent algorithm and then manually fit two regressions within the selected window.
- What is the problem of this approach?
- It is inconsistent with the bandwidth selector and often leads to larger biases.

Estimation in practice

- In practice, some scholars select the bandwidth using an existent algorithm and then manually fit two regressions within the selected window.
- What is the problem of this approach?
- It is inconsistent with the bandwidth selector and often leads to larger biases.
- One may add higher order terms of $(Z_i - c)$ into the regressions.

Estimation in practice

- In practice, some scholars select the bandwidth using an existent algorithm and then manually fit two regressions within the selected window.
- What is the problem of this approach?
- It is inconsistent with the bandwidth selector and often leads to larger biases.
- One may add higher order terms of $(Z_i - c)$ into the regressions.
- But for sharp RDD, linear regression has the optimal rate of convergence due to its nice performance on the boundary (Porter 2003).

Estimation in practice

- Never fit global polynomials to estimate the two intercepts (Gelman and Imbens 2019).

Estimation in practice

- Never fit global polynomials to estimate the two intercepts (Gelman and Imbens 2019).
- We are interested in local quantities rather than global fitness.

Estimation in practice

- Never fit global polynomials to estimate the two intercepts (Gelman and Imbens 2019).
- We are interested in local quantities rather than global fitness.
- Estimates of the intercepts may be driven by points that are far away from the cutoff if you use global polynomials.

Estimation in practice

- Never fit global polynomials to estimate the two intercepts (Gelman and Imbens 2019).
- We are interested in local quantities rather than global fitness.
- Estimates of the intercepts may be driven by points that are far away from the cutoff if you use global polynomials.
- It is always critical to draw plots in RDD.

Estimation in practice

- Never fit global polynomials to estimate the two intercepts (Gelman and Imbens 2019).
- We are interested in local quantities rather than global fitness.
- Estimates of the intercepts may be driven by points that are far away from the cutoff if you use global polynomials.
- It is always critical to draw plots in RDD.
- The result would not be convincing if we cannot see the jump of the outcome variable from the plot.

Test the assumption of continuity

- As the assumption of continuity is crucial for identification in RDD, we should test its validity in practice.

Test the assumption of continuity

- As the assumption of continuity is crucial for identification in RDD, we should test its validity in practice.
- A popular approach is the placebo test: variables that are not affected by the treatment should change smoothly over the cutoff.
- For example, we can run the estimator for a covariate X to see whether the result is significant.

Test the assumption of continuity

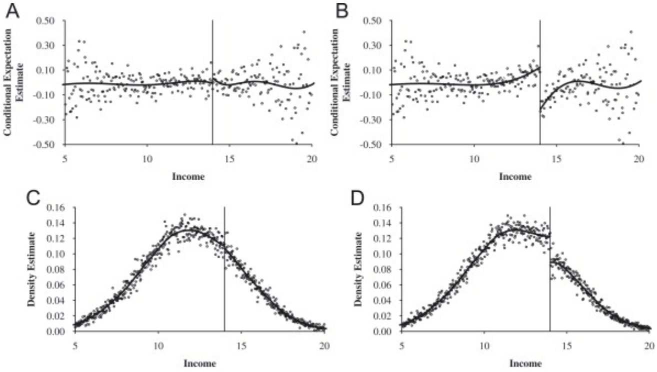
- As the assumption of continuity is crucial for identification in RDD, we should test its validity in practice.
- A popular approach is the placebo test: variables that are not affected by the treatment should change smoothly over the cutoff.
- For example, we can run the estimator for a covariate X to see whether the result is significant.
- See Meyersson (2014) for examples

Test the assumption of continuity

- The assumption is violated if units in the sample self-select into one side of the cutoff.

Test the assumption of continuity

- The assumption is violated if units in the sample self-select into one side of the cutoff.
- For example, students may cheat to meet the requirement of a scholarship.
- As a result, the density function $f(z)$ will not change smoothly across the cutoff.



Test the assumption of continuity

- McCrary (2008) develop the first formal test of the discontinuity in density.

Test the assumption of continuity

- McCrary (2008) develop the first formal test of the discontinuity in density.
- Essentially, we are testing whether

$$\theta = \ln \lim_{z \to 0^+} f(z) - \ln \lim_{z \to 0^-} f(z)$$

has a significant deviation from 0.

Test the assumption of continuity

- McCrary (2008) develop the first formal test of the discontinuity in density.
- Essentially, we are testing whether

$$\theta = \ln \lim_{z \to 0^+} f(z) - \ln \lim_{z \to 0^-} f(z)$$

  has a significant deviation from 0.
- Similarly, we need to estimate the two boundary points of $\ln f(z)$ using local regression (notice that Y_i is not needed).

Test the assumption of continuity

- McCrary (2008) develop the first formal test of the discontinuity in density.
- Essentially, we are testing whether

$$\theta = \ln \lim_{z \to 0^+} f(z) - \ln \lim_{z \to 0^-} f(z)$$

  has a significant deviation from 0.
- Similarly, we need to estimate the two boundary points of $\ln f(z)$ using local regression (notice that Y_i is not needed).
- Cattaneo, Jansson, and Ma (2020) provide an augmented algorithm for non-parametric density estimation.

An example

- We will work with the Meyersson (2014) paper: “Islamic Rule and the Empowerment of the Poor and the Pious”
- The paper shows a (local) result: the victory of Islamic parties in Turkey resulted in better outcomes for women.
- Running variable: the difference in vote share between the largest Islamic party and the largest secular party (not two party)
- Outcome that we’ll look at: high school education

Set up the data

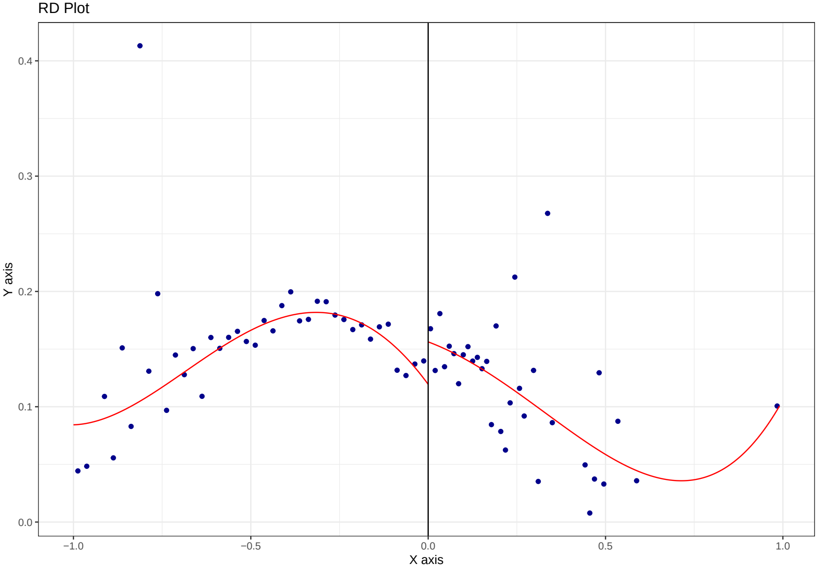
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -1.0000 -0.4600 -0.3102 -0.2786 -0.1061  0.9905     544
```

Estimation

```
## Call: rdrobust
##
## Number of Obs.                 2630
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 2315         315
## Eff. Number of Obs.             529         266
## Order est. (p)                    1           1
## Order bias  (q)                   2           2
## BW est. (h)                   0.172       0.172
## BW bias (b)                   0.286       0.286
## rho (h/b)                     0.603       0.603
## Unique Obs.                    2313         315
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional     0.030     0.014     2.116     0.034     [0.002 , 0.058]
##         Robust         -         -     1.776     0.076    [-0.003 , 0.063]
## =============================================================================
```

Plot it

```
rdplot(d$hischshr1520f, d$iwm94, p = 4)
```

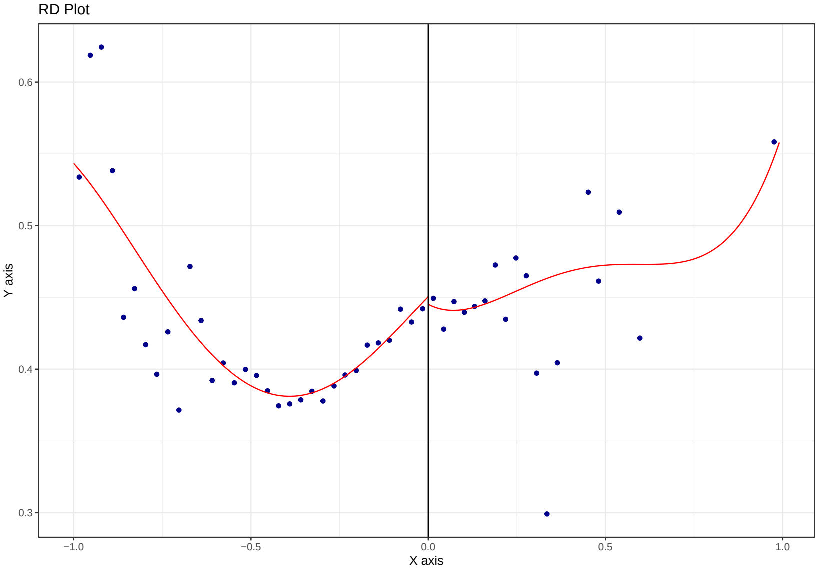


Placebo tests

- Do placebo tests on other covariates and other outcomes.

```
## $coef
##                        Coeff
## Conventional   0.004097863
## Bias-Corrected 0.008070629
## Robust         0.008070629
##
## $se
##                 Std. Err.
## Conventional   0.01227104
## Bias-Corrected 0.01227104
## Robust         0.01408919
```

Placebo plot



More Placebos

```
## $coef
##                     Coeff
## Conventional   0.01285314
## Bias-Corrected 0.01263466
## Robust         0.01263466
##
## $se
##                 Std. Err.
## Conventional   0.01403235
## Bias-Corrected 0.01403235
## Robust         0.01691543
```

```
## $coef
##                       Coeff
## Conventional   0.0039695120
## Bias-Corrected 0.0008440962
## Robust         0.0008440962
##
## $se
##                 Std. Err.
## Conventional   0.01537558
## Bias-Corrected 0.01537558
```

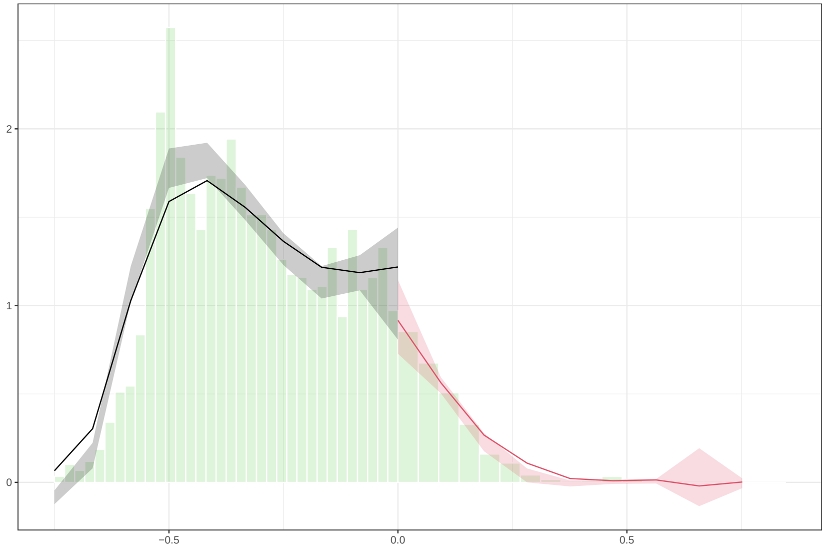
Sorting

- Density tests are also a good way to examine the possibility of sorting.

```
##
## Manipulation testing using local polynomial density est:
##
## Number of obs =        2660
## Model =                unrestricted
## Kernel =               triangular
## BW method =            estimated
## VCE method =           jackknife
##
## c = 0                  Left of c             Right of c
## Number of obs          2332                  328
## Eff. Number of obs     769                   314
## Order est. (p)         2                     2
## Order bias (q)         3                     3
## BW est. (h)            0.25                  0.282
##
```

Density Plot

```
rdplotdensity(rd_density, d$iwm94)
```



$Estl

Kink

```
## Call: rdrobust
##
## Number of Obs.                 2630
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 2315          315
## Eff. Number of Obs.             428          238
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   0.139        0.139
## BW bias (b)                   0.286        0.286
## rho (h/b)                     0.486        0.486
## Unique Obs.                    2313          315
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional     0.020     0.242     0.082     0.934    [-0.455 , 0.495]
##         Robust         -         -     0.205     0.838    [-0.689 , 0.849]
## =============================================================================
```

Bias of RDD estimation

- Let's introduce some notations for deriving the bias of RDD estimation.
- Denote $(Y_1, Y_2, \ldots, Y_N)$ as $\mathbf{Y}$, $(Z_1, Z_2, \ldots, Z_N)$ as $\mathbf{Z}$, and

$$\mathbf{R} = (\iota, \mathbf{Z})$$

$$\mathbf{W}_+ = \{\mathbf{1}\{Z_i \geq c\} K\left(\frac{Z_i - c}{h_N}\right)\}_{N \times N}$$

$$\mathbf{M} = (\mu(Z_1), \mu(Z_2), \ldots, \mu(Z_N))'$$

  where ι is a vector with N 1s and $\mathbf{W}$ is a diagonal matrix of weights.
- We use $\mu_+^{(k)}$ to denote the k-th order derivative of μ_+ (similar for μ_-) and $\sigma^2(z)$ to denote $Var[Y_i|Z_i = z]$.

Bias of RDD estimation

- Notice that $\mathbf{Y} = \mathbf{M} + \mathbf{e}$.

Bias of RDD estimation

- Notice that $\mathbf{Y} = \mathbf{M} + \mathbf{e}$.
- The estimate $\hat{\mu}_+$ equals to the first row of

$$
\begin{aligned}
&(\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}(\mathbf{R}'\mathbf{W}_+\mathbf{Y}) \\
=&(\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}(\mathbf{R}'\mathbf{W}_+\mathbf{M}) + (\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}(\mathbf{R}'\mathbf{W}_+\mathbf{e})
\end{aligned}
$$

Bias of RDD estimation

- Notice that $\mathbf{Y} = \mathbf{M} + \mathbf{e}$.
- The estimate $\hat{\mu}_+$ equals to the first row of

$$
\begin{aligned}
&(\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}(\mathbf{R}'\mathbf{W}_+\mathbf{Y}) \\
=&(\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}(\mathbf{R}'\mathbf{W}_+\mathbf{M}) + (\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}(\mathbf{R}'\mathbf{W}_+\mathbf{e})
\end{aligned}
$$

- Expectation of the second term is zero and we have the Taylor expansion for $\mu(Z_i)$:

$$
\mu(Z_i) = \mu_+ + \mu_+^{(1)}(0)Z_i + \frac{\mu_+^{(2)}(0)}{2}Z_i^2 + \nu_i
$$

Bias of RDD estimation

- Notice that $\mathbf{Y} = \mathbf{M} + \mathbf{e}$.
- The estimate $\hat{\mu}_+$ equals to the first row of

$$
\begin{aligned}
&(\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}(\mathbf{R}'\mathbf{W}_+\mathbf{Y}) \\
=&(\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}(\mathbf{R}'\mathbf{W}_+\mathbf{M}) + (\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}(\mathbf{R}'\mathbf{W}_+\mathbf{e})
\end{aligned}
$$

- Expectation of the second term is zero and we have the Taylor expansion for $\mu(Z_i)$:

$$\mu(Z_i) = \mu_+ + \mu_+^{(1)}(0)Z_i + \frac{\mu_+^{(2)}(0)}{2}Z_i^2 + \nu_i$$

- Hence,

$$\mathbf{M} = \mathbf{R}\begin{pmatrix}\mu_+ \\ \mu_+^{(1)}(0)\end{pmatrix} + \mathbf{S}_2\frac{\mu_+^{(2)}(0)}{2} + \nu$$

where $\mathbf{S}_2 = (Z_1^2, Z_2^2, \ldots, Z_N^2)$ and $\nu = (\nu_1, \nu_2, \ldots, \nu_N)$.

Bias of RDD estimation

- Now, the estimation bias of $\hat{\mu}_+$, $\mathbb{E}[\hat{\mu}_+] - \mu_+$, is the first row of

$$\mu_+^{(1)}(0) + (\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}\left(\mathbf{R}'\mathbf{W}_+\mathbf{S}_2\frac{\mu_+^{(2)}(0)}{2}\right) + (\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}\left(\mathbf{R}'\mathbf{W}_+\nu\right)$$

Bias of RDD estimation

- Now, the estimation bias of $\hat{\mu}_+$, $\mathbb{E}[\hat{\mu}_+] - \mu_+$, is the first row of

$$\mu_+^{(1)}(0) + (\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}\left(\mathbf{R}'\mathbf{W}_+\mathbf{S}_2\frac{\mu_+^{(2)}(0)}{2}\right) + (\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}\left(\mathbf{R}'\mathbf{W}_+\nu\right)$$

- The convergence rates of these two terms rely on the properties of the kernel.

Bias of RDD estimation

- Now, the estimation bias of $\hat{\mu}_+$, $\mathbb{E}[\hat{\mu}_+] - \mu_+$, is the first row of

$$\mu_+^{(1)}(0) + (\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}\left(\mathbf{R}'\mathbf{W}_+\mathbf{S}_2\frac{\mu_+^{(2)}(0)}{2}\right) + (\mathbf{R}'\mathbf{W}_+\mathbf{R})^{-1}\left(\mathbf{R}'\mathbf{W}_+\nu\right)$$

- The convergence rates of these two terms rely on the properties of the kernel.
- Via some cumbersome calculation, we can see that

$$\mathbb{E}[\hat{\mu}_+] - \mu_+ = C_1\mu_+^{(2)}(0)h^2 + o_p(h^2)$$

Bias of RDD estimation

- We can similarly derive the variance of $\hat{\mu}_+$ using the properties of regression:

$$\mathbb{V}[\hat{\mu}_+] = \frac{C_2}{Nh} \frac{\sigma_+^2(0)}{f_+(0)} + o_p(\frac{1}{Nh})$$

Bias of RDD estimation

- We can similarly derive the variance of $\hat{\mu}_+$ using the properties of regression:

$$\mathbb{V}[\hat{\mu}_+] = \frac{C_2}{Nh}\frac{\sigma_+^2(0)}{f_+(0)} + o_p(\frac{1}{Nh})$$

- Obviously, the bias and the variance of $\hat{\mu}_-$ have similar forms.

Bias of RDD estimation

- We can similarly derive the variance of $\hat{\mu}_+$ using the properties of regression:

$$\mathbb{V}[\hat{\mu}_+] = \frac{C_2}{Nh}\frac{\sigma_+^2(0)}{f_+(0)} + o_p(\frac{1}{Nh})$$

- Obviously, the bias and the variance of $\hat{\mu}_-$ have similar forms.
- More generally, we can estimate the k-th order derivative of μ_+ and μ_- with a p-th order local regression.
- The bias will be of order $p + 1$.

Bandwidth selection for optimizing MSE

- Before G. Imbens and Kalyanaraman (2012), the practice is to minimize the regression's MSE on the entire support using cross-validation.

Bandwidth selection for optimizing MSE

- Before G. Imbens and Kalyanaraman (2012), the practice is to minimize the regression's MSE on the entire support using cross-validation.
- G. Imbens and Kalyanaraman (2012) argue that we should select a bandwidth to minimize the MSE of estimation:

$$\begin{aligned} MSE(h_N) =& \mathbb{E}\left[\hat{\tau}_{SRD} - \tau_{SRD} | \mathbf{Z}\right]^2 \\ =& \left(\mathbb{E}\left[\hat{\tau}_{SRD} | \mathbf{Z}\right] - \tau_{SRD}\right)^2 + Var\left[\hat{\tau}_{SRD} | \mathbf{Z}\right] \\ =& Bias^2 + Variance. \end{aligned}$$

Bandwidth selection for optimizing MSE

- G. Imbens and Kalyanaraman (2012) show that in practice we can minimize the asymptotic MSE:

$$AMSE(h_N) = C_1 h_N^4 \left(\mu_+^{(2)}(0) - \mu_-^{(2)}(0)\right)^2 + \frac{C_2}{N h_N} \frac{\sigma^2(0)}{f(0)}$$

Bandwidth selection for optimizing MSE

- G. Imbens and Kalyanaraman (2012) show that in practice we can minimize the asymptotic MSE:

$$AMSE(h_N) = C_1 h_N^4 \left(\mu_+^{(2)}(0) - \mu_-^{(2)}(0)\right)^2 + \frac{C_2}{N h_N} \frac{\sigma^2(0)}{f(0)}$$

- From the expression we can solve the optimal bandwidth:

$$h_N^* = C \left(\frac{\frac{\sigma^2(0)}{f(0)}}{\left(\mu_+^{(2)}(0) - \mu_-^{(2)}(0)\right)^2} \right)^{\frac{1}{5}} N^{-\frac{1}{5}}.$$

Bandwidth selection for optimizing MSE

- G. Imbens and Kalyanaraman (2012) show that in practice we can minimize the asymptotic MSE:

$$AMSE(h_N) = C_1 h_N^4 \left(\mu_+^{(2)}(0) - \mu_-^{(2)}(0)\right)^2 + \frac{C_2}{Nh_N}\frac{\sigma^2(0)}{f(0)}$$

- From the expression we can solve the optimal bandwidth:

$$h_N^* = C \left(\frac{\frac{\sigma^2(0)}{f(0)}}{\left(\mu_+^{(2)}(0) - \mu_-^{(2)}(0)\right)^2} \right)^{\frac{1}{5}} N^{-\frac{1}{5}}.$$

- In practice, we can estimate h_N^* with a plug-in estimator.

Bias correction

- G. Imbens and Kalyanaraman (2012) do not prove the asymptotic normality of the RDD estimate.

Bias correction

- G. Imbens and Kalyanaraman (2012) do not prove the asymptotic normality of the RDD estimate.
- Calonico, Cattaneo, and Titiunik (2014) study the asymptotic distribution of the studentized RDD estimate:

$$\frac{\hat{\tau}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}}.$$

Bias correction

- G. Imbens and Kalyanaraman (2012) do not prove the asymptotic normality of the RDD estimate.
- Calonico, Cattaneo, and Titiunik (2014) study the asymptotic distribution of the studentized RDD estimate:

$$\frac{\hat{\tau}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}}.$$

- They show that the bandwidth selected via the previous algorithm is too wide to guarantee the the asymptotic normality of the studentized estimate.

Bias correction

- G. Imbens and Kalyanaraman (2012) do not prove the asymptotic normality of the RDD estimate.
- Calonico, Cattaneo, and Titiunik (2014) study the asymptotic distribution of the studentized RDD estimate:

$$\frac{\hat{\tau}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}}.$$

- They show that the bandwidth selected via the previous algorithm is too wide to guarantee the the asymptotic normality of the studentized estimate.
- We need $h_N = o_p(N^{-\frac{1}{5}})$ while the algorithm leads to $h_N = O_p(N^{-\frac{1}{5}})$.

Bias correction

- G. Imbens and Kalyanaraman (2012) do not prove the asymptotic normality of the RDD estimate.
- Calonico, Cattaneo, and Titiunik (2014) study the asymptotic distribution of the studentized RDD estimate:

$$\frac{\hat{\tau}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}}.$$

- They show that the bandwidth selected via the previous algorithm is too wide to guarantee the the asymptotic normality of the studentized estimate.
- We need $h_N = o_p(N^{-\frac{1}{5}})$ while the algorithm leads to $h_N = O_p(N^{-\frac{1}{5}})$.
- Consequently, the studentized estimate will be asymptotically biased.

Bias correction

- Intuitively,

$$\frac{\hat{\tau}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}} = \frac{\hat{\tau}_{SRD} - \mathbb{E}[\hat{\tau}_{SRD}]}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}} + \frac{\mathbb{E}[\hat{\tau}_{SRD}] - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}}.$$

Bias correction

- Intuitively,

$$\frac{\hat{\tau}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}} = \frac{\hat{\tau}_{SRD} - \mathbb{E}[\hat{\tau}_{SRD}]}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}} + \frac{\mathbb{E}[\hat{\tau}_{SRD}] - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}}.$$

- The first term is a weighted average of residuals and converges to $N(0, 1)$ by CLT.

Bias correction

- Intuitively,

$$\frac{\hat{\tau}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}} = \frac{\hat{\tau}_{SRD} - \mathbb{E}[\hat{\tau}_{SRD}]}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}} + \frac{\mathbb{E}[\hat{\tau}_{SRD}] - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}}.$$

- The first term is a weighted average of residuals and converges to $N(0,1)$ by CLT.
- We need to guarantee that the second term is $o_p(1)$.

Bias correction

- Intuitively,

$$\frac{\hat{\tau}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}} = \frac{\hat{\tau}_{SRD} - \mathbb{E}[\hat{\tau}_{SRD}]}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}} + \frac{\mathbb{E}[\hat{\tau}_{SRD}] - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}}.$$

- The first term is a weighted average of residuals and converges to $N(0,1)$ by CLT.
- We need to guarantee that the second term is $o_p(1)$.
- Remember that the numerator is $O_p(h^2)$ and the denominator is $O_p(\frac{1}{\sqrt{Nh}})$, thus the total bias is $O_p(\sqrt{Nh^5})$.

Bias correction

- Calonico, Cattaneo, and Titiunik (2014) provide a bias-correction estimator $\hat{\tau}^{bc}_{SRD}$.

Bias correction

- Calonico, Cattaneo, and Titiunik (2014) provide a bias-correction estimator $\hat{\tau}^{bc}_{SRD}$.
- Intuitively, we use another local regression estimator with bandwidth b_N to estimate the first-order derivative of μ_+ and μ_- and subtract them from $\hat{\mu}_+$ and $\hat{\mu}_-$.

Bias correction

- Calonico, Cattaneo, and Titiunik (2014) provide a bias-correction estimator $\hat{\tau}^{bc}_{SRD}$.
- Intuitively, we use another local regression estimator with bandwidth b_N to estimate the first-order derivative of μ_+ and μ_- and subtract them from $\hat{\mu}_+$ and $\hat{\mu}_-$.
- Of course, bias correction introduces extra uncertainty (from the extra local regression) into the estimate, hence the variance has to be adjusted accordingly.

Bias correction

- Calonico, Cattaneo, and Titiunik (2014) provide a bias-correction estimator $\hat{\tau}^{bc}_{SRD}$.
- Intuitively, we use another local regression estimator with bandwidth b_N to estimate the first-order derivative of μ_+ and μ_- and subtract them from $\hat{\mu}_+$ and $\hat{\mu}_-$.
- Of course, bias correction introduces extra uncertainty (from the extra local regression) into the estimate, hence the variance has to be adjusted accordingly.
- They propose two variance estimators, one based on regression analysis and the other based on the idea of nearest neighborhood matching (Abadie and Imbens 2006).

Bias correction

- Calonico, Cattaneo, and Titiunik (2014) prove that

$$\frac{\hat{\tau}^{bc}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}^{bc}_{SRD}]}} \to N(0,1)$$

as long as $N \min\{b_N^5, h_N^5\} \max\{b_N^2, h_N^2\} \to 0$.

Bias correction

- Calonico, Cattaneo, and Titiunik (2014) prove that
$$\frac{\hat{\tau}_{SRD}^{bc} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}_{SRD}^{bc}]}} \to N(0,1)$$
as long as $N \min\{b_N^5, h_N^5\} \max\{b_N^2, h_N^2\} \to 0$.
- Now the numerator is $O_p(h^3 + h^2 b^2)$ and the denominator is $O_p(\frac{1}{\sqrt{Nh}} + \frac{1}{\sqrt{Nh^4 b^5}})$.

Bias correction

- Calonico, Cattaneo, and Titiunik (2014) prove that

$$\frac{\hat{\tau}^{bc}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}^{bc}_{SRD}]}} \to N(0,1)$$

  as long as $N \min\{b_N^5, h_N^5\} \max\{b_N^2, h_N^2\} \to 0$.
- Now the numerator is $O_p(h^3 + h^2 b^2)$ and the denominator is $O_p(\frac{1}{\sqrt{Nh}} + \frac{1}{\sqrt{Nh^4 b^5}})$.
- In other words, we can still use the algorithm in G. Imbens and Kalyanaraman (2012) to select the bandwidth.

Bias correction

- Calonico, Cattaneo, and Titiunik (2014) prove that

$$\frac{\hat{\tau}^{bc}_{SRD} - \tau_{SRD}}{\sqrt{\mathbb{V}[\hat{\tau}^{bc}_{SRD}]}} \to N(0,1)$$

  as long as $N \min\{b_N^5, h_N^5\} \max\{b_N^2, h_N^2\} \to 0$.
- Now the numerator is $O_p(h^3 + h^2 b^2)$ and the denominator is $O_p(\frac{1}{\sqrt{Nh}} + \frac{1}{\sqrt{Nh^4 b^5}})$.
- In other words, we can still use the algorithm in G. Imbens and Kalyanaraman (2012) to select the bandwidth.
- We just need to modify the obtained estimate to ensure asymptotic normality.

RDD with a discrete running variable

- The classic theory developed in Calonico, Cattaneo, and Titiunik (2014) runs into difficulties when the running variable is discrete.

RDD with a discrete running variable

- The classic theory developed in Calonico, Cattaneo, and Titiunik (2014) runs into difficulties when the running variable is discrete.
- By definition, there cannot be more observations around the cutoff point as N increases.

RDD with a discrete running variable

- The classic theory developed in Calonico, Cattaneo, and Titiunik (2014) runs into difficulties when the running variable is discrete.
- By definition, there cannot be more observations around the cutoff point as N increases.
- Kolesár and Rothe (2018) provide a finite-sample confidence interval in this case.

RDD with a discrete running variable

- The classic theory developed in Calonico, Cattaneo, and Titiunik (2014) runs into difficulties when the running variable is discrete.
- By definition, there cannot be more observations around the cutoff point as N increases.
- Kolesár and Rothe (2018) provide a finite-sample confidence interval in this case.
- The CI does not rely on asymptotics and holds for any fixed N.

RDD with a discrete running variable

- The classic theory developed in Calonico, Cattaneo, and Titiunik (2014) runs into difficulties when the running variable is discrete.
- By definition, there cannot be more observations around the cutoff point as N increases.
- Kolesár and Rothe (2018) provide a finite-sample confidence interval in this case.
- The CI does not rely on asymptotics and holds for any fixed N.
- The intuition is to bound the curvature of $\mu(z)$ and consider the worst scenario.

RDD with a discrete running variable

- First notice that the local regression estimator is simply a weighted average of the outcome:

$$\hat{\tau}_{SRD} = \sum_{Z_i \geq 0} \hat{\gamma}_{+}(Z_i) Y_i - \sum_{Z_i \leq 0} \hat{\gamma}_{-}(Z_i) Y_i$$

  where $\hat{\gamma}_{i,+}$ and $\hat{\gamma}_{i,-}$ are weights that are only dependent on Z.
- In particular, if we use the rectangular kernel, we have

$$\hat{\gamma}_{+}(Z_i) = \frac{\sum_{0 \leq Z_j \leq h_N} Z_j^2 - Z_i \sum_{0 \leq Z_j \leq h_N} Z_j}{\sum_{0 \leq Z_j \leq h_N} Z_j^2 - (\sum_{0 \leq Z_j \leq h_N} Z_j)^2 / N^*}.$$

  where $N^* = \#\{j : 0 \leq Z_j \leq h_N\}$.

RDD with a discrete running variable

- Let's assume that $|\mu^{(2)}(z)| \leq C$, then the bias of $\hat{\tau}_{SRD}$ is bounded by

$$B = C \sum_{i=1}^{N} |\hat{\gamma}_{+}(Z_i) + \hat{\gamma}_{-}(Z_i)| Z_i^2$$

RDD with a discrete running variable

- Let's assume that $|\mu^{(2)}(z)| \leq C$, then the bias of $\hat{\tau}_{SRD}$ is bounded by

$$B = C \sum_{i=1}^{N} |\hat{\gamma}_{+}(Z_i) + \hat{\gamma}_{-}(Z_i)| Z_i^2$$

- Kolesár and Rothe (2018) show that the one-sided CI of $\hat{\tau}_{SRD}$ has the form $[\hat{c}, \infty]$, where

$$\hat{c} = \hat{\tau}_{SRD} - B - z_{1-\alpha}\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}$$

RDD with a discrete running variable

- Let's assume that $|\mu^{(2)}(z)| \le C$, then the bias of $\hat{\tau}_{SRD}$ is bounded by

$$B = C \sum_{i=1}^{N} |\hat{\gamma}_{+}(Z_i) + \hat{\gamma}_{-}(Z_i)| Z_i^2$$

- Kolesár and Rothe (2018) show that the one-sided CI of $\hat{\tau}_{SRD}$ has the form $[\hat{c}, \infty]$, where

$$\hat{c} = \hat{\tau}_{SRD} - B - z_{1-\alpha}\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}$$

- Comparing the result with that from Calonico, Cattaneo, and Titiunik (2014), the only difference is that we need to bound the bias using the "smooth parameter" C.

RDD with a discrete running variable

- Let's assume that $|\mu^{(2)}(z)| \leq C$, then the bias of $\hat{\tau}_{SRD}$ is bounded by

$$B = C \sum_{i=1}^{N} |\hat{\gamma}_{+}(Z_i) + \hat{\gamma}_{-}(Z_i)| Z_i^2$$

- Kolesár and Rothe (2018) show that the one-sided CI of $\hat{\tau}_{SRD}$ has the form $[\hat{c}, \infty]$, where

$$\hat{c} = \hat{\tau}_{SRD} - B - z_{1-\alpha}\sqrt{\mathbb{V}[\hat{\tau}_{SRD}]}$$

- Comparing the result with that from Calonico, Cattaneo, and Titiunik (2014), the only difference is that we need to bound the bias using the "smooth parameter" C.
- The problem here leads to some novel perspectives to understand RDD.

RDD as local randomization

- We have discussed that the classic RDD can be seen as a simple experiment when $Z = 0$.

RDD as local randomization

- We have discussed that the classic RDD can be seen as a simple experiment when $Z = 0$.
- Yet the conventional analysis is not a design-based approach as we do not try to model the assignment process.

RDD as local randomization

- We have discussed that the classic RDD can be seen as a simple experiment when $Z = 0$.
- Yet the conventional analysis is not a design-based approach as we do not try to model the assignment process.
- A recent perspective treats RDD as an experiment that is implemented where $Z \in [-h, h]$.

RDD as local randomization

- We have discussed that the classic RDD can be seen as a simple experiment when $Z = 0$.
- Yet the conventional analysis is not a design-based approach as we do not try to model the assignment process.
- A recent perspective treats RDD as an experiment that is implemented where $Z \in [-h, h]$.
- If we know h, classic approaches (regression, weighting, matching, etc.) can be applied to estimate the treatment effect.

RDD as local randomization

- We have discussed that the classic RDD can be seen as a simple experiment when $Z = 0$.
- Yet the conventional analysis is not a design-based approach as we do not try to model the assignment process.
- A recent perspective treats RDD as an experiment that is implemented where $Z \in [-h, h]$.
- If we know h, classic approaches (regression, weighting, matching, etc.) can be applied to estimate the treatment effect.
- Cattaneo, Frandsen, and Titiunik (2015) suggest that we should find h by balancing all the covariates.

RDD as local randomization

- We have discussed that the classic RDD can be seen as a simple experiment when $Z = 0$.
- Yet the conventional analysis is not a design-based approach as we do not try to model the assignment process.
- A recent perspective treats RDD as an experiment that is implemented where $Z \in [-h, h]$.
- If we know h, classic approaches (regression, weighting, matching, etc.) can be applied to estimate the treatment effect.
- Cattaneo, Frandsen, and Titiunik (2015) suggest that we should find h by balancing all the covariates.
- The perspective is criticized by Eckles et al. (2020): if the treatment is randomly assigned on $[-h, h]$, then $\mu(z)$ should be constant on the interval.

RDD as noise-induced randomization

- Eckles et al. (2020) argue that Z should be seen as a noisy measure of the true confounder U.

RDD as noise-induced randomization

- Eckles et al. (2020) argue that Z should be seen as a noisy measure of the true confounder U.
- For example, U_i is one's true capability and Z_i is her test score.

RDD as noise-induced randomization

- Eckles et al. (2020) argue that Z should be seen as a noisy measure of the true confounder U.
- For example, U_i is one's true capability and Z_i is her test score.
- We assume the measurement error is normally distributed

$$Z_i | U_i \sim \mathcal{N}(U_i, \nu^2)$$

and conditionally independent to the potential outcomes

$$\{Y_i(1), Y_i(0)\} \perp Z_i | U_i$$

RDD as noise-induced randomization

- Eckles et al. (2020) argue that Z should be seen as a noisy measure of the true confounder U.
- For example, U_i is one's true capability and Z_i is her test score.
- We assume the measurement error is normally distributed

$$Z_i | U_i \sim \mathcal{N}(U_i, \nu^2)$$

  and conditionally independent to the potential outcomes

$$\{Y_i(1), Y_i(0)\} \perp Z_i | U_i$$

- The assumptions imply that

$$\{Y_i(1), Y_i(0)\} \perp D_i | U_i$$

  and the only issue is that U_i is unobservable.

RDD as noise-induced randomization

- Eckles et al. (2020) claim that it is possible to estimate the causal effect using the noisy measure Z_i.

RDD as noise-induced randomization

- Eckles et al. (2020) claim that it is possible to estimate the causal effect using the noisy measure Z_i.
- Remember that when the confounders are observable, all we need to do is to balance the confounders across treatment groups.

RDD as noise-induced randomization

- Eckles et al. (2020) claim that it is possible to estimate the causal effect using the noisy measure Z_i.
- Remember that when the confounders are observable, all we need to do is to balance the confounders across treatment groups.
- Now the confounder is unobservable, hence we should balance a functional f of its noisy measure Z_i.

RDD as noise-induced randomization

- Eckles et al. (2020) claim that it is possible to estimate the causal effect using the noisy measure Z_i.
- Remember that when the confounders are observable, all we need to do is to balance the confounders across treatment groups.
- Now the confounder is unobservable, hence we should balance a functional f of its noisy measure Z_i.
- Z_i contains information of U_i and we know their relationship.

RDD as noise-induced randomization

- Eckles et al. (2020) claim that it is possible to estimate the causal effect using the noisy measure Z_i.
- Remember that when the confounders are observable, all we need to do is to balance the confounders across treatment groups.
- Now the confounder is unobservable, hence we should balance a functional f of its noisy measure Z_i.
- Z_i contains information of U_i and we know their relationship.
- If $f(Z_i)$ is balanced and f is properly chosen, we expect U_i to be balanced as well.

RDD as noise-induced randomization

- Eckles et al. (2020) claim that it is possible to estimate the causal effect using the noisy measure Z_i.
- Remember that when the confounders are observable, all we need to do is to balance the confounders across treatment groups.
- Now the confounder is unobservable, hence we should balance a functional f of its noisy measure Z_i.
- Z_i contains information of U_i and we know their relationship.
- If $f(Z_i)$ is balanced and f is properly chosen, we expect U_i to be balanced as well.
- Essentially, we are learning the information of unobservables from observable variables.

RDD as noise-induced randomization

- Eckles et al. (2020) claim that it is possible to estimate the causal effect using the noisy measure Z_i.
- Remember that when the confounders are observable, all we need to do is to balance the confounders across treatment groups.
- Now the confounder is unobservable, hence we should balance a functional f of its noisy measure Z_i.
- Z_i contains information of U_i and we know their relationship.
- If $f(Z_i)$ is balanced and f is properly chosen, we expect U_i to be balanced as well.
- Essentially, we are learning the information of unobservables from observable variables.
- This is an idea called "manifold learning."

RDD as noise-induced randomization

- Eckles et al. (2020) also focus on estimators with the following form:

$$\hat{\tau}_{SRD} = \sum_{Z_i \geq 0} \hat{\gamma}_{+}(Z_i) Y_i - \sum_{Z_i \leq 0} \hat{\gamma}_{-}(Z_i) Y_i$$

RDD as noise-induced randomization

- Eckles et al. (2020) also focus on estimators with the following form:
$$\hat{\tau}_{SRD} = \sum_{Z_i \geq 0} \hat{\gamma}_{+}(Z_i) Y_i - \sum_{Z_i \leq 0} \hat{\gamma}_{-}(Z_i) Y_i$$
- $\hat{\gamma}_{i,+}$ and $\hat{\gamma}_{i,-}$ need to balance the confounder U_i, thus we require

$$\int_{-\infty}^{0} \hat{\gamma}_{-}(z) dF(z) = 1, \int_{0}^{\infty} \hat{\gamma}_{+}(z) dF(z) = 1$$
$$\left| \int_{-\infty}^{0} \hat{\gamma}_{-}(z) \phi(z|u) dz - \int_{0}^{\infty} \hat{\gamma}_{+}(z) \phi(z|u) dz \right| \leq t \text{ for all } u$$

RDD as noise-induced randomization

- Eckles et al. (2020) also focus on estimators with the following form:

$$\hat{\tau}_{SRD} = \sum_{Z_i \geq 0} \hat{\gamma}_+(Z_i) Y_i - \sum_{Z_i \leq 0} \hat{\gamma}_-(Z_i) Y_i$$

- $\hat{\gamma}_{i,+}$ and $\hat{\gamma}_{i,-}$ need to balance the confounder U_i, thus we require

$$\int_{-\infty}^{0} \hat{\gamma}_-(z) dF(z) = 1, \int_{0}^{\infty} \hat{\gamma}_+(z) dF(z) = 1$$

$$\left| \int_{-\infty}^{0} \hat{\gamma}_-(z) \phi(z|u) dz - \int_{0}^{\infty} \hat{\gamma}_+(z) \phi(z|u) dz \right| \leq t \text{ for all } u$$

- Note that the classic local regression estimator is a special case of $\hat{\tau}_{SRD}$ where the weights are estimated based on OLS.

RDD as noise-induced randomization

- Nevertheless, the regression weights are not optimal for the MSE $\mathbb{E}[\hat{\tau}_{SRD} - \tau_{SRD}]^2$.

RDD as noise-induced randomization

- Nevertheless, the regression weights are not optimal for the MSE $\mathbb{E}[\hat{\tau}_{SRD} - \tau_{SRD}]^2$.
- Eckles et al. (2020) suggest that researchers should directly minimize the MSE under the constraints above:

$$\begin{aligned}
&\hat{\gamma}_+(), \hat{\gamma}_-() \\
&= \arg\max_{\gamma_+, \gamma_-} \left[\frac{\sigma^2}{N} \left(\int_{-\infty}^{0} \hat{\gamma}_-^2(z) dF(z) + \int_{0}^{\infty} \hat{\gamma}_+^2(z) dF(z) \right) + \frac{t^2}{M^2} \right]
\end{aligned}$$

where $\sigma^2 = Var[Y_i | Z_i]$ and M is a constant.

RDD as noise-induced randomization

- Nevertheless, the regression weights are not optimal for the MSE $\mathbb{E}[\hat{\tau}_{SRD} - \tau_{SRD}]^2$.
- Eckles et al. (2020) suggest that researchers should directly minimize the MSE under the constraints above:

$$\hat{\gamma}_+(), \hat{\gamma}_-()$$
$$= \arg \max_{\gamma_+, \gamma_-} \left[\frac{\sigma^2}{N} \left(\int_{-\infty}^{0} \hat{\gamma}_-^2(z) dF(z) + \int_0^{\infty} \hat{\gamma}_+^2(z) dF(z) \right) + \frac{t^2}{M^2} \right]$$

  where $\sigma^2 = Var[Y_i | Z_i]$ and M is a constant.
- It is a standard convex optimization problem and can be easily solved in statistical software.

RDD as noise-induced randomization

- Nevertheless, the regression weights are not optimal for the MSE $\mathbb{E}[\hat{\tau}_{SRD} - \tau_{SRD}]^2$.
- Eckles et al. (2020) suggest that researchers should directly minimize the MSE under the constraints above:

$$\hat{\gamma}_+(), \hat{\gamma}_-()$$
$$= \arg\max_{\gamma_+,\gamma_-} \left[\frac{\sigma^2}{N} \left(\int_{-\infty}^{0} \hat{\gamma}_-^2(z) dF(z) + \int_0^{\infty} \hat{\gamma}_+^2(z) dF(z) \right) + \frac{t^2}{M^2} \right]$$

where $\sigma^2 = Var[Y_i | Z_i]$ and M is a constant.

- It is a standard convex optimization problem and can be easily solved in statistical software.
- Eckles et al. (2020) show that the studentized $\hat{\tau}_{SRD}$ is asymptotically normal.

RDD as noise-induced randomization

- The approach in Eckles et al. (2020) is purely design-based and requires no assumption on the smoothness of the regression function.

RDD as noise-induced randomization

- The approach in Eckles et al. (2020) is purely design-based and requires no assumption on the smoothness of the regression function.
- It applies to cases with either a continuous running variable or a discrete running variable.

RDD as noise-induced randomization

- The approach in Eckles et al. (2020) is purely design-based and requires no assumption on the smoothness of the regression function.
- It applies to cases with either a continuous running variable or a discrete running variable.
- Inference is also straightforward in this method.

RDD as noise-induced randomization

- The approach in Eckles et al. (2020) is purely design-based and requires no assumption on the smoothness of the regression function.
- It applies to cases with either a continuous running variable or a discrete running variable.
- Inference is also straightforward in this method.
- However, it requires the knowledge of the measurement error, which might be unrealistic in practice.

RDD as noise-induced randomization

- The approach in Eckles et al. (2020) is purely design-based and requires no assumption on the smoothness of the regression function.
- It applies to cases with either a continuous running variable or a discrete running variable.
- Inference is also straightforward in this method.
- However, it requires the knowledge of the measurement error, which might be unrealistic in practice.
- The future direction is to learn the conditional distribution from data.

References I

Abadie, Alberto, and Guido W Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74 (1). Wiley Online Library: 235–67.

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell. 2020. "Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs." *The Econometrics Journal* 23 (2). Oxford University Press: 192–210.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6). Wiley Online Library: 2295–2326.

Card, David, David S Lee, Zhuan Pei, and Andrea Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink

Design." *Econometrica* 83 (6). Wiley Online Library: 2453–83.

Cattaneo, Matias D, Brigham R Frandsen, and Rocio Titiunik. 2015. "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the Us Senate." *Journal of Causal Inference* 3 (1). De Gruyter: 1–24.

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma. 2020. "Simple Local Polynomial Density Estimators." *Journal of the American Statistical Association* 115 (531). Taylor & Francis: 1449–55.

Cattaneo, Matias D, Luke Keele, Rocío Titiunik, and Gonzalo Vazquez-Bare. 2020. "Extrapolating Treatment Effects in Multi-Cutoff Regression Discontinuity Designs." *Journal of the American Statistical Association*. Taylor & Francis, 1–12.

Cattaneo, Matias D, Rocío Titiunik, Gonzalo Vazquez-Bare, and Luke Keele. 2016. "Interpreting Regression Discontinuity Designs with Multiple Cutoffs." *The Journal of Politics* 78 (4). University of

References III

Chicago Press Chicago, IL: 1229–48.

Eckles, Dean, Nikolaos Ignatiadis, Stefan Wager, and Han Wu. 2020. "Noise-Induced Randomization in Regression Discontinuity Designs." *arXiv Preprint arXiv:2004.09458*.

Fan, Jianqing, and Irene Gijbels. 1996. *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. Vol. 66. CRC Press.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37 (3). Taylor & Francis: 447–56.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1). JSTOR:

201–9.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3). Oxford University Press: 933–59.

Imbens, Guido, and Stefan Wager. 2019. "Optimized Regression Discontinuity Designs." *Review of Economics and Statistics* 101 (2). MIT Press: 264–78.

Kolesár, Michal, and Christoph Rothe. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *American Economic Review* 108 (8): 2277–2304.

Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in Us House Elections." *Journal of Econometrics* 142 (2).

References V

Elsevier: 675–97.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2). Elsevier: 698–714.

Meyersson, Erik. 2014. "Islamic Rule and the Empowerment of the Poor and Pious." *Econometrica* 82 (1). Wiley Online Library: 229–69.

Porter, Jack. 2003. "Estimation in the Regression Discontinuity Model." *Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison* 2003. Citeseer: 5–19.

Thistlethwaite, Donald L, and Donald T Campbell. 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology* 51 (6). American Psychological Association: 309.