Regression Discontinuity Design I

Ye Wang
University of North Carolina at Chapel Hill

Linear Methods in Causal Inference
POLI784

Review

- Assumptions underlying the instrumental variable method are hard to satisfy in observational studies.
- If an instrument is weak, then the finite-sample bias of the estimate will be large and the confidence interval will not have the correct coverage rate.
- We can use the Anderson-Rubin test or adjust the critical value of the t-statistic based on the value of the first-stage F-statistic.
- We can test exclusion restriction and monotonicity jointly.
- Under model-specification assumptions, we can charaterize the compliers and generalize the LATE to the ATE.

Some history

- Regression discontinuity design (RDD) became popular in political science since Lee (2008) applied this method to analyze congressional elections in the US.
- Thistlethwaite and Campbell (1960) first developed this idea in psychology.
- Hahn, Todd, and Van der Klaauw (2001) formally discussed the identification assumptions in RDD.
- Porter (2003) proposed the kernel regression estimator based on the work by Fan and Gijbels (1996).
- Imbens and Kalyanaraman (2012) introduced the first data-driven bandwidth selector.
- Calonico, Cattaneo, and Titiunik (2014) derived the asymptotic distributions for a group of RDD estimators.

Setup

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

$$E[Y_i(1) - Y_i(0) | Z_i = c].$$

- Notice that the causal parameter is conditional/local by definition.
- It only captures the effect on those with $Z_i = c$.

Ideal experiment behind sharp RDD

- This setup is slightly different from what we saw before.
- Positivity is violated everywhere but at the cutoff point.
- We can treat RDD as a randomized experiment conducted in the block defined by $Z_i = c$.
- Suppose there are $1,000$ congressional elections, in all of which both candidates win 50% of the votes.
- We then randomly select a winner for each election with a coin flip, which implies that $D_i \perp \{Y_i(1), Y_i(0)\} | Z_i = c$.
- In theory, all we need to do is to calculate the difference in means.
- The result reflects the causal effect of the winner's attributes (e.g., party affiliation).

Ideal experiment behind sharp RDD

- Unfortunately, we do not have that many elections with a 50-50 split in practice.
- We need structural restrictions on the outcome such that we can approximate the two average outcomes using data beyond the cutoff point.
- The RDD estimate is thus biased by default.
- Yet the bias diminishes to zero as the sample size increases, as we can use more observations that are close to the cutoff.
- Then, the estimate converges to the unbiased one under the ideal experiment.

Identification in sharp RDD

- Identification in the setting of sharp RDD requires the continuity of the expected outcome (Hahn, Todd, and Van der Klaauw 2001).
- Define $\mu(z) = E[Y_i|Z_i = z]$, $\mu_+ = \lim_{z \to c^+} \mu(z)$, and $\mu_- = \lim_{z \to c^-} \mu(z)$.
- Continuity means that $\mu_+ = E[Y_i(1)|Z_i = c]$ and $\mu_- = E[Y_i(0)|Z_i = c]$.
- If so, we have

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

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

Estimation in sharp RDD

- To estimate μ_+ and μ_-, the most common choice is the kernel regression estimator.
- Given a selected bandwidth h, we can fit the model within the windows $[c-h, c]$ and $[c, c+h]$, respectively:

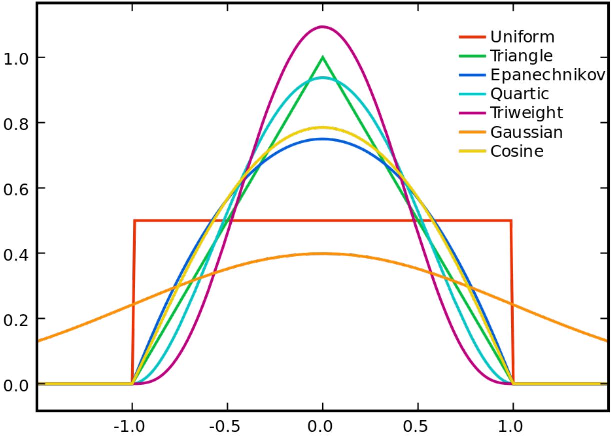
$$
\begin{aligned}
&(\hat{\mu}_+, \hat{\beta}_+) \\
&= \arg\min_{\mu,\beta} \sum_{i=1}^{N} \mathbf{1}\{Z_i \geq c\} \left(Y_i - \mu - \beta(Z_i - c)\right)^2 \mathbf{K}\left(\frac{Z_i - c}{h}\right)
\end{aligned}
$$

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

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

Estimation in sharp RDD

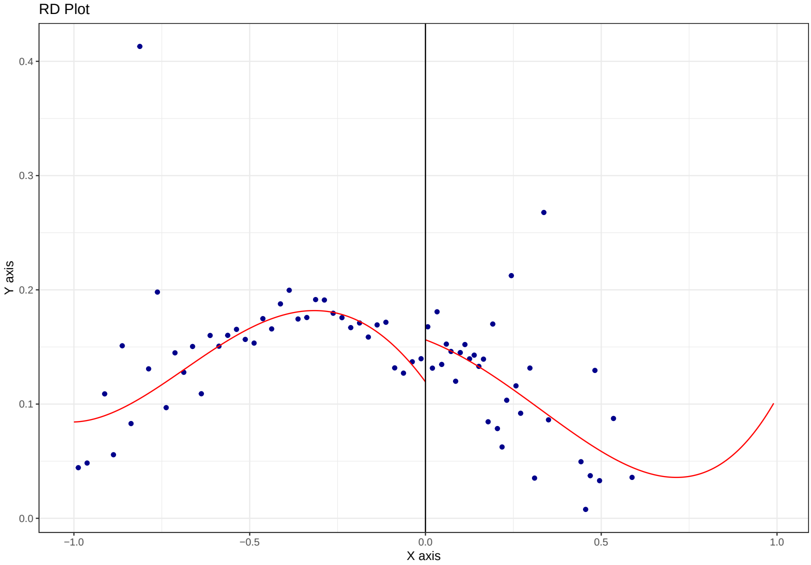
- Remember that to improve the precision of prediction, units that are closer to the cutoff should receive a larger weight.
- We usually use the triangular kernel in practice.



Sharp RDD: application

- We will use Meyersson (2014) as an illustrative example.
- The author investigated the consequences of the 1994 municipal elections in Turkey.
- Does the victory of Islamic candidates endanger the rights of women?
- In the raw data, we can find a negative correlation between the vote share of Islamic candidates and high school attainment of women across Turkish cities.
- But results from RDD tell a different story.
- The victory of Islamic mayors actually empowered "the poor and pious" and encouraged women to receive more education.

Sharp RDD: application



Sharp RDD: application

```
## Sharp RD estimates using local polynomial regression.
##
## 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|
```

Implement RDD in practice

- 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.
- Note that the plot is based on global polynomials rather than kernel regressions.
- It does not include all the data points.
- Never use 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.

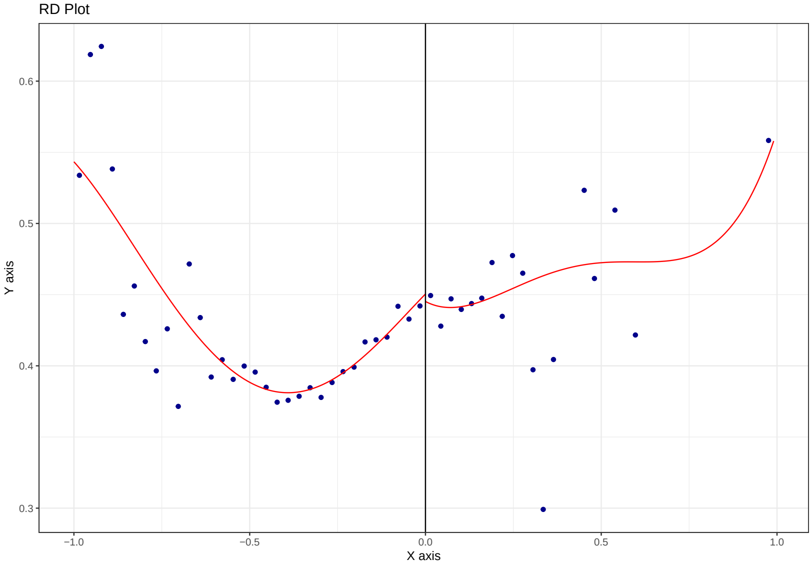
Implement RDD in practice

- In practice, some scholars select the bandwidth using *rdrobust* 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.
- Bias correction is ignored and inference may be problematic.
- 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).

Test the assumption of continuity

- As continuity is crucial for identification in sharp RDD, we should test its validity in practice.
- A common approach is to rely on placebo tests.
- Several ways to do this:
  - Apply the estimator to a covariate;
  - Apply the estimator to a placebo outcome;
  - Apply the estimator at a point other than the cutoff (placebo treatment).
- Meyersson (2014) did a really good job.

Tests in sharp RDD: application

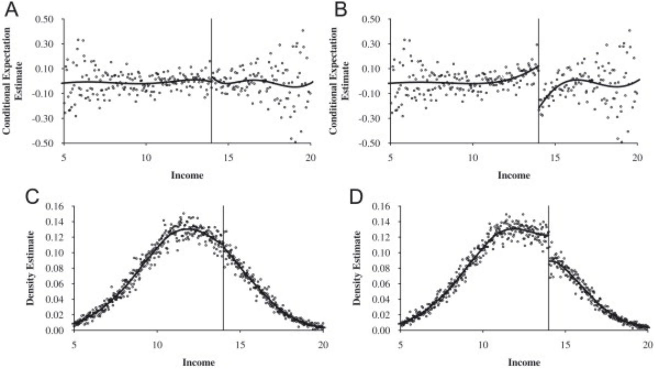


Tests in sharp RDD: application

```
##                   CI Lower   CI Upper
## Conventional   -0.01464976 0.04035605
## Bias-Corrected -0.01486825 0.04013756
## Robust         -0.02051897 0.04578828

##                   CI Lower   CI Upper
## Conventional   -0.02616607 0.03410509
## Bias-Corrected -0.02929148 0.03097967
## Robust         -0.03526899 0.03695719

##                   CI Lower   CI Upper
## Conventional   -0.02713105 0.02631536
## Bias-Corrected -0.02997479 0.02347162
## Robust         -0.03497856 0.02847539

##                   CI Lower   CI Upper
## Conventional   -0.01436241 0.03079918
## Bias-Corrected -0.01068460 0.03447699
## Robust         -0.01436412 0.03815651
```

Test the assumption of continuity

- The assumption is also violated when units in the sample self-select into one side of the cutoff (known as sorting).
- 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.



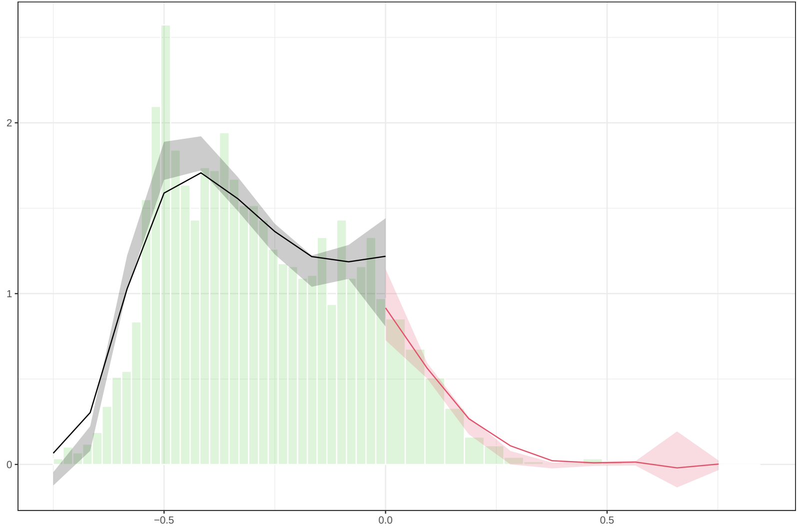
Test the assumption of continuity

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

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

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

Tests in sharp RDD: application



Bias in RDD estimation (*)

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

$$\mathbf{R} = (\iota, \mathbf{Z})$$
$$\mathbf{W}_+ = \left\{ \mathbf{1}\{Z_i \geq c\} K\left(\frac{Z_i - c}{h_N}\right) \right\}_{N \times N}$$
$$\mathbf{M} = (\mu(Z_1), \mu(Z_2), \dots, \mu(Z_N))'$$

  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 in RDD estimation (*)

- Notice that $\mathbf{Y} = \mathbf{M} + \varepsilon$ with $E[\varepsilon_i | Z_i] = 0$.
- The estimate $\hat{\mu}_+$ equals to the first row of

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

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

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

- Hence,

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

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

Bias in RDD estimation (*)

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

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

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

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

Bias in RDD estimation (*)

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

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

- 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 have an order of $p + 1$.

Bandwidth selection in RDD

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

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

Bandwidth selection for optimizing MSE

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

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

- 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

- 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:

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

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

Bias correction

- Intuitively,

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

- 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 sufficiently small.
- Remember that the numerator is $O_p(h^2)$ and the denominator is $O_p\left(\frac{1}{\sqrt{Nh}}\right)$, thus the total bias is $O_p(\sqrt{Nh^5})$ and does not decline to zero.

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 second-order derivative of μ_+ and μ_- and subtract them from $\hat{\mu}_+$ and $\hat{\mu}_-$.
- Bias correction introduces extra uncertainty (from the extra local regression) into the estimate, hence the variance has to be adjusted accordingly.
- The bias-corrected CI does not account for this extra uncertainty.
- 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

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

  as long as $N \min\{b^5, h^5\} \max\{b^2, h^2\} \to 0$.
- In other words, we can still use the algorithm in Imbens and Kalyanaraman (2012) to select the bandwidth.
- We just need to modify the obtained estimate to ensure asymptotic normality.

Bandwidth selection for the bias-correction estimator

- Notice that now we need two bandwidths, h and b.
- In practice, we start from two pilot bandwidths h_0 and b_0 selected by some rule of thumb.
- We use h_0 and b_0 to run the bias-correction estimator for the second-order derivatives and obtain b^* using the algorithm in Imbens and Kalyanaraman (2012).
- We then use h_0 and b^* to run the bias-correction estimator for the two intercepts and obtain h.

Covariates in RDD

- In theory, there should be no confounder in RDD as it approximates a simple experiment.
- But we can use the information contained in covariates to enhance the estimator's efficiency.
- Calonico et al. (2019) discuss how to include covariates in the estimation.
- The basic idea is just the FWL theorem for regression.

References I

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

Calonico, Sebastian, Matias D Cattaneo, Max H Farrell, and Rocio Titiunik. 2019. "Regression Discontinuity Designs Using Covariates." *Review of Economics and Statistics* 101 (3): 442–51.

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

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

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

References II

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): 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): 201–9.

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

Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in US House Elections." *Journal of Econometrics* 142 (2): 675–97.

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

References III

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

Porter, Jack. 2003. "Estimation in the Regression Discontinuity Model." *Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison* 2003: 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): 309.