How Far is Too Far? Generalization of a Regression Discontinuity Design Away from the Cutoff

Magdalena Bennett

January 9, 2020

Teachers College, Columbia University

My Research Agenda

- Use of new methods for impact evaluation in experiments and observational studies:
    - Spillover effects through network of peers.
    - Performance prediction on college admission using ML.
- Development and improvement of causal inference methods:
    - Representative template matching.
    - Generalization of Regression Discontinuity Design.

My Research Agenda

- Use of new methods for impact evaluation in experiments and observational studies:
    - Spillover effects through network of peers.
    - Performance prediction on college admission using ML.
- Development and improvement of causal inference methods:
    - Representative template matching.
    - Generalization of Regression Discontinuity Design.

Today's talk

Motivation

Generalized Regression Discontinuity Design

- Framework
- GRD in practice
- Simulations

Application: Free Higher Education in Chile

Conclusions

Today's talk

Motivation

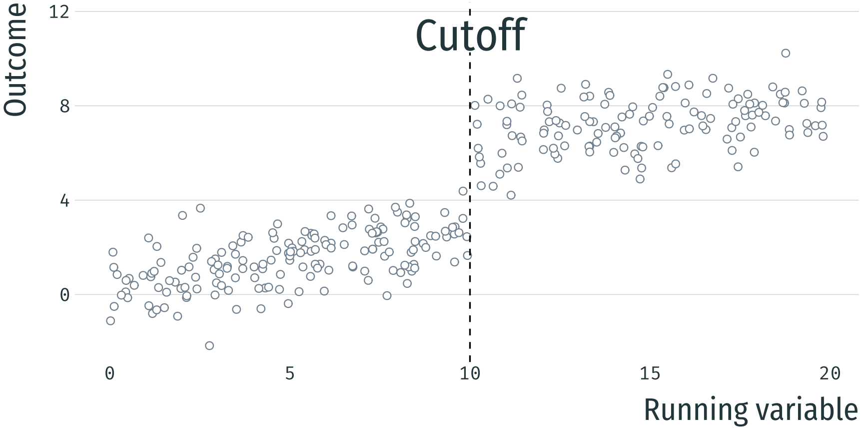
Generalized Regression Discontinuity Design

- Framework
- GRD in practice
- Simulations

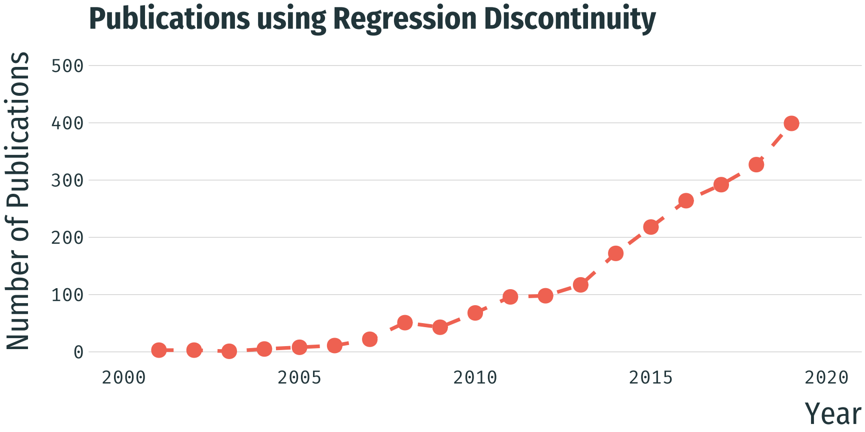
Application: Free Higher Education in Chile

Conclusions

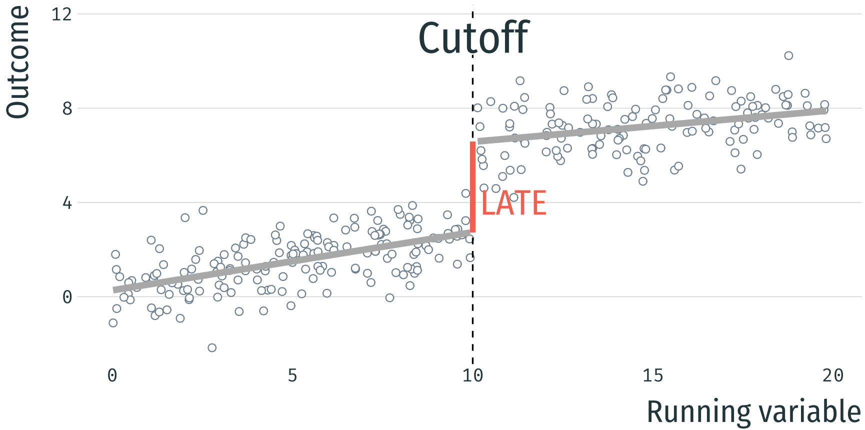
Regression discontinuity design



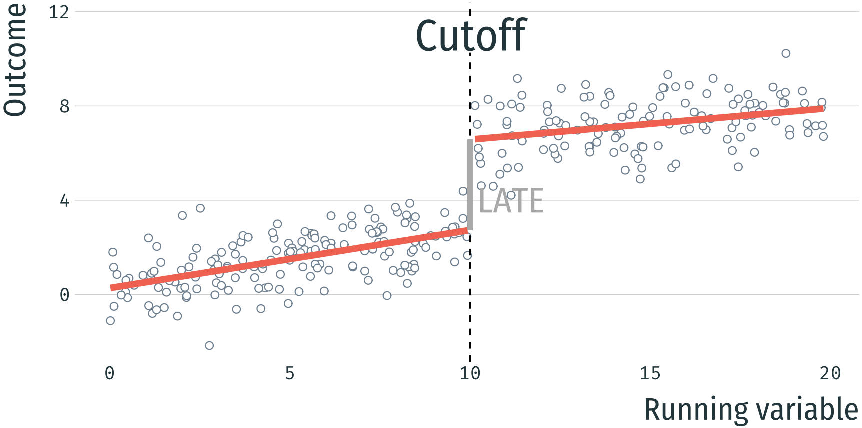
Regression discontinuity design: Increasingly popular



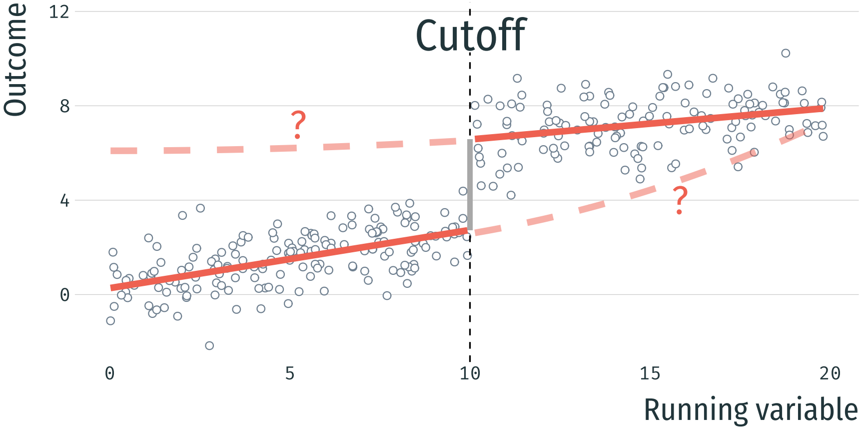
Regression discontinuity design: Strong interval validity



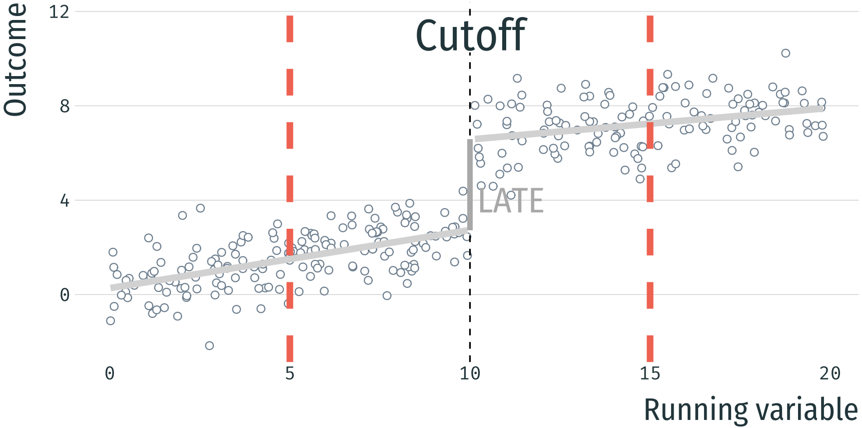
Regression discontinuity design: Limited external validity



Regression discontinuity design: Limited external validity



Regression discontinuity design: Generalization bandwidth?



This paper

Estimation of ATT for population within a generalization interval:

- Pre-intervention period (or other group) informs generalization interval

  (Wing & Cook, 2013; Keele, Small, Hsu, & Fogarty, 2019)

- Leverage the use of predictive covariates

  (Angrist & Rokkanen, 2015; Rokkanen, 2015; Keele, Titiunik, & Zubizarreta, 2015)

- Based on local randomization near the cutoff

  (Lee, 2008; Cattaneo, Frandsen, & Titiunik, 2015)

This paper

Main advantages:

- Gradual approach
  - No need for “All or Nothing”
  - Interval informed by the data (Cattaneo et al., 2015)
- No extrapolation of population characteristics
  - Compare like-to-like (Rosenbaum, 1987)
  - Makes overlap region explicit
- Generalization to population of interest
  - Use of representative template matching

    (Silber et al, 2014; Bennett, Vielma, & Zubizarreta, 2018)
- Sensitivity analysis to hidden bias (Rosenbaum, 2010; Keele et al., 2019)

Outline

Generalized Regression Discontinuity Design (GRD)

Setup:

- Two periods: pre- and post-intervention ($t = 0$ and $t = 1$)
- R determines assignment to Z in $t = 1$, e.g.:

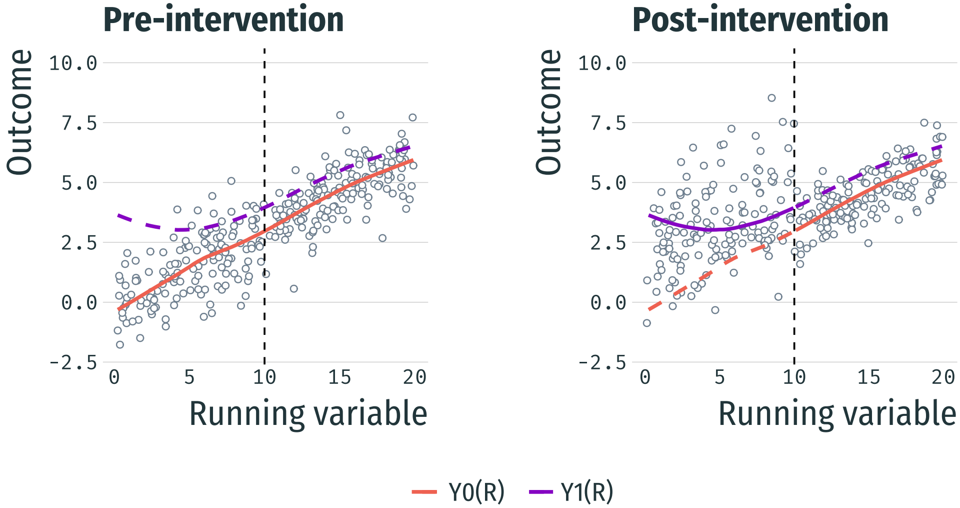
$$Z = \mathbb{I}(R < c)$$

- Potential outcomes under treatment $z = 0, 1$:

$$Y_{it}^{(z)} = g_z(\mathbf{X}_{it}, \mathbf{u}_{it}, r_{it}) + z_{it} \cdot \underbrace{\tau_{it}(\mathbf{X}_{it}, \mathbf{u}_{it}, r_{it})}_{\text{Treat. Effect}} + \underbrace{\alpha_t}_{\text{Period FE}}$$

  - $\mathbf{X}$: Predictive covariates
  - $\mathbf{u}$: Unobserved confounder
  - τ_i: individual causal effect

Two periods for GRD



GRD: A gradual approach

- Conditional expectations of potential outcomes:

$$Y_0^{(0)}(R) = \mathbb{E}[Y_{i0}^{(0)}|R] = \mu(R)$$

$$Y_0^{(1)}(R) = \mathbb{E}[Y_{i0}^{(1)}|R] = \underbrace{\mu(R)}_{\text{Avg. Outcome by R}} + \underbrace{\tau(R)}_{\text{Treat. Effect by R}}$$

- Identify generalization interval $H = [H_-, H_+]$ for $t = 0$:

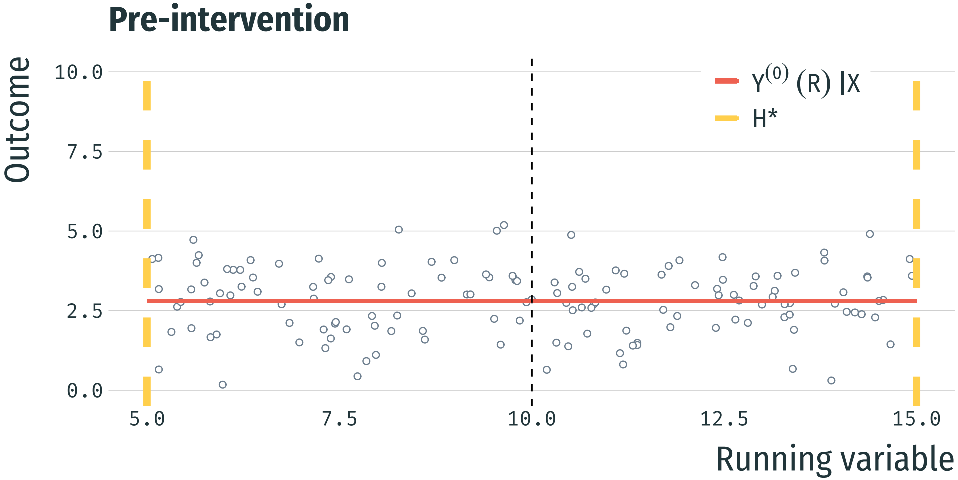
$$R_i = h(\mathbf{X}_i) + \eta_i \quad \forall\, R_i \in H$$

where $H^* = \max\{|H|\}$.

- If H^* exists, then for a set of covariates $\mathbf{X} = \mathbf{X}_T$:

$$Y_0^{(0)}(R')|\mathbf{X}_T = Y_0^{(0)}(R'')|\mathbf{X}_T \quad \text{for any } R', R'' \in H^*$$

Conditional Outcome within Generalization Interval



GRD: Assumptions for generalization to t=1

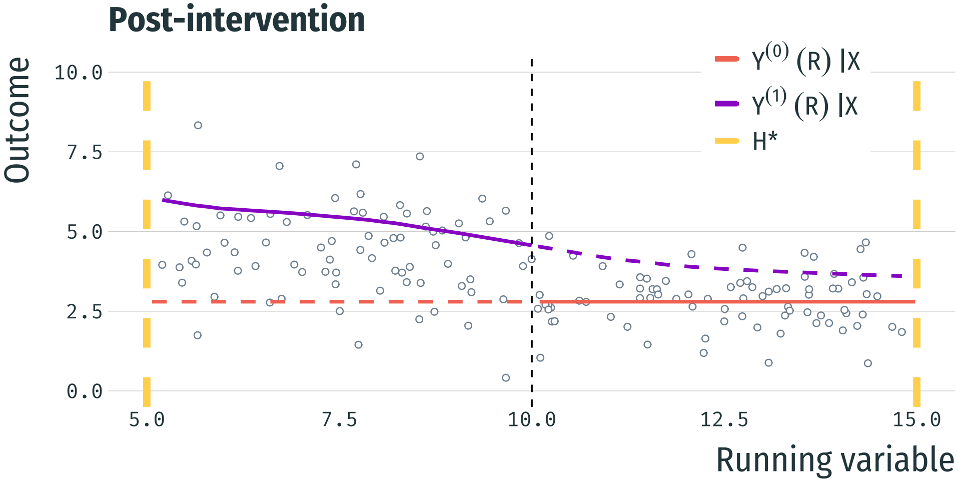
Assumption I: Conditional time-invariance under control

$$Y_0^{(0)}(R|\mathbf{X}) = Y_1^{(0)}(R|\mathbf{X}) + \alpha_t, \quad \forall\, R \in H^*$$

Assumption II: Heterogeneity only through τ

$$Y_1^{(1)}(R|\mathbf{X}) \perp \mathbf{u} \quad \forall R \in H^*$$

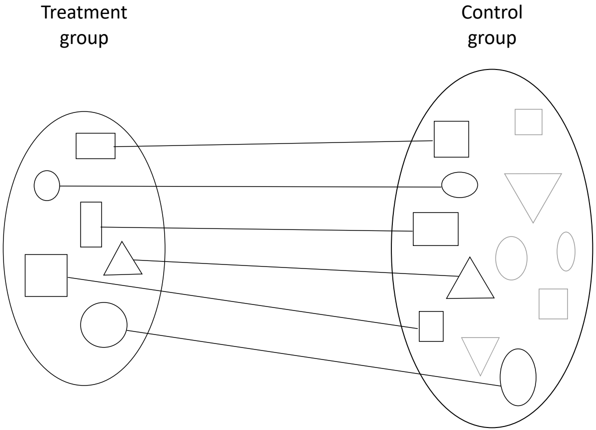
GRD: Estimating effects away from the cutoff



Outline

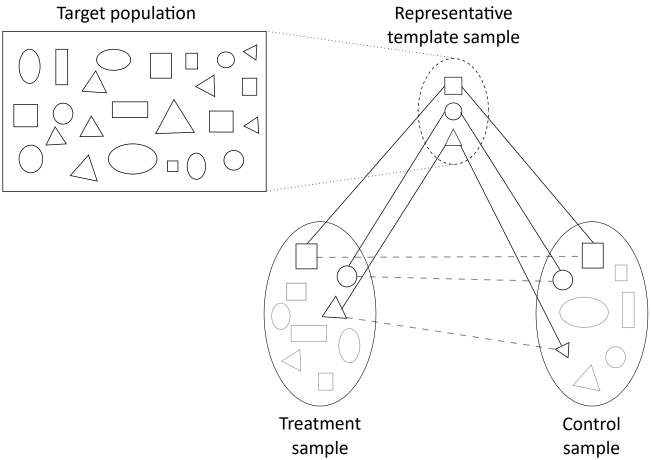
Overview: Representative Template Matching (Bennett, Vielma, & Zubizarreta, 2019)

Traditional Matching



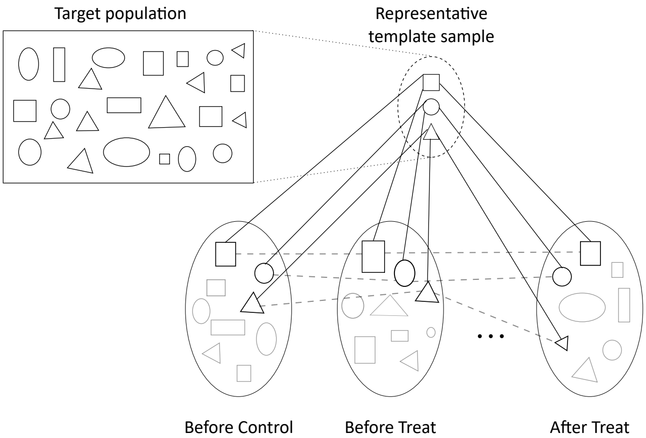
Overview: Representative Template Matching (Bennett, Vielma, & Zubizarreta, 2019)

Representative Template Matching for Two Groups

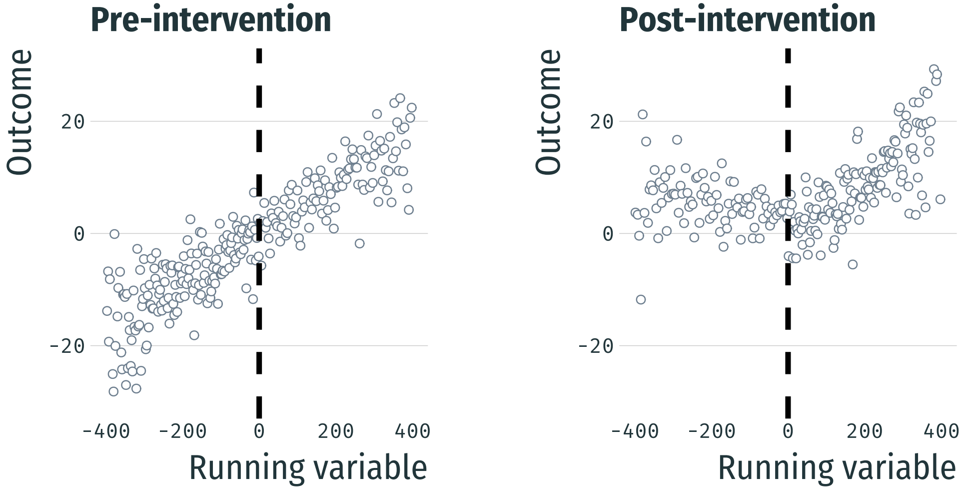


Overview: Representative Template Matching (Bennett, Vielma, & Zubizarreta, 2019)

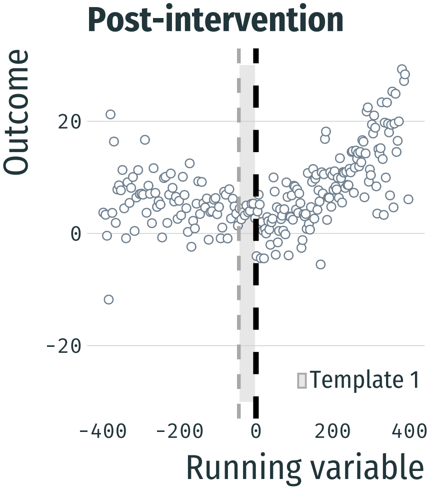
Representative Template Matching for Diff-in-Diff



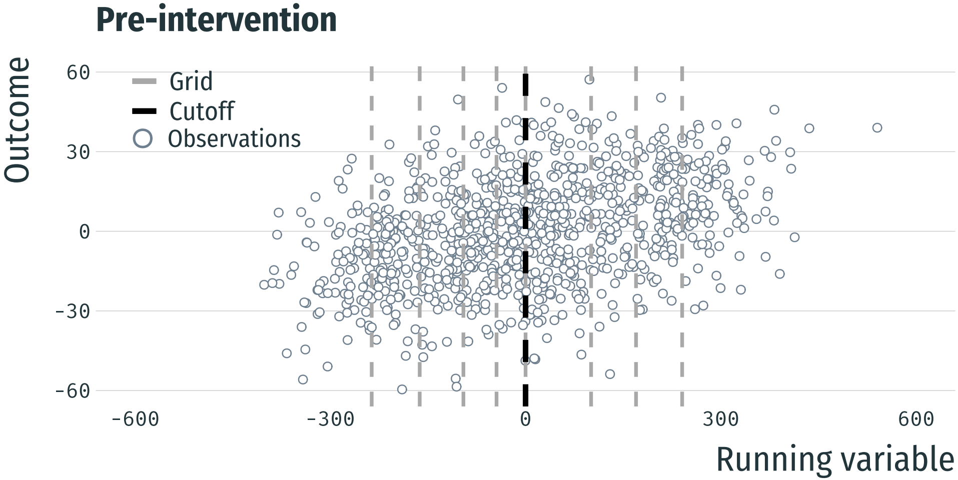
GRD: Start with two periods



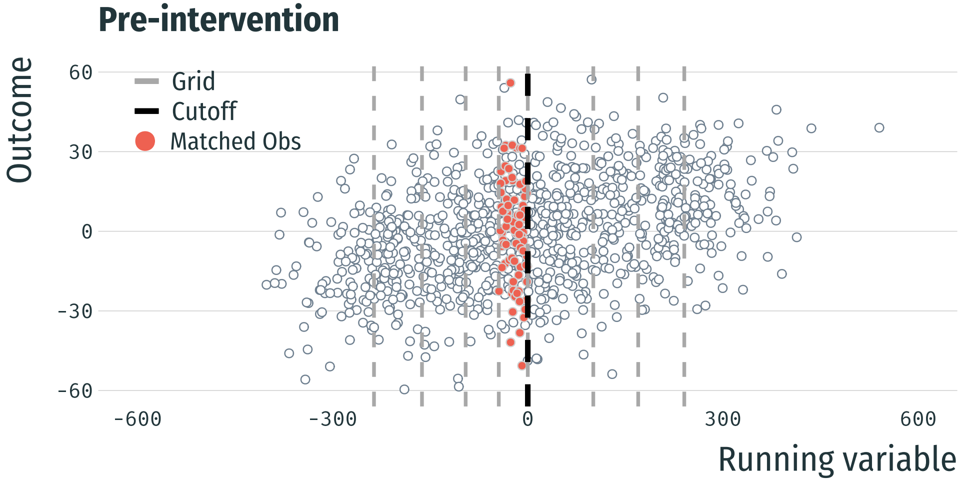
GRD: Select template sample from post-intervention



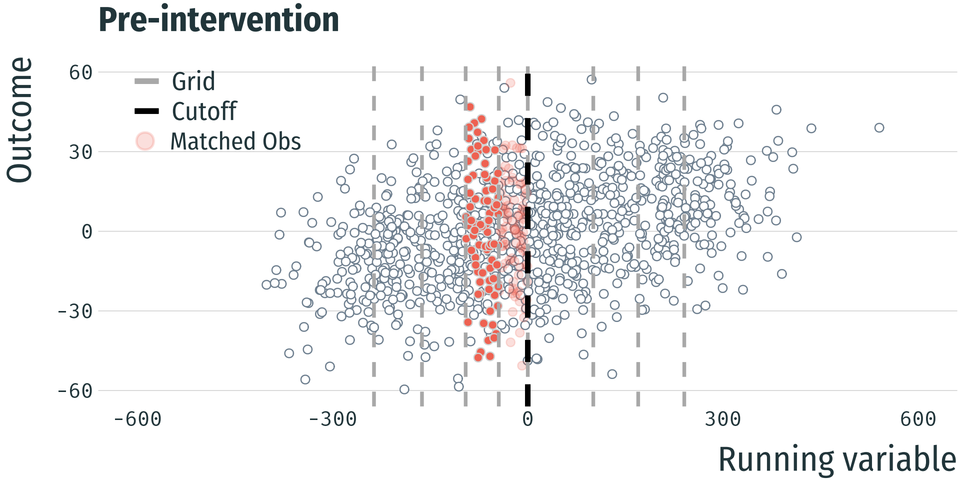
GRD: Divide pre-intervention into grid



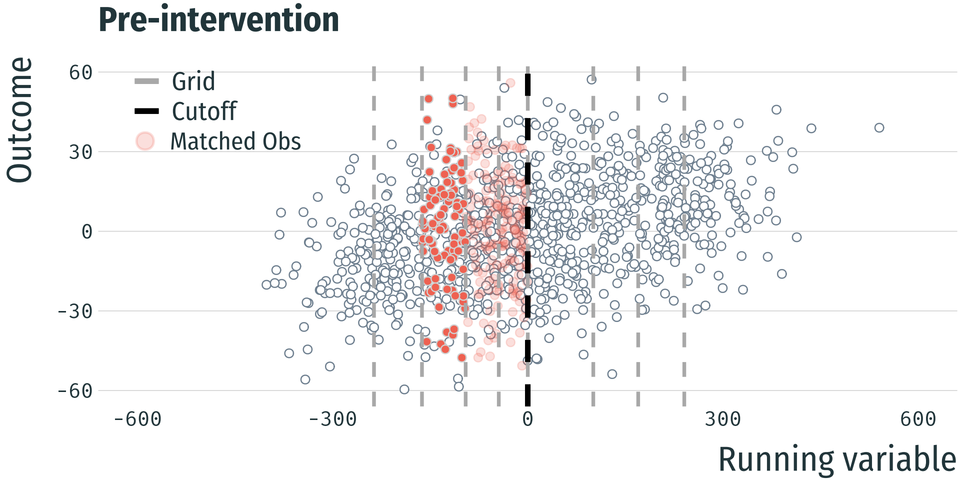
GRD: Match template to grid



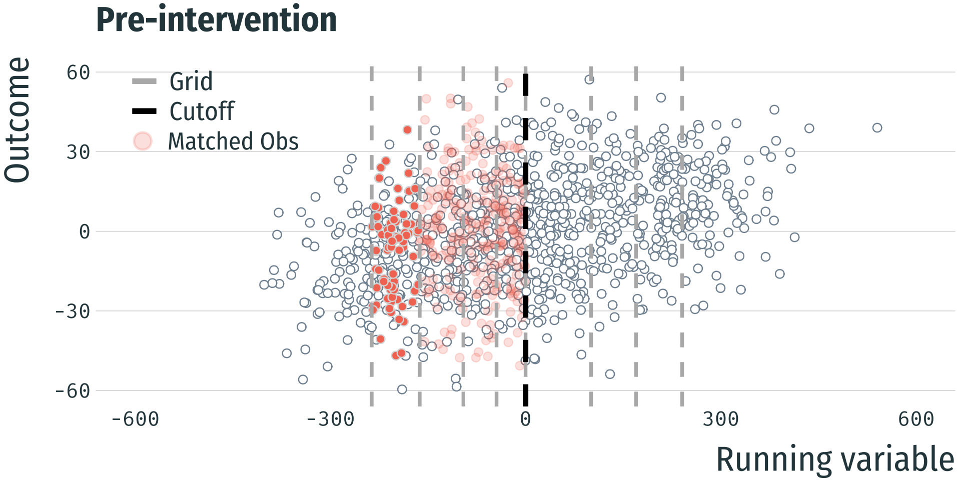
GRD: Match template to grid



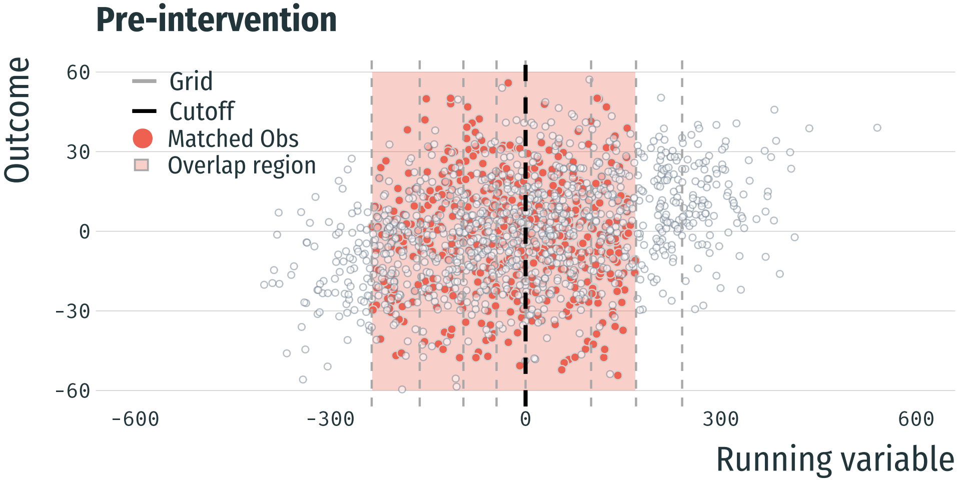
GRD: Match template to grid



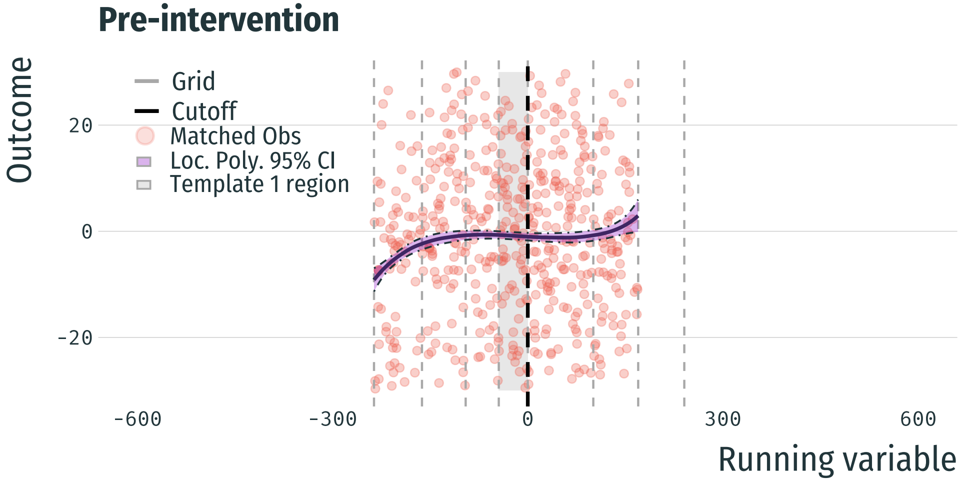
GRD: Match template to grid



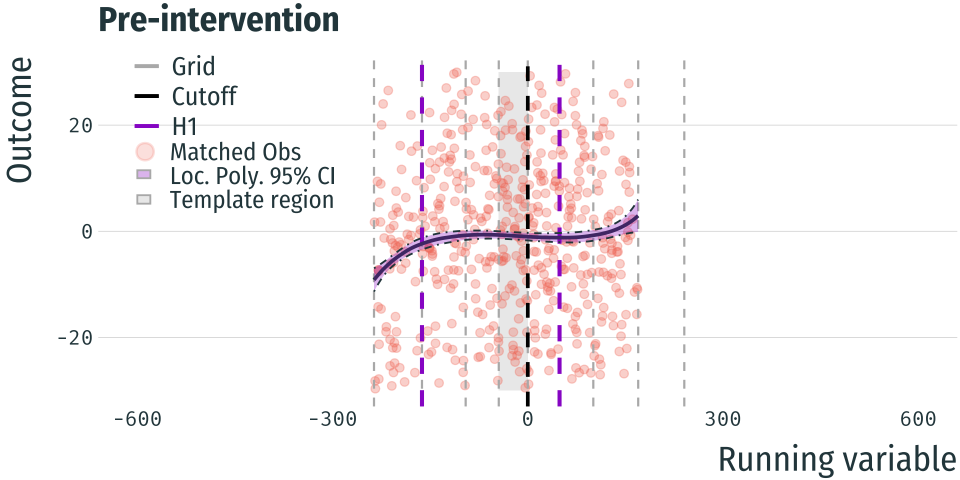
GRD: Explicit overlap region



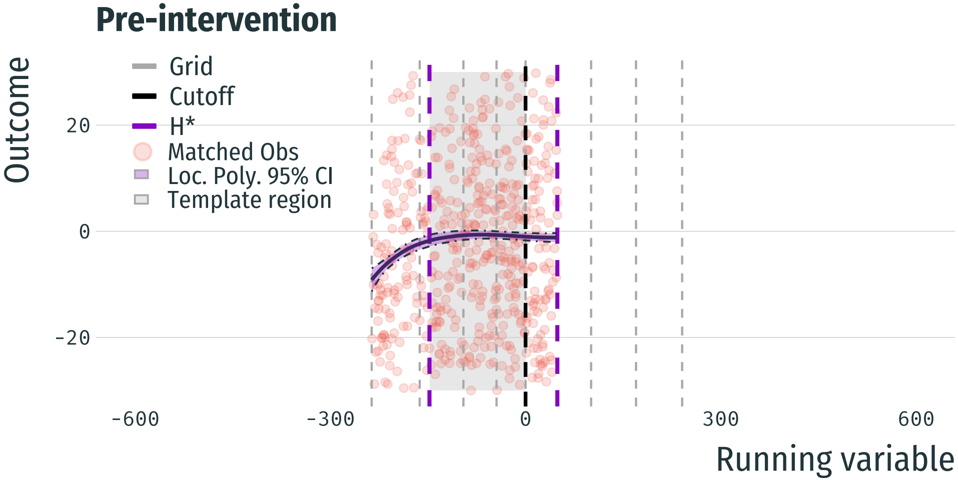
GRD: Estimate local polynomial on matched sample



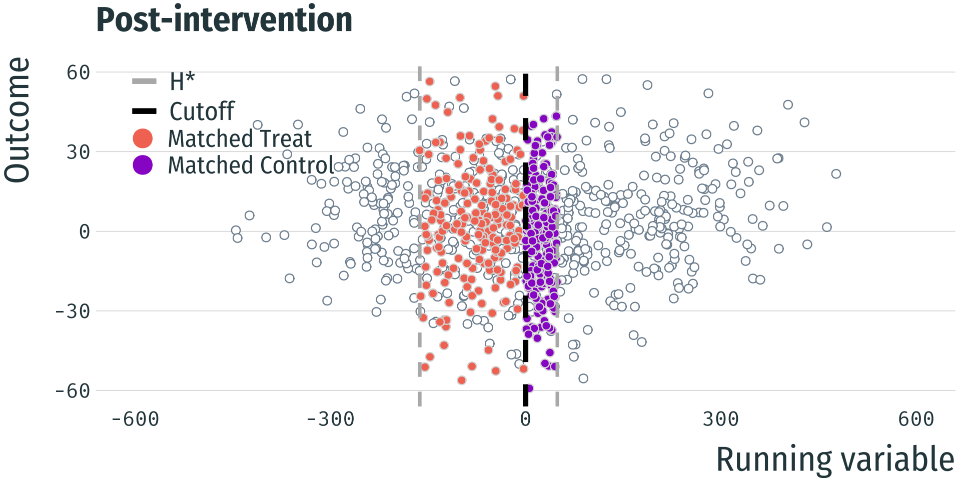
GRD: Identify generalization interval H_1



GRD: Repeat procedure until $H_j \subseteq T$



GRD: Match post-intervention period to the template



GRD: ATT Estimation

Straightforward estimation given matched sample:

- E.g. paired t-test:

$$\hat{\tau}_{ATT} = \sum_{k=1}^{N} \frac{Y_{k(1)1} - Y_{k(0)1} - (Y_{k(1)0} - Y_{k(0)0})}{N} = \sum_{k=1}^{N} \frac{d_k}{N}$$

$Y_{k(z)t}$: outcome within matched group k with treatment $z = \{0, 1\}$ for period $t = \{0, 1\}$

Outline

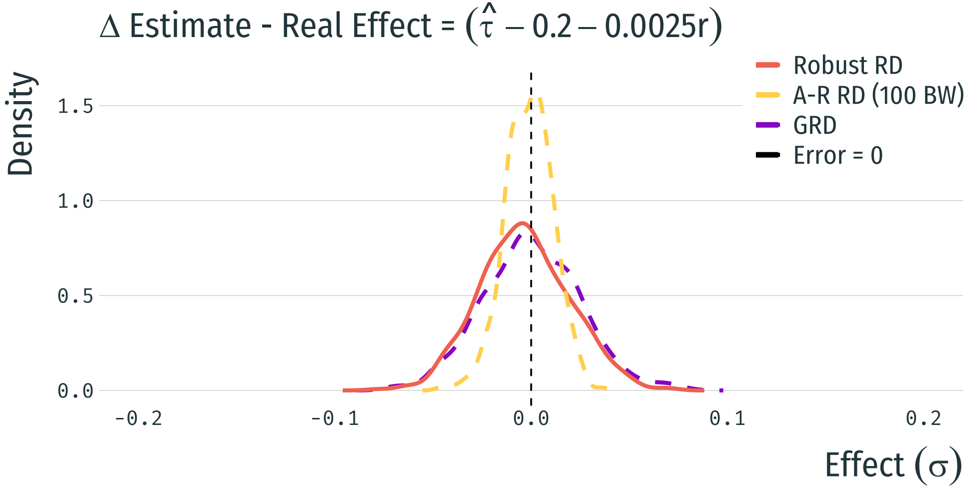
Simulations: Assess performance of GRD

- Compare GRD performance to `rdrobust()` (LATE) (Calonico et al., 2018) and A-R RD generalization (ATT) (Angrist & Rokkanen, 2015)
  $\rightarrow$ 500 simulations
- Simulations scenarios:
  - Low vs. high correlation:
    $$Corr(R, X) = \{0.33, 0.66\}$$
  - Constant vs. heterogeneous effects:
    $$\tau_{constant} = 0.2\sigma$$
    $$\tau_{linear} = 0.2\sigma + 0.0025\sigma \cdot R$$
    $$\tau_{quad} = 0.2\sigma + 0.0025\sigma \cdot R^2$$
  - Small vs. large samples:
    2,000 vs 20,000 obs

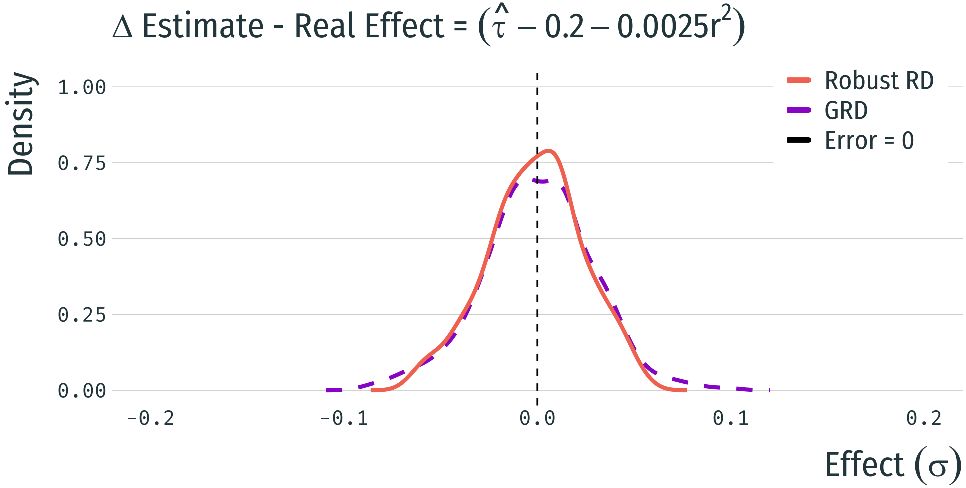
Simulation Results

- Similar performance for GRD and RD robust.
- A-R RD generalization performs better than GRD **if treatment effect is linear and tested within generalization interval**.
- Unlike GRD, A-R RD biased under other functional forms of treatment effect.
    - 23% simulations failed residual test in quadratic treatment effect.

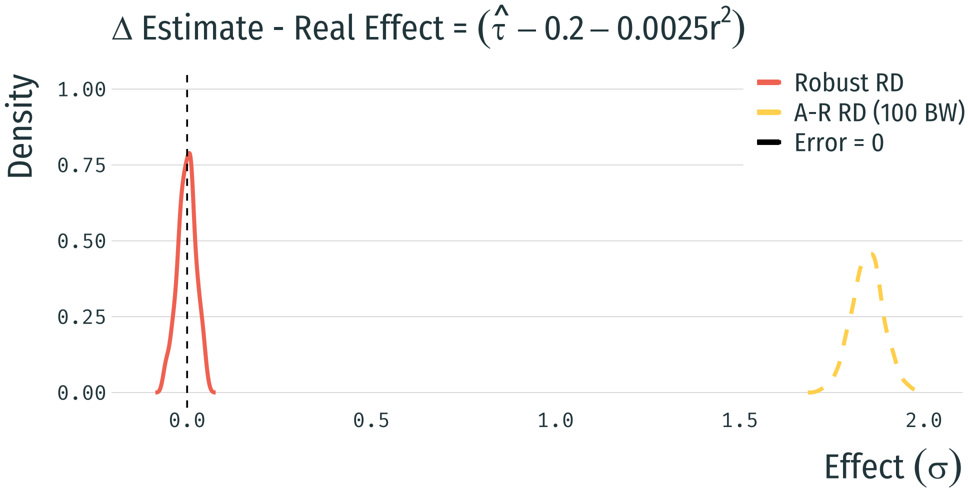
Simulation distribution: τ_{linear} (s: high corr & large sample)



Simulation distribution: τ_{quad} (s: high corr & large sample)



Simulation distribution: τ_{quad} (s: high corr & large sample)



Outline

Free Higher Education (FHE) in Chile

Higher education in Chile:

- Centralized admission system (deferred admission mechanism)
- Admission score: PSU score + GPA score + ranking score
- Before 2016: Scholarships + government-backed loans

Free higher education policy:

- Introduced in December 2015 (unanticipated)
- Eligibility: Lower 50% income distribution + admitted to eligible program

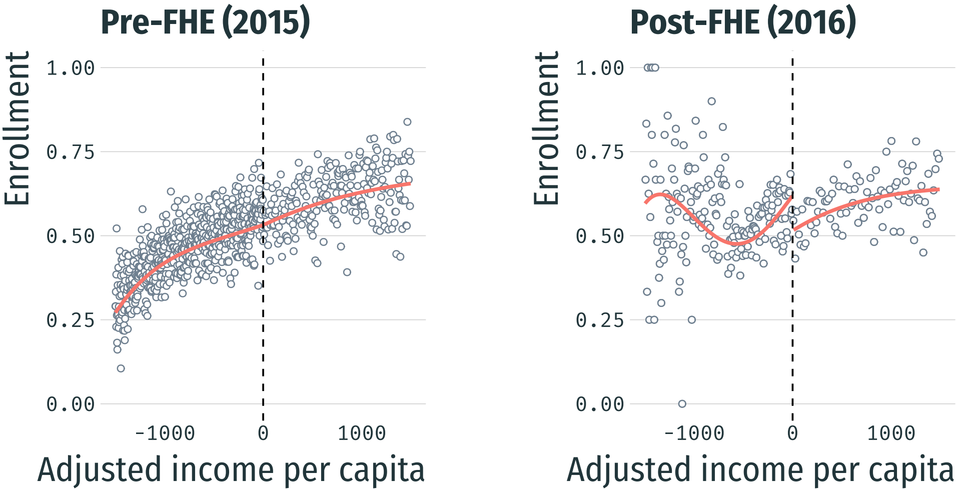
FHE: Research Question

- **Treatment:** SE eligibility for FHE
- **Two outcomes:** Application to university and enrollment
  - Lower-income students $\rightarrow$ financial constraints
  - Salience of policy
- Larger effects for students away from the cutoff?
  - Compare RD and GRD results

FHE: Data

- **3 Cohorts:** 2014, 2015, and 2016. ($\sim$ 200,000 students)
- **Rich baseline data:** Demographic and socioeconomic data at student level, 10th (8th) grade standardized scores, school characteristics.
- **Application data:** Scores by subject, application, enrollment.

FHE: How does the RDs look like?

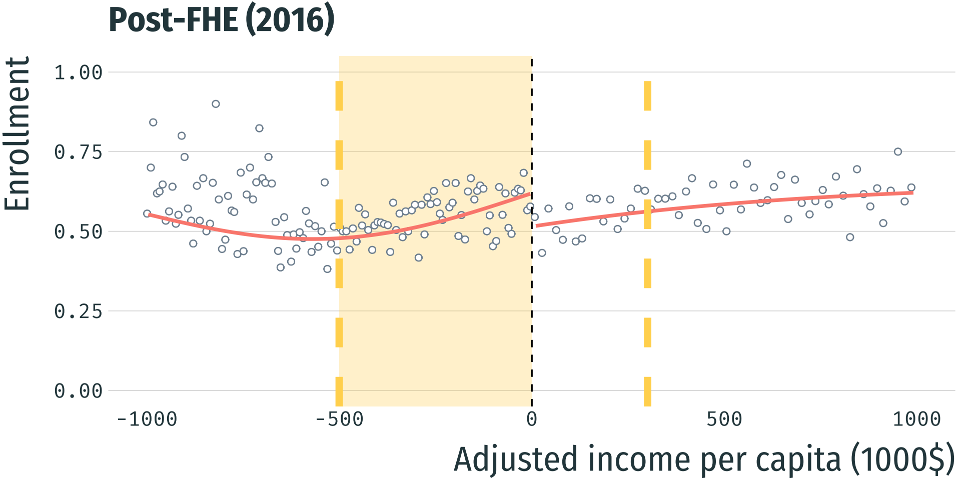


GRD for Free Higher Education

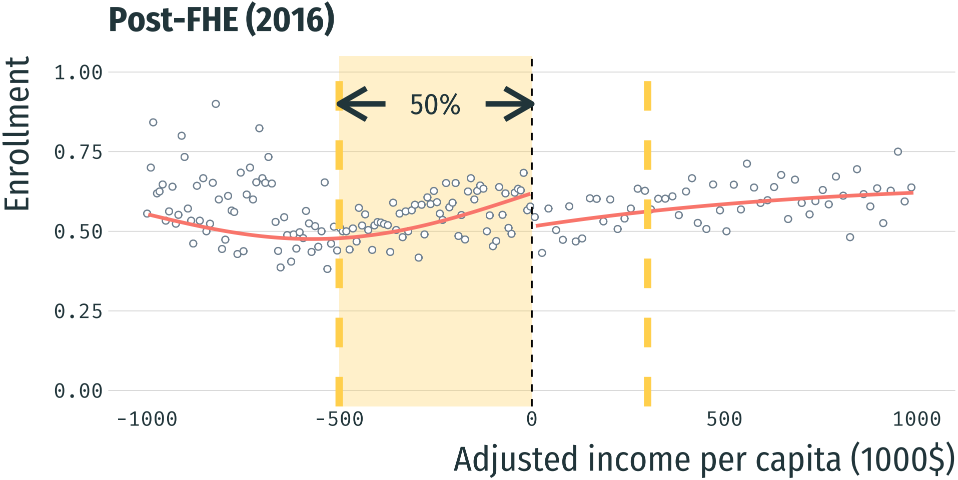
Steps for GRD:

- Select template size: $N = 1,000$
- 20 bins for grid
- MIP matching:
  - Restricted mean balance (0.05 SD):
    - Academic performance, school characteristics, demographic/socioeconomic variables.
  - Fine balance:
    - Gender, mother's and father's education (8 cat), PSU Language score (deciles), PSU math score (deciles), HS GPA (quintiles).
- Generalization interval: [-M$500.3, M$300.9]

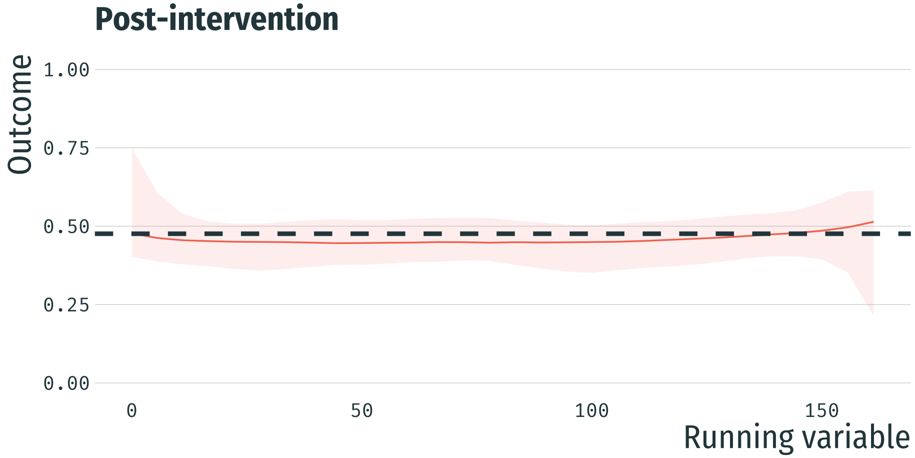
For what population are we generalizing for?



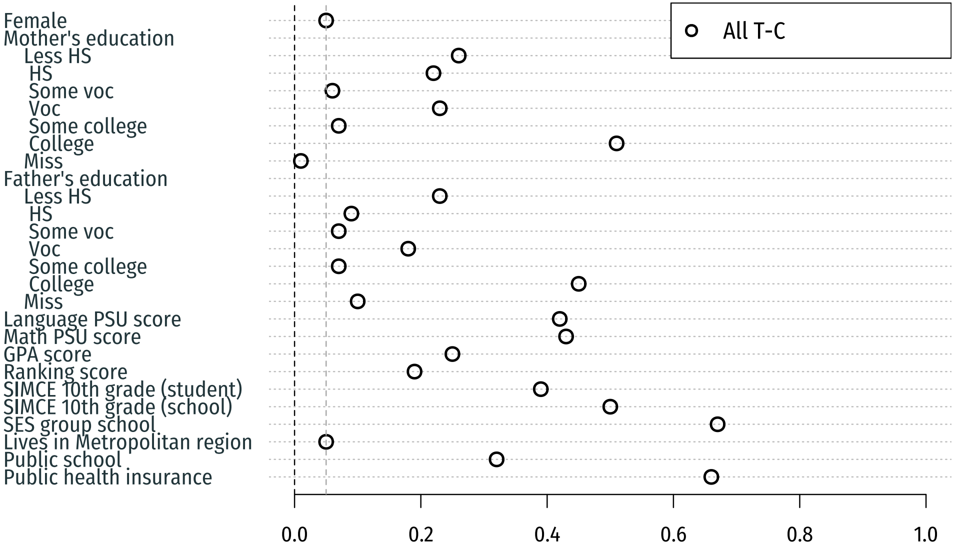
For what population are we generalizing for?



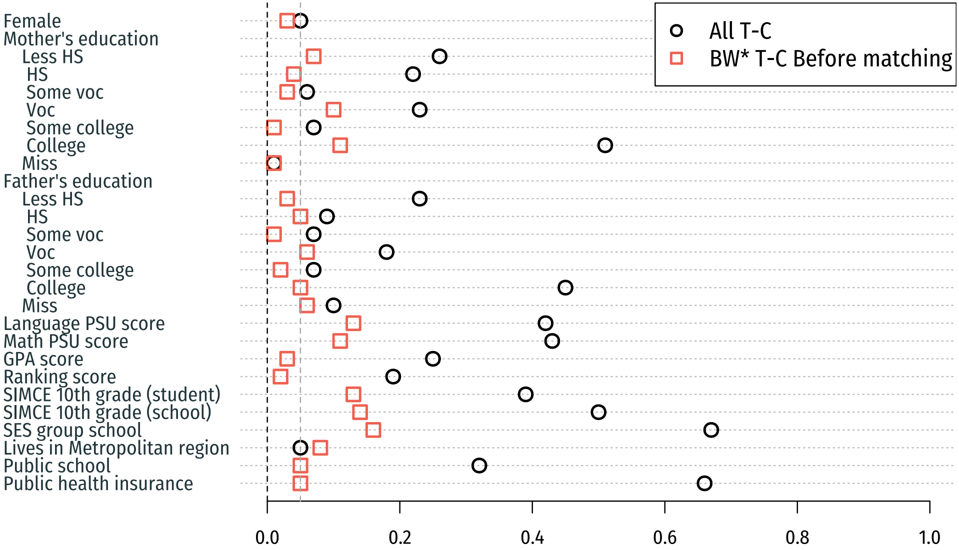
Local Polynomial for Control Outcome in t=1



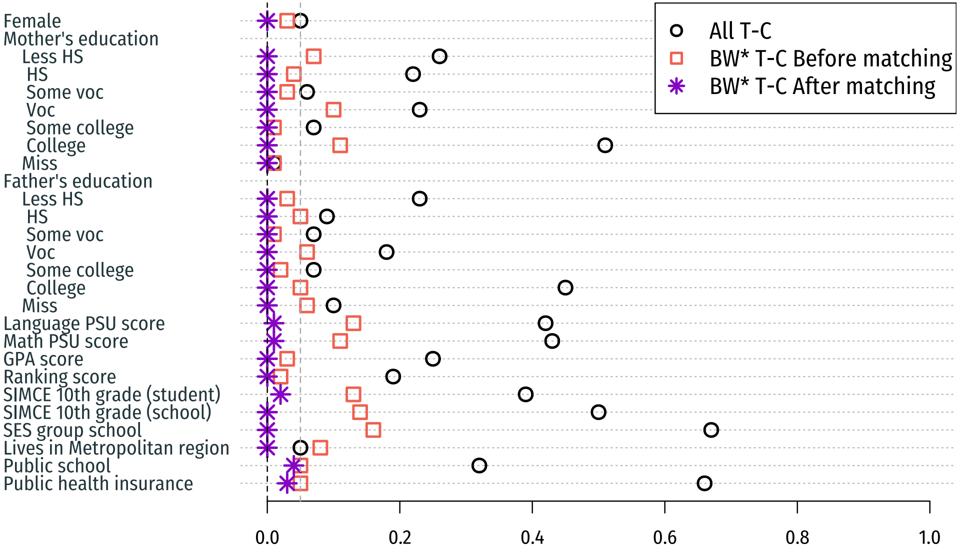
Balance: Entire sample



Balance: Within H^* before matching



Balance: Within H^* after matching



Effects of introduction of FHE: RD and GRD

	Robust RD results		GRD results	
	Application	Enrollment	Application	Enrollment
Effect	0.035	0.069**	0.052**	0.077***
	[-0.007, 0.077]	[0.026, 0.112]	[0.008, 0.096]	[0.029, 0.125]
Effective N Obs	6,588	6,458	2,000	2,000
Control Mean	0.606	0.515	0.568	0.472

Generalization interval [-M$500, M$301]

95% CI in brackets

Effects of introduction of FHE: Application

	Robust RD results		GRD results	
	Application	Enrollment	Application	Enrollment
Effect	0.035	0.069**	0.052**	0.077***
	[-0.007, 0.077]	[0.026, 0.112]	[0.008, 0.096]	[0.029, 0.125]
Effective N Obs	6,588	6,458	2,000	2,000
Control Mean	0.606	0.515	0.568	0.472

Generalization interval [-M$500, M$301]

95% CI in brackets

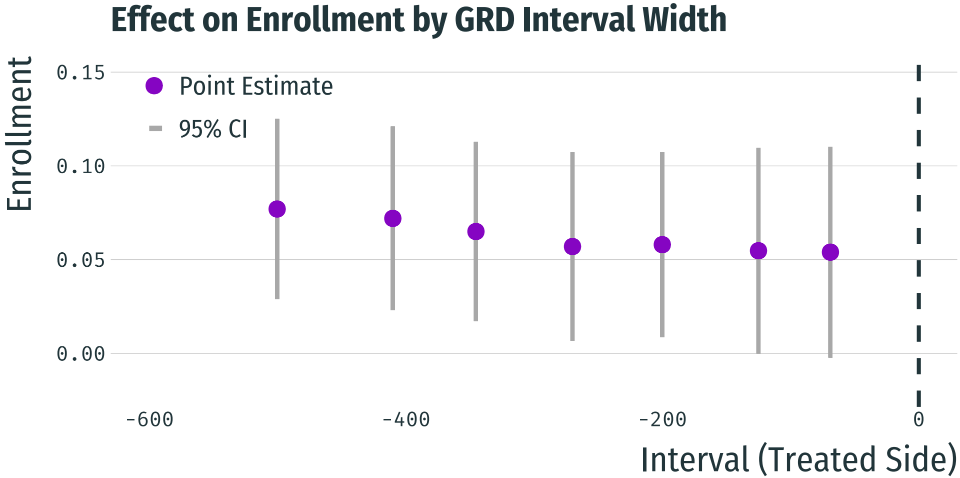
Effects of introduction of FHE: Enrollment

	Robust RD results		GRD results	
	Application	Enrollment	Application	Enrollment
Effect	0.035	0.069**	0.052**	0.077***
	[-0.007, 0.077]	[0.026, 0.112]	[0.008, 0.096]	[0.029, 0.125]
Effective N Obs	6,588	6,458	2,000	2,000
Control Mean	0.606	0.515	0.568	0.472

Generalization interval [-M$500, M$301]

95% CI in brackets

How does the effect change with interval width?



Sensitivity Analysis to Hidden Bias

- Quantify bias of unobserved confounder to change qualitative results of the study
- Adaptation of Keele et al. (2019) sensitivity analysis for Diff-in-Diff.
- Moderately sensitive to hidden bias: $\mathbf{\Gamma{=}1.6}$

$$\rightarrow \Pr(Z_{i1} = 1) = 0.62 \ \wedge \ \Pr(Z_{i1} = 0) = 0.38$$

Outline

Conclusions

- GRD as a gradual approach for generalization (not “all or nothing”)
- Use data to inform interval for generalization
- Use of matching to avoid extrapolation
- Limitations
    - More data: two periods
    - Conditional time invariance assumption for $t = 1$
- Multiple applications for DD-RD: e.g. geographic RDs.

How Far is Too Far?
Generalization of a Regression Discontinuity Design Away from the Cutoff

Magdalena Bennett

January 9, 2020

Teachers College, Columbia University

Data Generating Processes for Simulations

- Observed covariate: $X \sim \mathcal{N}(0, 10)$
- Unobserved confounder: $U \sim \mathcal{N}(0, 10)$
- Running variable for scenario s:

$$r_{it} = \alpha_{s,x} x_{it} + \alpha_{s,u} u_{it} + \varepsilon_{it}$$

- Observed outcome for scenario s:

$$y_{it} = \beta_{s,x} x_{it} + \beta_{s,u} u_{it} + \beta_{s,r} r_{it} + Z_{it} \tau_s + \nu_{it}$$

- True $H = [-200, 200]$

Simulations: Setup for GRD

- Distributional (fine) balance for X deciles
- Template size: 1,000 and 100
- Grid: Equally sized bins (20)
- Significance level for detecting GRD interval: 0.1

Comparison between treatment groups

	Treat group (All)	Treat group within H*
Female	0.55	0.55
Mother's education (years)	11.37	11.57
Father's education (years)	11.52	11.67
Language PSU score	504.08	510.20
Math PSU score	507.69	513.30
GPA score	554.88	558.11
Ranking score	579.84	583.11
SIMCE 10th grade (student)	274.90	276.95
SIMCE 10th grade (school)	266.91	268.52
SES group school	2.68	2.73
Lives in Metropolitan region	0.40	0.42
Public school	0.35	0.34
Public health insurance	0.82	0.79

Different Diff-in-Diff Scenarios