Empirical Methods for Policy Evaluation

Matteo Bobba

Toulouse School of Economics (TSE)

TSE PhD Program (MRes)
Fall 2025

Regression Discontinuity Designs (5 Classes)

1. RD methods
   - Local randomization approach
   - Continuity-based approach
   - Fuzzy RD
   - Extensions
2. Applications

   ⇒ Teacher sorting and student outcomes (Bobba et al, 2024)

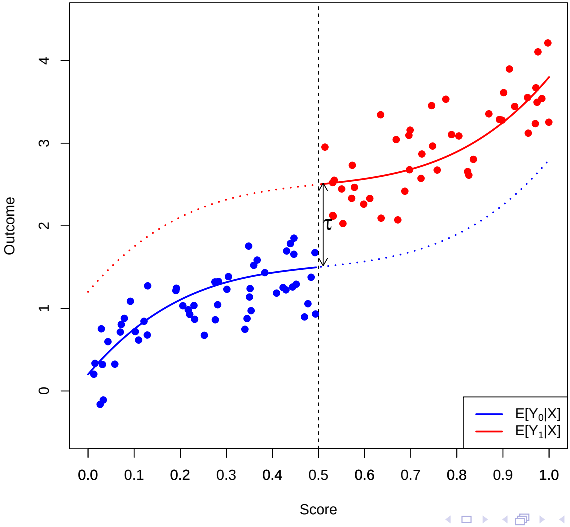
   ⇒ The equilibrium effect of a large-scale education reform (Khanna, JPE 2023)

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

$\Rightarrow$ Sharp design (will extend this later): $P[D_i = 1|X_i] = \mathbb{I}(X_i \geq c)$

Graphical Intuition



The local randomization approach

RD as a Randomized Experiment

- Idea: close enough to the cutoff, some units were "lucky"
- Treatment as if randomly assigned in window $W_0 \in [c-w, c+w]$ for $w > 0$:
  - $\Rightarrow$ Probability distribution of X_i is uncounfounded

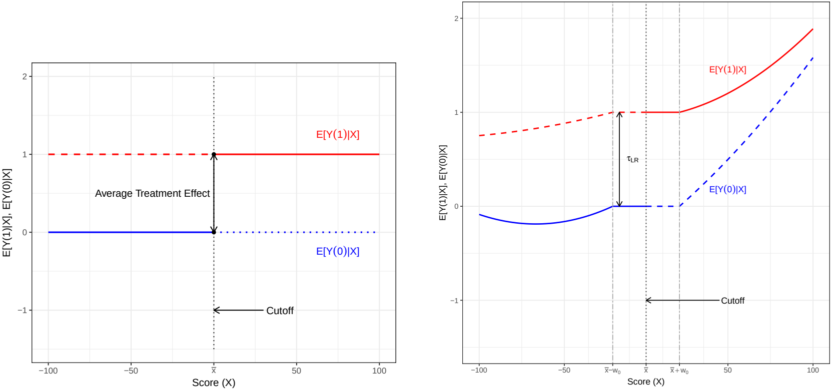
$$P[X_i \leq x | y_i(0), y_i(1)] = P[X_i \leq x], \quad \forall X_i \in W_0$$

  - $\Rightarrow$ Potential outcomes not affected by value of the score

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

- Example: $X_i \sim U[0,1]$, $D_i = \mathbb{I}(X_i \geq c)$, then $P[D_i = 1] = 1 - c$

RD as a Randomized Experiment



(a) Randomized Experiment

(b) RD Design

Window Selection

- 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

  $\Rightarrow$ 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

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

$\Rightarrow$ Alternative (and better) approach: randomization inference (!)

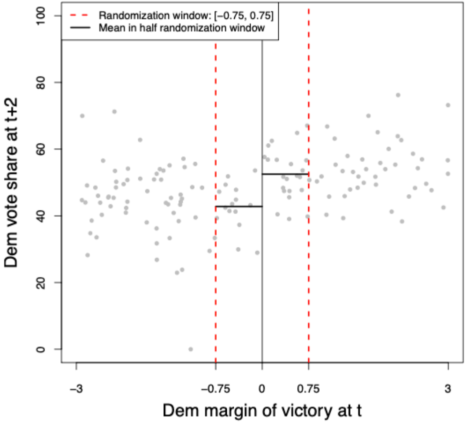
Software Implementations

⇒ `rdlocrand` package (R and Stata)

- `rdwinselect`: window selection
- `randinf`: randomization inference
- `rdsensitivity`: sensitivity analysis
- `rdrbounds`: Rosenbaum bounds

Example: Incumbency Advantage in U.S. Senate

- Y_i = election outcome at $t+2$ ($= 1$ 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

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 $x = c$

$$\tau_{\mathsf{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]$$

$\Rightarrow$ Treatment effect only nonparametrically identified at the cutoff

- We actually have zero observations at $X_i = c$
- The model is generally misspecified

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

1. Global Parametric
   - Estimate a polynomial on full sample
   - Sensitive to misspecification
   - Erratic behavior at boundary points
2. Local parametric
   - Estimate a polynomial within an ad-hoc bandwidth
   - Sensitive to misspecification and bandwidth choice
3. Local non-parametric
   - $\Rightarrow$ Data-driven bandwidth selection
   - $\Rightarrow$ Accounts for misspecification when performing inference

Global Parametric Approach

- Parametric assumption on conditional expectations, e.g.

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

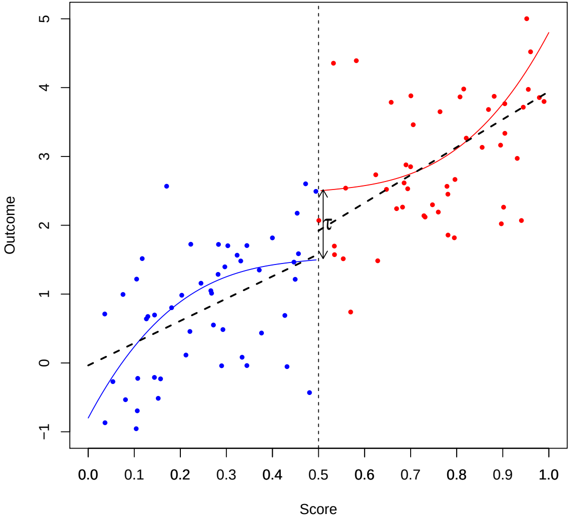
$\Rightarrow$ This implies

$$Y_i = \alpha_0 + (\alpha_1 - \alpha_0)D_i + \beta_0(X_i - c) + (\beta_1 - \beta_0)(X_i - c)D_i + u_i$$

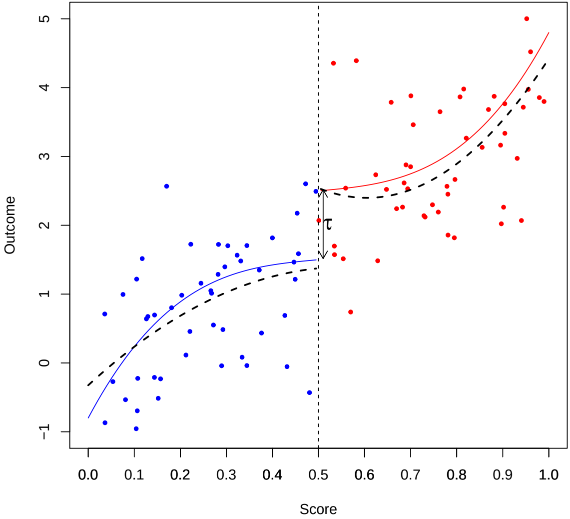
- Easily estimated by OLS on full sample

$\Rightarrow$ Coefficient $\alpha_1 - \alpha_0$ recovers the treatment effect at the cutoff

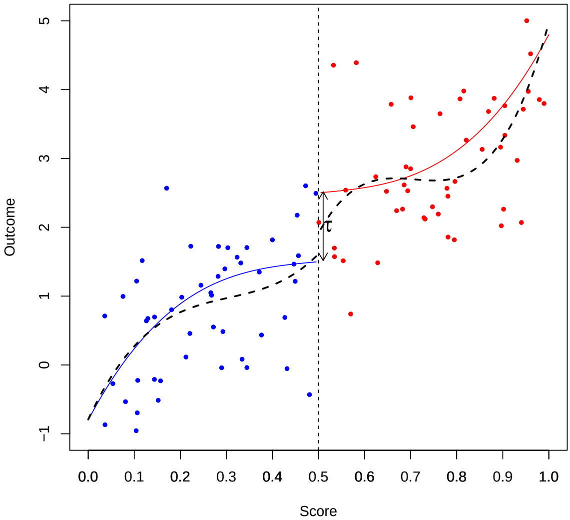
Global Parametric Approach: $p = 1$



Global Parametric Approach: $p = 2$



Global Parametric Approach: $p = 3$



Local Linear (Polynomial) Regression

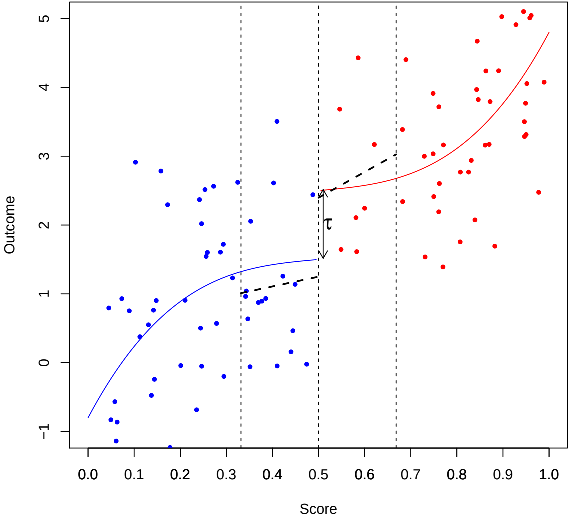
- Suppose $c = 0$ (otherwise, use $X_i - c$)
- Choose some bandwidth $h > 0$ and estimate by OLS:

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

$\Rightarrow$ 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:

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

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

- $\bar{X}_h, \bar{Y}_h$ are local means

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

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

- Symmetrically for $\hat{\beta}^{-}, \hat{\alpha}^{-}$

Local Linear Regression: Bias

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

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

$\Rightarrow$ $\mu_1(x) = \mathbb{E}[Y_i(1)|X_i = x]$

$\Rightarrow$ $\mathcal{B}_+$ is a constant that depends on:

1. The curvature of $\mu_1(x)$
2. The kernel function
3. The order of polynomial, p

$\Rightarrow$ Smaller h implies small bias but fewer observations: more variance

Variance

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

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

⇒ $\mathcal{V}_+$ is a constant that depends on:

1. $\mathbb{V}[Y_i(1)|X_i = 0]$
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

- Mean-squared error (MSE):

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

- Where:

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

Bandwidth Selection

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

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

Inference

- Under regularity and rate restrictions on $h \to 0$

$$\sqrt{nh}(\hat{\tau} - \tau - h^2\mathcal{B}) \to_{\mathcal{D}} \mathcal{N}(0, \Omega)$$

- Where $\Omega \propto \frac{\mathcal{V}_+ + \mathcal{V}_-}{f}$ can be estimated using residuals $\epsilon_{+,i} = Y_i - \widehat{\mu}_+(X_i)$

- Two main approaches to handle the leading bias term $\mathcal{B}$:

  $\Rightarrow$ Undersmoothing: use a "smaller" bandwidth

  $\Rightarrow$ Bias correction

Undersmoothing

- Choose a small enough bandwidth and proceed as if $\mathcal{B} \approx 0$

$$\sqrt{nh}(\hat{\tau} - \tau - h^2\mathcal{B}) = \sqrt{nh}(\hat{\tau} - \tau) + o_p(1) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \Omega)$$

$\Rightarrow$ But recall that $h^*_{\text{MSE}} \propto n^{1/5}$ so this approximation will have a bias

Robust Bias Correction

- Consider the following de-biasing approach:

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

$\Rightarrow$ Bias correction:

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

- But the bias is unknown, use $\mathcal{B}_p \propto \mu_+^{(p+1)} - \mu_-^{(p+1)}$ with pilot bandwidth b_n

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

- Σ accounts for the estimation of the bias

Density discontinuity tests

- RDDs can be invalid if individuals manipulate X_i
- Manipulation can imply sorting on one side of the cutoff

$\Rightarrow$ Test whether the density of X_i is continuous around c

Continuity in covariates / placebo outcomes

- Some variables should reveal no treatment effect:
  - $\Rightarrow$ Outcomes not targeted by treatment (placebo outcomes)
  - $\Rightarrow$ Exogenous or predetermined covariates
- Estimate an RD effect on these variables

$\Rightarrow$ Finding a non-zero effect suggests an invalid RDD

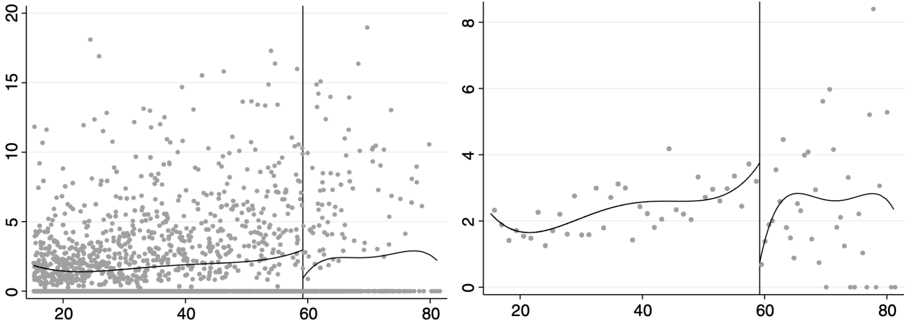
- Existence of other (unobserved) treatments at the cutoff
- Selection

Software Implementations

- `rdrobust` package (R, Stata, and other softwares)
  - `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, $J_- = 37$, $J_+ = 38$

Impact of Head Start on Child Mortality: Falsification

⇒ 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.

Impact of Head Start on Child Mortality: RD Estimates

- Local parametric RD
  - $\Rightarrow$ $\hat{\tau}_{\{p=4,\text{full sample}\}} = -3.065$, p-value $= 0.005$
  - $\Rightarrow$ $\hat{\tau}_{\{p=1,h=18\}} = -1.198$, p-value $= 0.071$
  - $\Rightarrow$ $\hat{\tau}_{\{p=1,h=9\}} = -1.895$, p-value $= 0.055$
- Local non-parametric RD
  - $\Rightarrow$ $\hat{\tau}_{\{p=0,\hat{h}_{MSE}=3.24\}} = -2.114$, robust p-value $= 0.037$
  - $\Rightarrow$ $\hat{\tau}_{\{p=1,\hat{h}_{MSE}=6.81\}} = -2.409$, robust p-value $= 0.042$

Fuzzy RD

Sharp Vs. 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 wrt treatment assignment
  - ⇒ 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

- 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 usual continuity assumption

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

$\Rightarrow$ ITT can be ≈ 0 even if τ is large

The Monotonicity Assumption

- Rule out the presence of defiers

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

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

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

$\Rightarrow$ $X_i \geq c$ does not decrease the probability of treatment

First Stage

- Effect of treatment assignment on treatment status

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

- Under monotonicity

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

$\Rightarrow$ First stage identifies the proportion of compliers at the cutoff

Recovering the ATE on Compliers

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

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

$\Rightarrow$ 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/packages to this case
  - ⇒ Data driven bandwidth selection
  - ⇒ Local polynomial estimation
  - ⇒ Robust bias-corrected inference

Extensions

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
3. Multiple scores:
   - ⇒ Treatment assigned based on multiple running variables

RD with Multiple Cutoffs

- Common empirical approach: pooling

  $\Rightarrow$ Discrete cutoffs: $\mathcal{C} = \{c_0, c_1, ..., c_J\}$

  $\Rightarrow$ Re-centered running variable: $\tilde{X}_i = X_i - C_i$

  $\Rightarrow$ 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, then

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

$\Rightarrow$ Where the weights reflect the densities at each cutoff

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

Exploiting Multiple Cutoffs

- Two features of the pooling approach:
  - Combines TEs for different populations
  - Discards variation that can identify parameters of interest

⇒ Potential CEFs:

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

⇒ CATE

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

RD with Cumulative Cutoffs

- 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 assumption

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

- `rdc`: robust bias-corrected techniques for pooled or cutoff-specific RD effects

RD with Multiple Scores (Boundary RD)

- Bivariate score: $\mathbf{X}_i = (X_{1i}, X_{2i})$
- Treat/control regions $\mathcal{A}_1, \mathcal{A}_0$ and boundary $\mathcal{B} = \text{bd}(\mathcal{A}_1) \cap \text{bd}(\mathcal{A}_0)$
- Multidimensional RD parameter:

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

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

RD with Multiple Scores (Boundary RD)

1. Univariate distance to boundary
   - ⇒ Define distance measure, e.g. $d(\mathbf{X}_i, \mathbf{b}) = \sqrt{(X_{1i} - b_1)^2 + (X_{2i} - b_2)^2}$
   - ⇒ Run RD on (unidimensional) normalized running variable $\tilde{d}_i$
   - ⇒ Misspecification bias when $\mathcal{B}$ is non-smooth (e.g. near a kink, next example)

2. Bivariate location relative to boundary
   - ⇒ Additional (mild) regularity condition on $\mathcal{B}$ is needed
   - ⇒ Estimator is well behaved at any point of $\mathcal{B}$
   - ⇒ Identification, estimation, and inference are standard (`rd2d` package for `R`)

Location-based Multiple RD: Identification

- Continuity assumption, for all $\mathbf{b} \in \mathcal{B}$

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

$\Rightarrow$ $\mathcal{A}_t(\mathcal{A}_c) =$ treated (control) region

- Aggregated ATE along the boundary $\mathcal{B}$ is

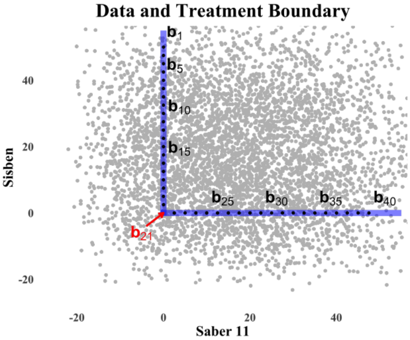
$$ATE(\mathcal{B}) = \frac{\int_{\mathcal{B}} \tau(\mathbf{b}) w(\mathbf{b}) d\mathbf{b}}{\int_{\mathcal{B}} w(\mathbf{b}) d\mathbf{b}}$$

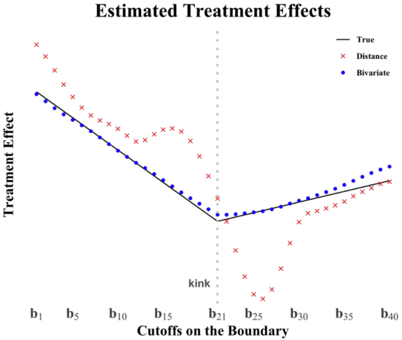
$\Rightarrow$ Which admits a generic plug-in estimator $\widehat{\tau}(\mathbf{b})$ for $\mathbf{b} = (\mathbf{b}_1, ..., \mathbf{b}_j)$ over $\mathcal{B}$

Boundary RD: Example

⇒ Ser Pilo Paga: Colombian scholarship program; students $i = 1, ..., n$

- $X_i = (SABER11_i, SISBEN_i)$; exam score + poverty score
- $\mathcal{B} = \{SABER11_i \geq 0 \& SISB_i = 0\} \cup \{SISB_i \geq 0 \& SABER11_i = 0\}$



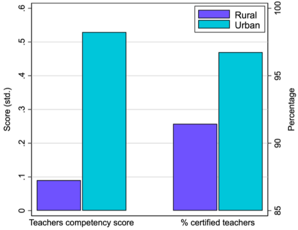


Teacher sorting and student outcomes (Bobba et al, 2024)

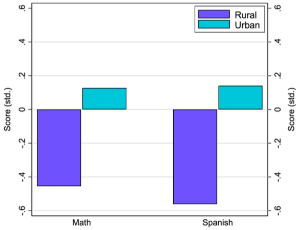
Teacher Compensation and Structural Inequality: Evidence from Centralized Teacher School Choice in Peru

1. Descriptives on nation-wide allocation of public teachers in Peru
   - $\Rightarrow$ Document large urban-rural gaps in teacher quality and student test scores
2. RD-based evidence of teacher wage bonuses in rural locations
   - $\Rightarrow$ Teacher competency $\uparrow$ by 0.39σ + student test scores $\uparrow$ by 0.23-0.32σ
3. Model of teacher school choice/value added to study aggregate policy effects
   - $\Rightarrow$ Framework to design cost-effective wage policy for equity/efficiency objectives

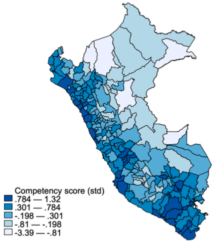
Inequality of Education Inputs and Output



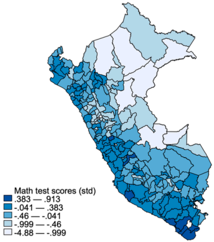
a) **Teacher Competency by Urban/Rural**



b) **Student Achievement by Urban/Rural**

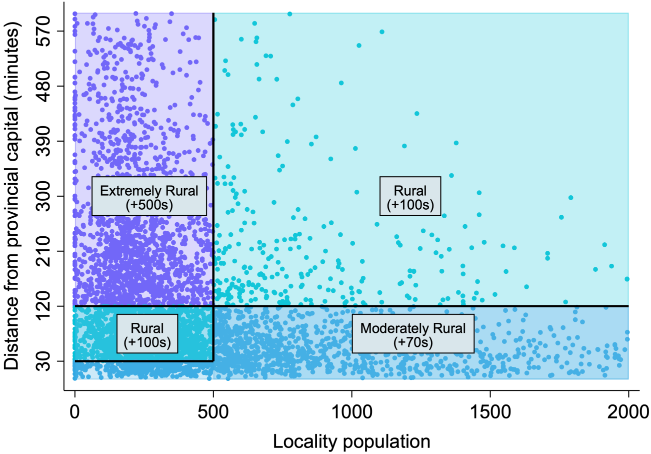


c) **Teacher Competency by Province**



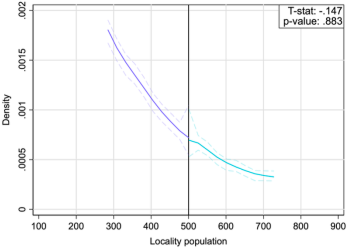
d) **Student Achievement (Math) by Province**

The Rural Wage Bonus Policy

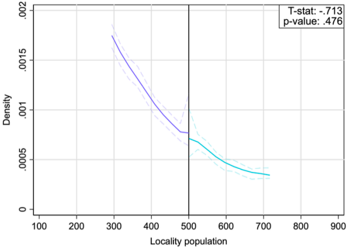


⇒ From Rural to Extremely Rural wage increase by ≈ $1/4$ of base salary

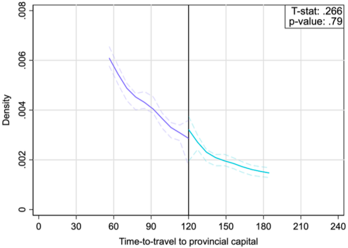
Density Tests Around Extremely Rural Cutoff

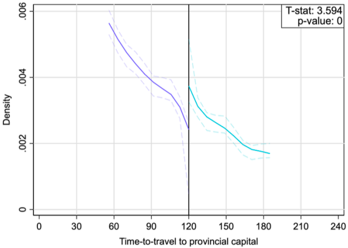


a. Population (2016)



b. Population (2018)





Sharp RD Along Population Cutoff

- We rely on pop-based assignment rule for rural schools with dist>30min

⇒ Weighted average increase in wages of 11%

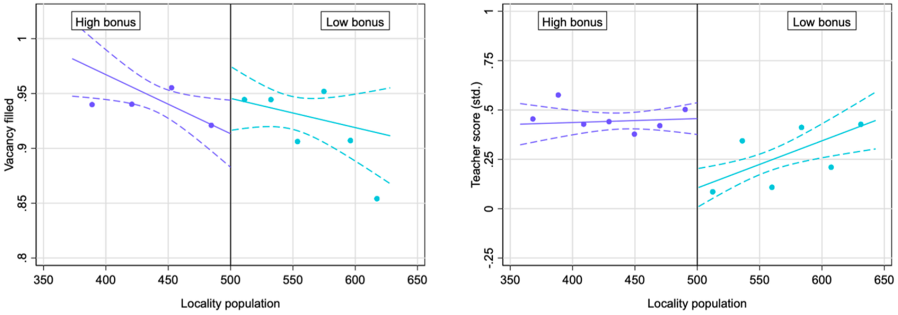
- Given continuity of potential outcomes around the population cutoff

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

  - $g(\cdot)$: 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



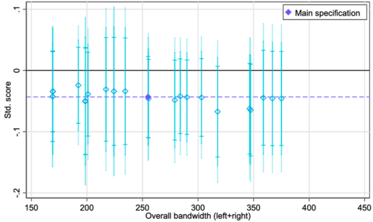
a) Vacancy Filled **b) Competency Score**

	(1) Vacancy filled	(2) Preferences	(3) Teacher Score (Std.)
High Bonus	-0.043	0.103	0.386
	(0.040)	(0.035)	(0.137)
Bandwidth	127.521	157.452	141.447
Schools	715	850	764
Observations	1851	2080	1870

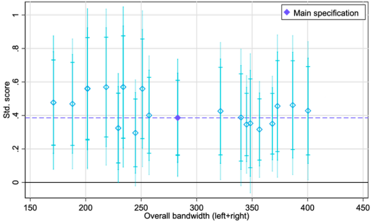
Rural Bonus and Student Achievement

Panel A: Dependent Variable is Spanish Test (z-score)			
	(1)	(2)	(3)
	Vacancy	No vacancy	All
High Bonus	0.395	-0.004	0.232
	(0.152)	(0.127)	(0.088)
Bandwidth	107.818	148.920	105.822
Schools	264	451	832
Observations	4635	6773	16681
Panel B: Dependent Variable is Math Test (z-score)			
	(1)	(2)	(3)
	Vacancy	No vacancy	All
High Bonus	0.579	0.067	0.317
	(0.193)	(0.143)	(0.105)
Bandwidth	85.848	155.174	95.638
Schools	220	470	764
Observations	3939	7039	15363

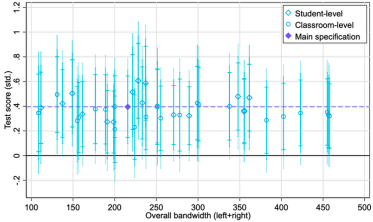
RD Robustness



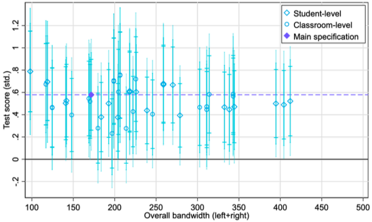
a. Vacancy filled



b. Teacher scores (std)



c. Spanish scores



d. Math scores

What is the Rationale for the Model?

- The RD evidence is limited to the local effect of the rural wage bonus

⇒ What is the effect of the policy on urban-rural gaps in education outcomes?

⇒ Can we characterize more effective teacher-school allocations?

⇒ Can we achieve those with alternative wage schedules?

An Empirical Matching Model of Teachers and Schools

1. A discrete choice framework recovers teachers' preferences
   - ⇒ Flexibly parametrize pref. distributions over wages and non-wage amenities
2. A matching equilibrium that maps preferences into assignment outcomes
3. A value-added model that maps teacher sorting into student achievement
   - ⇒ Teacher VA varies with students' prior achievement and demographics
   - ⇒ Teacher VA parameters correlate with teachers' preferences

Teacher Preferences

- Teacher i utility from school j (off-platform $j = p$) + outside option $j = 0$:

$$U_{ijt} = \underbrace{w_{jt}}_{\text{wage}} + \underbrace{\alpha_i^{-1}(u(a_{jt}, x_{it}) + \epsilon_{ijt})}_{\text{non-pecuniary amenities}}, \qquad U_{i0t} = \alpha_i^{-1}(\beta_i + \epsilon_{i0t})$$

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

$$\mu_w^*(i,t) = \arg\max_{j \in \Omega(s_{it})} U_{ijt}$$

Teacher Value Added

- Student l potential outcome when matched with teacher i:

$$Y_{lij} = + \underbrace{c_{jt}\beta}_{\text{school/classroom effect}} + \underbrace{z_{lt}\bar{\delta}}_{\text{student ability}} + \underbrace{z_{lt}(\delta_i - \bar{\delta})}_{\text{teacher ATE + match effects}} + \nu_{lij}$$

- We allow value-added to correlate with preferences $\theta_i = (\log \alpha_i, \beta_i)$

$$(\theta_i, \delta_i)|x_{it} \sim \mathcal{N}\left[\begin{pmatrix} x'_{1it}\gamma^{\theta} \\ x'_{2it}\gamma^{\delta} \end{pmatrix}, \begin{pmatrix} \Sigma_{\theta,\theta} & \Sigma_{\theta,\delta} \\ \Sigma_{\delta,\theta} & \Sigma_{\delta,\delta} \end{pmatrix}\right]$$

$\Rightarrow \Sigma_{\theta,\theta}$: how teachers value non-wage benefits/outside option against wages

$\Rightarrow \Sigma_{\delta,\theta}$: selection on teacher quality resulting from changes in wages

Identification

- We identify the achievement prod. function using teacher-classroom data
  - Estimate teacher effectiveness as fixed effects δ_i
  - Use variation in observables x_{2it} to recover $(\gamma^{\delta}, \Sigma_{\delta,\delta})$
- We identify choice parameters using panel data on choices/choice sets
  - Repeated choice data help identify the distribution of random coefficients θ_i
  - Wages vary only with observables $\Rightarrow$ residual variation is RD effect
- We identify $\Sigma_{\theta,\delta}$ by linking assignments with teacher-classroom data
  - Conditional on knowing $\Sigma_{\delta,\delta}$ we can recover $\Sigma_{\theta,\delta}$

Estimation

- We flexibly parametrize the non-wage component of the choice model as:

$$u(a_{jt}, x_{it}, \boldsymbol{\theta}) = \underbrace{x'_{it}\Gamma_1 q_{jt}}_{\text{amenities}} + \underbrace{x'_{it}\Gamma_2 d_{ijt}}_{\text{moving costs}} + \underbrace{x'_{it}\Gamma_3 m_{ij}}_{\text{match effects}} + \underbrace{\kappa_j}_{\text{unobs. amenities}}$$

- Estimation in two steps (see Appendix D.2 for details)
  1. Estimate the parameters of the achievement production function
  2. Estimate $(\boldsymbol{\Gamma}, \gamma, \Sigma)$ by maximizing the log-likelihood function:

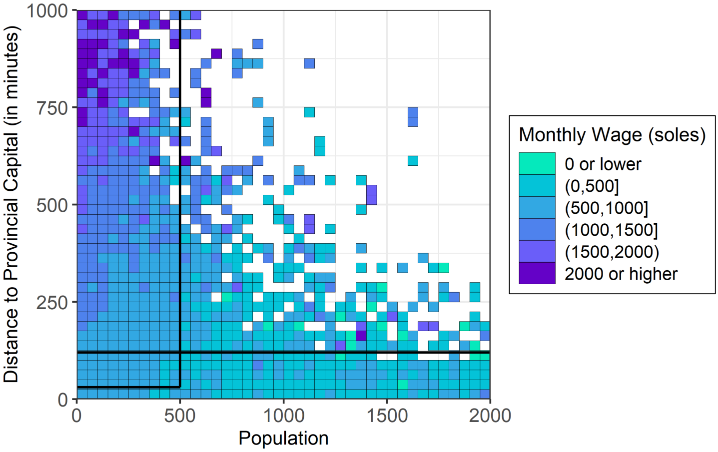
$$L(\boldsymbol{\Gamma}, \gamma, \Sigma) = \sum_{i=1}^{n_w} \sum_{t:\{\mu^*(i,t)\neq\emptyset\}} \log \mathbb{P}\left((\mu^*(i,t))_{t=1}^{T}, \hat{\delta}_i | x_i, \boldsymbol{w}, \boldsymbol{a}, \Omega(s_{it}) \right),$$

Monthly Willingness to Pay for Non-Wage Characteristics

	Mean		10% Quantile		90% Quantile	
	Soles	% Wage	Soles	% Wage	Soles	% Wage
	(1)	(2)	(3)	(4)	(5)	(6)
Amenities, Infrastructure and Remoteness						
Amenity/Infrastructures	200	10	30	2	440	22
Closer to Home by 1km						
$0 \leq$ Distance < 20	200	10	33	2	443	22
$20 \leq$ Distance < 100	113	6	23	1	243	12
Distance ≥ 100	20	1	3	0	43	2
Ethnolinguistic Proximity						
Same Language: Spanish	2,777	139	393	20	6,180	309
Same Language: Quechua	986	49	303	15	1,929	96
Same Language: Aymara	3,264	163	656	33	6,976	349
Teaching Conditions						
No Border	406	20	-97	-5	1,122	56
No Multigrade	962	48	147	7	2,121	106
No Single Teacher	1,758	88	120	6	4,123	206

$\Rightarrow$ Non-wage attributes induce vertical+horizontal differentiation across schools

Rural vs. Urban Non-Pecuniary Utility Differences



⇒ Utility differences are merely compensated by the wage bonus policy

Standard Deviation of TVA Coefficients

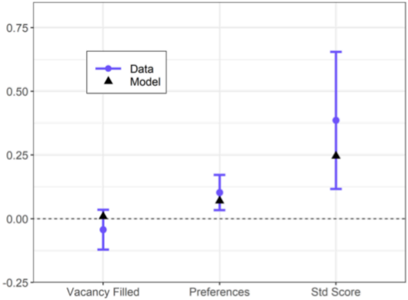
	Math	Spanish
	(1)	(2)
ATE	0.465 (0.006)	0.408 (0.006)
Lagged Score	0.145 (0.005)	0.150 (0.005)
Lagged Score^2	0.049 (0.004)	0.061 (0.003)
Female	0.098 (0.010)	0.083 (0.013)
Quechua - Aymara	0.040 (0.030)	0.067 (0.019)
Age	0.115 (0.007)	0.110 (0.008)

$\Rightarrow$ One SD increase in TVA $\Rightarrow \uparrow$ in test scores by 0.44-0.50 SD

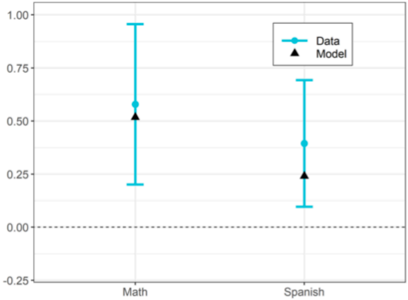
$\Rightarrow$ Significant match effects on lagged measures of student achievement

$\Rightarrow$ 12-18% of variance in TVA explained by teachers' comparative advantage

Model Fit



a. Sorting



b. Value Added

- Estimated model replicates the RD evidence induced by the rural wage bonus
- Good fit on moments away from the pop. threshold (urban-rural gaps, etc.)

Counterfactual 1: Aggregate Effects of the Rural Bonus

- Predict teachers' choices over schools with and without rural bonus
  - Simulate U_{ijt} from estimated parameters and a random draw of ϵ_{ijt} and θ_i
- Compute the stable matching eq. using the teacher-proposing DA algorithm
- Predict the distribution of teacher value-added without and with rural bonus
  - Use the mean of the posterior distribution of δ_i (see Appendix D.3)

Counterfactual 1: Aggregate Effects of the Rural Bonus

	Status Quo (1)	No Rural Bonus (2)	Policy Effect (3)
Panel A: Total Value Added			
Urban-Rural Gap	0.077	0.164	-0.087
Urban	0.024	0.059	-0.036
Rural	-0.053	-0.105	0.052
Moderately Rural	-0.033	-0.055	0.022
Rural	-0.111	-0.049	-0.063
Extremely Rural	0.067	-0.099	0.166
Panel B: Match Effects			
Urban	-0.007	0.002	-0.009
Rural	0.008	0.001	0.007

$\Rightarrow$ Small average effects on TVA, mostly concentrated in very remote schools

$\Rightarrow$ Rural bonus does not induce sorting based on comparative advantages

Counterfactual 2: Gains from Teachers' Reallocation

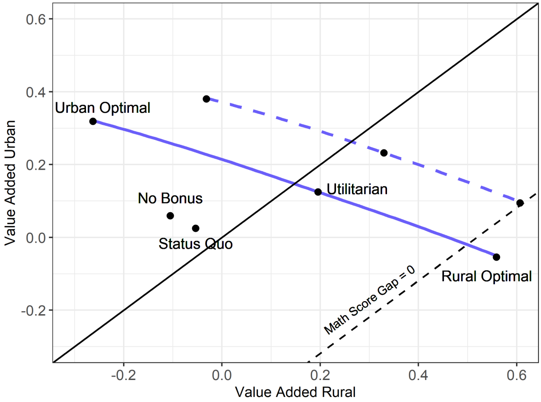
- We consider the following the linear program:

$$\max_{\mu} \sum_{i=1}^{n_w} \sum_{j=1}^{n_m} \pi_j \bar{z}_j' \hat{\delta}_i \mathbb{I}\{\mu(i) = j\}$$

- $\bar{z}_j{}' \hat{\delta}_i$ is the predicted (shrunken) average TVA for teacher i in school j

- Solution $\mu^{\star}(\pi_j)$ depends on weight put on students in school j

Counterfactual 2: Gains from Teachers Reallocation



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

⇒ No trade-off between equity and efficiency

Counterfactual 3: Optimal Wage Policy

- Policymaker can set priorities and wages in each school such that:

$$\min_w \sum_j w_j, \text{ s.t. } \begin{cases} \max_{i\in\mu(j)} z'_{lt}\delta_i \geq c_j, \ \forall j \in \mathcal{S} & \text{(C1)} \\ \mu \text{ is stable given } w \text{ and using } z'_{lt}\delta_i \text{ as priorities} & \text{(C2)} \end{cases}$$

- For a fixed wage, schools strictly rank teachers according to $z'_{lt}\delta_i$
- Otherwise, the allocation with the lower wage is always strictly preferred
- A stable set of contracts always exists in this counterfactual economy
  - Each school $j \in \mathcal{S}$ bids upward until (C1) is satisfied
  - Outcome is (μ, w) that satisfies (C1)-(C2) while minimizing total wage bill

Counterfactual 3: Optimal Wage Policy

	Status Quo	Teacher Value Added Threshold				
		$c = -0.4$	$c = -0.3$	$c = -0.2$	$c = -0.1$	$c = 0$
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Teacher Value Added						
Urban	0.055	0.036	0.035	0.019	-0.009	-0.058
Rural	-0.048	0.015	0.076	0.133	0.197	0.258
Moderately Rural	0.025	0.007	0.058	0.040	0.127	0.203
Rural	-0.154	-0.060	0.034	0.094	0.117	0.199
Extremely Rural	-0.022	0.080	0.131	0.225	0.296	0.357
Panel B: Match Effects						
Urban	0.019	0.017	0.018	0.018	0.013	0.022
Rural	0.040	0.063	0.111	0.137	0.180	0.191
Moderately Rural	0.008	0.002	0.031	0.022	0.065	0.089
Rural	0.039	0.085	0.141	0.107	0.154	0.161
Extremely Rural	0.070	0.106	0.168	0.218	0.247	0.300
Panel C: Monthly Total Cost (in Soles)						
% Base Wage	0.111	0.086	0.140	0.234	0.379	0.621
Mean Bonus per School	223	171	279	467	759	1,242
SD Bonus per School	220	407	576	839	1,184	1,698

- It's possible to close the urban-rural gap in TVA at a small cost
- Optimal policy induces teachers to sort on their comparative advantage

RDD $\Leftrightarrow$ Economic Model

- RD evidence is local in nature and may neglect indirect/equilibrium effects
- An estimated school choice model allows us to
  - Quantify aggregate effect of the policy reform
  - Characterize gains from alternative teacher assignments

$\Rightarrow$ RD variation is embedded in the model, enhancing its credibility

The equilibrium effect of a large-scale education reform (Khanna, JPE 2023)

Large-scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India

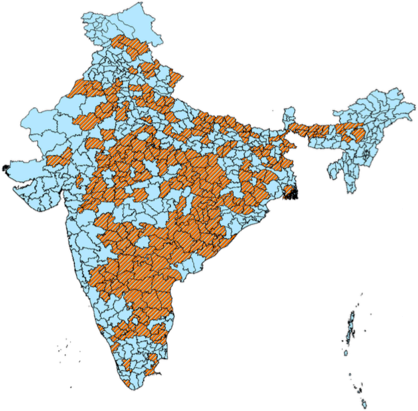
- RD+GE model of large-scale expansion of public schooling in India
- GE effects dampen the returns to skill by 33%
  - $\Downarrow$ Increase in the relative supply of skilled workers
  - $\Uparrow$ Adoption of skill-biased capital
- Large distributional effects from skilled to unskilled (young) workers
- Higher welfare due to lower costs of education and more economic output

The Schooling Expansion Policy

- Improve student access to and retention in primary education
- Donor assisted program giving funds to 271 (out of 600) districts in India
- 160,000 new schools, trained 1.1 million teachers, $\uparrow$ funds by 17-20%

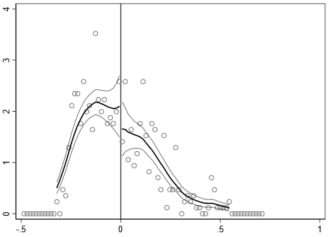
$\Rightarrow$ Any district below the national average female literacy rate was eligible

The District Primary Education Program (DPEP)

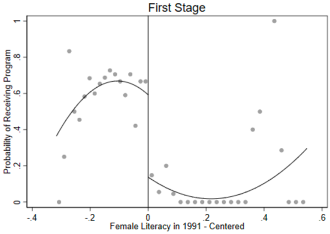


Orange and shaded districts received DPEP, whereas blue-unshaded districts did not.

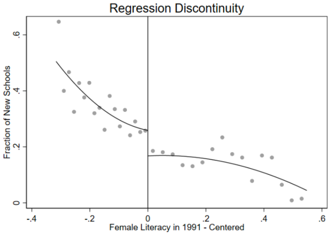
RD Results (I)



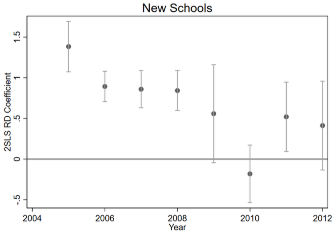
(a) District Level McCrary Test



(b) First Stage of DPEP



(c) Fraction of All Schools Built Post 1993



(d) Frac New Schools (2SLS) Over Time

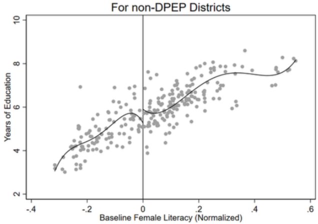
RD Results (II)

Panel A: First Stage				
Years of Education	Young	Old	Young	Old
RD Estimate	0.720	-0.0856	0.698	0.100
	(0.199)***	(0.218)	(0.173)***	(0.188)
Observations	10,175	11,293	14,277	16,007
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Finished Upper-primary	Young	Old	Young	Old
RD Estimate	0.0743	-0.0169	0.0733	0.000212
	(0.0197)***	(0.0184)	(0.0169)***	(0.0158)
Observations	9,045	7,729	10,175	9,920
Bandwidth selection procedure	CCT	CCT	I and K	I and K

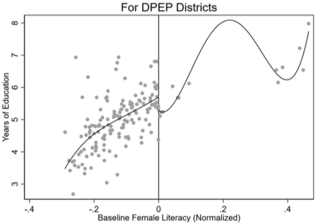
RD Results (III)

Panel B: Reduced Form				
Log Earnings	Young	Old	Young	Old
RD Estimate	0.112	-0.0114	0.145	0.0432
	(0.0312)***	(0.0372)	(0.0269)***	(0.0318)
Observations	10,175	11,293	14,277	16,007
Bandwidth selection procedure	CCT	CCT	I and K	I and K
Panel C: 2SLS IV-LATE Conventional Method Returns				
Log Earnings	Young	Old	Young	Old
Years of Education	0.155	0.129	0.208	0.442
	(0.0427)***	(0.303)	(0.0460)***	(0.666)
Observations	10,175	7,994	14,277	8,627
Bandwidth selection procedure	CCT	CCT	I and K	I and K

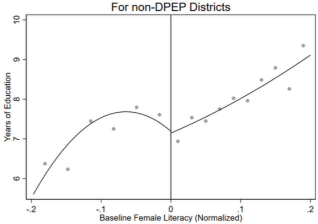
RD Checks: Imperfect Compliance and External Validity



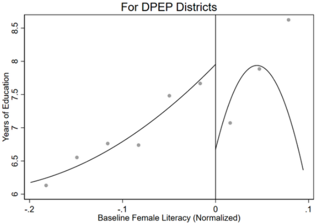
(a) Full sample, $D = 0$



(b) Full sample, $D = 1$



(c) Binned in bandwidth, $D = 0$



(d) Binned in bandwidth, $D = 1$

What is the Rationale for the Model?

- The ratio of 0.112 log earnings and 0.72 years: return of about 15.5%
- The 2SLS-LATE return is a weighted average of partial eq and GE effects

$$\underbrace{log\frac{w_{y,D=1}}{w_{y,D=0}}}_{\text{Avg earnings}} = \underbrace{\ell_{sy,D=1}}_{\text{Skilled}} \underbrace{log\frac{w_{sy,D=1}}{w_{sy,D=0}}}_{\neq 0} + \underbrace{\ell_{uy,D=1}}_{\text{Unskilled}} \underbrace{log\frac{w_{uy,D=1}}{w_{uy,D=0}}}_{\neq 0} + \underbrace{\Delta\ell_{sy}}_{\text{Compliers}} \underbrace{log\frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{ys,D=0}}$$

$\Rightarrow$ Need a model to unpack these two

Economic Production and Labor Supply

- District-level production function

$$Y_d = L_d^{\rho} K_d^{1-\rho}, \text{ where } L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}} \right)^{\frac{\sigma_E}{\sigma_E - 1}}$$

$\Rightarrow$ σ_E: elasticity of substitution across skill groups

$\Rightarrow$ θ_{sd}: skill productivity, with $\theta'_{sd}(k_{sd}) > 0$

- Aggregate labor supply by skill level s

$$L_{sd} = \left(\sum_a \psi_a l_{asd}^{\frac{\sigma_A - 1}{\sigma_A}} \right)^{\frac{\sigma_A}{\sigma_A - 1}}$$

$\Rightarrow$ σ_A: elasticity of substitution across age cohorts

$\Rightarrow$ ψ_a: productivity of a given cohort a

Wages and Returns to Skill

- Differences in average earnings across labor markets

$$\log\left(\frac{w_{asd}}{w_{a's'd'}}\right) = \underbrace{\log\left(\frac{\theta_{sd}}{\theta_{s'd'}}\right)}_{\text{productivity}} + \underbrace{\log\left(\frac{\psi_a}{\psi_{a'}}\right)}_{\text{cohort}}$$
$$+ \underbrace{\frac{1}{\sigma_E}\log\left(\frac{Y_d}{Y_{d'}}\right)}_{\text{output}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)\log\frac{L_{sd}}{L_{s'd'}}}_{\text{skill distribution}} \underbrace{-\frac{1}{\sigma_A}\log\frac{\ell_{asd}}{\ell_{a's'd'}}}_{\text{skill-cohort distribution}}$$

$\Rightarrow$ σ_A and σ_E shape the effect of labor supply on wages (across/within cohorts)

- Returns to skill depend on local labor market conditions

$$log\frac{w_{asd}}{w_{aud}} = log\frac{\theta_{sd}}{\theta_{ud}} + \left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right) log\frac{L_{sd}}{L_{ud}} - \frac{1}{\sigma_A} log\frac{\ell_{asd}}{\ell_{aud}} \equiv \beta_{asd}$$

Schooling Decision

- PDV of expected log earnings

$$\mathbb{E}[\log\ w_{aid}(s_{id})] = \mathbb{E}[\gamma_d + \gamma_a] + \mathbb{E}[\tilde{\beta}_{asd}]s_{id} + \log\ \epsilon_i$$

- Quadratic schooling costs

$$\log\ r_{id} + r_{id}s_{id} + \frac{1}{2}\Gamma s_{id}^2$$

- Ability ϵ_i is negatively correlated with marginal cost $r_{id} = -\Psi A_d + p_d + \eta_i$

- Where A_d is chosen by the district given government resources R_d

$$A_d(R_d, \mathbf{p_m}) = \mathbf{R_d} \prod_{\mathbf{m}} \left(\frac{\alpha_{\mathbf{m}}}{\mathbf{p_m}}\right)^{\alpha_{\mathbf{m}}}$$

Education Market Equilibrium

- Overall demand and supply of schooling in district d

$$S_d^{Dd} = \int \frac{\mathbb{E}[\tilde{\beta}_d] + \Psi A_d - p_d - \eta_i}{\Gamma} dH(\eta) = \frac{\mathbb{E}[\tilde{\beta}_d] + \Psi A_d - p_d - \bar{\eta}_d}{\Gamma},$$

$$S_d^{Sy} = \frac{\bar{\theta}_d}{z_{2d}}[p_d\bar{\theta}_d - \mathbb{E}_d(z_{1j}|z_{1j} < p_d\bar{\theta}_d)] + A_d$$

- Optimal years of education

$$S_d^* = \phi_1 \mathbb{E}[\tilde{\beta}_d] + \phi_2 R_d - \frac{\eta_d}{\Gamma} \ ,$$

$\Rightarrow$ Schooling depends on expected returns + government spending - costs

Using the RD to Estimate the Model

- Define treatment as $D_d \in \{0, 1\}$, such that $\phi D_d \equiv \phi_1 \mathbb{E}[\tilde{\beta}_{asd}] + \phi_2 R_d$

$\Rightarrow$ In the neighborhood of the policy cutoff: $\mathbb{E}[\eta_d \mid D_d = 0] = \mathbb{E}[\eta_d \mid D_d = 1]$

- Skill premium of older cohort o isolates GE effect of the policy

$$\underbrace{log\frac{w_{so,D=1}}{w_{so,D=0}} - log\frac{w_{uo,D=1}}{w_{uo,D=0}}}_{\text{GE effects on all cohorts}} = \underbrace{\left(log\frac{\theta_{s,D=1}}{\theta_{u,D=1}} - log\frac{\theta_{s,D=0}}{\theta_{u,D=0}}\right)}_{\text{Skill biased capital}} + \underbrace{\left(\frac{1}{\sigma_A} - \frac{1}{\sigma_E}\right)\left[log\frac{L_{s,D=1}}{L_{u,D=1}} - log\frac{L_{s,D=0}}{L_{u,D=0}}\right]}_{\text{Aggregate skill distribution}}$$

- Additional GE effect on young y after differencing out the skill premium of o

$$\underbrace{\left[log\frac{w_{sy,D=1}}{w_{sy,D=0}} - log\frac{w_{uy,D=1}}{w_{uy,D=0}}\right] - \left[log\frac{w_{so,D=1}}{w_{so,D=0}} - log\frac{w_{uo,D=1}}{w_{uo,D=0}}\right]}_{\text{Additional GE on young}} = \underbrace{-\frac{1}{\sigma_A}\left[log\frac{\ell_{ys,D=1}}{\ell_{yu,D=1}} - log\frac{\ell_{ys,D=0}}{\ell_{yu,D=0}}\right]}_{\text{Age specific skill distribution}}$$

Using the RD to Estimate the Model

- Use RD on young cohort to estimate returns to skill in $D = 1$ and $D = 0$

$$log\frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=1}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=1}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=0}}{w_{uy,D=0}}}_{\beta_{ys,D=0}}$$

$$log\frac{w_{y,D=1}}{w_{y,D=0}} = \ell_{sy,D=0}log\frac{w_{sy,D=1}}{w_{sy,D=0}} + \ell_{uy,D=0}log\frac{w_{uy,D=1}}{w_{uy,D=0}} + \Delta\ell_{sy}\underbrace{log\frac{w_{sy,D=1}}{w_{uy,D=1}}}_{\beta_{ys,D=1}}$$

- Average increase in years of education across cutoff

$$\begin{aligned}\Delta S &= (\ell_{sy,D=1}s_1 + \ell_{uy,D=1}s_0) - (\ell_{sy,D=0}s_1 + \ell_{sy,D=0}s_0) \\ &= \Delta\ell_{sy}s_1 + \Delta\ell_{uy}s_0 = \Delta\ell_{sy}(s_1 - s_0)\ ,\end{aligned}$$

Returns to Skill without and with GE

	Fraction Switched	Change in Returns $\Delta\tilde{\beta}$	
Estimate	0.171	-0.066	
SE	(0.045)	(0.030)	
	Returns without GE $\tilde{\beta}_{y,D=0}$	Returns with GE $\tilde{\beta}_{y,D=1}$	% Change in returns
Estimate	0.199	0.134	-33%
Bootstrapped p-val	[0.00]	[0.00]	
	Change for older cohorts	Additional on Young	% Change on young
$\Delta\tilde{\beta}$	-0.0097	-0.057	85.3%

Outcomes and Economic Benefits

- Labor market benefits for young always-skilled (infra-marginal): $log\frac{w_{ys,D=1}}{w_{ys,D=0}}$
  - $\uparrow$ Reductions in education costs
  - $\downarrow$ GE effects for all cohorts + GE effects for the young
- Labor market benefit for always unskilled (old and young): $log\frac{w_{au,D=1}}{w_{au,D=0}}$
- For the young compliers:

$$log\frac{w_{ys,D=1}}{w_{yu,D=0}} = log\frac{w_{ys,D=0}}{w_{yu,D=0}} + log\frac{w_{ys,D=1}}{w_{ys,D=0}} = \beta_{ys,D=0} + log\frac{w_{ys,D=1}}{w_{ys,D=0}}$$

Labor Market Benefits

Change in Yearly Labor Market Benefits for

(1) Young, Induced into getting more Skill

With GE	Without GE	% Change	Fraction of cohort
0.120	0.157	-23.8%	0.17

(2) Always Skilled (Young)

With GE	Without GE	% Change	Fraction of cohort
-0.037	0	-	0.39

(3) Always Unskilled (Young)

With GE	Without GE	% Change	Fraction of cohort
0.014	0	-	0.44

Transfer in Yearly Benefits from Skilled to Unskilled

Among Old with GE	Among Old without GE	Among Young with GE	Among Young without GE
0.007	0	0.052	0

Change in Lifetime Welfare for Induced Students

Costs	Benefits	Net	% Change (due to GE)
5.15	6.55	1.4	-23.8%

RDD ⇔ Economic Model

- Minimal structure around RDD to unpack channels of policy impact

⇒ Marschak's Maxim: only a subset of structural parameters are needed

- This puts less constraints on the data by focusing on a few key parameters

- But difficult to extrapolate away from the RD cutoff without further structure