Causal Inference with Observational Data

5. Regression Discontinuity

Marc F. Bellemare

May 2018

Introduction

As with almost everything else we will do in this class, the question of interest is this:

> *What is the effect of a treatment (often some program) D on a given outcome y?*

The difference between this topic and the other topics covered in this class is that I do not have first-hand experience with regression discontinuity (RD) designs. As such, these lecture notes were first drafted by my doctoral student Camilo Bohorquez-Peñuela, who gave the RD lecture in my class as a guest lecturer two years ago.

Introduction

In an RD design, broadly speaking, we know the “rule” assigning an observation to treatment or control, and we have some score s that determines that assignment.

For instance, you may be interested in the effect of a private Catholic education D on wage y. If admission to private Catholic schools is determined by an admissions test whose score is s and you know the admission threshold (e.g., 60 percent),

Introduction

An RD estimates the treatment effect using a more precise knowledge of assignment rules:

- Treatment is determined—at least in part—by a continuous assignment variable s.
- There is a discontinuity in treatment status (i.e., D goes from 0 to 1) at the cutoff point $s = c$.
- Observations around the threshold are assumed to be identical except for treatment status—observable characteristics x do not vary around c, like a balancing test in an RCT.
- Units cannot manipulate the assignment mechanism.

Perfect Compliance

Treatment is such that $D = 1$ if $s \geq c$ and $D = 0$ if $s < c$. With full compliance, we have what's called a sharp RD:

1. $P(D = 1|s \geq c) = 1$ and $P(D = 1|s < c) = 0$.
2. Treatment is a deterministic function of the assignment (or running, or forcing) variable, and
3. There is no s for which some $D_i = 1$ and some $D_i = 0$.

For instance, grad school funding might be determined only by GRE scores, and we can look at whether grad school funding has an impact on student performance.

Imperfect Compliance

In this case, we have what's called a fuzzy RD.

1. $0 < P(D = 1|s \geq c) = 1 - P(D = 1|s < c) = 0$ for a given c.
2. Treatment status is not completely determined by the assignment variable. A fuzzy RD exploits the discontinuity in the probability of treatment, conditional on covariates.

For instance, grad school funding is determined by GRE scores and unobservable factors (e.g., admissions committee preferences, letters of recommendation content, etc.), and we can look at whether grad school funding has an impact on student performance.

Sharp RD

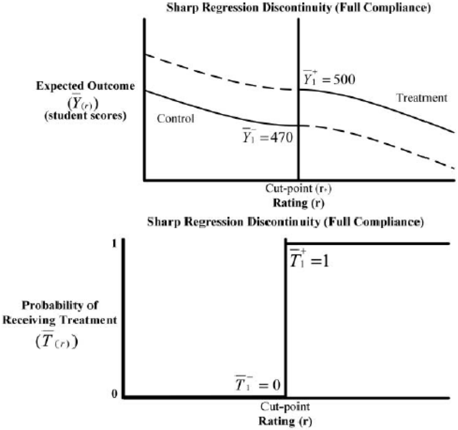


Figure: Source: Bloom (2012).

Fuzzy RD

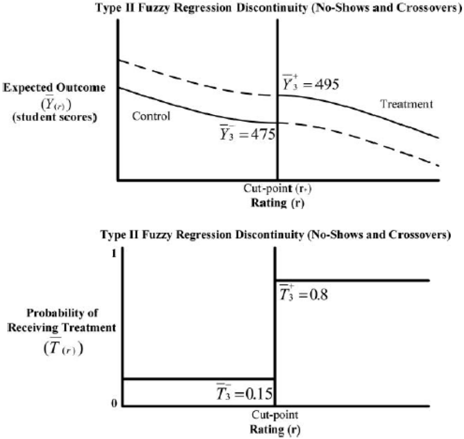


Figure: Source: Bloom (2012).

Sharp RD

A sharp RD is a kind of selection on observables design.

Validity, however, is not given by an overlap of covariate values among treatment and control groups (at least in the vicinity of the cutoff point), as in matching.

A rich set of controls can help control for endogeneity from other sources.

Sharp RD—Interpreting Estimates

Discontinuity of the outcome at the cutoff point implies an average effect of the ITT.

In a sharp RD, however, assignment to treatment is the same as receiving the treatment.

Thus, estimates around the threshold correspond to ATE for the treated and imply a LATE.

The assumption here is that the expected outcome is a continuous function of the running variable around the cutoff point.

Sharp RD—Parametric Estimation

Suppose the outcome can be described by the following linear, constant effect model:

$$E(y_{0i}|s_i) = \alpha + \beta s_i \tag{1}$$

and

$$y_{1i} = y_{0i} + \rho, \tag{2}$$

where ρ represents the treatment effect. Then, estimation of the following by OLS yields the causal effect ρ:

$$y_i = \alpha + \beta s_i + \rho D_i + \epsilon_i. \tag{3}$$

Sharp RD—Parametric Estimation

What if $E(y_i|s_i)$ is nonlinear? Then,

1. Estimate $y_i = f(s_i) + \rho D_i + \epsilon_i$, where $f(s_i)$ is a continuous function of s_i at the cutoff point.
2. $f(s_i)$ is an N^{th}-order polynomial.
3. $y_i = \alpha + \sum \beta_j s_i^j + \rho D_i + \epsilon_i$.

If the functional form differs on both sides of the cutoff point, you can introduce interaction terms between D and distance from the cutoff.

Sharp RD—Parametric Estimation

Note, however, that in a recent article, Gelman and Imbens (*JBES*, forthcoming), write:

> *We argue that controlling for global high-order polynomials in regression discontinuity analysis is a flawed approach with three major problems: it leads to noisy estimates, sensitivity to the degree of the polynomial, and poor coverage of confidence intervals. We recommend researchers instead use estimators based on local linear or quadratic polynomials or other smooth functions.*

Sharp RD—Parametric Estimation

The estimation of a sharp RD thus generalizes to

$$E(y_i|s_i) = \underbrace{E(y_{0i}|s_i)}_{\text{Average Behavior}} + \underbrace{\{E(y_{1i}|s_i) - E(y_{0i}|s_i)\}\, D_i,}_{\text{Average Effect}} \qquad (4)$$

and so

$$y_i = \beta_{01}\widetilde{s}_i + ... + \beta_{0j}\widetilde{s}_i^j + \rho D_i + \beta_1^* D_i \widetilde{s}_i + ... + \beta_j^* D_i \widetilde{s}_i^j + \epsilon_i, \qquad (5)$$

where ρ is the treatment effect at the cutoff and $\widetilde{s}_i = s_i - c$.

Sharp RD—Parametric Estimation

In the preceding slide:

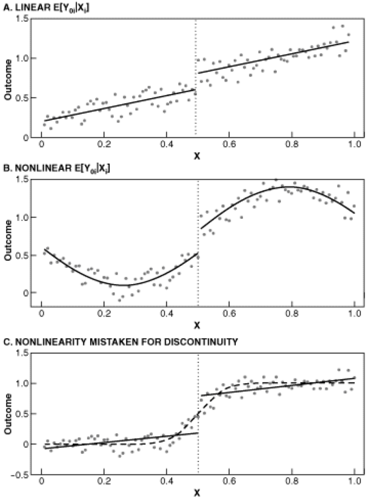
- The treatment effect at the cutoff is ρ.
- The treatment effect at point s is $\rho + \beta_1^* s + ... + \beta_j^* s$.

Sharp RD—Parametric Estimation

Under a parametric (i.e., OLS) approach, validity comes from the adequate selection of the functional form representing the causal effect.

- No room for agnostic approaches regarding the functional form,
- Nonparametric approaches could ease the problem, but at the cost of less efficiency ("the curse of dimensionality"),
- Importance of visual evidence.

Sharp RD—Parametric Estimation


A. LINEAR $E[Y_{0i}|X_i]$

B. NONLINEAR $E[Y_{0i}|X_i]$

C. NONLINEARITY MISTAKEN FOR DISCONTINUITY

Figure: Different Functional Forms.

Sharp RD—Nonparametric Estimation

Compare averages in a small neighborhood of the discontinuity:

- Approaches: Kernel, local linear weighted regression
- Problem: Incorrect bandwidths imply biased estimators
- Another problem: Small bandwidths imply less efficiency
- There are some techniques to select the “right” bandwidth; see Lee and Lemieux (2010)
- Anyway, they help to provide visual evidence

Fuzzy RD

Imperfect compliance allows exploiting treatment assignment as an IV for treatment, unlike deterministic assignment under a sharp RD.

- Hahn et al. (2001), Imbens and Lemieux (2008), Angrist and Pischke (2009), and Lee and Lemieux (2010) explain and demonstrate that under fuzzy RD, full exogeneity of the running variable is no longer required.
- That external manipulation of the running variable resembles a randomized experiment.
- Still, an assumption is made regarding balancing across observations around the cutoff point remains.
- Moreover, we still assume that units cannot manipulate the running variable themselves.

Fuzzy RD—Parametric Estimation

$$P(D_i = 1|s_i) = \begin{array}{l} g_1(s_i) \text{ if } s_i \geq c \\ g_0(s_i) \text{ if } s_i < c \end{array} \tag{6}$$

such that

$$E(D_i|s_i) = g_0(s_i) + [g_1(s_i) - g_0(s_i)]\, T_i \tag{7}$$

where $T_i = I(s_i \geq c)$. $E(D_i|s_i)$ is the first-stage equation of the 2SLS system where T_i is the IV. Interpreting the parameter, we get a LATE only at $s_i = c$, and only for the compliers. Pretty limited!

Tips

- Determine whether you actually have an RD.
- Determine whether your RD is sharp or fuzzy.
- Graph y vs. s with a smoothing tool, and visually inspect it.
- Also graph a histogram of s and inspect any discontinuity around the threshold to assess whether there was manipulation of s.
- Remember that RD is invalid if units can manipulate s in order to receive or avoid treatment!

Tips

- Begin with nonparametric methods
- Nonparametrics are not the solution—they provide good visual evidence about which nonlinearities to incorporate in parametric estimations
- Use the simplest specification possible, and exploit complexity when conducting robustness checks.
- Conduct a parallel RD analysis on baseline covariates, like a balancing test.
- Sharp RD yields LATE at $s_i = 0$ with full compliance; fuzzy RD yields late at $s_i = 0$ *only*, and only for compliers.
- All versions of RD have low external validity.

Miller et al. (2013)

Question: What is the effect of a subsidized health care regime on financial risk protection, use of health care services, and self-reported health care conditions among Colombian households?

Subsidized regime (SR) eligibility based on household well-being conditions measured by a proxy means test.

- SISBEN score: Categorization of households not only based on their income, but also on socio-demographic characteristics (e.g., dwelling, durable goods, employment, schooling)
- SISBEN is in $(0, 100)$, with a low SISBEN meaning a poorer household.
- SISBEN also determines eligibility for other publicly provided welfare programs.

Miller et al. (2013)

Participation in the SR is subject to self-selection and political manipulation.

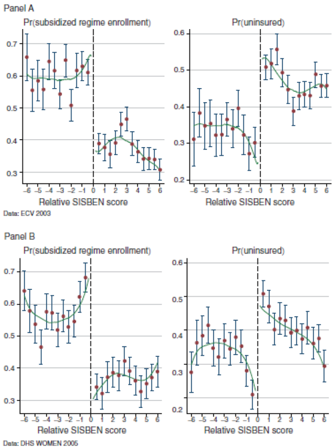
- Researchers cannot observe SISBEN, but know the algorithm, so they estimate a synthetic score using independent data (2003 Living Standards Survey and 2005 DHS)
- This synthetic score is not subject to manipulation by the households, but may suffer from measurement error
- Due to financial constraints, some municipalities used different cutoffs for determining eligibility so the authors follow Chay et al. (2005) in order to estimate municipality-specific thresholds

Miller et al. (2013)

Identification strategy:

- IV is a dummy variable for whether the household's SISBEN is below the cutoff
- IV seeks to correct for self-selection and for manipulation of the assignment process by politicians (unobserved heterogeneity).
- Reverse causality seems not to be a problem

Miller et al. (2013)



Data: ECV 2003

Data: DHS WOMEN 2005

Figure 1. Subsidized Regime Enrollment and 'Uninsurance' across SISBEN Eligibility Thresholds

Figure: Probability of Enrollment in the SR Around the Cutoff.

Miller et al. (2013)

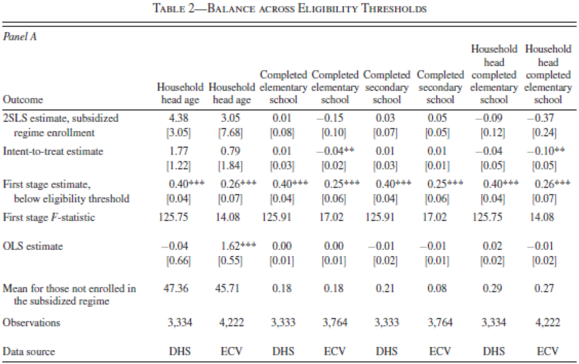


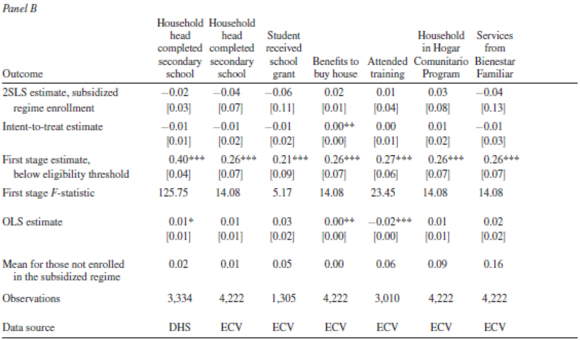
Table 2—Balance across Eligibility Thresholds

Panel A

Outcome	Household head age	Household head age	Completed elementary school	Completed elementary school	Completed secondary school	Completed secondary school	Household head completed elementary school	Household head completed elementary school
2SLS estimate, subsidized regime enrollment	4.38	3.05	0.01	–0.15	0.03	0.05	–0.09	–0.37
	[3.05]	[7.68]	[0.08]	[0.10]	[0.07]	[0.05]	[0.12]	[0.24]
Intent-to-treat estimate	1.77	0.79	0.01	–0.04**	0.01	0.01	–0.04	–0.10**
	[1.22]	[1.84]	[0.03]	[0.02]	[0.03]	[0.01]	[0.05]	[0.05]
First stage estimate, below eligibility threshold	0.40***	0.26***	0.40***	0.25***	0.40***	0.25***	0.40***	0.26***
	[0.04]	[0.07]	[0.04]	[0.06]	[0.04]	[0.06]	[0.04]	[0.07]
First stage *F*-statistic	125.75	14.08	125.91	17.02	125.91	17.02	125.75	14.08
OLS estimate	–0.04	1.62***	0.00	0.00	–0.01	–0.01	0.02	–0.01
	[0.66]	[0.55]	[0.01]	[0.01]	[0.02]	[0.01]	[0.02]	[0.02]
Mean for those not enrolled in the subsidized regime	47.36	45.71	0.18	0.18	0.21	0.08	0.29	0.27
Observations	3,334	4,222	3,333	3,764	3,333	3,764	3,334	4,222
Data source	DHS	ECV	DHS	ECV	DHS	ECV	DHS	ECV

Figure: Balance Across Eligibility Thresholds.

Miller et al. (2013)



Panel B

Outcome	Household head completed secondary school	Household head completed secondary school	Student received school grant	Benefits to buy house	Attended training	Household in Hogar Comunitario Program	Services from Bienestar Familiar
2SLS estimate, subsidized regime enrollment	−0.02 [0.03]	−0.04 [0.07]	−0.06 [0.11]	0.02 [0.01]	0.01 [0.04]	0.03 [0.08]	−0.04 [0.13]
Intent-to-treat estimate	−0.01 [0.01]	−0.01 [0.02]	−0.01 [0.02]	0.00** [0.00]	0.00 [0.01]	0.01 [0.02]	−0.01 [0.03]
First stage estimate, below eligibility threshold	0.40*** [0.04]	0.26*** [0.07]	0.21*** [0.09]	0.26*** [0.07]	0.27*** [0.06]	0.26*** [0.07]	0.26*** [0.07]
First stage *F*-statistic	125.75	14.08	5.17	14.08	23.45	14.08	14.08
OLS estimate	0.01* [0.01]	0.01 [0.01]	0.03 [0.02]	0.00** [0.00]	−0.02*** [0.00]	0.01 [0.01]	0.02 [0.02]
Mean for those not enrolled in the subsidized regime	0.02	0.01	0.05	0.00	0.06	0.09	0.16
Observations	3,334	4,222	1,305	4,222	3,010	4,222	4,222
Data source	DHS	ECV	ECV	ECV	ECV	ECV	ECV

Figure: Balance Across Eligibility Thresholds (cont'd).

Miller et al. (2013)

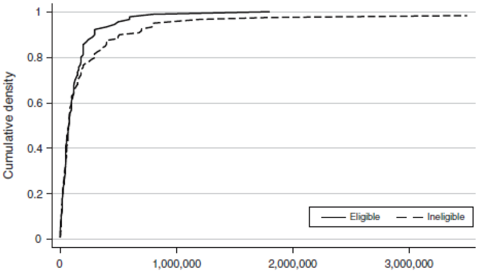


Figure 3. Individual Inpatient Medical Spending
(Among Those within Two Points of County-Specific Eligibility Thresholds)

Figure: Discontinuity of Outcomes.

Miller et al. (2013)

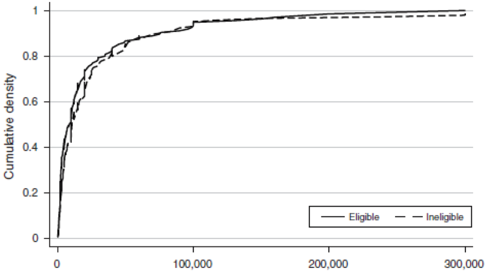


Figure 4. Individual Outpatient Medical Spending
(Among Those within Two Points of County-Specific Eligibility Thresholds)

Figure: Discontinuity of Outcomes (con'd)

Miller et al. (2013)

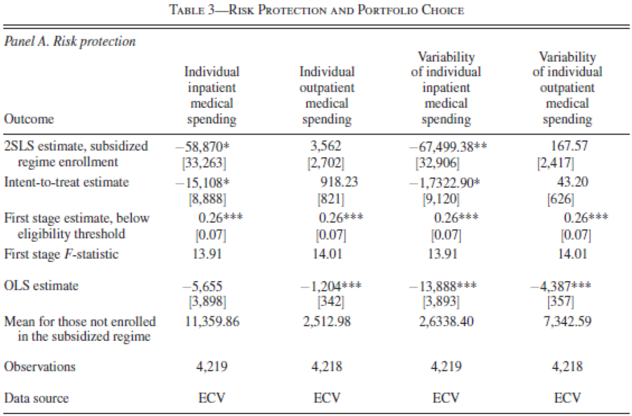
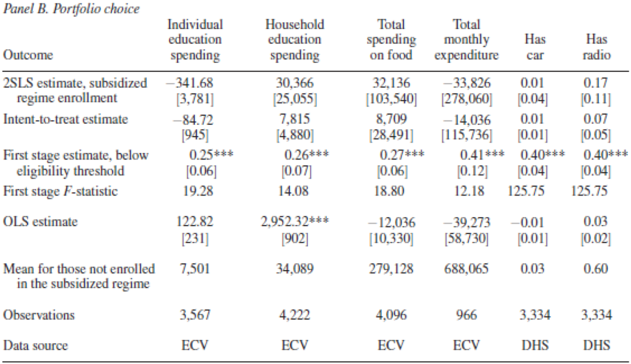


Table 3—Risk Protection and Portfolio Choice

Panel A. Risk protection

Outcome	Individual inpatient medical spending	Individual outpatient medical spending	Variability of individual inpatient medical spending	Variability of individual outpatient medical spending
2SLS estimate, subsidized regime enrollment	−58,870*	3,562	−67,499.38**	167.57
	[33,263]	[2,702]	[32,906]	[2,417]
Intent-to-treat estimate	−15,108*	918.23	−1,7322.90*	43.20
	[8,888]	[821]	[9,120]	[626]
First stage estimate, below eligibility threshold	0.26***	0.26***	0.26***	0.26***
	[0.07]	[0.07]	[0.07]	[0.07]
First stage *F*-statistic	13.91	14.01	13.91	14.01
OLS estimate	−5,655	−1,204***	−13,888***	−4,387***
	[3,898]	[342]	[3,893]	[357]
Mean for those not enrolled in the subsidized regime	11,359.86	2,512.98	2,6338.40	7,342.59
Observations	4,219	4,218	4,219	4,218
Data source	ECV	ECV	ECV	ECV

Miller et al. (2013)



Panel B. Portfolio choice

Outcome	Individual education spending	Household education spending	Total spending on food	Total monthly expenditure	Has car	Has radio
2SLS estimate, subsidized regime enrollment	−341.68	30,366	32,136	−33,826	0.01	0.17
	[3,781]	[25,055]	[103,540]	[278,060]	[0.04]	[0.11]
Intent-to-treat estimate	−84.72	7,815	8,709	−14,036	0.01	0.07
	[945]	[4,880]	[28,491]	[115,736]	[0.01]	[0.05]
First stage estimate, below eligibility threshold	0.25***	0.26***	0.27***	0.41***	0.40***	0.40***
	[0.06]	[0.07]	[0.06]	[0.12]	[0.04]	[0.04]
First stage *F*-statistic	19.28	14.08	18.80	12.18	125.75	125.75
OLS estimate	122.82	2,952.32***	−12,036	−39,273	−0.01	0.03
	[231]	[902]	[10,330]	[58,730]	[0.01]	[0.02]
Mean for those not enrolled in the subsidized regime	7,501	34,089	279,128	688,065	0.03	0.60
Observations	3,567	4,222	4,096	966	3,334	3,334
Data source	ECV	ECV	ECV	ECV	DHS	DHS

Miller et al. (2013)

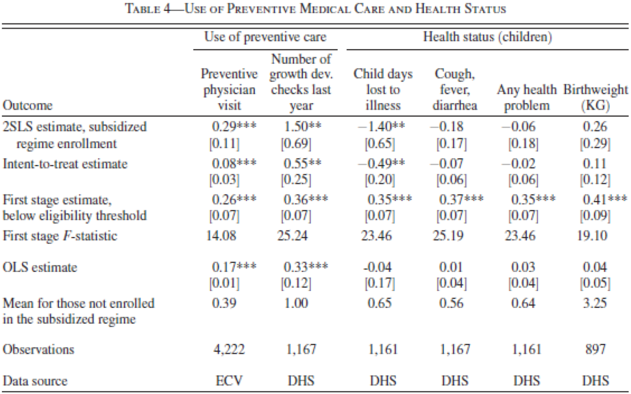


Table 4—Use of Preventive Medical Care and Health Status

	Use of preventive care		Health status (children)			
Outcome	Preventive physician visit	Number of growth dev. checks last year	Child days lost to illness	Cough, fever, diarrhea	Any health problem	Birthweight (KG)
2SLS estimate, subsidized regime enrollment	0.29***	1.50**	−1.40**	−0.18	−0.06	0.26
	[0.11]	[0.69]	[0.65]	[0.17]	[0.18]	[0.29]
Intent-to-treat estimate	0.08***	0.55**	−0.49**	−0.07	−0.02	0.11
	[0.03]	[0.25]	[0.20]	[0.06]	[0.06]	[0.12]
First stage estimate, below eligibility threshold	0.26***	0.36***	0.35***	0.37***	0.35***	0.41***
	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.09]
First stage F-statistic	14.08	25.24	23.46	25.19	23.46	19.10
OLS estimate	0.17***	0.33***	-0.04	0.01	0.03	0.04
	[0.01]	[0.12]	[0.17]	[0.04]	[0.04]	[0.05]
Mean for those not enrolled in the subsidized regime	0.39	1.00	0.65	0.56	0.64	3.25
Observations	4,222	1,167	1,161	1,167	1,161	897
Data source	ECV	DHS	DHS	DHS	DHS	DHS

Miller et al. (2013)

Robustness checks:

- Different bandwidths
- Higher-order polynomials of the simulated SISBEN score
- Excluding municipality FEs
- Nonparametric local linear regressions

Parameter signs are robust to different specifications, with changes in magnitudes and precision.

Miller et al. (2013)

External validity:

- Miller et al. acknowledge that their results lack external validity
- However, they exploit heterogeneity in thresholds (remember, they calculated them at the municipality level) in order to run second-stage equation including interaction between absolute SISBEN score and SR enrollment
- They found no statistical significance—standard errors on the interaction terms are large