Empirical Methods for Policy Evaluation

Second Part

Matteo Bobba

Toulouse School of Economics (TSE)

TSE PhD Program (MRes)
Fall 2024

Outline and Readings for this Section (3 Classes)

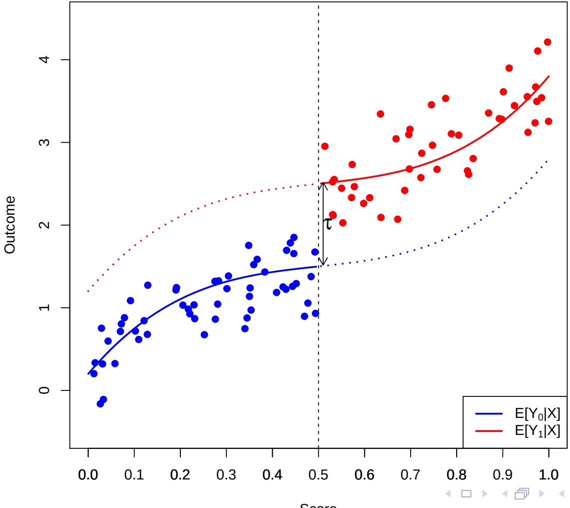
- Regression Discontinuity Designs
  - Local randomization approach (Cattaneo-Idrobo-Titiunik: Book 2, Section 2)
  - Continuity-based approach (CIT: Book 1)
  - RD extensions (CIT: Book 2, Sections 3, 4 and 5)
- RD and Empirical Matching Models
  - **Bobba, Ederer, Leon-Ciliotta, Neilson, and Nieddu (2024)**

The Local Randomization Approach

Setup and Notation

- Potential outcomes: $Y_i(1), Y_i(0)$, with $\tau_i = Y_i(1) - Y_i(0)$
- Continuous running variable (score): X_i
- Treatment indicator: $D_i = D_i(X_i) = 1$ if treated, 0 otherwise
- Observed outcome: $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$
- RD exploits a discontinuity in $P[D_i = 1|X_i]$ at some cutoff c
- Sharp design (will extend this later): $P[D_i = 1|X_i] = \mathbb{I}(X_i \geq c)$

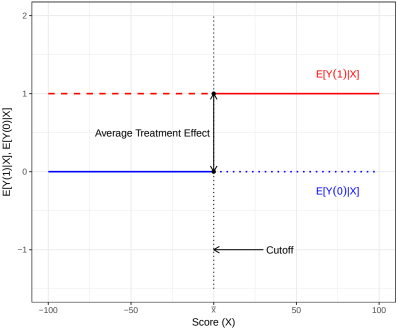
Graphical Intuition



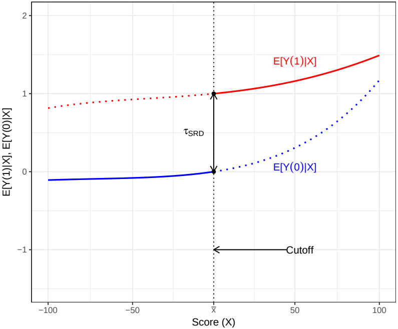
RD as a Randomized Experiment

- Idea: close enough to the cutoff, some units were "lucky"
- Treatment as if randomly assigned in a window around c if:
  - Units do not have **exact** control of their score
  - There is a random chance element in score each unit receives
  - Probability of this random "error" does not jump at the cutoff
- Example: each unit assigned a score $X_i \sim U[0,1]$, $D_i = \mathbb{I}(X_i \geq c)$, then $P[D_i = 1] = 1 - c$

RD as a Randomized Experiment



(a) Randomized Experiment



(b) RD Design

Local Randomization Approach to RD

- There is a window $W_0 = [c - w, c + w]$ in which:
  1. Probability distribution of X_i is unrelated to individual characteristics

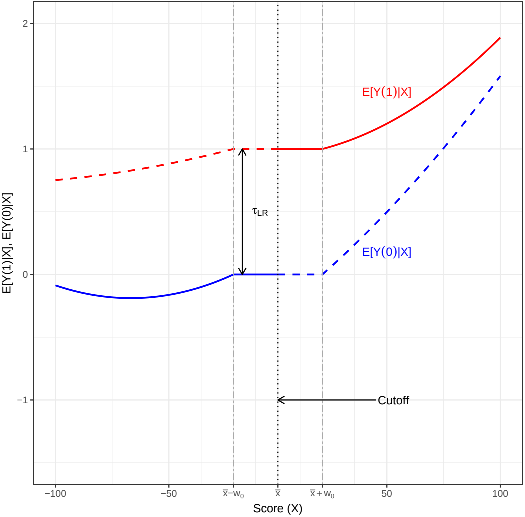
     $$P[X_i \leq x | X_i \in W_0] = F_0(x), \quad \forall\, i$$

  2. Potential outcomes not affected by value of the score:

     $$Y_i(d, x) = Y_i(d)$$

- Potential outcomes are a constant function of the score

Local Randomization RD: Intuition



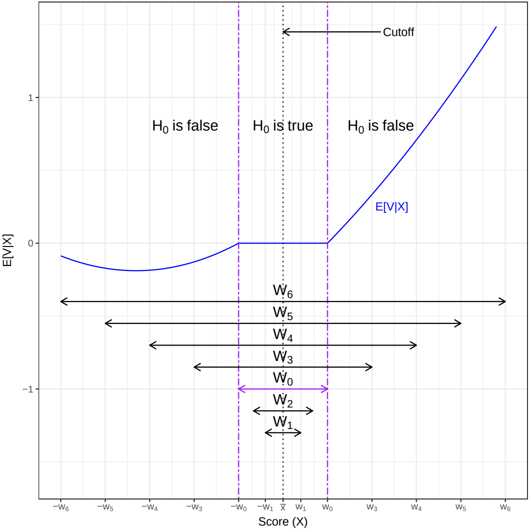
Window Selection: A Data Driven Method

- Under random assignment, covariates should be balanced:

$$P[V_i \leq v | D_i = 1] = P[V_i \leq v | D_i = 0]$$

- Can use this idea as a windows selection criterion:
  - Find window in which all covariates are balanced
- Iterative procedure:
  1. Choose a test statistic (diff. means, Kolmogorov-Smirnov,...)
  2. Choose an initial "small" window $W_0^{(1)} = [c - w_{(1)}, c + w_{(1)}]$
  3. Test null that covariates are balanced above and below c
  4. Enlarge slightly the window and repeat until null hypothesis is rejected

Window Selection Procedure



Estimation and inference

- Once W_0 is found, proceed as in a randomized experiment

$$\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$$

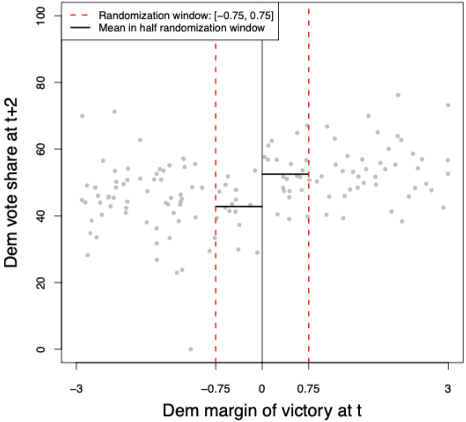
- Covariate-balance criterion may yield windows with few obs
- Inference based on large-sample approximations may not be reliable
- Alternative approach: randomization inference

Software Implementations

- Cattaneo, Titiunik and Vazquez-Bare (Stata Journal, 2016)
- `rdlocrand` package: statistical inference and graphical procedures for RDD employing local randomization methods
  - `rdwinselect`: window selection
  - `randinf`: randomization inference
  - `rdsensitivity`: sensitivity analysis
  - `rdrbounds`: Rosenbaum bounds

Example: Incumbency Advantage in U.S. Senate

- Y_i = election outcome at $t+1$ ($=1$ if party wins)
- D_i = election outcome at t ($=1$ if party wins)
- X_i = margin of victory at t ($c=0$)



The Continuity-based Approach

Identification (Hahn, Todd, and Van der Klauw, 2001)

1. (sharp design): $D_i = \mathbb{I}(X_i \geq c)$
2. (smoothness): $\mathbb{E}[Y_i(0)|X_i = x]$, $\mathbb{E}[Y_i(1)|X_i = x]$ continuous at $x = c$

Then,

$$\mathbb{E}[\tau_i|X_i = c] = \lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]$$

- Treatment effect only (nonparametrically) identified at the cutoff
  - Only point of overlap (in the limit)
  - We actually have zero observations at $X_i = c$

Identification

- Naive difference in means:

$$\begin{aligned}\Delta(h) &= \mathbb{E}\{Y_i \mid X_i \in [c, c+h]\} - \mathbb{E}\{Y_i \mid X_i \in [c-h, c)\} \\ &= \mathbb{E}\{Y_i(1) \mid X_i \in [c, c+h]\} - \mathbb{E}\{Y_i(0) \mid X_i \in [c-h, c)\} \\ &= \mathbb{E}\{\tau_i \mid X_i \in [c, c+h]\} + \text{Bias}(h)\end{aligned}$$

  where $\text{Bias}(h) = E\{Y_i(0) \mid X_i \in [c, c+h]\} - E\{Y_i(0) \mid X_i \in [c-h, c)\}$

- If $\mathbb{E}[Y_i(d)|X_i = x]$ is continuos at $x = c$ for $d = 0, 1$, then:

$$\lim_{h \downarrow 0} \Delta(h) = \mathbb{E}[\tau_i | X_i = c]$$

Estimation: Overview

1. Global:
    - Estimate a p-th-order polynomial on full sample
    - Sensitive to misspecification
    - Erratic behavior at boundary points
2. "Flexible parametric":
    - Estimate a polynomial within an ad-hoc bandwidth
    - Sensitive to misspecification and bandwidth choice
3. Nonparametric local polynomial:
    - Data-driven bandwidth selection
    - Accounts for misspecification when performing inference

Global Parametric Approach

- Parametric assumption on conditional expectations, e.g.

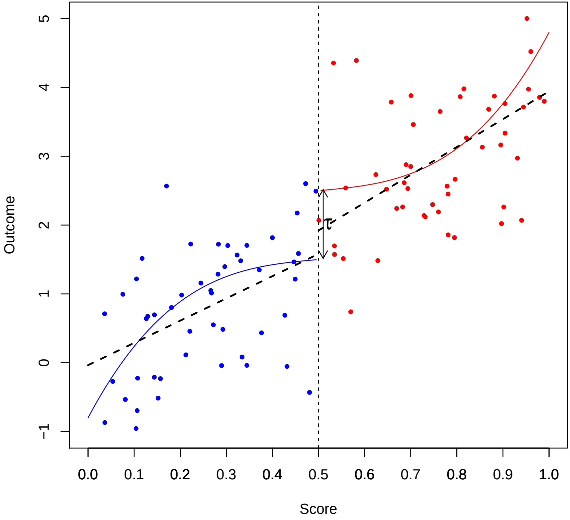
$$\mathbb{E}[Y_i(d)|X_i] = \alpha_d + \beta_d(X_i - c)$$

- This implies

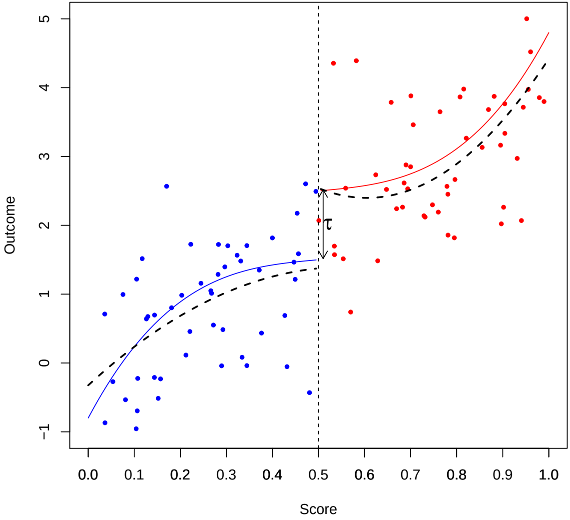
$$\mathbb{E}[Y_i|X_i] = \alpha_0 + (\alpha_1 - \alpha_0)D_i + \beta_0(X_i - c) + (\beta_1 - \beta_0)(X_i - c)D_i + u_i$$

- Easily estimated by OLS on full sample
- Coefficient $\alpha_1 - \alpha_0$ recovers the treatment effect at the cutoff

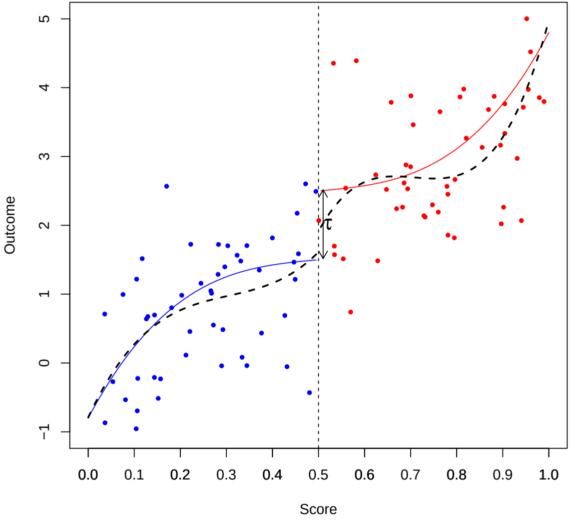
Global Parametric Approach: $p = 1$



Global Parametric Approach: $p = 2$



Global Parametric Approach: $p = 3$



Local Linear Regression

- Suppose $c = 0$ (otherwise, use $X_i - c$)
- Choose some bandwidth $h > 0$ and estimate by OLS:

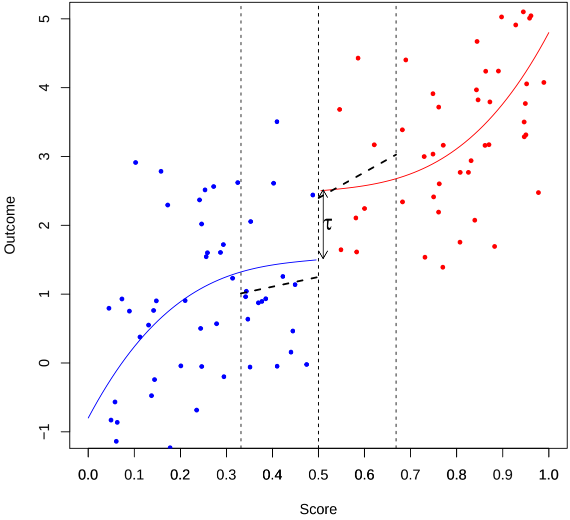
$$(\hat{\alpha}^{+}, \hat{\beta}^{+}) = \underset{(\alpha,\beta)}{\text{argmin}} \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2 \mathbb{I}(0 \leq X_i \leq h)$$

$$(\hat{\alpha}^{-}, \hat{\beta}^{-}) = \underset{(\alpha,\beta)}{\text{argmin}} \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2 \mathbb{I}(-h \leq X_i < 0)$$

- Estimated treatment effect at the cutoff:

$$\hat{\tau} = \hat{\alpha}^{+} - \hat{\alpha}^{-}$$

Local Linear Regression: Graphical Intuition



Local Linear Regression: OLS Estimands

- By standard OLS algebra:

$$\hat{\beta}^{+} = \frac{\sum_{i=1}^{n} Y_i (X_i - \bar{X}_h)\mathbb{I}(0 \leq X_i \leq h)}{\sum_{i=1}^{n} X_i (X_i - \bar{X}_h)\mathbb{I}(0 \leq X_i \leq h)}$$

$$\hat{\alpha}^{+} = \bar{Y}_h - \bar{X}_h \hat{\beta}^{+}$$

where

$$\bar{X}_h = \frac{\sum_{i=1}^{n} X_i \mathbb{I}(0 \leq X_i \leq h)}{\sum_{i=1}^{n} \mathbb{I}(0 \leq X_i \leq h)}$$

$$\bar{Y}_h = \frac{\sum_{i=1}^{n} Y_i \mathbb{I}(0 \leq X_i \leq h)}{\sum_{i=1}^{n} \mathbb{I}(0 \leq X_i \leq h)}$$

Local Linear Regression: Bias

- It can be shown that (analogous result for $E[\hat{\alpha}^-|\mathbf{X}]$):

$$E[\hat{\alpha}^+|\mathbf{X}] = \mu_1(0) + h^2\mathcal{B}_+ + o_p(h^2)$$

- $\mu_1(x) = E[Y_i(1)|X_i = x]$
- $\mathcal{B}_+$ is a constant that depends on:
  1. The curvature of $\mu_1(x)$
  2. The kernel function
  3. The order of polynomial, p
- If $h = 0$ the estimator would be unbiased
- But with $h = 0$ there are no observations within the bandwidth!

Variance

- Similarly, it can be shown that (analogous result for $V[\hat{\alpha}^{-}|\mathbf{X}]$):

$$V[\hat{\alpha}^{+}|\mathbf{X}] = \frac{\mathcal{V}_{+}}{nh} + o_p(h)$$

- $\mathcal{V}_{+}$ is a constant that depends on:
  1. $V[Y_i(1)|X_i = 0]$
  2. The density of the score variable at the cutoff
  3. The kernel function
  4. The order of polynomial, p

- Decreasing the variance requires $nh \to \infty$

MSE

- Therefore:

$$E[\hat{\tau}|\mathbf{X}] - \tau = h^2\mathcal{B} + o_p(h^2)$$
$$V[\hat{\tau}|\mathbf{X}] = \frac{\mathcal{V}}{nh} + o_p(h)$$

- Mean-squared error (MSE):

$$\begin{aligned}\text{MSE}(\hat{\tau}) &= Bias(\hat{\tau})^2 + V[\hat{\tau}] \\ &= h^4\mathcal{B}^2 + \frac{\mathcal{V}}{nh}\end{aligned}$$

Bandwidth Selection

- Trade-off in bandwidth selection: bias vs variance
- MSE-optimal bandwidth:

$$\begin{aligned} h^*_{\mathsf{MSE}} &= \underset{h}{\operatorname{argmin}}\, \mathsf{MSE}(\hat{\tau}) \\ &= \left(\frac{\mathcal{V}}{4\mathcal{B}^2}\right)^{1/5} n^{-1/5} \end{aligned}$$

- MSE-optimal bandwidth is proportional to $n^{-1/5}$

Inference

- In this case $V[\hat{\tau}|\mathbf{X}] = O_p(n^{-1}h^{-1})$ so:

$$\sqrt{nh}(\hat{\tau} - \tau) \to_{\mathcal{D}} \mathcal{N}(0, \Omega)$$

- But recall that $h^*_{\text{MSE}} \propto n^{-1/5}$ so the Normal approximation will have a bias
- Two alternatives:
  - Undersmoothing: use a "smaller" bandwidth
  - Bias correction

Bias Correction (Calonico, Cattaneo and Titiunik, ECMA 2014)

- CCT propose the following de-biasing approach:

$$\sqrt{nh}(\hat{\tau} - \tau) = \sqrt{nh}(\hat{\tau} - \mathbb{E}[\hat{\tau}|\mathbf{X}]) + \sqrt{nh}B_n$$

- Bias correction:

$$\sqrt{nh}(\hat{\tau} - \tau - B_n) = \sqrt{nh}(\hat{\tau} - \mathbb{E}[\hat{\tau}|\mathbf{X}]) \to_{\mathcal{D}} \mathcal{N}(0, \Omega)$$

- But the bias is unknown, so we need to estimate it:

$$\sqrt{nh}(\hat{\tau} - \tau - \hat{B}_n) \to_{\mathcal{D}} \mathcal{N}(0, \Omega + \Sigma)$$

where Σ accounts for the estimation of the bias

Assessing the Validity of the Continuity-based Approach

- Density discontinuity tests
- Continuity away from the cutoff
- Testing for discontinuities in covariates / placebo outcomes

Density discontinuity tests

- RDDs can be invalid if individuals manipulate X_i
- Manipulation can imply sorting on one side of the cutoff
- Test whether the density of X_i is continuous around c
- McCrary (2008), Cattaneo, Jansson and Ma (2018)

Continuity away from the cutoff

- Identification relies on continuity of $\mathbb{E}[Y_i(d)|X_i]$
- Can estimate $\mathbb{E}[Y_i(0)|X_i]$ for controls, $\mathbb{E}[Y_i(1)|X_i]$ for treated
- Check continuity away from the cutoff (graphically or formally)

Continuity in covariates / placebo outcomes

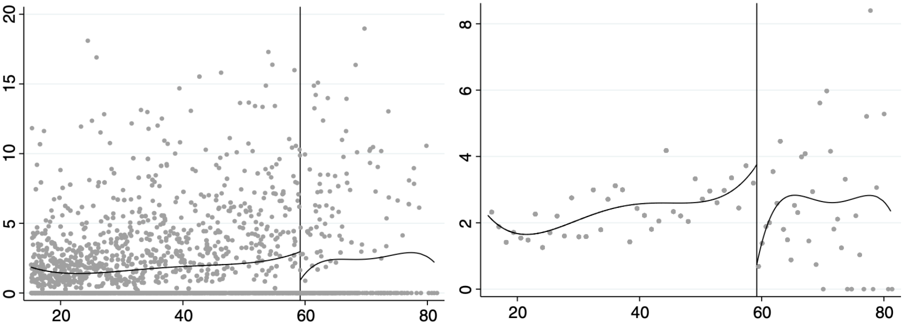
- Some variables should reveal no treatment effect:
  - Outcomes not targeted by treatment (placebo outcomes)
  - Exogenous or predetermined covariates
- Estimate an RD effect on these variables
- Finding a non-zero effect suggests an invalid RDD:
  - Existence of other (unobserved) treatments at the cutoff
  - Selection

Software Implementations

- Calonico, Cattaneo, Farrell, and Titiunik (Stata Journal, 2017)
- `rdrobust` package: estimation, inference and graphical analysis
  - `rdbwselect`: bandwidth selection procedures for local polynomial RD
  - `rdplot`: data-driven regression discontinuity plots
  - `rddensity`: manipulation testing
  - `rdpower`: power and sample size calculations for RD designs

Example: Impact of Head Start on Child Mortality

- Federal program that provides health and social services for children aged 5-9
  - HS assistance for 300 counties based on poverty index ($X_i \geq 59.19$)
  - Y_i = county-level mortality rates per 100,000



(a) Scatter Plot, Raw Data, $N^- = 2,455$, $N^+ = 290$

(b) RD Plot, ES, and MV, $J_- = 37$, $J_+ = 38$

Example: Impact of Head Start on Child Mortality

- Running variable manipulation falsification approach
  - Non-parametric test for continuity of the PDF of X_i near the cutoff

	Density tests				
	h_-	h_+	N_W^-	N_W^+	p-value
Method					
Unrestricted, 2-h	10.151	9.213	351	221	0.788
Unrestricted, 1-h	9.213	9.213	316	221	0.607
Restricted (1-h)	13.544	13.544	482	255	0.655

Notes: (i) Cutoff is $\bar{r} = 59.1984$ and $W = [\bar{r} - h,\ \bar{r} + h]$ denotes the symmetric window around the cutoff used for each choice of bandwidth; (ii) Density test p-values are computed using Gaussian distributional approximation to bias-corrected local-linear polynomial estimator with triangular kernel and robust standard errors; (iii) column "Method" reports unrestricted inference with two distinct estimated bandwidths ("U, 2-h"), unrestricted inference with one common estimated bandwidth ("U, 1-h"), and restricted inference with one common estimated bandwidth ("R, 1-h"). See Cattaneo, Jansson, and Ma (2016a, 2016b) for methodological and implementation details.

Example: Impact of Head Start on Child Mortality

- Ludwig and Miller (QJE, 2007): flexible parametric RD
  - $\hat{\tau}_{\{p=4,\text{full sample}\}} = -3.065$, p-value $= 0.005$
  - $\hat{\tau}_{\{p=1,h=18\}} = -1.198$, p-value $= 0.071$
  - $\hat{\tau}_{\{p=1,h=9\}} = -1.895$, p-value $= 0.055$
- Cattaneo, et al (JPAM, 2017): robust bias-corrected non-parametric RD
  - $\hat{\tau}_{\{p=0,\hat{h}_{MSE}=3.24\}} = -2.114$, robust p-value $= 0.037$
  - $\hat{\tau}_{\{p=0,h=9\}} = -1.059$, robust p-value $= 0.048$
  - $\hat{\tau}_{\{p=1,\hat{h}_{MSE}=6.81\}} = -2.409$, robust p-value $= 0.042$

RD Extensions

Fuzzy RD

- Sharp RD: score perfectly determines treatment status
  - All units scoring above the cutoff receive the treatment
  - $D_i = \mathbb{I}(X_i \geq c)$
  - Probability of treatment jumps from 0 to 1 at c
- Fuzzy RD: imperfect compliance
  - Some units below c may be treated or vice versa
  - Jump in probability at c may be < 1 (but > 0)

Intention-to-treat (ITT) parameter

- ITT: effect of being assigned to treatment
- Sharp RD design on the treatment assignment variable

$$\tau_{\text{ITT}} = \lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x]$$

- Under some continuity assumptions,

$$\tau_{\text{ITT}} = \mathbb{E}[(\underbrace{Y_i(1) - Y_i(0)}_{\tau_i})(\underbrace{D_{1i} - D_{0i}}_{\substack{= \ \ 1 \text{ for compliers} \\ = -1 \text{ for defiers} \\ = \ \ 0 \text{ for always/never takers}}})|X_i = c]$$

The Monotonicity Assumption

- We will rule out the presence of defiers:

$$P[\text{defier}|X_i = c] = 0$$

- This assumption is called *monotonicity*, since it implies that:

$$D_{1i} \geq D_{0i}, \quad \forall\, i$$

- Intuition: $X_i \geq c$ does not decrease the probability of treatment

Intention-to-treat (ITT) Parameter

- $D_{1i} - D_{0i} = 1$ for compliers, 0 for always-takers and never-takers
- Then

$$\tau_{\text{ITT}} = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|X_i = c, D_{1i} > D_{0i}]}_{\text{ATE on compliers: LATE}} \times \underbrace{P[D_{1i} > D_{0i}|X_i = c]}_{\text{prop of compliers}}$$

- ITT can be ≈ 0 even if LATE is large
- But still a policy relevant parameter:
  - Effect of offering the treatment

First Stage

- First stage: effect of treatment assignment on treatment status:

$$\tau_{\textsf{FS}} = \lim_{x \downarrow c} \mathbb{E}[D_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[D_i | X_i = x]$$

- Under monotonicity,

$$\tau_{\textsf{FS}} = P[D_{1i} > D_{0i} | X_i = c] = P[\text{complier} | X_i = c]$$

- First stage identifies the proportion of compliers at the cutoff

Recovering the ATE on Compliers

- Instrument D_i with $\mathbb{I}(X_i \geq c)$

$$\mathbb{E}[Y_i(1) - Y_i(0)|X_i = c, D_{1i} > D_{0i}] = \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[D_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[D_i|X_i = x]}$$

- Fuzzy RD parameter is "doubly local":
  - At the cutoff
  - On the subpopulation of compliers

Estimation in Fuzzy Designs

- ITT and FS are sharp RD estimators
- The FRD parameter can be estimated using two-stage least squares
- Can adapt all previous tools to this case
  - Data driven bandwidth selection
  - Local polynomial estimation
  - Robust bias-corrected inference

Multicutoff and Multiscore RD

1. Multiple cutoffs:
   - Cutoffs change across regions, time periods, etc
   - All units receive the same treatment when they exceed their cutoff
2. Cumulative cutoffs:
   - Treatment is multivalued
   - Different dosage of treatment depending on value of X_i
   - E.g. $D_i = \mathbb{I}(X_i \leq c_1) + 2\mathbb{I}(c_1 < X_i \leq c_2)$
3. Multiple scores:
   - Treatment assigned based on multiple running variables
   - E.g. scholarship if both math and language scores above a cutoff

RD with Multiple Cutoffs

- Common empirical approach: pooling
  - $C_i \in \mathcal{C}$ (random) cutoff faced by unit i
  - Discrete cutoffs: $\mathcal{C} = \{c_0, c_1, ..., c_J\}$
  - Re-centered running variable: $\tilde{X}_i = X_i - C_i$
  - Pooled estimand:

$$\tau^p = \lim_{x \downarrow 0} \mathbb{E}[Y_i | \tilde{X}_i = x] - \lim_{x \uparrow 0} \mathbb{E}[Y_i | \tilde{X}_i = x]$$

Identification under the Pooling Approach

- If the CEFs and $f_{X|C}(x|c)$ are continuous at the cutoffs,

$$\tau^p = \sum_{c \in \mathcal{C}} \mathbb{E}[Y_i(1) - Y_i(0) | X_i = c, C_i = c] \omega(c)$$

- Where

$$\omega(c) = \frac{f_{X|C}(c|c) P[C_i = c]}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c) P[C_i = c]}$$

Exploiting Multiple Cutoffs

- Two drawbacks of the pooling approach:
  - Policy relevance: combines TEs *for different populations*
  - Discards variation that can identify parameters of interest
- What are the parameters of interest in this context?
- Potential CEFs:

$$\mu_d(x,c) = \mathbb{E}[Y_i(d)|X_i = x, C_i = c], \qquad d \in \{0,1\}$$

- (Conditional) ATE:

$$\tau(x,c) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x, C_i = c] = \mu_1(x,c) - \mu_0(x,c)$$

RD with Cumulative Cutoffs: Parameters

- Multivalued treatment $D_i \in \{d_1, d_2, \ldots, c_J\}$
- Effect of switching to one dosage to the next one:

$$\tau_j = \mathbb{E}[Y_i(d_j) - Y_i(d_{j-1})|X = c_j]$$

- Under continuity assumptions,

$$\tau_j = \lim_{x \downarrow c_j} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c_j} \mathbb{E}[Y_i|X_i = x]$$

RD with Cumulative Cutoffs: Estimation and Inference

- Can use robust bias-corrected techniques cutoff by cutoff
- Unit i is "control" for some units, "treated" for others
- Bandwidth selection:
  - Ensure bandwidths are non-overlapping or
  - Joint estimation accounting for overlap

RD with Multiple Scores

- Bivariate score: $\mathbf{X}_i = (X_{1i}, X_{2i})$
- Suppose treatment is assigned if both scores exceed a cutoff:

$$D_i = \mathbb{I}(X_{1i} \geq b_1)\mathbb{I}(X_{2i} \geq b_2)$$

- Multidimensional RD parameter:

$$\tau(\mathbf{b}) = \mathbb{E}[Y_i(1) - Y_i(0)|\mathbf{X}_i = \mathbf{b}], \quad \mathbf{b} \in \mathcal{B}$$

  - ATE at each point in the boundary set $\mathcal{B}$

RD with Multiple Scores: Identification

- Under continuity assumptions,

$$\tau(\mathbf{b}) = \lim_{\substack{d(\mathbf{x},\mathbf{b})\to 0 \\ \mathbf{x}\in\mathcal{B}_t}} \mathbb{E}[Y_i|\mathbf{X}_i = \mathbf{x}] - \lim_{\substack{d(\mathbf{x},\mathbf{b})\to 0 \\ \mathbf{x}\in\mathcal{B}_c}} \mathbb{E}[Y_i|\mathbf{X}_i = \mathbf{x}]$$

  - $\mathcal{B}_t$ = treated region
  - $\mathcal{B}_c$ = control region

- Need to define a notion of distance $d(\mathbf{x}, \mathbf{b})$

RD with Multiple Scores: Estimation

- Estimating a whole curve of $\tau(\mathbf{b})$ may not be feasible
- Alternative approach: pooling
  - Define distance measure $d(\cdot, \cdot)$
  - Normalize running variable as distance to closest boundary point
  - Run RD on (unidimensional) normalized running variable $\tilde{X}_i$

Bobba, Ederer, Leon-Ciliotta, Neilson, and Nieddu (2024)

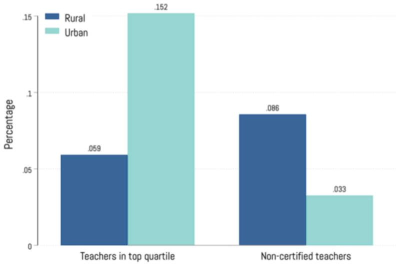
Teacher Compensation and Structural Inequality: Evidence from Centralized Teacher School Choice in Peru

- Centralized labor clearinghouses are increasingly used for entry-level positions in professional marketplaces
  - Both sides of the market express preferences over each other while a matching algorithm clears the market

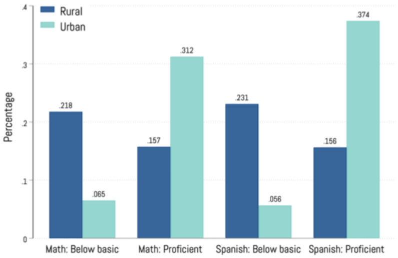
- Ideal setting to study the provision of services in the public sector
  - Rigid wage schedules lead workers to sort on non-pecuniary aspects of job

- How teacher wage policies can alleviate inequality of opportunities for students?
  - Wide learning gaps between urban and rural areas
  - Teachers matter for both education and life-cycle outcomes

Inequality of Education Inputs and Output



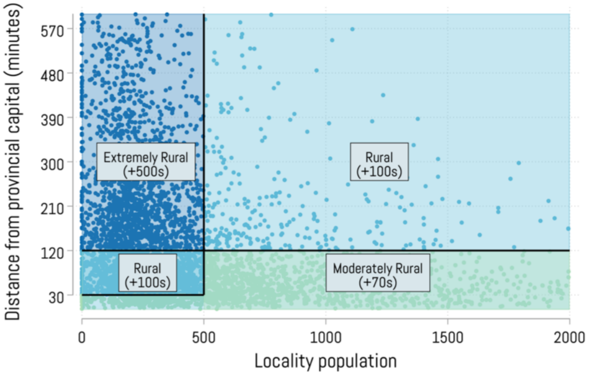
a) Teachers



b) Students

- Similar gaps across a broader set of indicators
- High spatial correlation between competent teachers and proficient students

The Rural Wage Bonus Policy



- From Rural to Extremely Rural wages increase by $\approx 1/3$ of base salary

Sharp RD Assumptions

- We compare the characteristics of teachers who choose and are assigned to a position at a high- vs. low-paying school and the test scores between the students of newly recruited teachers who are offered high- vs low-compensation

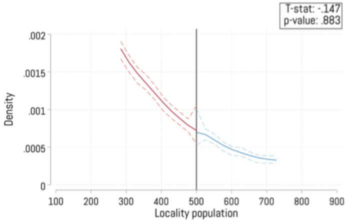
- The validity of the sharp RD design relies on two assumptions:

  Cont $E(Y_i(d) \mid X_i = x)$ is continuos in x, $\forall d \in \{0, 1\}$

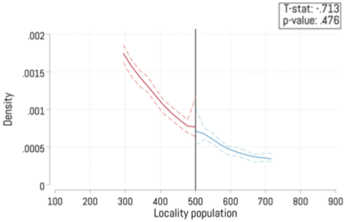
  SUTVA $E(Y_i(d) \mid D_{-i}, X_i = x) = E(Y_i(d) \mid X_i = x)$, $\forall d \in \{0, 1\}$

- The plausibility of both assumptions (should) can be checked in the data

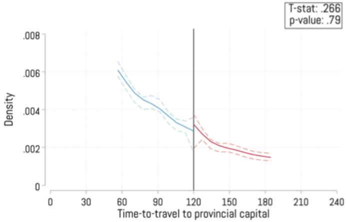
Density Tests Around Extremely Rural Cutoff



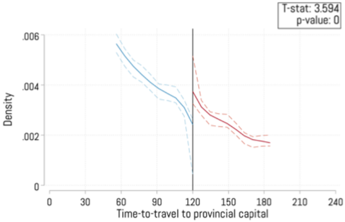
a. Population (2015)



b. Population (2017)

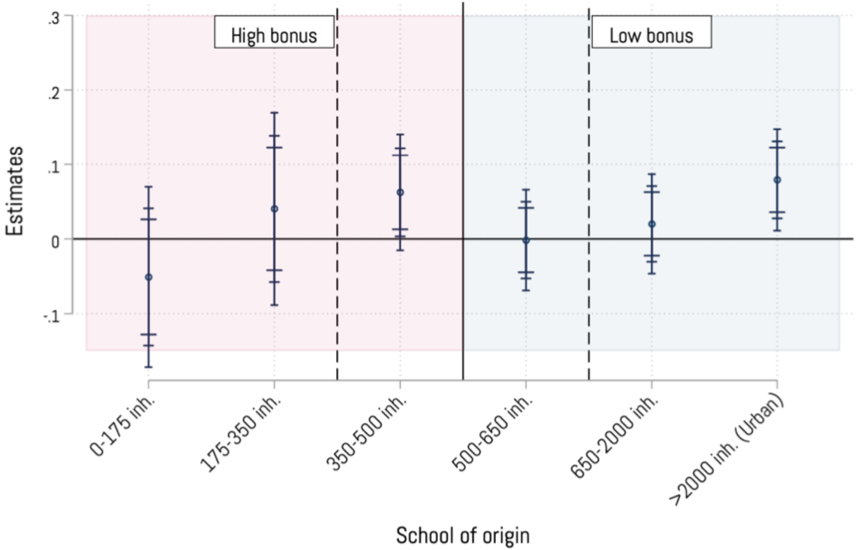


c. Time-to-travel (2015)



d. Time-to-travel (2017)

Rural Bonus and the Origin of Newly Recruited Teachers



Sharp RD Along Population Cutoff

- We solely rely on population-based assignment rule for rural schools with time-to-travel>30min

  ⇒ Weighted average increase in wages of 11%

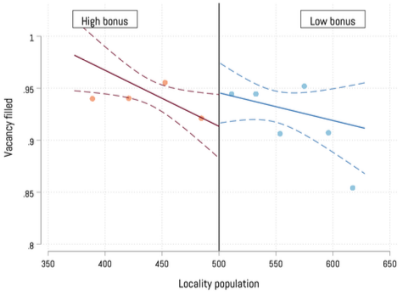
- Given continuity and SUTVA of potential outcomes around the cutoff

$$y_{ijt} = \gamma_0 + \gamma_1 \mathbf{1}(pop_{jt} < pop_c) + g(pop_{jt}) + \delta_t + u_{ijt}$$

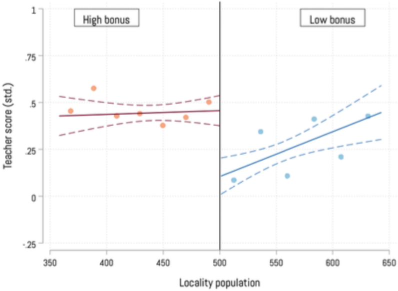
  - $g(\cdot)$: flexible polynomial on population of the locality of school j
  - δ_t: indicator for year of assignment
  - u_{jt}: error term, clustered at the school-year level

⇒ Estimate γ_1 non-parametrically within MSE-optimal bandwidths

Rural Bonus and Teacher Choices over Job Postings



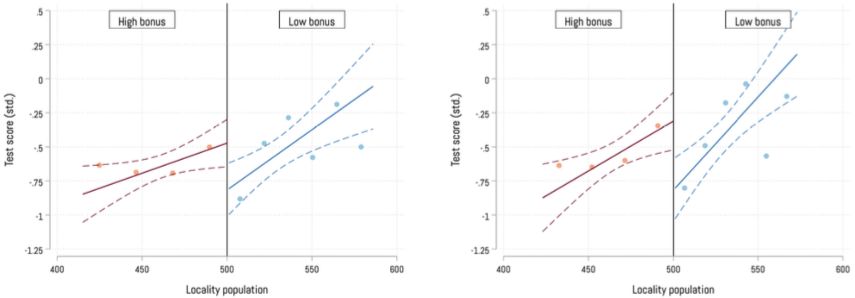
a) Vacancy Filled



b) Competency Score

NOTES. This figure shows how applicants' preferences and quality vary based on the difference between the 500-inhabitants cutoff and the population of the community where the school is located. In Panel A the outcome is an indicator variable that is equal to one if a vacancy was filled by a contract teacher during the centralized assignment, while Panel B uses the standardized score obtained in the centralized test by the newly-assigned contract teacher. Each marker indicates the average of the outcome variable within each bin, defined following the IMSE-optimal evenly spaced method by Calonico et al. (2015). Solid lines represent the predictions from linear regressions estimated separately for observations to the left and to the right of the cutoff. Dashed lines are 95% asymptotic confidence intervals.

Rural Bonus and Student Achievement



a. Spanish **b. Math**

NOTES. Panel A reports the relative shares of students by decile of the distribution of the average score in Spanish and math, separately for schools located to the right (Low Bonus) and left (High Bonus) of the population cutoff. Bars and vertical lines depicted in Panel B indicates the corresponding bias-corrected regression-discontinuity estimates of crossing the population threshold and the associated confidence intervals at the 90% level (Calonico et al., 2014). The sample includes schools with open positions for contract teachers.

Takeaways from RD Analysis

1. Rural bonus shifted supply of teachers towards disadvantaged areas
   - No effect on the probability of creating new matches, but inflow of more competent teachers (0.39 SD)
2. Students perform better in schools that pay higher wages
   - Only in schools that attracted better quality teachers (0.4-0.5 SD)
   - No effort response to higher wages for incumbent teachers

A Model of Teacher Sorting and Student Achievement

- RD evidence is by definition local
- What are the **global effect** of the wage bonus policy?
- Can we characterize **alternative teacher-school allocations**?
- Can we achieve those with **alternative wage schedules**?
- To tackle these questions, **we need a model** of teacher school choice and student outcomes

Preferences, Outcomes and Equilibrium

- Teacher i's utility from being matched with school j + outside option:

$$U_{ijt} = \underbrace{w_{jt}}_{\text{wage}} + \underbrace{\alpha_i^{-1}(u(a_{jt}, x_{it}) + \epsilon_{ijt})}_{\text{non-pecuniary utility}},$$

$$U_{i0t} = \beta_i + \epsilon_{i0t}$$

- Student l's potential outcome when matched with teacher i:

$$Y_{lij} = \underbrace{z_l'\bar{\delta}}_{\text{student ability}} + \underbrace{x_j'\beta}_{\text{school/classroom effect}} + \underbrace{z_l'(\delta_i - \bar{\delta})}_{\text{teacher ATE + match effects}} + \nu_{lij},$$

- Serial dictatorship $\Rightarrow$ discrete choice with observed choice sets

$$\mu_w^{\star}(i,t) = \text{arg max}_{j \in \Omega(s_{it})} U_{ijt},$$

Identification

- Assumptions
  1. u bounded + ϵ_{ijt} iid and Gumbel distributed
  2. $\mathbb{E}[\nu_{lij}|d_{ij}, x_j, z_l, i = \mu_s(l,t)] = 0$
  3. $(\log \alpha_i, \beta_i, \delta_i)|x_i \sim \mathcal{N}(\gamma(x_i), \Sigma)$
- Characterize mapping between preferences and equilibrium sorting:

$$\mathbb{P}(\mu^*(i,t) = j|x_{it}) = \frac{\exp\{\alpha_i w_{jt} + u(a_{jt}, x_{it})\}}{\exp\{\beta_i\} + \sum_{k \in \Omega(s_{it})} \exp\{\alpha_i w_{kt} + u(a_{kt}, x_{it})\}}$$

Identification

- We identify $\mathbb{P}(\mu^*(i,t) = j | x_{it})$ using panel data on realized matches as well as observed amenities, wages, teacher characteristics and choice sets
  - Repeated choice data to identify the distribution of random coefficients
  - Wages vary **only** with observables
- We then use the panel of matched teacher-classroom data to identify the parameters of the student achievement production function
  - Residualize wrt to school characteristics $x_j'\beta$
  - Estimate teacher effectiveness as fixed effect δ_i
  - $\delta_i \sim \mathcal{N}(x_{2i}'\gamma, \Sigma_{\delta,\delta})$

Estimation

- We parameterize $u(a_{jt}, x_{it})$ as a flexible function of a wide range of schools' and teachers' characteristics:

$$u(a_{jt}, x_{it}, \boldsymbol{\theta}) = x_{it}'\Gamma_1 z_{jt} + x_{it}'\Gamma_2 d_{ijt} + x_{it}'\Gamma_3 m_{ij} + \eta_j,$$

- Estimation in two steps
  1. Conditional pdf of $\hat{\delta}_i$ writes follows:

$$f(\hat{\delta}_i | x_{2i}) = \phi(\hat{\delta}_i | x_{2i}'\gamma^{\delta}, \Sigma_{\delta,\delta} + \hat{\Sigma}_i)$$

  2. Estimate $(\boldsymbol{\theta}, \gamma, \Sigma)$ by maximizing the following log-likelihood function:

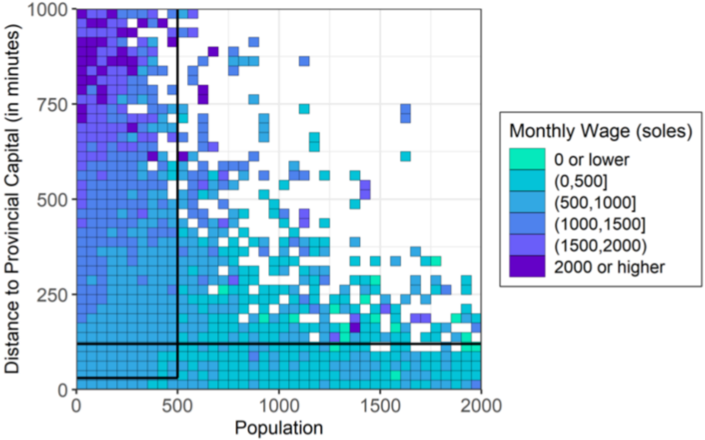
$$L(\boldsymbol{\theta}, \gamma, \Sigma) = \sum_{i=1}^{n_w} \sum_{t:\{\mu^*(i,t)\neq\emptyset\}} \log \mathbb{P}\left((\mu^*(i,t))_{t=1}^{T}, \hat{\delta}_i | x_i, \boldsymbol{w}, \boldsymbol{a}, \Omega(s_{it}) \right),$$

Result 1: Willingness to Pay Non-Wage Characteristics

- Teachers would be willing to give up on average 40% of their base salary to switch from a locality with bottom 5% to top 5% of amenities
- The average willingness to pay to avoid difficult teaching conditions ranges from 20% to 88% of the base wage
- Quechua teachers would be willing to give up 188% of their base wage to teach in a Quechua school
- Teachers would be on average willing to pay 2X base wage to switch from a school 20km away from home to home location
- These willingness to pay estimates are highly heterogeneous across teachers

$\Rightarrow$ Non wage attributes thus induce substantial vertical and horizontal differentiation across schools and locations

Result 2: Rural vs. Urban Non-Pecuniary Utility Diff



⇒ Utility differences are merely compensated by the wage bonus policy

Result 3: Standard Deviation of Value Added Coefficients

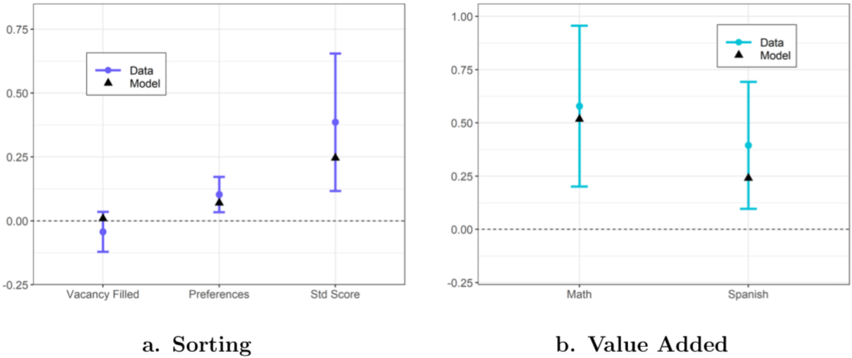
	Math	Spanish
	(1)	(2)
ATE	0.465 (0.006)	0.408 (0.006)
Lagged Score	0.145 (0.005)	0.150 (0.005)
Lagged Score^2	0.049 (0.004)	0.061 (0.003)
Female	0.098 (0.010)	0.083 (0.013)
Quechua - Aymara	0.040 (0.030)	0.067 (0.019)
Age	0.115 (0.007)	0.110 (0.008)

NOTES. This Table displays the estimates of the standard deviation of the population distribution of the teacher value added coefficients. Standard errors are in parentheses.

$\Rightarrow$ One SD increase in teacher value added corresponds to a 0.50 SD increase in students' math test score and a 0.44 SD increase in Spanish score

$\Rightarrow$ Students with past achievement 1 SD below average can experience test score gains of 0.58 SD when being matched with a teacher in top 5% vs. bottom 5%.

Model Fit



⇒ Slightly under-predict RD effects on pref. for rural schools and teacher score

Average Effect of the Wage Bonus Policy

	Status Quo (1)	No Bonus (2)	(1) - (2) (3)
Panel A: Math			
Urban Rural Gap	0.093	0.184	-0.091
Urban	0.037	0.058	-0.021
Rural	-0.056	-0.126	0.070
Moderately Rural	0.057	0.050	0.007
Rural	-0.151	-0.099	-0.052
Extremely Rural	-0.004	-0.175	0.171
Panel B: Spanish			
Urban Rural Gap	0.128	0.210	-0.082
Urban	0.035	0.050	-0.015
Rural	-0.093	-0.160	0.067
Moderately Rural	0.019	0.021	-0.002
Rural	-0.178	-0.134	-0.044
Extremely Rural	-0.052	-0.205	0.153

NOTES. This Table displays the estimates of the preference parameter θ. Standard errors are in parentheses.

$\Rightarrow$ Small average effects. Mostly concentrated in extremely rural schools

$\Rightarrow$ Limited negative consequences for urban areas $\Rightarrow$ Good teachers substituting away from the outside option

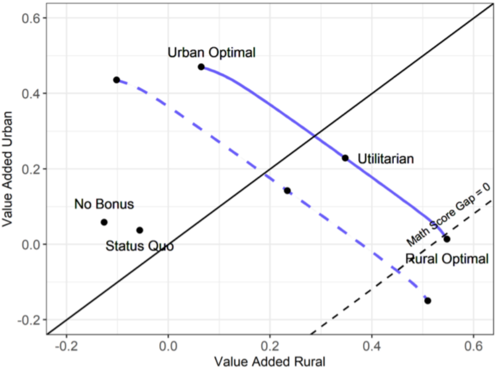
Equity and Efficiency Gains from Reallocation

- We define the following objective function:

$$\max_{\mu} S(\mu, \pi) = \sum_{i=1}^{n_w} \sum_{j=1}^{n_m} \pi_j \bar{Y}_{ij} \mathbf{1}\{\mu(i) = j\}$$

- $\bar{Y}_{ij}$ is the average predicted test score in school j
- π_j is the weight put on students in school j
- $\pi_j = 1$ if $j \in$ Urban, and $\pi_k = 1 + x$ if $k \in$ Rural
- $x = 0$ is the **Utilitarian Allocation**
- $\pi_j = 0$ is the **Rural Optimal Allocation**
- $\pi_k = 0$ is the **Urban Optimal Allocation**

Equity and Efficiency Gains from Reallocation



⇒ Match effects loom large for efficiency (esp. drawing from outside option)

⇒ No Trade-off between equity and efficiency

Optimal Wage Policy

- Develop framework to **design wage policy**
  - **Matching with contracts:** let schools bid for teachers (Hatfield & Milgrom, 2005)
- Compute wage bill of **increasing equity/efficiency** under different scenarios
  - **Perfect information** of policy maker about **teacher value added**
  - Information only on **absolute advantage**
  - Information only on **competency score**

Optimal Wage Policy

- **Goal:** solve following problem

$$\min_{w} \sum_{j} w_{jt}, \text{ s.t.} \begin{cases} Y_{\mu(j)j} > c, \ \forall j \in \mathcal{S} \\ \mu \text{ is } \textbf{stable} \text{ given } w \text{ and using } Y_{ij} \text{ as } \textbf{priorities} \end{cases}$$

Y_{ij}: outcome of interest (student score, teacher competency score)

- **Result:** school-proposing Generalized DA converges to solution
    - Each school $j \in \mathcal{S}$ bids upward until $Y_{\mu(j)j} > c$
    - Outcome is a pair (μ, w) that satisfies problem above while **minimizing total wage bill**

Wrapping up

- **Unique setting** and very rich **nation-wide** admin data + survey data
    - Local **policy evaluation** of wage bonus policy using **RD design**
    - Identify and estimate **joint model** of **teacher school choice** and **student outcomes**
    - Build framework to design **optimal wage policies**

- Take-aways:
    - Teacher compensation design can alleviate structural inequalities
    - Large efficiency and equity gains from optimal policy design
    - But implementing cost effective policy requires information

Market Design and Policy Design

- Inferring heterogenous preferences is crucial to design effective policies

  $\Rightarrow$ Current wage bonus policy is largely misallocated

  $\Rightarrow$ Alternative wage bonuses increase both equity and efficiency

- Pricing strategies that effectively take into account demand and supply in real time are increasingly common in the private sector

  $\Rightarrow$ Leveraging this approach for the public sector may be promising

  $\Rightarrow$ Leveling the playing field in the access to fundamental services such as health, educ...