Chapter 9: Regression Discontinuity

Peter Hull

Applied Econometrics II
Brown University
Spring 2024

Outline

1. RD Identification
2. RD Estimation
3. Application: Abdulkadiroglu et al. (2014)
4. Bonus Track 1: RD Extrapolation
5. Bonus Track 2: Design-Based RD

Sharp RD

RD starts from a fundamentally different place than the design-based and model-based ID strategies we've so far encountered: deterministic rules

In a classic "sharp" RD, the treatment of interest obeys $D_i = \mathbf{1}[X_i \geq c]$

- X_i is the observed *running variable*: for now, continuously distributed
- c is the known *cutoff*: without loss we can normalize it to zero

Unlike the earlier strategies, there are no values of the covariate X_i where we observe both $D_i = 0$ and $D_i = 1$

- In RD, the name of the game is about *extrapolating* across covariate values instead of making comparisons conditional on X_i
- To formalize the validity of such extrapolations, we need a new kind of identifying assumption...

Continuity

Let $D_i = \mathbf{1}[X_i \geq 0]$ and, as usual, $Y_i = Y_i(0)(1 - D_i) + Y_i(1)D_i$. Suppose:

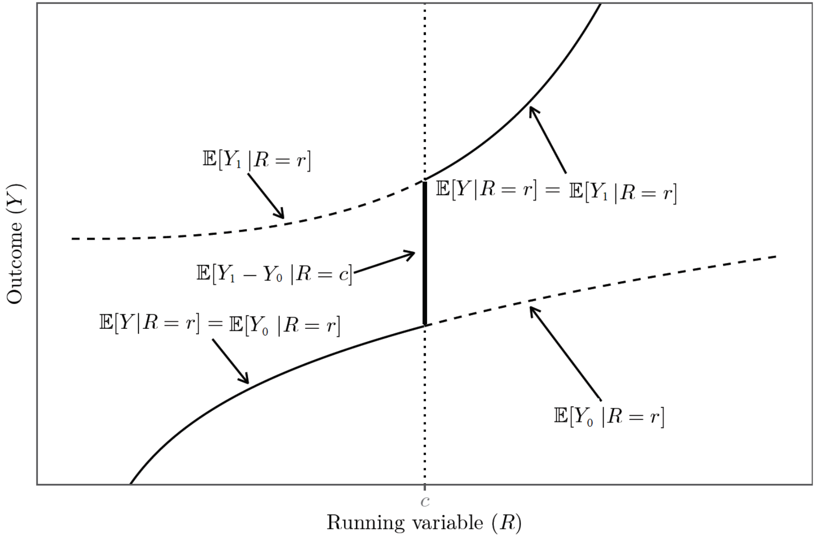
$$\lim_{x \downarrow 0} E[Y_i(1) \mid X_i = x] = E[Y_i(1) \mid X_i = 0]$$
$$\lim_{x \uparrow 0} E[Y_i(0) \mid X_i = x] = E[Y_i(0) \mid X_i = 0]$$

I.e. that potential outcome CEFs are (right-/left-)continuous at zero

Since $E[Y_i \mid X_i \geq 0] = E[Y_i(1) \mid X_i \geq 0]$ & $E[Y_i \mid X_i < 0] = E[Y_i(0) \mid X_i < 0]$ such continuity gives us identification of a particular CATE:

$$\begin{aligned}
&\lim_{x \downarrow 0} E[Y_i \mid X_i = x] - \lim_{x \uparrow 0} E[Y_i \mid X_i = x] \\
=&\lim_{x \downarrow 0} E[Y_i(1) \mid X_i = x] - \lim_{x \uparrow 0} E[Y_i(0) \mid X_i = x] \\
=&E[Y_i(1) \mid X_i = 0] - E[Y_i(0) \mid X_i = 0] \\
=&E[Y_i(1) - Y_i(0) \mid X_i = 0]
\end{aligned}$$

Continuity, Illustrated



(With apologies for the slight changes in notation)

Fuzzy RD

We might instead have a $Z_i = \mathbf{1}[X_i \geq 0]$ which can induce effects in Y_i through a non-deterministic (and possibly non-binary) treatment D_i

- This yields "fuzzy" RD, which can be seen as a form of IV

Continuity here is required in both the reduced form and first stage:

$$\lim_{x \downarrow 0} E[W_i(1) \mid X_i = x] = E[W_i(1) \mid X_i = 0]$$

$$\lim_{x \uparrow 0} E[W_i(0) \mid X_i = x] = E[W_i(0) \mid X_i = 0]$$

for $W_i \in \{Y_i, D_i\}$, with potentials defined in terms of Z_i

This gives identification of the ratio of CATEs:

$$\frac{\lim_{x \downarrow 0} E[Y_i \mid X_i = x] - \lim_{x \uparrow 0} E[Y_i \mid X_i = x]}{\lim_{x \downarrow 0} E[D_i \mid X_i = x] - \lim_{x \uparrow 0} E[D_i \mid X_i = x]} = \frac{E[Y_i(1) - Y_i(0) \mid X_i = 0]}{E[D_i(1) - D_i(0) \mid X_i = 0]}$$

Under exclusion + monotonicity, this ratio identifies a "LATE" as usual

Parametric RD

If we knew $\mu_1(x) \equiv E[Y_i(1) \mid X_i = x]$ and $\mu_0(x) \equiv E[Y_i(0) \mid X_i = x]$ were contained in some parametric classes, then RD becomes pretty easy

- Base case $\mu_d(x) = p(x)'\gamma_d$ where $p(x)$ is a vector of polynomials in x
- E.g. $p(x) = [1, x, x^2]$ so $\mu_1(x)$ and $\mu_0(x)$ are at most quadratic

Since $E[Y_i \mid X_i] = E[Y_i(0) \mid X_i] + E[Y_i(1) - Y_i(0) \mid X_i]D_i$ in sharp RD, this yields $E[Y_i \mid X_i] = p(X_i)'\gamma_0 + p(X_i)'(\gamma_1 - \gamma_0)D_i$ which regression can fit

- E.g. $Y_i = \alpha + \beta D_i + \gamma X_i + \delta X_i^2 + \phi D_i X_i + \psi D_i X_i^2 + \varepsilon_i$ where $\beta = E[Y_i(1) - Y_i(0) \mid X_i = 0]$, etc.
- In a fuzzy RD, we'd instead instrument D_i with Z_i controlling for X_i, X_i^2, $Z_i X_i$, and $Z_i X_i^2$ (think in terms of RF and FS)

Parametric RD identifies effects *everywhere*, not just at the threshold

- But of course we'd like to avoid making functional form restrictions, so we typically do non-parametric estimation...

Illustration: Lee (2008)

To estimate the effects of party incumbency, Lee uses the fact that election winners are determined by $D_i = \mathbf{1}[X_i \geq 0]$ for vote share margin X_i

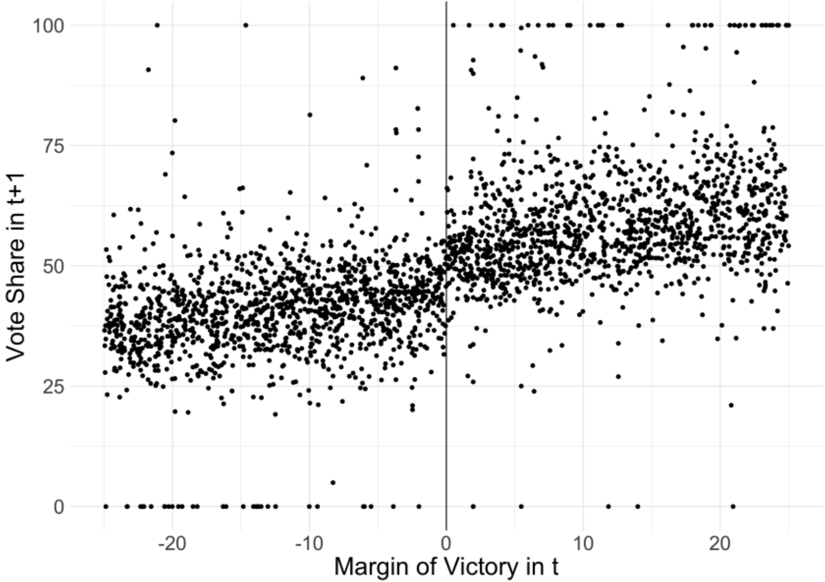
- Results are striking: incumbency appears to raise party re-election probabilities by around 40 percentage points

The credibility of this "close election RD" is bolstered by several checks:

- The RD jump is visually clear, across different binning strategies ✓
- There's no RD jump in pre-treatment outcomes (balance test) ✓
- There's no "bunching" at the cutoff (McCrary sorting test) ✓
- The RD estimate is similar as we restrict to data near the cutoff ✓

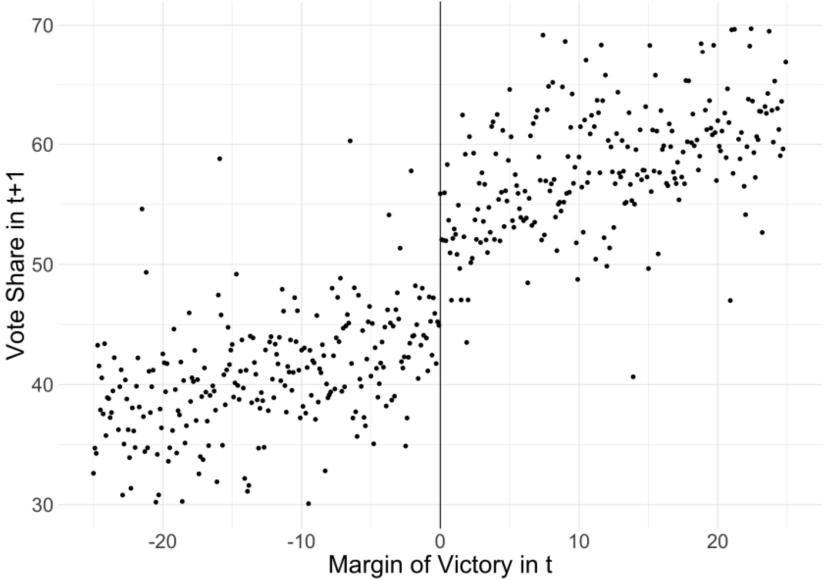
If you find yourself doing RD, it's a good idea to follow this "checklist"...

Raw Lee (2008) Data (Vote Share Outcome)



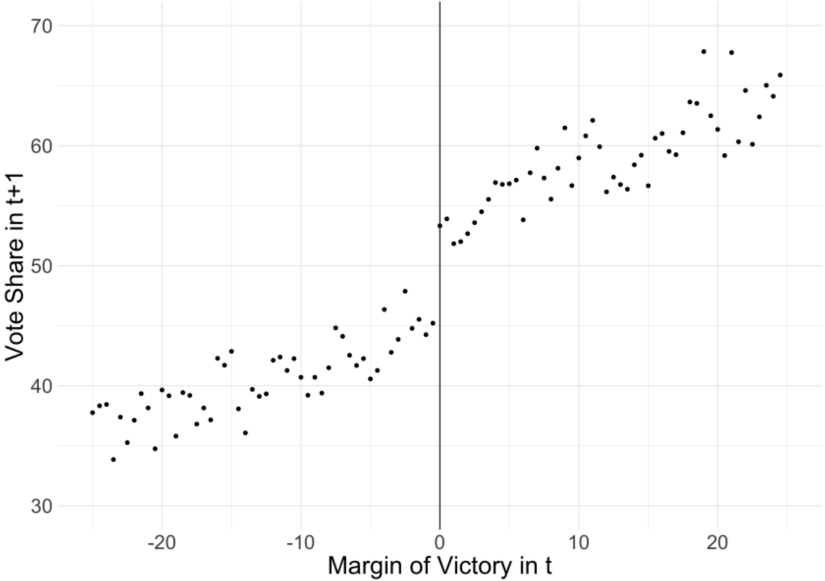
Seems like there's a jump at $X_i = 0$, but it's a bit hard to see how big it is

Lee RD with a 0.1 Percent Binscatter



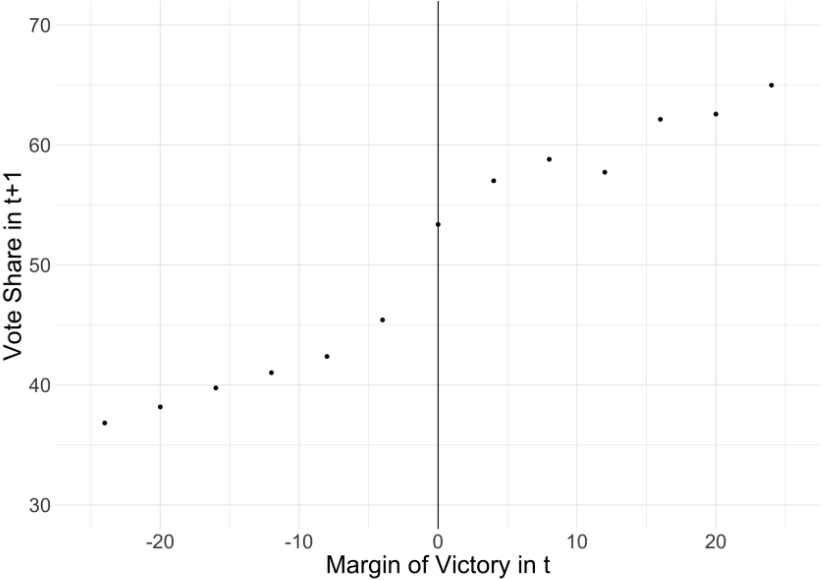
Jump is clearer, though we've lost some of the variability (note the y-axis!)

Lee RD with a 0.5 Percent Binscatter



Jump is clearer still ... but does it now seem artificially clean?

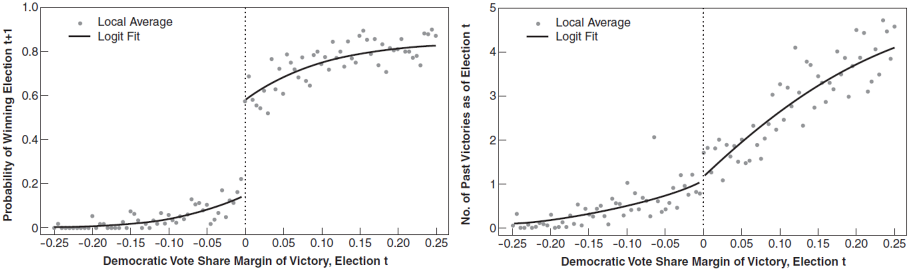
Lee RD with a 4 Percent Binscatter



See Cattaneo et al. (2020) and Korting et al. (2023) for viz. discussions

Balance Test

As usual, we should check balance using the same spec as we estimate:



Canay and Kamat (2018) propose a more complete permutation test, which checks whether covariates are approximately identically distributed on each side of the cutoff (under a slightly stronger identifying assumption)

Bunching Test

Conceptually, there are two reasons RD continuity might fail:

1. The cutoff is set *systematically*, such that confounding factors change discontinuously (e.g. lots of policies change at state borders)
2. The running variable is *manipulable* (e.g. people can manipulate test scores to get just above/below a passing grade)

Both of these are likely to show up as balance failures, but manipulation can also be detected by "bunching" of the running variable

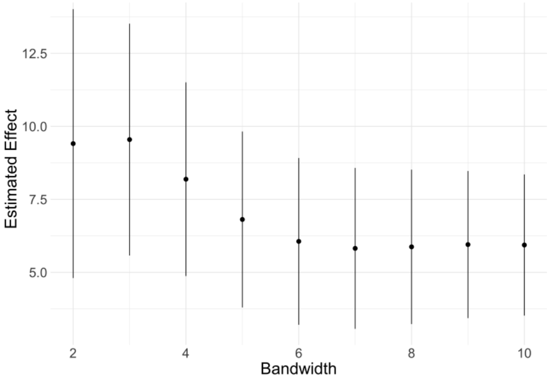
- Canonical check here is due to McCrary (2008), which checks whether the density of the running variable changes discontinuously
- In Stata/R, run with the *rddensity* package

Robustness

As we'll next see, there are many ways to estimate RD jumps

- Showing robustness across different ways is often a good way to convince a reader/audience member of your findings
- Not a great strategy: "Don't worry, I did this the optimal way!"

A nice way to present robustness is with a graph; e.g.:



Outline

1. RD Identification✓
2. RD Estimation
3. Application: Abdulkadiroglu et al. (2014)
4. Bonus Track 1: RD Extrapolation
5. Bonus Track 2: Design-Based RD

Non-Parametric RD

We saw $\tau \equiv E[Y_i(1) - Y_i(0) \mid X_i = 0]$ is ID'd by $\lim_{x \downarrow 0} \mu_1(x) - \lim_{x \uparrow 0} \mu_0(x)$ under continuity of $\mu_1(x) \equiv E[Y_i(1) \mid X_i = x]$ & $\mu_0(x) \equiv E[Y_i(0) \mid X_i = x]$

- Can we get a good estimate of τ without more restrictions on $\mu_d(x)$?

Intuitive approach: average observed $Y_i(1)$ and $Y_i(0)$ in bins near $X_i = 0$

- Smaller bins $\rightarrow$ less bias: $E[Y_i(1) \mid 0 \leq X_i < \varepsilon] \approx E[Y_i(1) \mid X_i = 0]$
- Bigger bins $\rightarrow$ more precision: $Var(\widehat{E}[Y_i(1) \mid 0 \leq X_i < \varepsilon]) \approx 0$
- If we can characterize MSE, we can pick the bin size ("bandwidth") to asymptotically minimize it — accounting for this trade-off

But how should we do the averaging? This will "shape" the tradeoff...

- In simple weighted avg's (i.e. Nadaraya-Watson / "locally constant regression") the bias for estimating boundary points turns out to be too large asymptotically. So we need a slightly fancier approach ...

Local Linear Regression

Current best practice: $\hat{\tau} = \hat{\alpha}_1 - \hat{\alpha}_0$ where

$$(\hat{\alpha}_1, \hat{\beta}_1) = \arg\min_{\alpha,\beta} \sum_{i:X_i \geq 0} (Y_i - \alpha - \beta X_i)^2 K(X_i/h_1)$$

$$(\hat{\alpha}_0, \hat{\beta}_0) = \arg\min_{\alpha,\beta} \sum_{i:X_i \leq 0} (Y_i - \alpha - \beta X_i)^2 K(X_i/h_0)$$

(h_1, h_0) are bandwidths and $K(\cdot)$ is a kernel function

- Triangle kernel is a natural choice for estimating boundary points

MSE approximation (Cattaneo et al. 2020):

$$MSE(\hat{\alpha}_d) \approx B_{d,K}^2 h_d^4 + \frac{V_{d,K}}{N \cdot h_d}$$

where $B_{d,K}$ is a term capturing bias and $V_{d,K}$ is a term capturing variance

- Optimal bandwdith: $h_d^* = \left(\frac{V_{d,K}}{4B_{d,K}^2 \cdot N}\right)^{1/5}$, so $o(h_d^*) = N^{-1/5}$
- A feasible $\hat{h}_d^*$ plugs in first-step estimates of $B_{d,K}$ and $V_{d,K}$

Accounting for Bias

With the optimal bandwidth we get asymptotic bias by design:

$$\sqrt{Nh_d^*}(\hat{\alpha}_d - \alpha_d) \Rightarrow \mathrm{N}(B_{d,K}, V_{d,K})$$

One way to avoid this is *undersmoothing*: picking a smaller-than-MSE-optimal bandwidth to make bias negligible (at the cost of variance)

Calonico, Cattaneo, and Titiunik (2015) propose an alternative path: use a higher-order approximation (e.g. local quadratic) to estimate the bias

- Intuitively, if we find a high degree of local curvature we know our linear approximation is worse $\rightarrow$ can adjust to minimize bias
- The details here are a bit tricky, but in practice folks usually apply the defaults in their *rdrobust* package and check robustness

rdrobust Example Output

```
Number of Obs.                 2763
BW type                       mserd
Kernel                   Triangular
VCE method                       NN

Number of Obs.                1376
Eff. Number of Obs.            490
Order est. (p)                   1
Order bias  (q)                  2
BW est. (h)                  8.422
BW bias (b)                 13.990
rho (h/b)                    0.602
Unique Obs.                   1344

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional     5.876     1.322     4.444     0.000     [3.284 , 8.468]
Bias-Corrected     5.542     1.322     4.191     0.000     [2.950 , 8.134]
        Robust     5.542     1.537     3.606     0.000     [2.530 , 8.554]
=============================================================================
```

Order est. (p) — Local linear for estimating ATE

Order bias (q) — Local quadratic for estimating bias

BW est. (h) — Bandwidth for estimating ATE

BW bias (b) — Bandwidth for estimating bias

- Conventional: standard local linear estimates + SEs
- Bias-Corrected: adjusted coefficient, standard SEs
- Robust: adjusted coefficient, SEs adjusted for bias estimation

Quick Aside: Discrete Running Variables

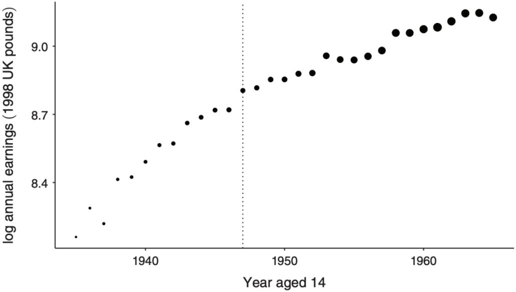


FIGURE 2. AVERAGE OF NATURAL LOGARITHM OF ANNUAL EARNINGS BY YEAR AGED 14

Notes: Vertical line indicates the year 1947, in which the minimum school-leaving age changed from 14 to 15. Volume of dots is proportional to share of workers in the full data with the corresponding age.

Sometimes we can't shrink bandwidths beyond some point, even as $N \to \infty$

- Still, we might imagine a true/latent continuous running variable

Kolesar and Rothe (2018) show how we can bound τ by placing restrictions on the second derivative of the true/latent $\mu_d(x)$'s

- This "honest" RD approach also works with continuous running variables (Armstrong and Kolesar 2018); see their *rdhonest* package

Outline

1. RD Identification✓
2. RD Estimation✓
3. Application: Abdulkadiroglu et al. (2014)
4. Bonus Track 1: RD Extrapolation
5. Bonus Track 2: Design-Based RD

Motivation: Is "Elite" Schooling Worth It?

Much money and effort is spent every year by parents to try to get their kids into schools with high-achieving (and whiter) peers

- Observe how house prices jump at schooling boundaries (Black 1999)
- But are these more elite schools actually higher "value-added"?

Of course, to answer this question we must overcome selection bias

- Families that succeed at enrolling their kids in elite schools likely differ in a number of important ways that affect later achievement
- Unlike charters, there are usually no admissions lotteries

Abdulkadiroglu et al. (2014; AAP) use an RD strategy to overcome selection for three flagship elite schools in both Boston and NYC

- We'll focus on Boston Latin School, Boston Latin Academy, and O'Bryant (also in Boston)

Identification Strategy: Test-Based Admissions

Boston 7th and 9th graders interested in an exam school take an entrance exam (ISEE) in the fall prior to applying

- They submit school preferences into a centralized (DA) mechanism, which incorporates score cutoffs and other priority groups

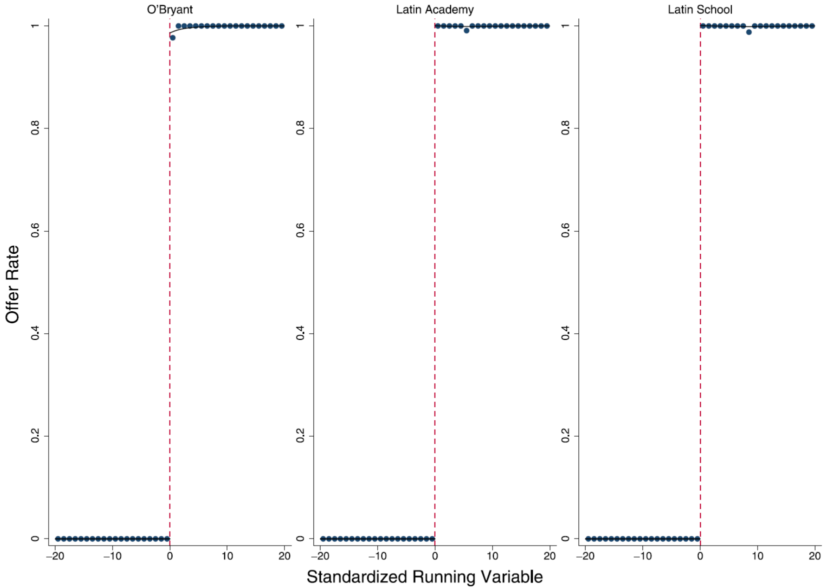
AAP first identify "sharp" samples of students whose ISEE scores make them marginal for different exam school offers

- This requires knowledge of the DA mechanism; Abdulkadiroglu et al. (2017, 2022) refine the strategy using propensity score methods

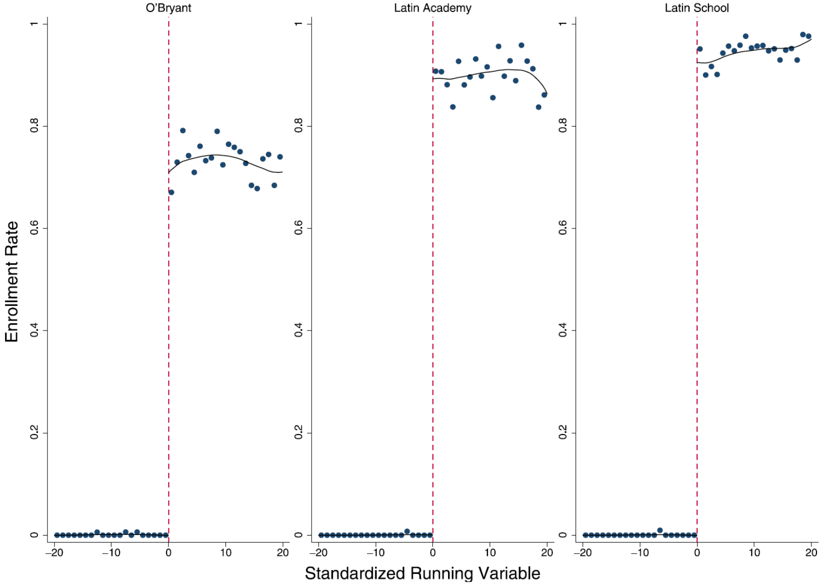
Within sharp samples, they have sharp RDs for school offers

- As a first step, we can look at a bunch of "reduced form" effects

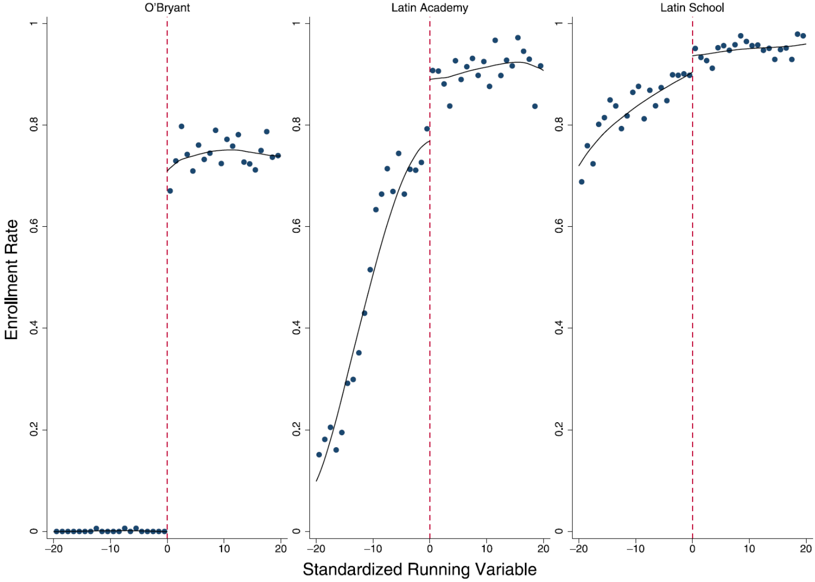
RD for Admission Offers (“Zeroth Stage”)



RD for School-Specific Enrollment



RD for Any Elite School Enrollment



Applicant Enrollments, Given Offers

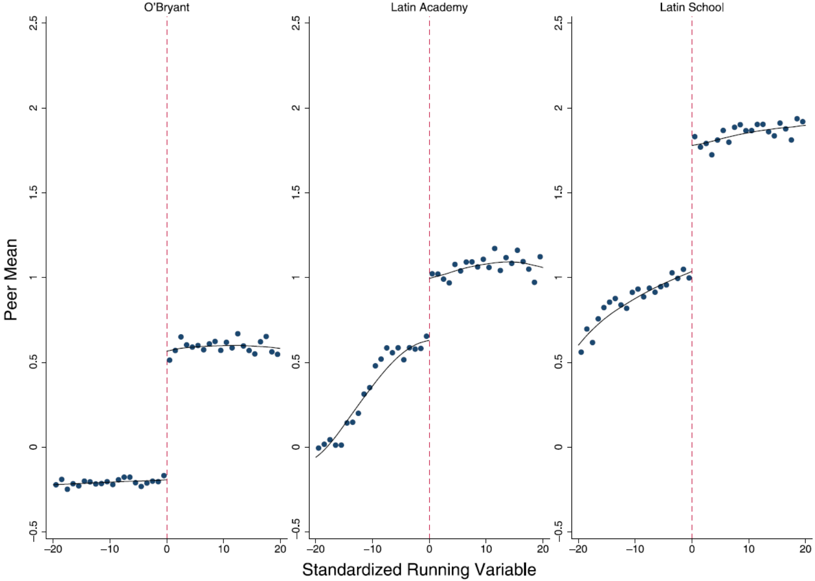
	All Applicants						Compliers		
	$Z=0$	$Z=1$	$Z=0$	$Z=1$	$Z=0$	$Z=1$	$Z=0$	$Z=0$	$Z=0$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Boston 7th Grade Applicants									
	O'Bryant		Latin Academy		Latin School		O'Bryant	Latin Academy	Latin School
Traditional Boston public schools	1.00	0.28	0.22	0.09	0.08	0.06	1.00	0.15	0.03
O'Bryant	...	0.72	0.77	...	0.06	...	...	0.84	0.05
Latin Academy	...	...	...	0.91	0.86	0.01	...	...	0.93
Latin School	...	...	...	...	...	0.92	...	...	...
Panel B. Boston 9th Grade Applicants									
	O'Bryant		Latin Academy		Latin School		O'Bryant	Latin Academy	Latin School
Traditional Boston public schools	1.00	0.32	0.27	0.14	0.14	0.04	1.00	0.13	0.10
O'Bryant	...	0.68	0.73	...	0.01	...	...	0.85	...
Latin Academy	...	...	...	0.87	0.86	0.02	...	...	0.91
Latin School	...	...	...	−0.01	...	0.94	...	0.01	...

School "fallbacks" vary by school eliteness

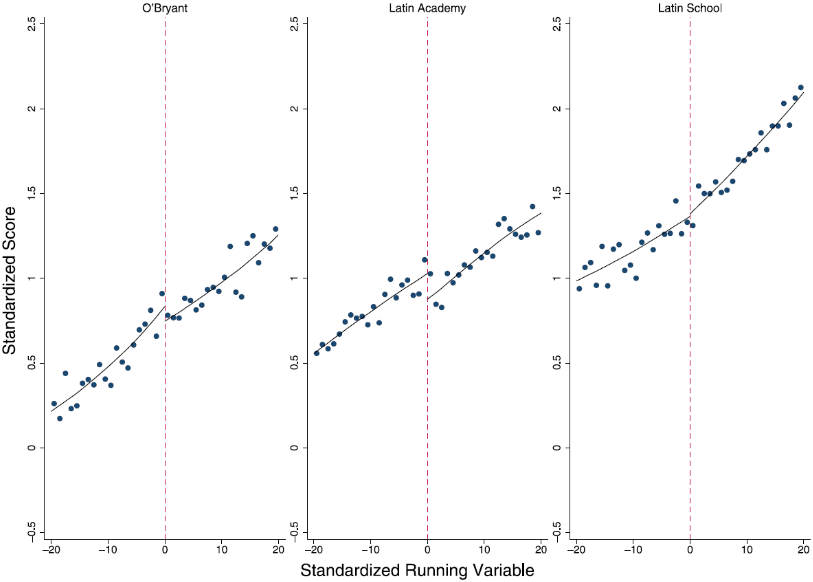
Complier fallbacks are estimated here by the "kappa-weighting" method of Abadie (2003), but AAP could also have just done IV

- Specifically, fuzzy RD with an outcome of (e.g.) $\mathbf{1}$[enroll in O'Bryant] and a treatment of $\mathbf{1}$[enroll in BLA] should yield a LATE of ≈ 0.84

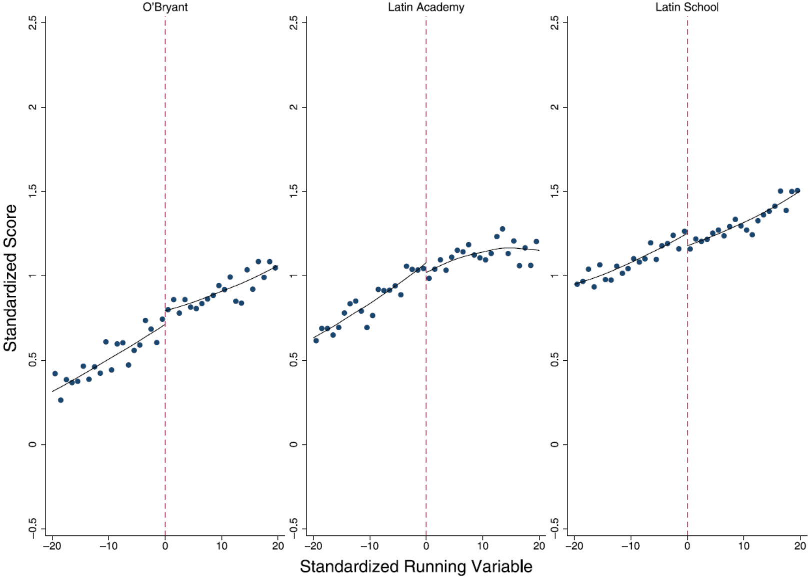
RD for Baseline Peer Achievement



RD for Middle School Math Test Outcomes



RD for High School Math Test Outcomes



Fuzzy RD with Multiple Treatments

AAP estimate second-stage equations of the form:

$$Y_{itk} = p_{it}'\theta + X_{itk}'\gamma + \varepsilon_{itk}$$

for student i applying in year t finding themselves in sharp sample k

- Y_{itk} is a later test score (math or ELA)
- p_{it} is peer achievement and/or composition post-application
- X_{it} includes RD controls (school-specific running variables and sharp sample FE) as well as year/grade FE

They instrument p_{it} with school-specific offer dummies, yielding fuzzy RD

- Interacted with year FE, to boost power (so 2SLS-weighted average)
- Possible exclusion concerns (i.e. unmodeled mediators left out of p_{it}), though given the zero reduced forms IV is a foregone conclusion...

IV Estimates

	Math					English				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	2SLS Estimates (Models With Cohort Interactions)									
Peer mean	−0.038		0.064	−0.035		0.006		0.044	−0.047	
	(0.032)		(0.080)	(0.044)		(0.030)		(0.064)	(0.051)	
Proportion nonwhite		0.145	0.421		0.160		−0.014	0.141		0.063
		(0.110)	(0.279)		(0.137)		(0.102)	(0.218)		(0.134)
Years in exam school				−0.003	0.006				0.045	0.027
				(0.036)	(0.030)				(0.034)	(0.025)
	First-Stage F-Statistics (Models With Cohort Interactions)									
Peer mean	65.8		9.1	50.0		39.8		5.7	22.8	
Proportion nonwhite		65.8	17.6		60.0		52.3	12.4		41.2
Years in exam school				12.0	16.2				10.6	15.8
N	31,911	33,313	31,911	31,911	33,313	31,222	32,185	31,222	31,222	32,185

Takeaways

This is probably the “cleanest” RD you’ll ever hope to see

- No need for any fancy bandwidth selection, de-biasing, etc... A picture here is worth a thousand words / LaTeX symbols
- The lack of a reduced-form effect on later test scores makes the IV task easier, though the authors are careful with possible treatments

Even so, important policy/economic questions remain

- RD is inherently local; are exam school effects different once we move away from the cutoff?

Outline

1. RD Identification ✓

2. RD Estimation ✓

3. Application: Abdulkadiroglu et al. (2014) ✓

4. Bonus Track 1: RD Extrapolation

5. Bonus Track 2: Design-Based RD

Parametric Extrapolation

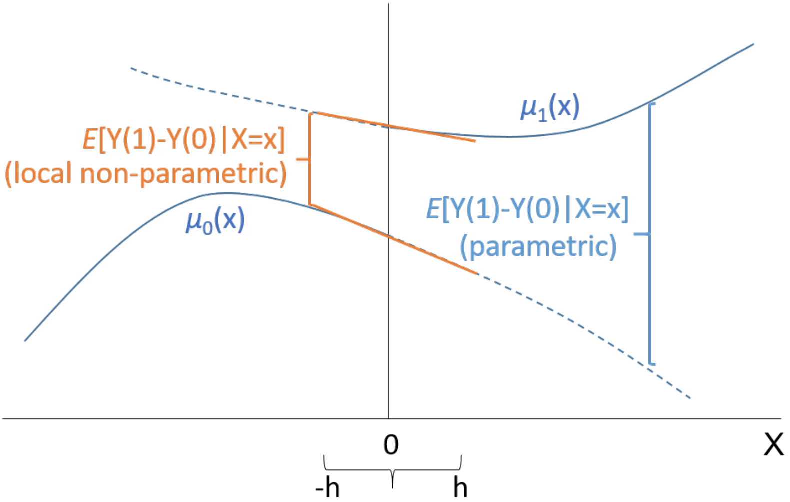
Recall: parameterizing $\mu_d(x)$ ID's effects everywhere, not just at $x = 0$

- E.g. w/quadratic $Y_i = \alpha + \beta D_i + \gamma X_i + \delta X_i^2 + \phi D_i X_i + \psi D_i X_i^2 + \varepsilon_i$, we can estimate $E[Y_i(1) - Y_i(0) \mid X_i = x] = \beta + \phi x + \psi x^2$
- Of course, this only works well when we've parameterized accurately

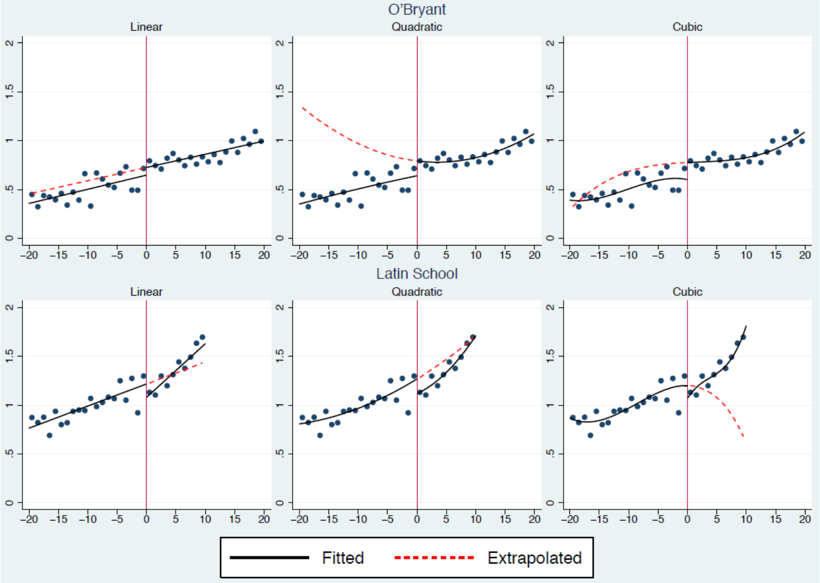
Dong and Lewbel (2015) propose a local version of this approach

- E.g. w/local linear $Y_i = \alpha + \beta D_i + \gamma X_i + \phi D_i X_i + \varepsilon_i$ in $X_i \in [-h, h]$, we can estimate $E[Y_i(1) - Y_i(0) \mid X_i = x \in [-h, h]] \approx \beta + \phi x$
- This will work well locally, assuming we've set h well
- The performance can be improved w/higher-order approximations

Extrapolation Near and Far



Extrapolation of 10th Grade Exam School Effects



Note: 9th grade applicants

Extrapolation with Covariates

Angrist and Rokkanen (2014) propose a different approach which makes use of a pre-treatment covariate W_i (e.g. middle school GPA)

Key conditional independence assumption (CIA):

$$E[Y_i(d) \mid X_i, W_i] = E[Y_i(d) \mid W_i], \;\; d \in \{0,1\}$$

E.g.: conditional on middle school GPA, the exam school running variable (ISEE score) is as-good-as-randomly assigned

- I.e. $ISEE = GPA + noise$ (see Mika's JMP for common factor version)
- AR '15 also consider this assumption in a bandwidth of $x = 0$

The CIA identifies counterfactuals:

$$\begin{aligned} E[Y_i(0) \mid D_i = 1] &\overset{LIE}{=} E[E[Y_i(0) \mid X_i \geq 0, W_i] | X_i \geq 0] \\ &\overset{CIA}{=} E[E[Y_i(0) \mid X_i < 0, W_i] | X_i \geq 0] \\ &= E[E[Y_i \mid D_i = 0, W_i] | D_i = 1] \end{aligned}$$

and similarly for $E[Y_i(1) \mid D_i = 0]$

Testing the CIA

Knowing $D_i = \mathbf{1}[X_i \geq 0]$ lets us test CIA. We should have:

$$E[Y_i(1) \mid X_i, W_i, D_i = 1] = E[Y_i(1) \mid W_i, D_i = 1]$$
$$E[Y_i(0) \mid X_i, W_i, D_i = 0] = E[Y_i(0) \mid W_i, D_i = 0]$$

I.e. the running variable should not predict Y_i when controlling flexibly for W_i on the left and the right of the treatment cutoff

The CIA implies conditional effect ignorability (CEI):

$$E[Y_i(1) - Y_i(0) \mid X_i, W_i] = E[Y_i(1) - Y_i(0) \mid W_i]$$

Under CEI, we can intuitively reweight conditional non-parametric RDs

- But unlike CIA, CEI is not directly testable: we can only identify $E[Y_i(1) - Y_i(0) \mid X_i = x, W_i = w]$ at $x = 0$
- CIA has much more power by using variation away from the cutoff

Using the CIA

AR '15 consider two approaches for using conditional independence:

1. Propensity score weighting: estimate $E[D_i \mid W_i]$ and inversely weight
2. Kline '11 linear reweighting (recall the TVA application in Chapter 1)

Main interest: ATU for O'Bryant and ATT for BLS

CIA fails its test for 7th grade applicants, but passes for 9th grade

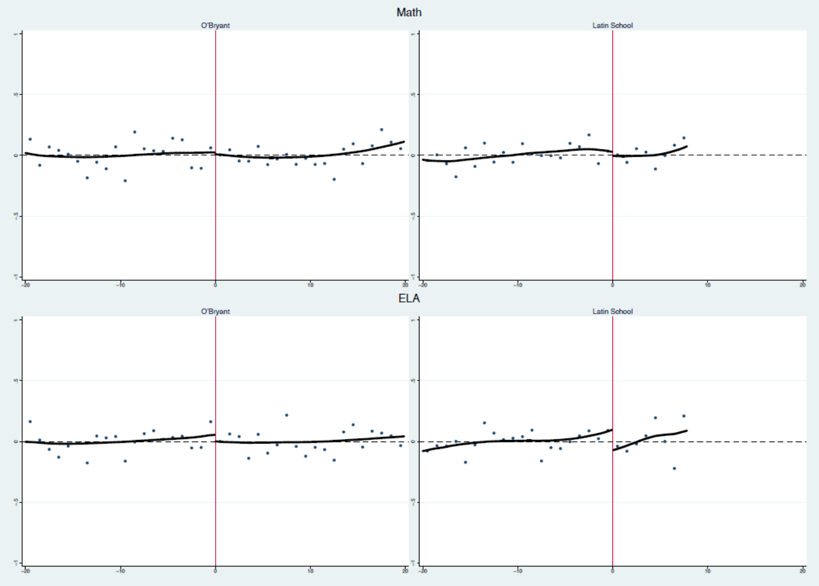
- Consider extrapolations in different windows of ISEE scores
- Effects seem positive at O'Bryant but nill at BLS

Testing CIA in 10th Grade

	Math				ELA			
	O'Bryant		Latin School		O'Bryant		Latin School	
	D = 0	D = 1	D = 0	D = 1	D = 0	D = 1	D = 0	D = 1
Window	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A. 7th Grade Applicants							
20	0.022***	0.015***	0.008***	0.014***	0.015***	0.006	0.013***	0.018***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.004)	(0.005)	(0.003)	(0.003)
	838	618	706	748	840	621	709	750
15	0.023***	0.015***	0.010***	0.012***	0.014**	0.006	0.007	0.015***
	(0.006)	(0.005)	(0.003)	(0.003)	(0.005)	(0.006)	(0.005)	(0.005)
	638	587	511	517	638	590	514	519
10	0.030***	0.016**	0.010*	0.007	0.024**	0.001	0.012	0.012
	(0.009)	(0.008)	(0.006)	(0.005)	(0.010)	(0.009)	(0.010)	(0.008)
	419	445	335	347	421	447	338	348
	Panel B. 9th Grade Applicants							
20	0.002	0.005	0.008**	0.018	0.003	0.002	0.006	0.055
	(0.004)	(0.003)	(0.003)	(0.028)	(0.004)	(0.004)	(0.005)	(0.053)
	513	486	320	49	516	489	320	50
15	0.010	0.000	0.006	0.018	0.009	-0.000	0.000	0.055
	(0.006)	(0.005)	(0.006)	(0.028)	(0.006)	(0.006)	(0.007)	(0.053)
	375	373	228	49	376	374	229	50
10	0.003	-0.001	0.007	0.018	0.014	-0.004	0.014	0.055
	(0.011)	(0.009)	(0.009)	(0.028)	(0.011)	(0.010)	(0.015)	(0.053)
	253	260	142	49	253	261	142	50

Note: Running variable coefficients

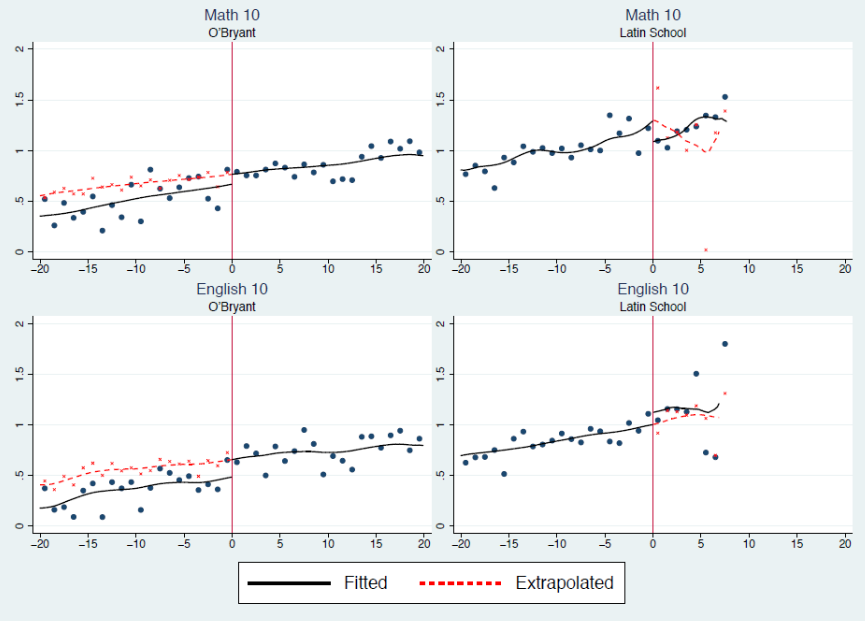
Visual CIA Test Results, 7th Grade Applicants



Applying CIA, 7th Grade Applicants

	Linear Reweighting				Propensity Score Weighting			
	Math		ELA		Math		ELA	
	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School
Window	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
20	0.156***	-0.031	0.198***	0.088	0.148***	-0.028	0.251***	0.054
	(0.039)	(0.094)	(0.041)	(0.084)	(0.052)	(0.192)	(0.090)	(0.207)
N untreated	513	320	516	320	509	320	512	320
N treated	486	49	489	50	482	49	485	50
15	0.129***	-0.080	0.181***	0.051	0.116**	-0.076	0.202***	0.018
	(0.043)	(0.055)	(0.047)	(0.088)	(0.052)	(0.161)	(0.069)	(0.204)
N untreated	375	228	376	229	373	228	374	229
N treated	373	49	374	50	370	49	371	50
10	0.091*	-0.065	0.191***	-0.000	0.123*	-0.093	0.186**	-0.052
	(0.054)	(0.054)	(0.055)	(0.097)	(0.070)	(0.249)	(0.073)	(0.356)
N untreated	253	142	253	142	253	142	253	142
N treated	260	49	261	50	258	49	259	50

Visualizing CIA-Based Counterfactuals, 7th Grade Apps



Outline

1. RD Identification✓
2. RD Estimation✓
3. Application: Abdulkadiroglu et al. (2014)✓
4. Bonus Track 1: RD Extrapolation✓
5. Bonus Track 2: Design-Based RD

A Different View of RD

So far our analysis of RD has been strictly "model-based"

- I.e. leveraging restrictions on $\mu_d(x) = E[Y_i(d) \mid X_i = x]$, without any explicit link to the quasi-experimental "design" of observed shocks

Consider the alternative "local randomization" view (Cattaneo et al. 2015)

- D_i is as-good-as-randomly assigned within a bandwidth of $X_i = 0$
- Formally: $D_i \perp (Y_i(1), Y_i(0)) \mid X_i \in [-h, h]$
- Like AR (2015)'s CIA, with $W_i = \mathbf{1}[-h \leq X_i \leq h]$

Rather intuitive, and can lead to a broader RD toolkit

- E.g. brings RD into the Borusyak and Hull (2022) framework, allowing treatments/instruments with complex/multiple thresholds

Example: Kott (2022)

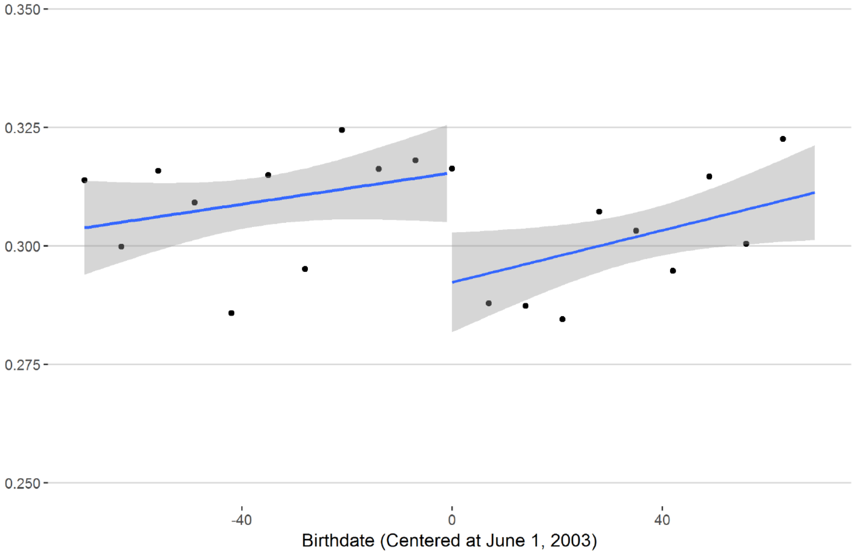
Assaf's JMP studies a 2003 policy change in Israel which reduced public child allowances discontinuously in the birthdate of some family's children

- Pre-reform, larger families got more generous allowances
- The reform implemented a new flat schedule, but only to children born after June 1, 2003
- Families with children born near this date thus saw payments change discontinuously in the pivotal child's birthdate, but in a complex way

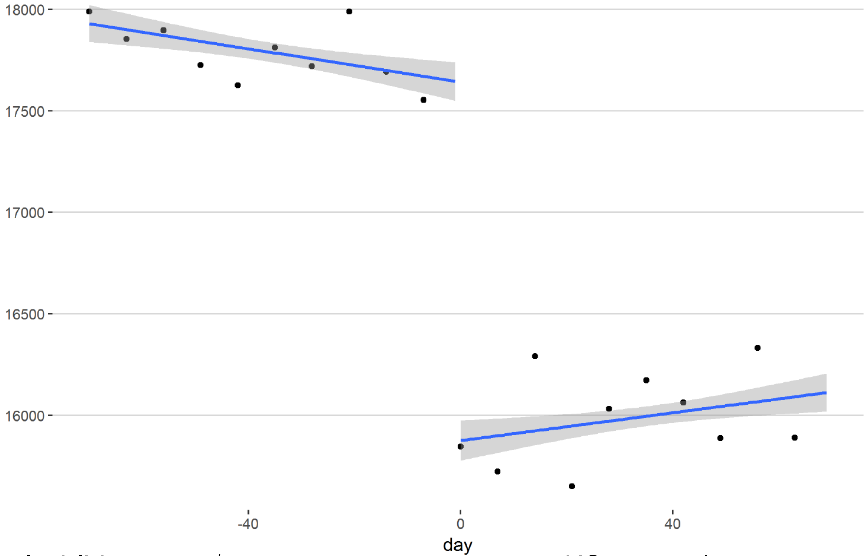
He shows large effects on high school matriculation rates (reduced form) and allowances (first stage) in a conventional non-parametric fuzzy RD

- He then confirms these effects in a local randomization setup, before showing how this setup allows him to recenter allowances themselves
- Specifically $\widetilde{CA}_i = CA_i - \frac{1}{2}(CA_i^+ + CA_i^-)$ where CA_i is the realized allowance and CA_i^+ (CA_i^-) is the predicted allowance if the pivotal child were born after (before) June 1, 2003

Reduced Form: RD in High School Matriculation



First Stage: RD in Allowance Amount



Implied IV: -0.024 / -1,600 $\approx$ 1.5pp increase in HS matriculation per $\$1,000$

Recentered Child Allowance Balance Test

	Dependent variable:	
	Child Allowance	Recentered Child Allowance
	(1)	(2)
Arab	400***	−32
	(74)	(44)
Mother Schooling	−6	2
	(11)	(6)
Father Schooling	13	−1
	(9)	(5)
Family Size	2,190***	12
	(32)	(15)
Birth Spacing	−152***	11
	(24)	(14)
Immigrant Parent	87	−4
	(67)	(39)
Ultraorthodox	260***	−28
	(84)	(46)
Birth Order	−164***	−1
	(26)	(11)
Year of Birth	326***	4
	(10)	(5)
Boy	−16	3
	(40)	(19)
Constant	−645,303***	−7,435
	(19,371)	(9,594)
Joint F-test	3428	0.74
p-value	0	0.64
Number of Students	9,869	
Number of Families	2,998	

Recentered Child Allowance Balance Test

	Dependent Variable: High School Matriculation				
	OLS			2SLS	
	(1)	(2)	(3)	(4)	(5)
CA by Age 18 (in 1,000 USD)	-0.024^{***}	-0.024^{***}	0.005	0.015^{**}	0.018^{***}
	(0.0005)	(0.001)	(0.004)	(0.006)	(0.007)
Controls	No	No	Yes	Yes	No
Adjusted R^2	0.045	0.044	0.311	0.311	0.050
Bandwidth (in days)	150	21	21	21	21

Virtually identical IV estimate as with conventional fuzzy RD

- In theory, recentering should yield more power via non-random exposure to the RD "shock" (in practice, the two SEs are similar)

Shortcomings of Local Randomization

Though intuitive, design-based RD can be seen as less elegant than conventional (continuity-based) RD

- No asymptotic guidance for picking the design bandwidth
- Local randomization of X_i implies locally flat $\mu_d(x)$ (testable!)

Eckles et al. (2020) propose an alternative design-based framework

- Key idea: observed running variable is measured with exogenous noise
- When the measurement error distribution is known, a weighted marginal treatment effect is identified by an integral equation
- Estimation and inference are a bit involved, and not clear this is appropriate for all RD settings (e.g. close elections, boundaries...)
- Suffice to say, more work to be done in this area!