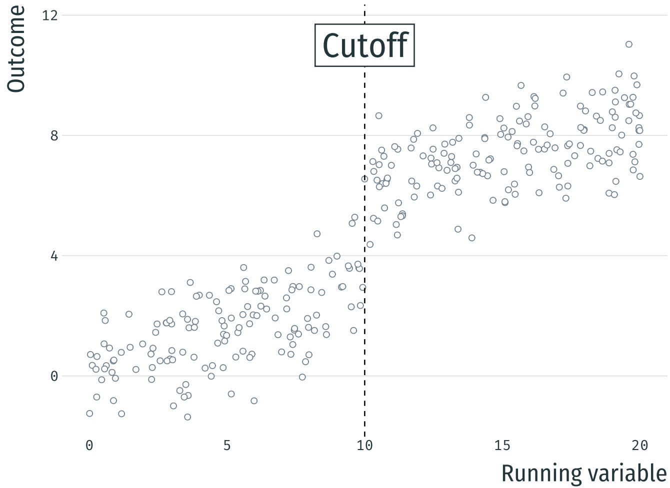
How Far is Too Far? Generalization of a Regression Discontinuity Design Away from the Cutoff

Magdalena Bennett

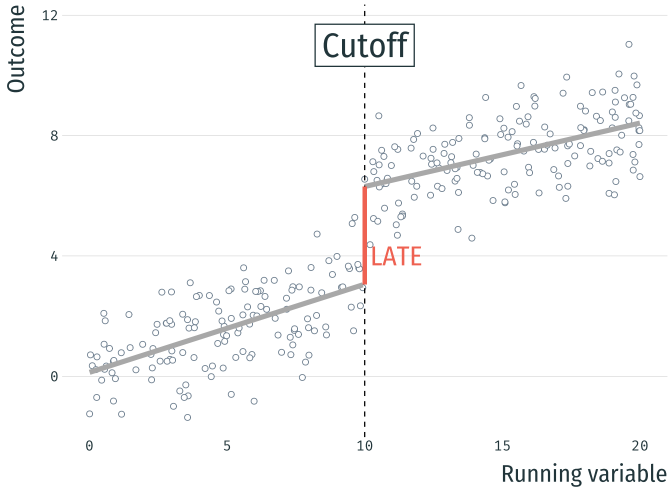
December 5, 2019

Teachers College, Columbia University

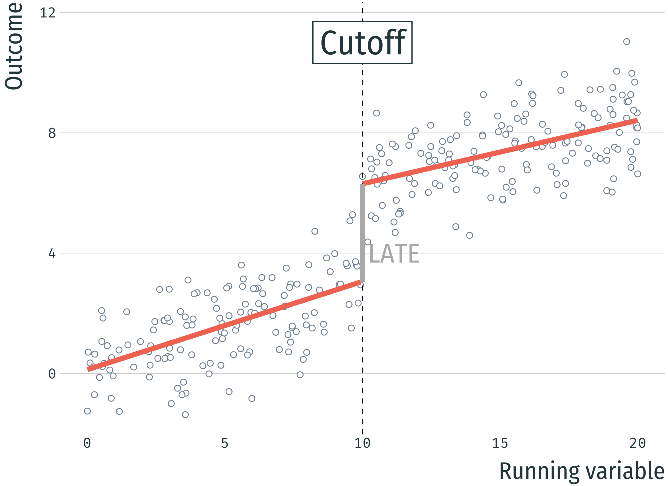
Regression discontinuity design: Strong interval validity



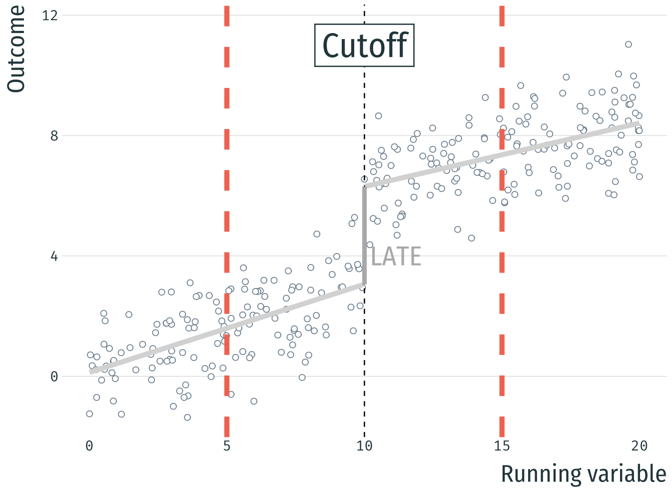
Regression discontinuity design: Strong interval validity



Regression discontinuity design: Limited external validity



Regression discontinuity design: Generalization bandwidth?



This paper

Estimation of TOT for population within a generalization interval:

- Based on idea of local randomization near the cutoff (Lee, 2008; Cattaneo, Frandsen, & Titiunik, 2015)
- Predictive covariates to explain correlation between running variable and outcome (Angrist & Rokkanen, 2015; Rokkanen, 2015; Keele, Titiunik, & Zubizarreta, 2015)
- Use of pre-intervention period to inform generalization interval (Wing & Cook, 2013; Keele, Small, Hsu, & Fogarty, 2019)

This paper

Main advantages:

- Gradual approach
  - No need to “All or Nothing”!
  - Bandwidth informed by the data (Cattaneo et al., 2015)
- No extrapolation
  - 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

- Framework
- GRD in practice
- Simulations

Application: Free Higher Education in Chile

Conclusions

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.:

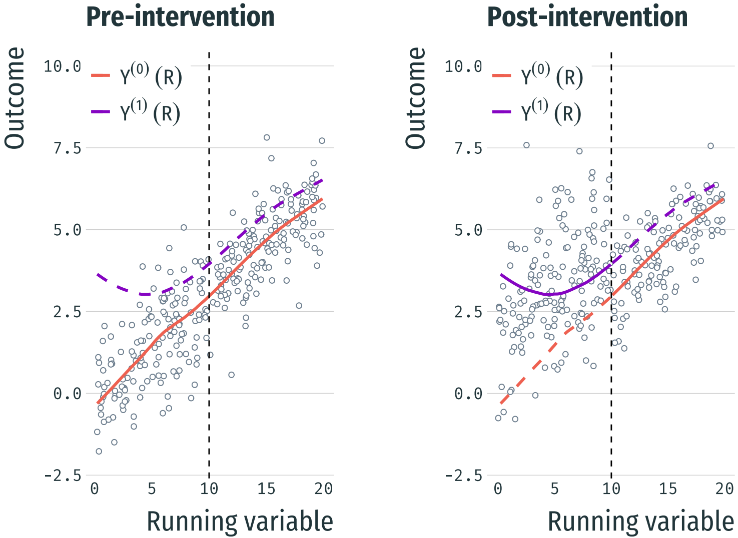
$$Z = \mathbb{I}(R < c)$$

- Predictive covariates $\mathbf{X}$ and unobserved confounders $\mathbf{u}$.
- Potential outcomes under treatment $z = 0, 1$:

$$Y_{it}^{(z)} = g_z(\mathbf{X}_i, \mathbf{u}_i, r_i) + z \cdot \tau_i$$

where τ_i: individual causal effect.

Two periods for GRD



GRD: A gradual approach

- Conditional expectations of potential outcomes:

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

$$Y_0^{(1)}(R) = \mathbb{E}[Y_{i0}^{(1)}|R] = \mu(R) + \tau(R)$$

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

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

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

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

$$Y_0^{(0)}(R_1)|\mathsf{X}_T = Y_0^{(0)}(R_2)|\mathsf{X}_T \quad \text{for any } R_1, R_2 \in H^{*}$$

GRD: Assumptions for generalization to t=1

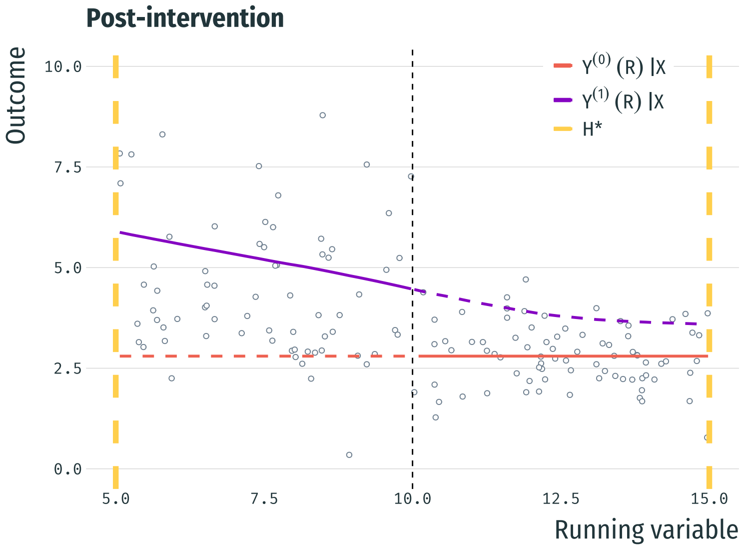
Assumption I: Conditional time-invariance under control

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

Assumption II: Heterogeneity only through τ

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

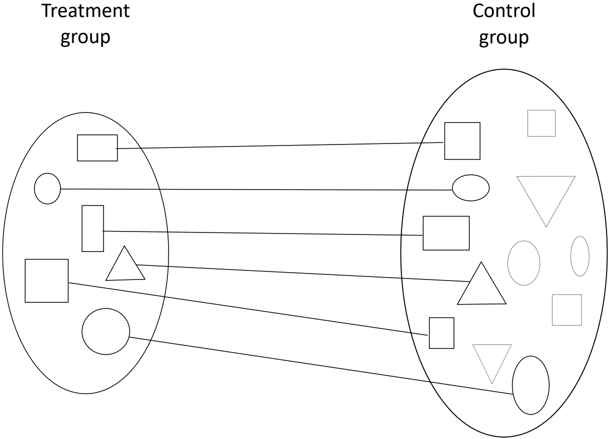
GRD: Estimating effects away from the cutoff



Outline

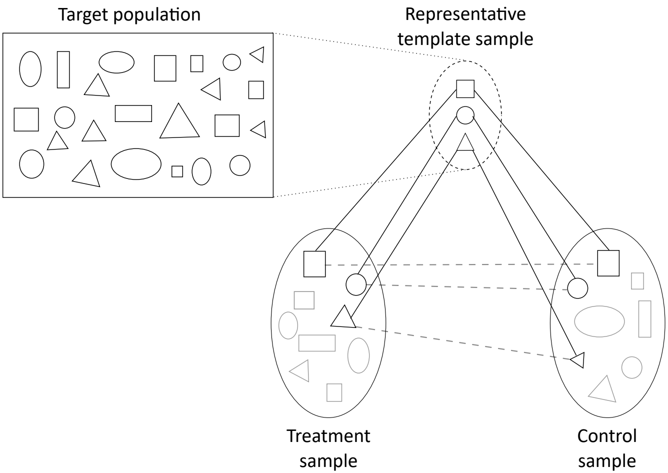
Overview: Representative Template Matching

Traditional Matching



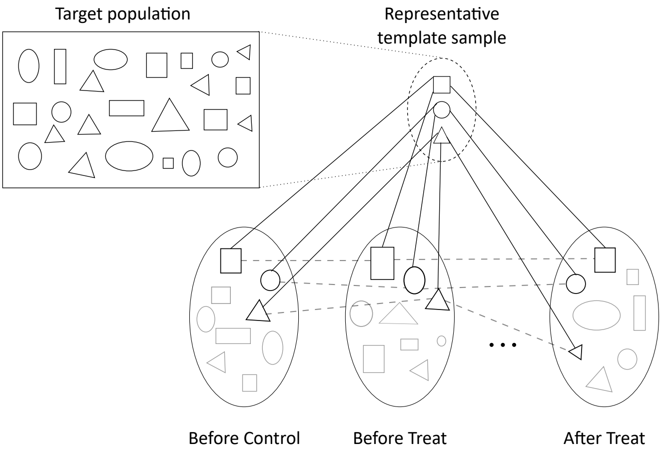
Overview: Representative Template Matching

Representative Template Matching for Two Groups

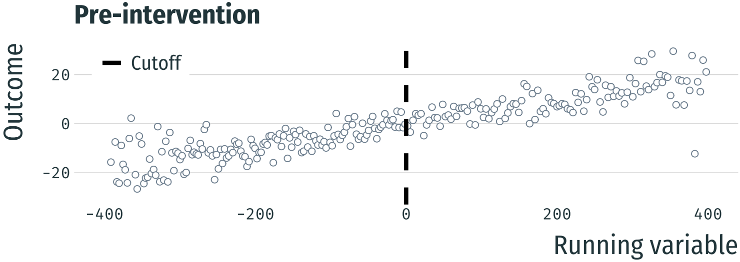


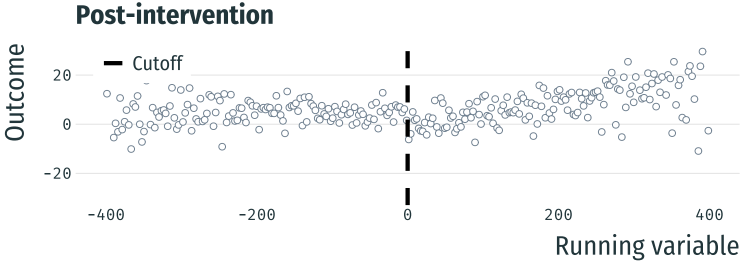
Overview: Representative Template Matching

Representative Template Matching for Diff-in-Diff

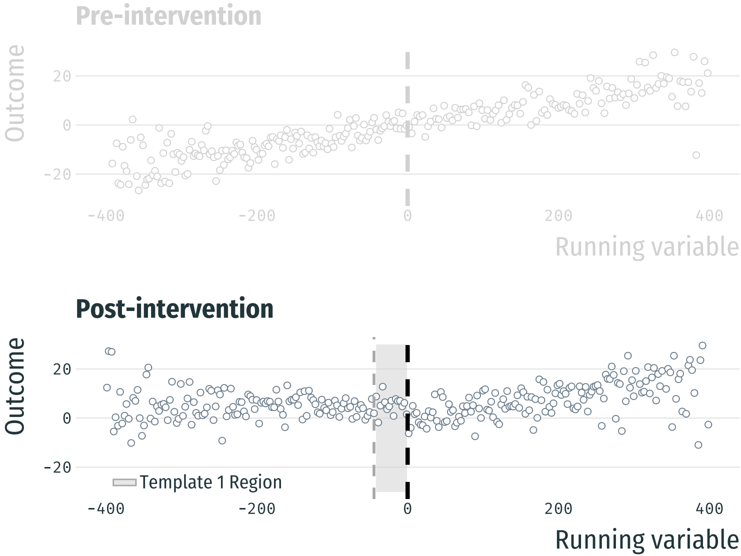


GRD: Start with two periods

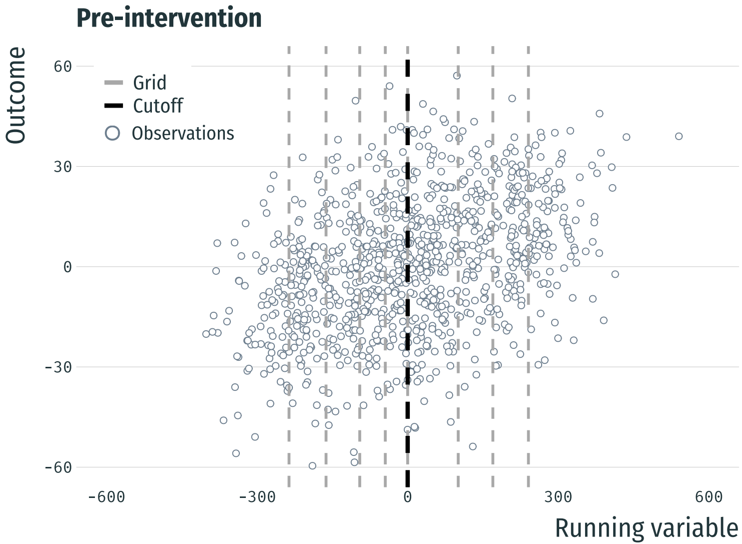




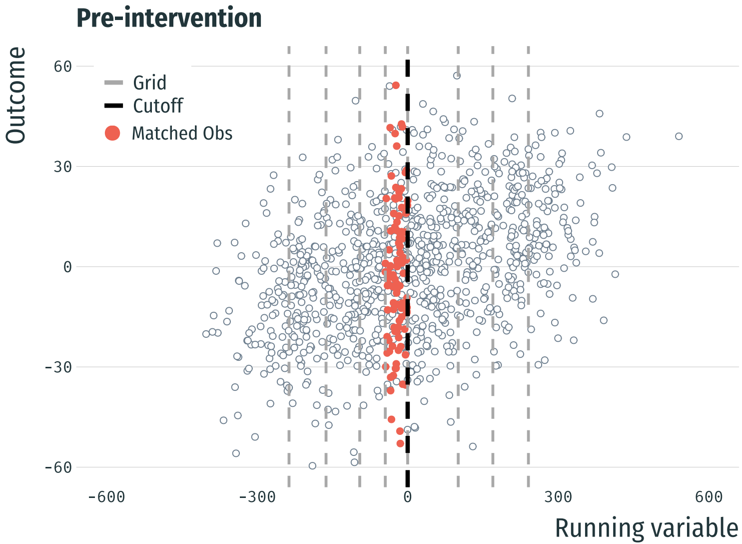
GRD: Select template sample from post-intervention



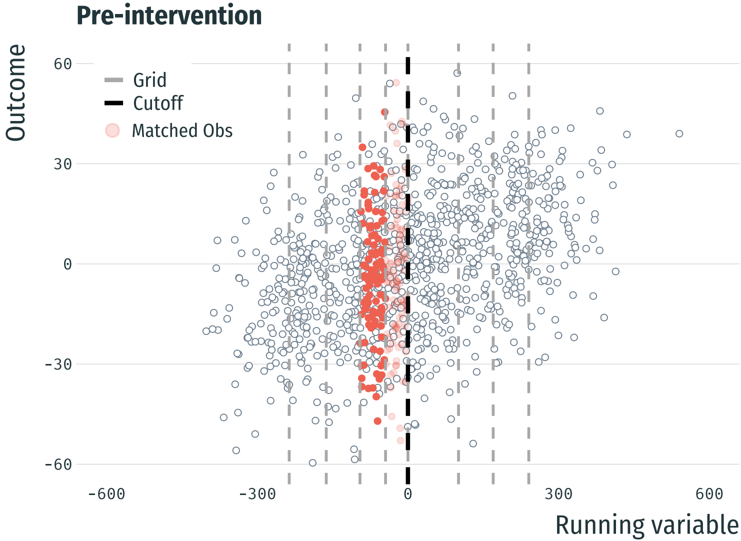
GRD: Divide pre-intervention into grid



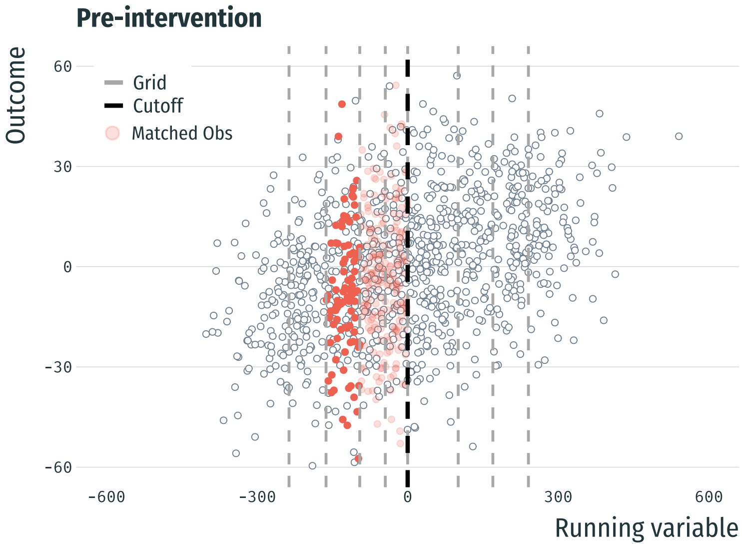
GRD: Match template to grid



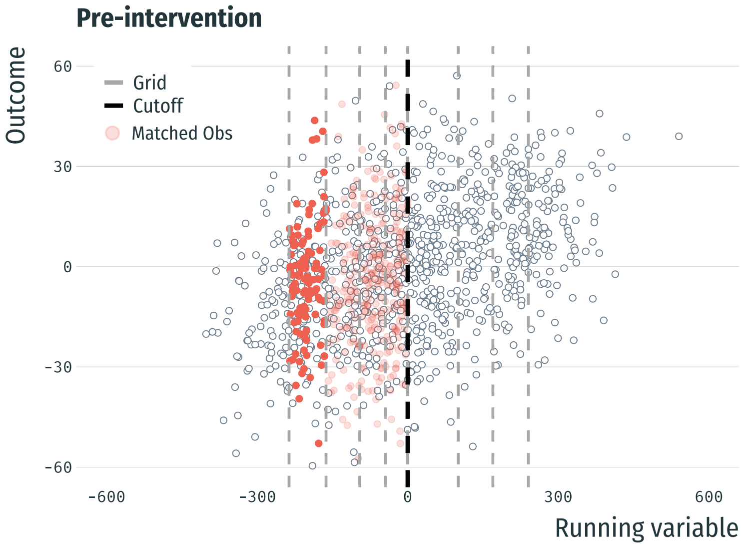
GRD: Match template to grid



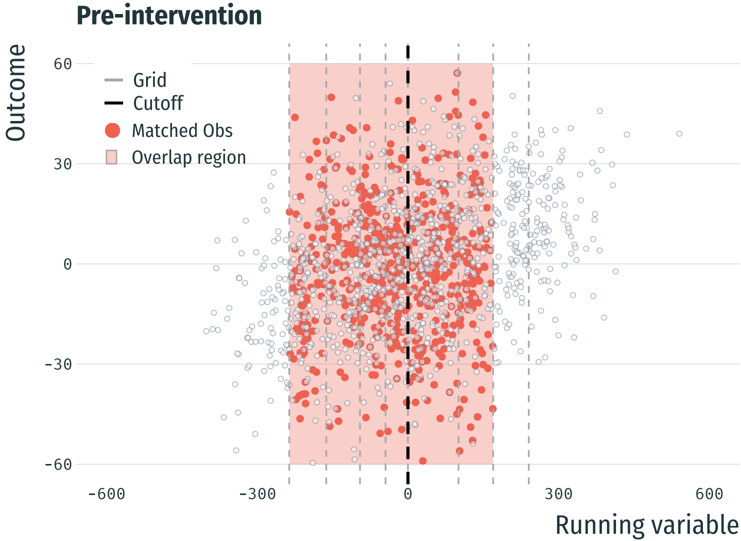
GRD: Match template to grid



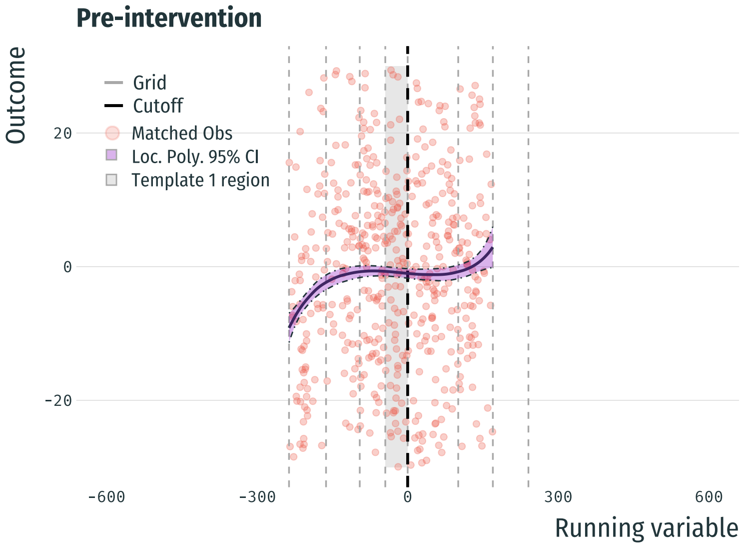
GRD: Match template to grid



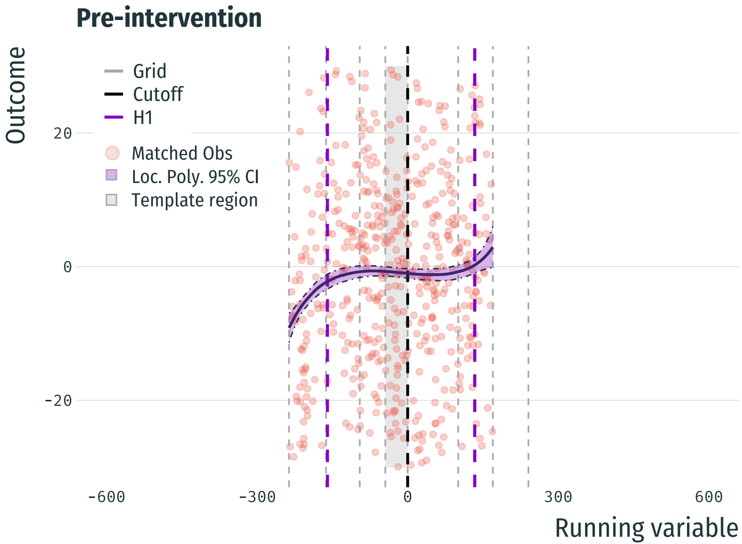
GRD: Explicit overlap region



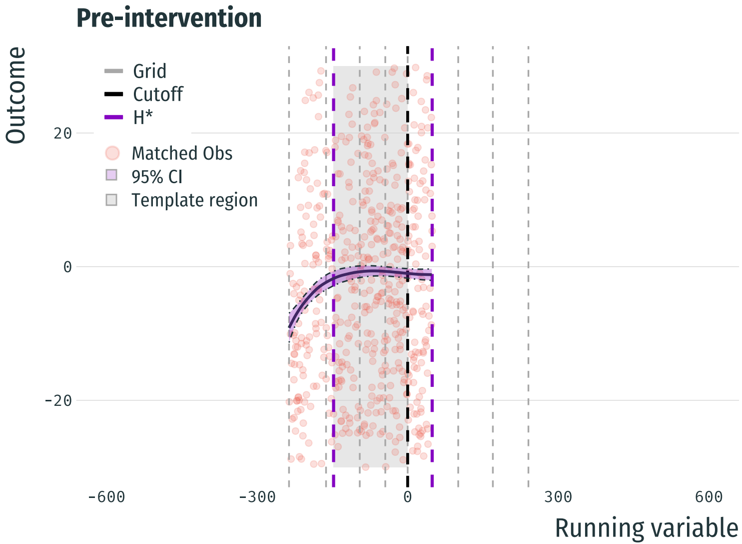
GRD: Estimate local polynomial on matched sample



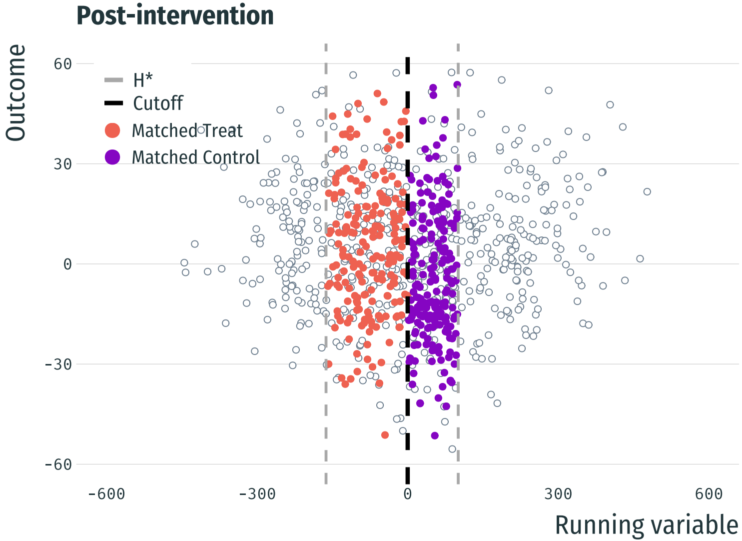
GRD: Identify generalization interval H_1



GRD: Repeat procedure until $H_j \subseteq T$



GRD: Match post-intervention period to the template



GRD: TOT Estimation

Straightforward estimation given matched sample:

- E.g. t-stat paired test:

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

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

$$Var(\hat{\tau}_{TOT}) = \sum_{k=1}^{N} \frac{(d_k - \hat{\tau}_{TOT})^2}{N-1}$$

Outline

Simulations: Assess performance of GRD

- Compare GRD performance to `rdrobust()` (Calonico et al., 2018) $\longrightarrow$ 500 simulations
- Simulations scenarios:
  - Low vs. high correlation:

    $$Corr(R, X) = \{0.33, 0.66\}$$

  - Constant vs. heterogeneous effects:

    $$\tau_{constant} = 0.2\sigma$$
    $$\tau_{heter} = 0.2\sigma + 0.0025\sigma \cdot R$$

  - Small vs. large samples:

    2,000 vs 20,000 obs

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:

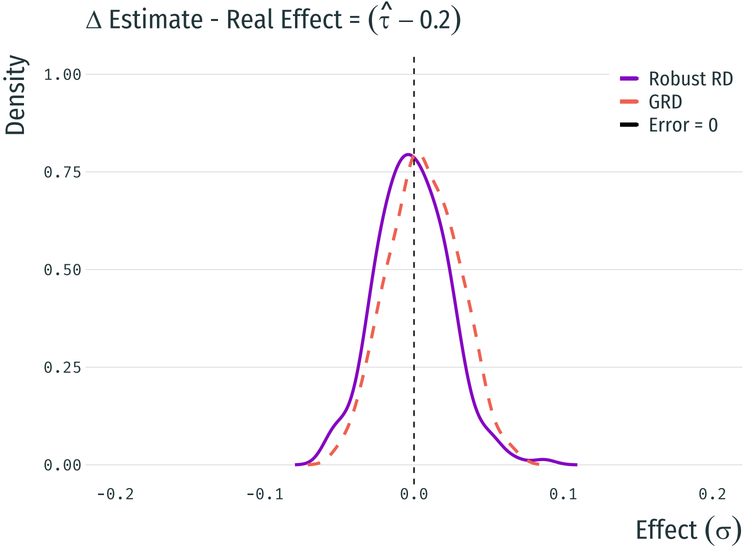
$$y_{it} = \beta_{s,x} x_{it} + \beta_{s,u} u_{it} + \beta_{s,r} r_{it} + Z_{it} \tau_s + \nu_{it}$$

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

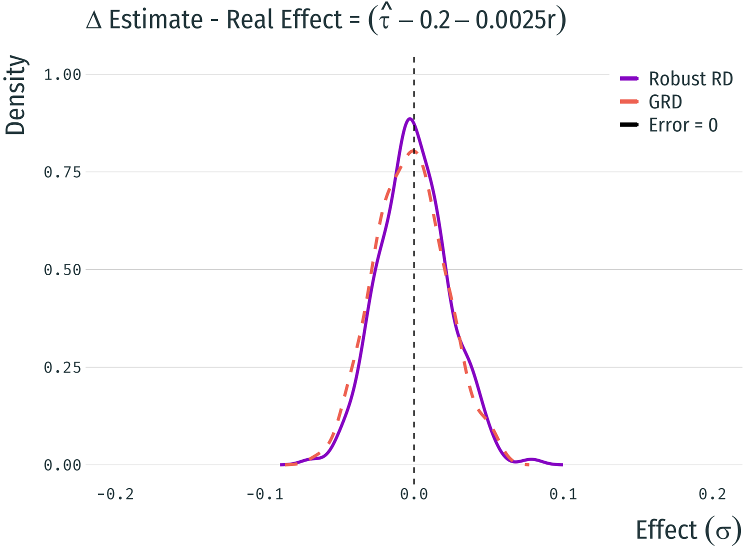
Simulations: Setup for GRD

- 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

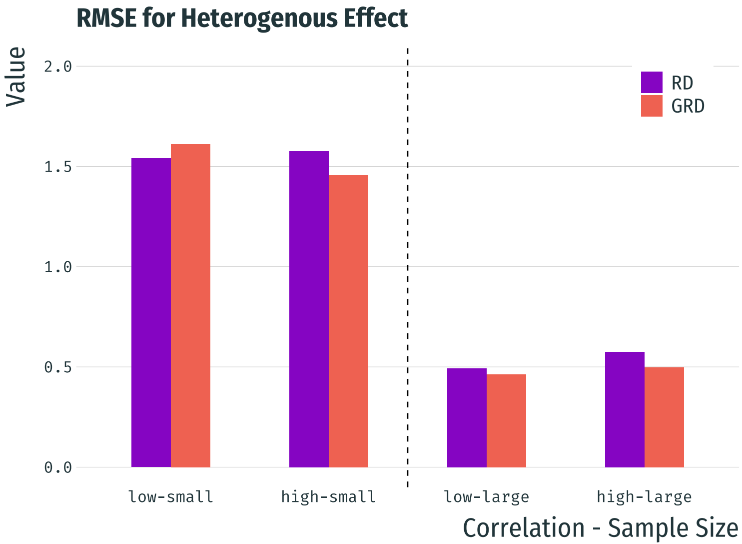
Simulation distribution (s: high corr & large sample)



Simulation distribution (s: high corr & large sample)



Simulation results: Root Mean Square Error



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

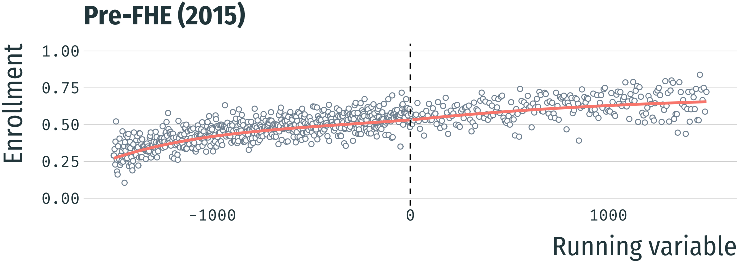
FHE: Research Question

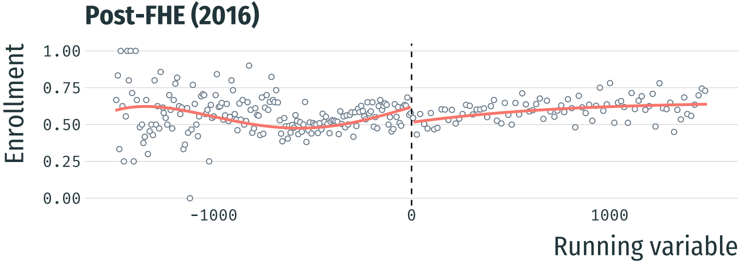
- Two outcomes: Application to university and enrollment
  - Lower-income students $\rightarrow$ financial constraints
  - Salience of policy
- Larger effects for students away from the cutoff?
  - Compare RD and GRD results

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?



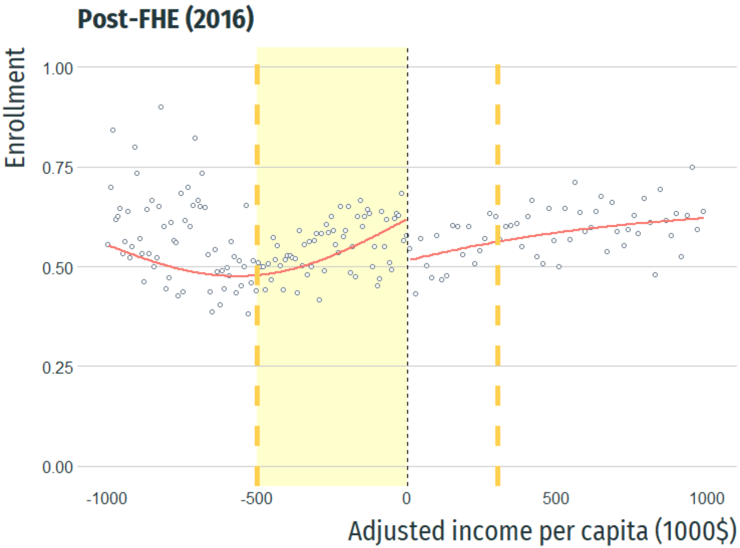


GRD for Free Higher Education

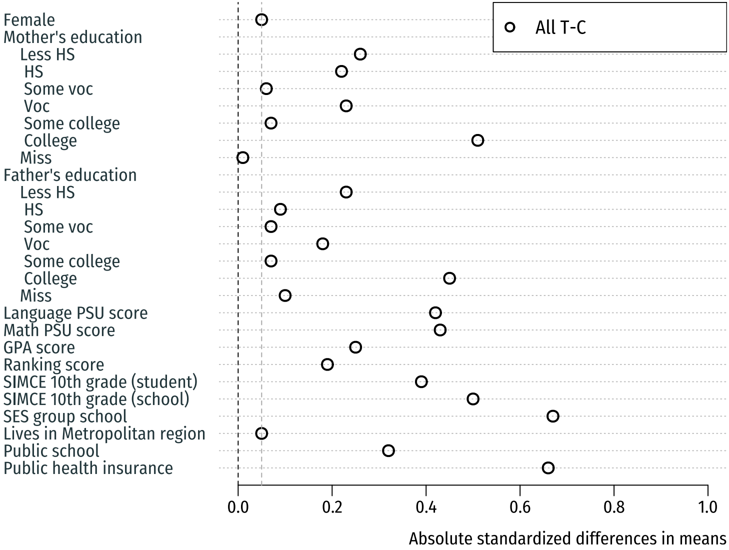
Steps for GRD:

- 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]**

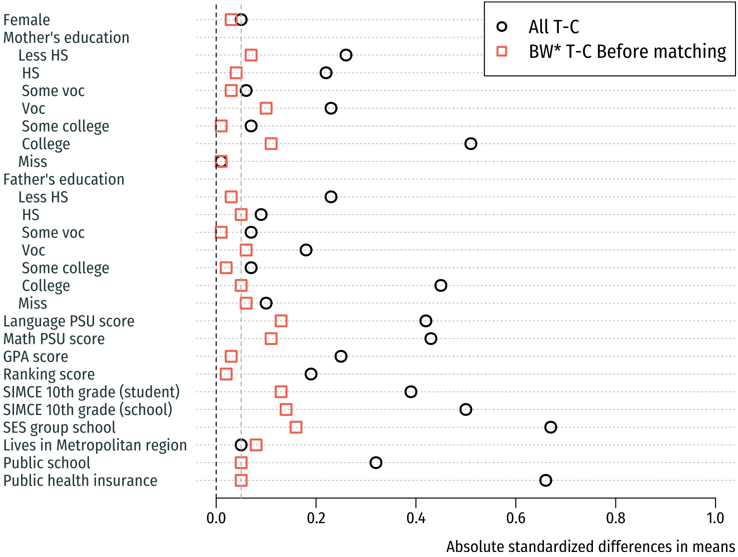
For what population are we generalizing for?



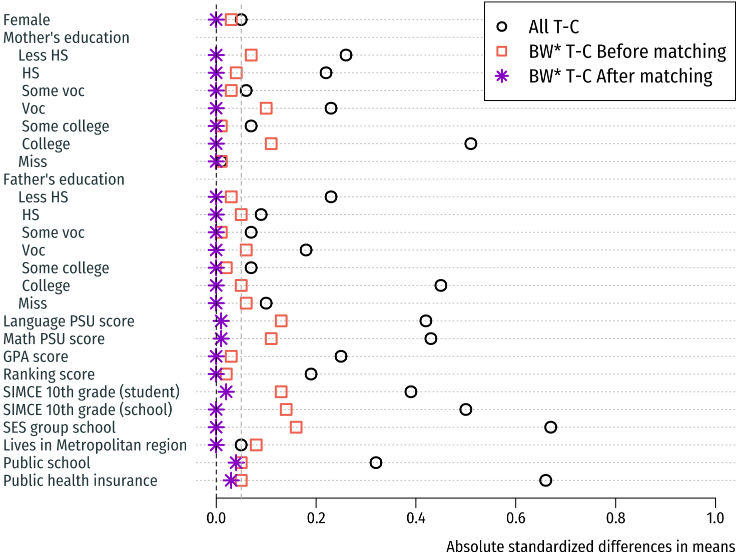
Balance: Entire sample



Balance: Within H^* before matching



Balance: Within H^* after matching



Effects of introduction of FHE: RD and GRD

(a) Robust RD results

	Application	Enrollment
Effect	0.035	0.069**
	[-0.007, 0.077]	[0.026, 0.112]
Effective N Obs	6,588	6,458
Mean control	0.606	0.515

(b) GRD Results

	Application	Enrollment
Effect	0.052**	0.077***
	[0.008, 0.096]	[0.029, 0.125]
N Obs	2,000	2,000
Mean control	0.568	0.472

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

95% CI in squared parenthesis.

Effects of introduction of FHE: Application

(a) Robust RD results

	Application	Enrollment
Effect	0.035	0.069**
	[-0.007, 0.077]	[0.026, 0.112]
Effective N Obs	6,588	6,458
Mean control	0.606	0.515

(b) GRD Results

	Application	Enrollment
Effect	0.052**	0.077***
	[0.008, 0.096]	[0.029, 0.125]
N Obs	2,000	2,000
Mean control	0.568	0.472

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

95% CI in squared parenthesis.

Effects of introduction of FHE: Enrollment

(a) Robust RD results

	Application	Enrollment
Effect	0.035	0.069**
	[-0.007, 0.077]	[0.026, 0.112]
Effective N Obs	6,588	6,458
Mean control	0.606	0.515

(b) GRD Results

	Application	Enrollment
Effect	0.052**	0.077***
	[0.008, 0.096]	[0.029, 0.125]
N Obs	2,000	2,000
Mean control	0.568	0.472

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

95% CI in squared parenthesis.

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: $\boldsymbol{\Gamma}$=**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-GRD: e.g. geographic RDs.

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

Magdalena Bennett

December 5, 2019

Teachers College, Columbia University