Week 10: More on RD

JJ Chen

March 20, 2015

Today's Plan

1. Reading RD papers
   - How good are internal validity and external validity?
   - Card and Giuliano (2014)
   - Angrist, Battistin, and Vuri (2014)
2. RD estimation methods
   - Local linear regression
3. Fuzzy "RD" with kink and regression kink design

Some General Questions to Ask Ourselves

- What are the running variable, the treatment, and the outcomes?
- What kinds of people have incentives to manipulate the running variable?
- Is there a density test? Covariates balanced Tests? How good are they?
    - If there are multiple cutoffs, we should expect multiple tests
    - Often time we would like to see as many covariates balanced tests as possible. And beware of whether the covariate are "relevant"
- For Fuzzy RD, do the assumptions for the LATE theorem hold?
    - What're those assumptions again?
- Are we interested in behavior or outcomes of those marginal people around the cutoff?

Card and Giuliano (2014): Introduction

- Card and Giuliano (2014) evaluate gifted classroom programs and show heterogeneous program effects for students.
  - Plan A: Gifted students based on IQ scores of 130
  - Plan B: Gifted students based on IQ scores of 116
  - Plan C: Gifted students score highest among their cohorts
- They show that assignments based on IQ thresholds has no impact on student achievement, whereas assignment based on achievement rank sees significant gains in reading and math and the gain is persistent at least for another grade.

Card and Giuliano (2014): Framework

- They are interested in the impact of gifted program on test scores
- But once a kid is "classified" as gifted students, other things happen
  1. Gifted classroom and teacher
  2. Gifted peers
  3. Individual gifted services
- A simple structural model:

$$\mathrm{Y} = \beta_1 \mathrm{D}_{\text{class}} + \beta_2 \mathrm{Q}_{\text{peer}} + \beta_3 \mathrm{D}_{\text{gifted}} + \mathrm{X}'\boldsymbol{\gamma} + \lambda \mathrm{IQ}^* + \eta$$

  - IQ^* is the running variable, * means we have perfect measure

Card and Giuliano (2014): ID Assumptions

- The model:

$$\mathrm{Y} = \beta_1 \mathrm{D_{class}} + \beta_2 \mathrm{Q_{peer}} + \beta_3 \mathrm{D_{gifted}} + \mathrm{X}'\boldsymbol{\gamma} + \lambda \mathrm{IQ}^* + \eta$$

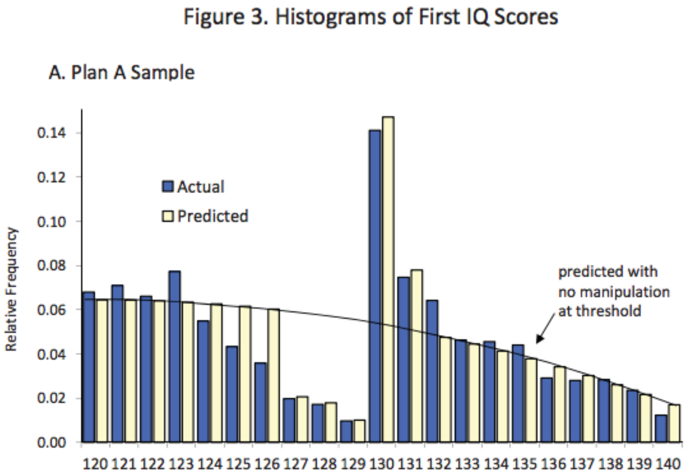
- Taking expectations conditional on observed IQ:

$$\begin{aligned} \mathrm{E}(\mathrm{Y}|\mathrm{IQ}) =& \beta_1 \Pr(\mathrm{D_{class}} = 1 \, |\mathrm{IQ}) + \beta_2 \, \mathrm{E}(\mathrm{Q_{peer}} \, |\mathrm{IQ}) + \\ & \beta_3 \Pr(\mathrm{D_{gifted}} = 1 \, |\mathrm{IQ}) + \mathrm{E}(\mathrm{X}'\boldsymbol{\gamma} + \lambda \mathrm{IQ} + \eta \, |\mathrm{IQ}) \end{aligned}$$

- Suppose the gifted threshold is T, what assumptions do we need to make in order to get

$$\begin{aligned} \mathrm{Dis}(\mathrm{Y}) &= \lim_{\mathrm{IQ} \to \mathrm{T}^+} \mathrm{E}(\mathrm{Y} \, |\mathrm{IQ}) - \lim_{\mathrm{IQ} \to \mathrm{T}^-} \mathrm{E}(\mathrm{Y} \, |\mathrm{IQ}) \\ &= \beta_1 \mathrm{Dis}(\mathrm{P_{class}}) + \beta_2 \mathrm{Dis}(\mathrm{Q_{peer}}) + \beta_3 \mathrm{Dis}(\mathrm{P_{gifted}}) \end{aligned}$$

Card and Giuliano (2014): Manipulation



Card and Giuliano (2014): Biases

- If the ID assumptions we made are fine, the RF, $\mathrm{Dis}(\mathrm{Y})$, is unbiased
    - What does the RF mean? Is it policy relevant?
- Suppose we want to know more about the impact of gifted classroom on test scores, we can rescale the RF

$$\frac{\mathrm{Dis}(\mathrm{Y})}{\mathrm{Dis}(\mathrm{P}_{\mathrm{class}})} = \beta_1 + \beta_2 \frac{\mathrm{Dis}(\mathrm{Q}_{\mathrm{peer}})}{\mathrm{Dis}(\mathrm{P}_{\mathrm{class}})} + \beta_3 \frac{\mathrm{Dis}(\mathrm{P}_{\mathrm{gifted}})}{\mathrm{Dis}(\mathrm{P}_{\mathrm{class}})}$$

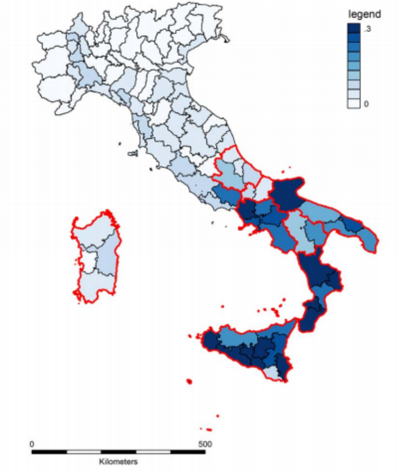
    - How would you interpret the descaled RF?
- In general, what do we learn from this setting?

Angrist, Battistin, and Vuri (2014): Introduction

- Angrist, Battistin, and Vuri (2014) documents a relationship between class size and test score manipulation in southern Italy
  - Background: Relate to Angrist and Lavy (1999) and Urquiola and Verhoogen (2009)
- Usually we focus on manipulation of the running variable
  - But manipulating outcome variables seem problematic too
  - They show score manipulation is pos corr with CS
- Ignoring other institutional details might lead to misleading results

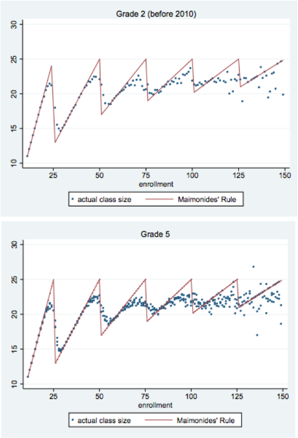
Angrist, Battistin, and Vuri (2014): Score Manipulations

Figure 1: Manipulation Rates by Province



Angrist, Battistin, and Vuri (2014): 1st Graph

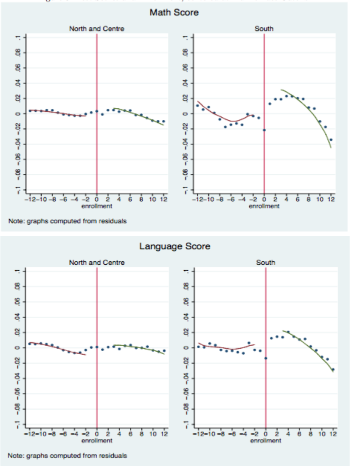
Figure 2: Class Size by Enrollment in Pre-reform Years



Notes: The figure shows actual class size and as predicted by Maimonides' Rule in pre-reform years

Angrist, Battistin, and Vuri (2014): RF Graph

Figure 5: Test Scores and Enrollment, centered at Maimonides Cutoffs



Angrist, Battistin, and Vuri (2014): RD Table

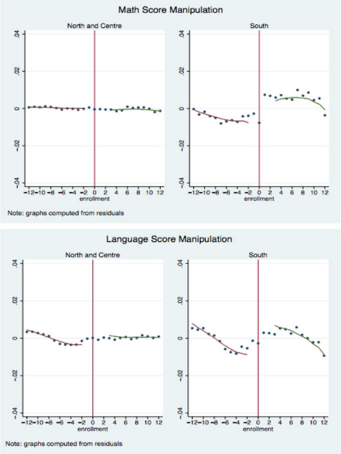
Table 2: OLS and IV/2SLS Estimates of the Effect of Class Size on Test Scores

	OLS			IV/2SLS					
	Italy (1)	North/Centre (2)	South (3)	Italy (4)	North/Centre (5)	South (6)	Italy (7)	North/Centre (8)	South (9)
A. Math									
Class size	-0.0078	-0.0224***	0.0091	-0.0519***	-0.0436***	-0.0957***	-0.0609***	-0.0417**	-0.1294**
	(0.0070)	(0.0067)	(0.0146)	(0.0134)	(0.0115)	(0.0362)	(0.0196)	(0.0171)	(0.0507)
Enrollment	✗	✗	✗	✗	✗	✗	✗	✗	✗
Enrollment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
Interactions							✗	✗	✗
N	140,010	87,498	52,512	140,010	87,498	52,512	140,010	87,498	52,512
B. Language									
Class size	0.0029	-0.0188***	0.0328***	-0.0395***	-0.0313***	-0.0641**	-0.0409***	-0.0215	-0.0937**
	(0.0055)	(0.0053)	(0.0114)	(0.0106)	(0.0092)	(0.0289)	(0.0155)	(0.0136)	(0.0403)
Enrollment	✗	✗	✗	✗	✗	✗	✗	✗	✗
Enrollment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
Interactions							✗	✗	✗
N	140,010	87,498	52,512	140,010	87,498	52,512	140,010	87,498	52,512

Notes: Columns 1-3 report OLS estimates of the effect of class size on scores. Columns 4-9 report 2SLS estimates using Maimonides' Rule as an instrument. The unit of observation is the class. Class size coefficients show the effect of 10 students. Models with interactions allow the quadratic running variable control to differ across windows of ±12 students around each cutoff. Robust standard errors, clustered on school and grade, are shown in parentheses. Control variables include: % female students, % immigrants, % fathers at least high school graduate, % employed mothers, % unemployed mothers, % mother NILF, grade and year dummies, and dummies for missing values . All regressions include sampling strata controls (grade enrollment at institution, region dummies and their interactions). * significant at 10%; ** significant at 5%; *** significant at 1%.

Angrist, Battistin, and Vuri (2014): Socre manipulation

Figure 6: Score Manipulation and Enrollment, centered at Maimonides Cutoffs



Notes: The solid line shows a one-sided LLR fit.

Angrist, Battistin, and Vuri (2014): Solution

- To address score manipulation, they model it as another endogenous variable

$$\mathrm{Y}_{igkt} = \rho_0(t, g) + \beta_1 \mathrm{S}_{igkt} + \beta_2 \mathrm{M}_{igkt} + \rho_1 f(\mathrm{E}_{gkt}) + \eta_{ijkt}$$

  - E_{gkt} is the running variable: number of enrollment
  - S_{igkt} is endogenous class size, instrumented by Maimonides' Rules
  - M_{igkt} is endogenous score manipulation, instrumented by randomly assigned school monitors
- How should we interpret β_1 and β_2?

Angrist, Battistin, and Vuri (2014): Results

Table 7: IV/2SLS Estimates of the Effect of Class Size and Score Manipulation on Test Scores

	IV/2SLS			IV/2SLS (overidentified)			IV/2SLS (overidentified-interacted)		
	Italy (1)	North/Centre (2)	South (3)	Italy (4)	North/Centre (5)	South (6)	Italy (7)	North/Centre (8)	South (9)
					A. Math				
Class size	0.0075	-0.0029	0.0062	0.0024	-0.0113	0.0133	0.0116	0.0136	0.0473
	(0.0213)	(0.0298)	(0.0441)	(0.0190)	(0.0251)	(0.0378)	(0.0316)	(0.0482)	(0.0675)
Score manipulation	3.82***	7.33***	2.88***	3.82***	7.02***	2.87***	4.10***	9.21**	3.33***
	(0.19)	(0.79)	(0.16)	(0.19)	(0.73)	(0.16)	(0.96)	(4.41)	(0.86)
Class size * Score manipulation							-0.1464	-1.2700	-0.2273
							(0.4814)	(2.1598)	(0.4304)
Overid test [P-value]				[0.914]	[0.600]	[0.541]	[0.914]	[0.475]	[0.476]
N	139,996	87,491	52,505	139,996	87,491	52,505	139,996	87,491	52,505
					B. Language				
Class size	0.0121	0.0049	0.0127	0.0218	0.0109	0.0491	0.0325	0.0098	0.1337*
	(0.0173)	(0.0196)	(0.0385)	(0.0153)	(0.0174)	(0.0329)	(0.0308)	(0.0320)	(0.0800)
Score manipulation	3.29***	4.50***	2.80***	3.21***	4.34***	2.74***	3.59***	4.31*	4.18***
	(0.18)	(0.45)	(0.18)	(0.18)	(0.42)	(0.18)	(1.03)	(2.25)	(1.30)
Class size * Score manipulation							-0.2130	-0.0029	-0.7058
							(0.4980)	(1.0898)	(0.6214)
Overid test (P-value)				[0.129]	[0.796]	[0.036]	[0.216]	[0.844]	[0.109]
N	140,003	87,493	52,510	140,003	87,493	52,510	140,003	87,493	52,510

Introduction

- In the homework we mainly use linear and quadratic regression to get RD estimates
- There are many more complicated methods
    - But the results we get from different methods should not be too different
    - And people make judgments based on RD pictures
- Here are three general methods used for estimating the discontinuity
    1. Parametric
    2. Parametric with bandwidth
    3. Non-parametric
- The first two methods have been addressed a lot in class, so I will just briefly mention the third one (more on it next semester)

Non-paramatric Estimation: LLR

- Local linear regression (LLR) is a non-parametric method for estimating a functional form
- When to do LLR? When the exact functional form is a big deal and you don't want to do just linear or quadratic approximation
  - RD really requires we model the functional form correctly
- LLR Steps:
  1. At each $\mathrm{X}_i = x_i$, we estimate a linear regression within a neighborhood $[x_i - h, x_i + h]$ (thus "local linear regression"):

$$\mathrm{Y} = \alpha + \gamma \mathrm{X}_i, [x_i - h, x_i + h]$$

  2. No local linear regression at the cutoff

LLR with Kernel smoothing

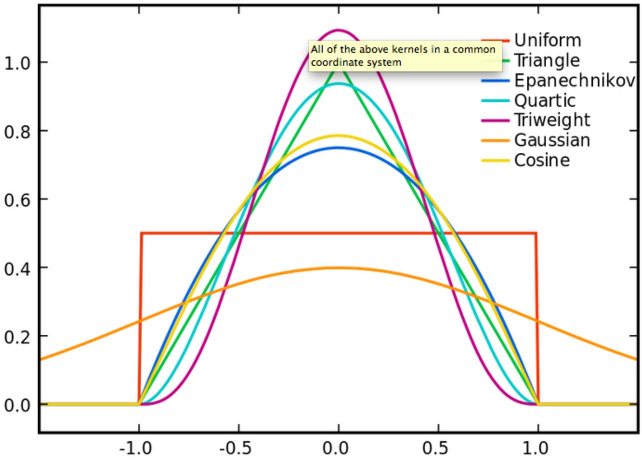
- Just make the LLR match the data even better, we can use kernel weighting
- That is, in the LLR

$$\mathrm{Y} = \alpha + \gamma \mathrm{X}_i, [x_i - h, x_i + h],$$

$\mathrm{X}_i = x_i$ get the most weight, and $x_i - h$ and $x_i + h$ get the least weight

- There are many different weights: Uniform, triangle, Epanechnikov, Quartic, Triweight, Gaussian, Cosine... And there are many different bandwidths (see Imbens and Kalyanaraman (2011) for optimal bandwidth)

Weights

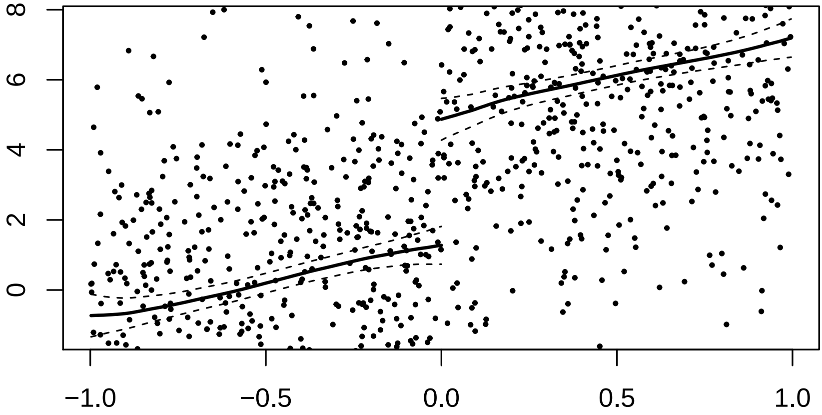

Uniform
Triangle
Epanechnikov
Quartic
Triweight
Gaussian
Cosine
All of the above kernels in a common coordinate system
1.0
0.8
0.6
0.4
0.2
0.0
-1.0
-0.5
0.0
0.5
1.0

R code

- There are packages that implement LLR for RD, such as `rd` in Stata or `rdd` in R
  - Stata: `ssc install rd`
  - R: `install.packages("rdd")`
- Here is a simple simulation of sharp RD data using R

```
x <- runif(1000, -1, 1)
covar <- rnorm(1000)
y <- 1 + 2 * x + 3 * covar + 4 * (x >= 0) + rnorm(1000)
library("rdd")
RD <- RDestimate(y ~ x | covar)
# plot(RD)
```

RD: LLR plot



Introduction

- In this section we will consider two extensions to RD design:
  1. Fuzzy RD with a kink instead of a jump
  2. Regression kink design (or RK from now)
- They share some similarities with RD and are relatively new techniques, which means perhaps there are still opportunities for arbitrage (I always wish everyone can stick to the principle of no arbitrage)
- I will just present the key ideas. To be more serious about ID assumptions or estimations, here are some relevant papers Nielsen, Sørensen, and Taber (2008), Saez (2010), Card et al. (2015), Dong (2014), Ando (2013)

Fuzzy "RD" with a Kink: Introduction

- Recall that fuzzy RD uses a *jump* in the probability of binary treatment D_i $(P_+ - P_- \neq 0)$

$$\rho = \lim_{\Delta \to 0} \frac{\mathrm{E}(\mathrm{Y}_i \mid x_0 < x_i < x_0 + \Delta) - \mathrm{E}(\mathrm{Y}_i \mid x_0 - \Delta < x_i < x_0)}{\mathrm{E}(\mathrm{D}_i \mid x_0 < x_i < x_0 + \Delta) - \mathrm{E}(\mathrm{D}_i \mid x_0 - \Delta < x_i < x_0)}$$
$$= \frac{E_+ - E_-}{P_+ - P_-}$$

- What if there is no jump but just a *kink*? $(P'_+ - P'_- \neq 0)$

$$\rho = \frac{E'_+ - E'_-}{P'_+ - P'_-}$$

Fuzzy "RD" with a Kink: 1st

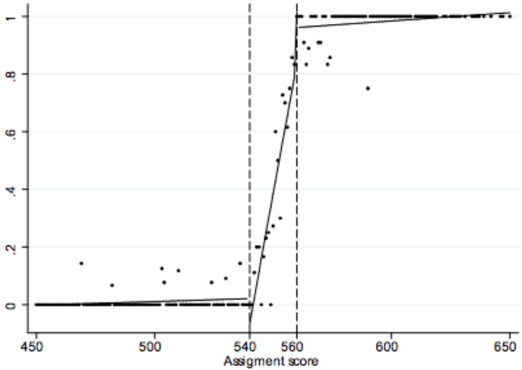


Figure 2: Assignment score and the probability of attending an elite school for males

Fuzzy "RD" with a Kink: RF

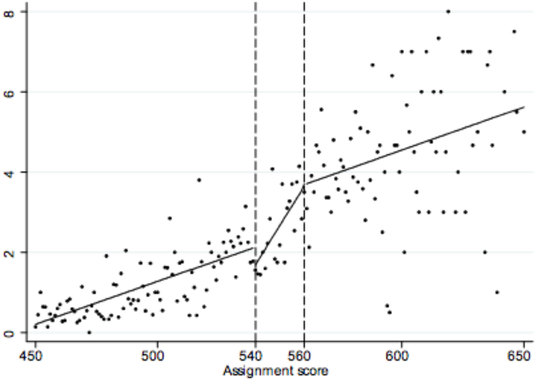


Figure 3: Assignment score and years of post-secondary education for males

RK: Introduction

- Consider a constant-effect model

$$\mathrm{Y}_i = \rho \mathrm{T}_i + f(\mathrm{X}_i) + \eta_i,$$

- Instead of a binary treatment, consider a continuous treatment variable T_i which is a deterministic function of a running variable X_i, $\mathrm{T}_i = g(\mathrm{X}_i)$ with a kink at $\mathrm{X}_i = x_0$
  - Example: Progressive taxation – Marginal tax rates are determined by the income level
  - Example: Unemployment insurance – The amount of unemployment benefit or the duration of getting the benefit can be based on income in previous year

RK: Introduction

- If $g(\mathrm{X}_i)$ and $\mathrm{E}(\eta \,|\mathrm{X}_i = x_0)$ have derivatives that are continuous in X_i at $\mathrm{X}_i = x_0$, then we have

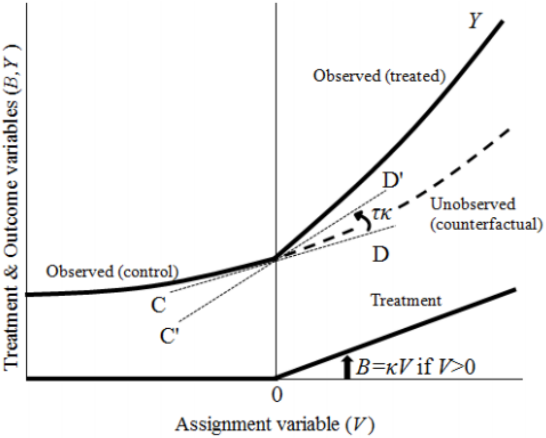
$$\rho = \frac{\lim\limits_{x_0 \to 0^+} \frac{d\,\mathrm{E}(\mathrm{Y}_i \,|\mathrm{X}_i = x_0)}{d\mathrm{X}_i} |_{x_0} - \lim\limits_{x_0 \to 0^-} \frac{d\,\mathrm{E}(\mathrm{Y}_i \,|\mathrm{X}_i = x_0)}{d\mathrm{X}_i} |_{x_0}}{\lim\limits_{x_0 \to 0^+} \frac{dg(\mathrm{X}_i)}{d\mathrm{X}_i} |_{x_0} - \lim\limits_{x_0 \to 0^-} \frac{dg(\mathrm{X}_i)}{d\mathrm{X}_i} |_{x_0}}$$
$$= \frac{E'_+ - E'_-}{\mathrm{T}'_+ - \mathrm{T}'_-},$$

which identify the causal parameter ρ

- Note that the denominator is still the treatment variable, instead of the probability of getting treatment
- In a nutshell: we have jumps in the first derivatives

RK Graph

Figure 1. Stylized feature of the Regression Kink Design



References I

Ando, Michihito. 2013. "How Much Should We Trust Regression-Kink-Design Estimates?" *Available at SSRN 2365086*.

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule To Estimate The Effect Of Class Size On Scholastic Achievement." *The Quarterly Journal of Economics* 114 (2): 533–75. `http://ideas.repec.org/a/tpr/qjecon/v114y1999i2p533-575.html`.

Angrist, Joshua D., Erich Battistin, and Daniela Vuri. 2014. *In a Small Moment: Class Size and Moral Hazard in the Mezzogiorno*. Working Paper 20173. Working Paper Series. National Bureau of Economic Research. doi:10.3386/w20173.

References II

Card, David, and Laura Giuliano. 2014. *Does Gifted Education Work? For Which Students?* Working Paper 20453. Working Paper Series. National Bureau of Economic Research. doi:10.3386/w20453.

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

Dong, Yingying. 2014. "Jump or Kink? Identification of Binary Treatment Regression Discontinuity Design Without the Discontinuity." *Unpublished Manuscript*.

Imbens, Guido, and Karthik Kalyanaraman. 2011. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies*. Oxford University Press, rdr043.

References III

Nielsen, Helena Skyt, Torben Sørensen, and Christopher R. Taber. 2008. *Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform*. Working Paper 14535. Working Paper Series. National Bureau of Economic Research. doi:10.3386/w14535.

Saez, Emmanuel. 2010. “Do Taxpayers Bunch at Kink Points?” *American Economic Journal: Economic Policy*. JSTOR, 180–212.

Urquiola, Miguel, and Eric Verhoogen. 2009. “Class-Size Caps, Sorting, and the Regression-Discontinuity Design.” *American Economic Review* 99 (1): 179–215. doi:10.1257/aer.99.1.179.