Lecture 14: Regression Discontinuity II

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

March 31, 2021

Last time:

- Basics of sharp regression discontinuity (RD) designs.
- Robust approximation of the (univariate) CEF to make predictions at cutpoint.

Today:

- "Fuzzy" RD.
- "Kink" design.
- Multiway & geographic RD.
- Threats to validity of RD studies.

“Fuzzy” RD

- Forcing variable, X_i, may not *deterministically* affect D_i.
- Rather induces discontinuous change in expected value at x_0.
- Such cases provide “fuzzy” RD designs.
- Then, $1(X_i \geq x_0)$ is an *instrument* for the treatment.

Motivating Examples

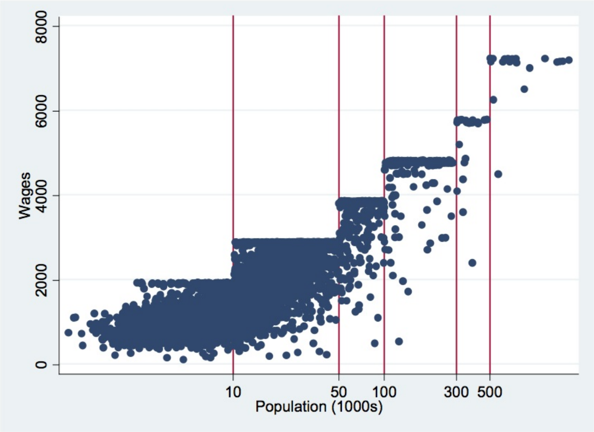


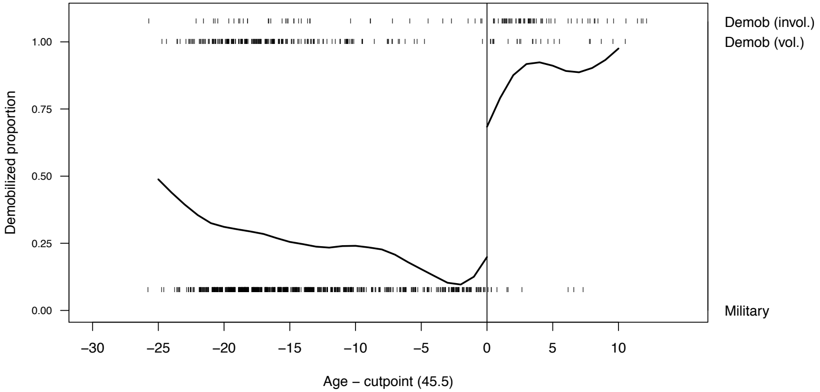
FIGURE 1: LEGISLATORS' SALARIES BY POPULATION

Notes: Figure shows legislators' salaries by population (in log scale). The vertical lines denote the various cutoff points.

(Ferraz and Finan, 2011)

Motivating Examples

Figure 2: **Demobilization rates by age (centered at eligibility cut-point of 45.5 years old)**



(Samii, 2013)

"Fuzzy" RD

- Formally

$$\Pr[D_i = 1|X_i = x] = \begin{cases} g_1(x) & \text{if } x \geq x_0 \\ g_0(x) & \text{if } x < x_0 \end{cases}, \text{ where } g_1(x_0) \neq g_0(x_0).$$

- $g_d(x_i)$ is conditional mean of D_i.
- $g_1(x_0) \neq g_0(x_0)$ implies a discontinuity at x_0.
- Without loss of generality, suppose $g_1(x_0) > g_0(x_0)$. Then,

$$\mathrm{E}[D_i|X_i] = \Pr[D_i = 1|X_i] = g_0(X_i) + [g_1(X_i) - g_0(X_i)]T_i,$$

where $T_i = 1(X_i \geq x_o)$.

“Fuzzy” RD

Consider the conditions for IV identification:

“Fuzzy” RD

Consider the conditions for IV identification:

- T_i and affects D_i in a non-trivial way (valid first stage).

"Fuzzy" RD

Consider the conditions for IV identification:

- T_i and affects D_i in a non-trivial way (valid first stage).
- T_i is a deterministic functions of X_i. (unconfounded instrument)

“Fuzzy” RD

Consider the conditions for IV identification:

- T_i and affects D_i in a non-trivial way (valid first stage).
- T_i is a deterministic functions of X_i. (unconfounded instrument)
- Need to assume that nothing else changes discontinuously at x_0 other than T_i and then then D_i and Y_i as functions of T_i. (exclusion).

"Fuzzy" RD

Consider the conditions for IV identification:

- T_i and affects D_i in a non-trivial way (valid first stage).
- T_i is a deterministic functions of X_i. (unconfounded instrument)
- Need to assume that nothing else changes discontinuously at x_0 other than T_i and then then D_i and Y_i as functions of T_i. (exclusion).
- Interpretation as LATE follows from monotonicity—namely, that T_i cannot cause both some units to *take up* while others to *reject* the treatment.

"Fuzzy" RD

Consider the conditions for IV identification:

- T_i and affects D_i in a non-trivial way (valid first stage).
- T_i is a deterministic functions of X_i. (unconfounded instrument)
- Need to assume that nothing else changes discontinuously at x_0 other than T_i and then then D_i and Y_i as functions of T_i. (exclusion).
- Interpretation as LATE follows from monotonicity—namely, that T_i cannot cause both some units to *take up* while others to *reject* the treatment.
- With these conditions met, T_i is a valid instrument for estimating the effect of D_i on Y_i.

"Fuzzy" RD

Polynomial approximation

- For outcomes, we have as before,

$$\begin{aligned} Y_i = &\alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \ldots + \beta_{0p}\tilde{X}_i^p \\ &+ \rho D_i + \beta_{11}^* D_i\tilde{X}_i + \beta_{12}^* D_i\tilde{X}_i^2 + \ldots + \beta_{1p}^* D_i\tilde{X}_i^p + \eta_i, \end{aligned}$$

  where $\tilde{X}_i = X_i - x_0$.

- Similarly, for the treatment we have,

$$\begin{aligned} D_i = &\gamma_{00} + \gamma_{01}\tilde{X}_i + \gamma_{02}\tilde{X}_i^2 + \ldots + \gamma_{0p}\tilde{X}_i^p \\ &+ \pi T_i + \gamma_{11}^* T_i\tilde{X}_i + \gamma_{12}^* T_i\tilde{X}_i^2 + \ldots + \gamma_{1p}^* T_i\tilde{X}_i^p + \nu_i, \end{aligned} \tag{1}$$

  where ν_i accounts for any residual nonlinearity to ensure $D_i = 0, 1$.

- As before, we will use local linear approximation (so, $p = 1$) within a bandwidth.

“Fuzzy” RD

Local linear approximation approach extended to IV:

"Fuzzy" RD

Local linear approximation approach extended to IV:

- Within bandwidth, Δ, for reduced form we have,

$$\lim_{\Delta \to 0} \mathrm{E}\left[Y_i | x_0 \leq X_i < x_0 + \Delta\right] - \mathrm{E}\left[Y_i | x_0 - \Delta < X_i < x_0\right] = \rho \pi$$

“Fuzzy” RD

Local linear approximation approach extended to IV:

- Within bandwidth, Δ, for reduced form we have,

$$\lim_{\Delta \to 0} \mathrm{E}\left[Y_i | x_0 \leq X_i < x_0 + \Delta\right] - \mathrm{E}\left[Y_i | x_0 - \Delta < X_i < x_0\right] = \rho \pi$$

- For first stage we have,

$$\lim_{\Delta \to 0} \mathrm{E}\left[D_i | x_0 \leq X_i < x_0 + \Delta\right] - \mathrm{E}\left[D_i | x_0 - \Delta < X_i < x_0\right] = \pi$$

“Fuzzy” RD

Local linear approximation approach extended to IV:

- Within bandwidth, Δ, for reduced form we have,

$$\lim_{\Delta \to 0} \mathrm{E}[Y_i | x_0 \leq X_i < x_0 + \Delta] - \mathrm{E}[Y_i | x_0 - \Delta < X_i < x_0] = \rho \pi$$

- For first stage we have,

$$\lim_{\Delta \to 0} \mathrm{E}[D_i | x_0 \leq X_i < x_0 + \Delta] - \mathrm{E}[D_i | x_0 - \Delta < X_i < x_0] = \pi$$

- “Wald estimator” for average treatment effect:

$$\lim_{\Delta \to 0} \frac{\mathrm{E}[Y_i | x_0 \leq X_i < x_0 + \Delta] - \mathrm{E}[Y_i | x_0 - \Delta < X_i < x_0]}{\mathrm{E}[D_i | x_0 \leq X_i < x_0 + \Delta] - \mathrm{E}[D_i | x_0 - \Delta < X_i < x_0]} = \rho.$$

- `rdrobust` function in Stata estimates this directly using analogues of Δ^o that we saw for sharp RD.

"Fuzzy" RD

- Analogous regression implementation,

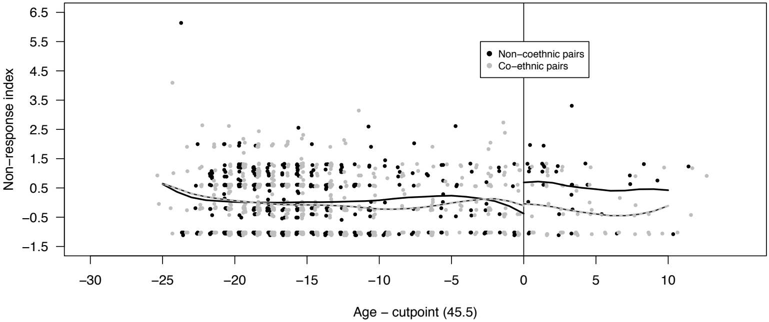
$$Y_i = \beta_0 + \rho D_i + \beta_1 \tilde{X}_i + \beta_2 D_i \tilde{X}_i + \eta_i,$$

$$D_i = \alpha_0 + \pi T_i + \alpha_1 \tilde{X}_i + \alpha_2 T_i \tilde{X}_i + \eta_i,$$

  for $\{i : X_i \in [x_0 - \Delta, x_0 + \Delta]\}$.
- 2SLS with T_i and $T_i \tilde{X}_i$ as instruments for D_i and $D_i \tilde{X}_i$
- Apply 2SLS inference with adjustments for bandwidth selection, with further adjustments as needed for heteroskedasticity or clustering. Again see `rdrobust` package.

“Fuzzy” RD

Figure 3: **Effects on prejudice**



The figure plots respondents’ non-response index scores over their age, with age centered at the 45.5 cut-point. The gray dots are for respondents that had coethnic enumerators, and the black dots are for respondents with non-coethnic enumerators. The gray and black curves are from local linear regression smoother fits (5-year bandwidth) to the coethnic and non-coethnic points, respectively. The local linear regression smoothers are fit on either side of the cut-point, demarcated by the vertical black line.

(Samii, 2013–NB: reduced form shown)

“Fuzzy” RD

Table 1: **Effects on prejudice**

	(1) Non-resp. index	(2) Non-resp. index	(3) Non-resp. index	(4) Non-resp. index
Integrated	-0.25	-0.00	-0.36	-0.04
	(0.88)	(0.58)	(0.48)	(0.67)
IntegratedXNon-coeth.	-1.04**	-1.00**	-0.74**	-1.11**
	(0.53)	(0.41)	(0.32)	(0.48)
Non-coeth. pair	0.93***	0.96***	0.94***	0.91***
	(0.33)	(0.28)	(0.23)	(0.33)
Age-45.5	-0.16	-0.05	-0.21	-0.05
	(0.17)	(0.12)	(0.15)	(0.13)
(Age<45.5)X(Age-45.5)	0.08	-0.03	0.08	-0.04
	(0.10)	(0.10)	(0.16)	(0.12)
$(\text{Age-45.5})^2$			0.02	
			(0.02)	
$(\text{Age}<45.5)\text{X}(\text{Age-45.5})^2$			-0.04*	
			(0.02)	
Constant	0.17	-0.06	0.21	-0.06
	(0.50)	(0.37)	(0.34)	(0.42)
Observations	141	161	265	150

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Weighted two-stage least squares estimates with standard error estimates that account for clustering by interview location/barrack. Model (1) uses an Imbens-Kalyanaraman optimal bandwidth of 4 years, and models (2) and (3) use 5-year and 10-year bandwidths, respectively. Model (4) uses the 5-year bandwidth and restricts the sample to Tutsi respondents.

(Samii, 2013)

“Fuzzy” RD

- Straightforward generalization to non-binary treatments (MHE, p. 263).
- Estimate the same way, only now D_i is multivalued/continuous.
- Extension to multiple cut-points:
  - Dummy variables for the different cut-points (general).
  - Step function predictions as instrument (Ferraz & Finan example).

“Fuzzy” RD

Let the function f_i^{cap} denote the maximum wage a legislator in municipality i can receive, specifically,

$$
\begin{aligned}
f_i^{cap} &= 1927.1 \times 1\{P_i \leq 10,000\} + 2890.6 \times 1\{P_i \in (10,000, 50,000]\} \\
&+ 3854.2 \times 1\{P_i \in (50,000, 100,000]\} + 4817.7 \times 1\{P_i \in (100,000, 300,000]\} \\
&+ 5781.2 \times 1\{P_i \in (300,000, 500,000]\} + 7226.6 \times 1\{P_i > 500,000\}
\end{aligned}
$$

where P_i denotes the population of municipality i. We estimate the following TSLS model:

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 w_i + g(P_i) + x_i'\delta + \varepsilon_i \qquad (5) \\
w_i &= \alpha_0 + \alpha_1 f_i^{cap} + g(P_i) + x_i'\theta + \nu_i
\end{aligned}
$$

where the function $g(\cdot)$ is a flexible function of population.

(Ferraz and Finan, 2011)

“Fuzzy” RD

Table 5: The Effects of Wages on Legislative Performance

Dependent variable:	Number of Bills Submitted		Number of Bills Approved		Functioning Commission		Public events	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: IV estimates								
Wages	0.807	0.672	0.584	0.515	0.065	0.062	0.074	0.06
	[0.238]***	[0.230]***	[0.125]***	[0.122]***	[0.025]***	[0.026]**	[0.033]**	[0.034]*
Panel B: Reduced-form estimates								
Salary caps	0.72	0.621	0.487	0.429	0.043	0.04	0.034	0.026
	[0.220]***	[0.211]***	[0.109]***	[0.105]***	[0.020]**	[0.021]*	[0.029]	[0.029]
R-squared	0.18	0.2	0.15	0.17	0.02	0.03	0.03	0.04
Municipal characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Observations	3544	3544	3544	3544	5093	5093	5093	5093

Notes: The table reports the TSLS and reduced-form estimates for the effects of wages on legislative performance for the 2005/2008 legislature. Municipal Characteristics include Log household income per capita, % urban population, Gini coefficient, % households with energy, % literate population, average wage in private and public sector in municipality, the number of hours the legislature functions per week and assistants per legislator. All regressions include a 3rd order polynomial in population along with a quadratic spline on the first cutoff. Wages and salary caps have been divided by 1000. * indicates statistical significance at the 10% level, ** at the 5% level and *** at the 1% level. Robust standard errors are reported in brackets. The excluded instrument is the salary caps.

(Ferraz and Finan, 2011)

"Fuzzy" RD

Table 7. The Effects of Wages on Political Selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Dependent variable	Years of schooling	No formal schooling	Some primary school	Primary school	Some high school	High school	Some college	College	High skilled occupation
Wages	0.495	-0.023	-0.016	-0.014	0.009	0.004	0.021	0.017	0.043
	[0.155]***	[0.008]***	[0.015]	[0.012]	[0.008]	[0.016]	[0.007]***	[0.013]	[0.018]**
Observations	5091	5093	5093	5093	5093	5093	5093	5093	5093
Panel B: Dependent variable	Average terms of experience	1 term of experience	2 terms of experience	3 terms of experience	4 terms of experience	5 terms of experience	6 terms of experience	7 terms of experience	Male
Wages	0.154	-0.047	-0.007	0.03	0.021	0.005	0.003	0.000	-0.005
	[0.056]***	[0.019]**	[0.015]	[0.012]**	[0.008]**	[0.005]	[0.002]	[0.003]	[0.010]
Observations	5093	5092	5092	5093	5092	5093	5093	5093	5093
Municipal characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports the TSLS estimates of the effects of wages on political selection of 2005/2008 legislature. Municipal Characteristics include Log household income per capita, % urban population, Gini coefficient, % households with energy, % literate population, average wage in private and public sector in municipality, the number of hours the legislature functions per week and assistants per legislator. All regressions include a 3rd order polynomial in population along with a quadratic spline on the first cutoff. Wages have been divided by 1000. * indicates statistical significance at the 10% level, ** at the 5% level and *** at the 1% level. Robust standard errors are reported in brackets. The excluded instrument is the salary caps.

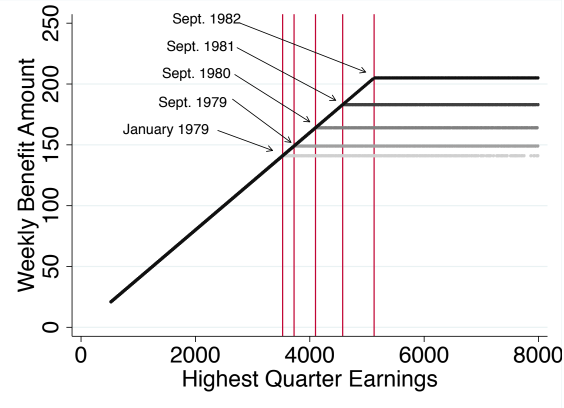
(Ferraz and Finan, 2011)

Kink design

- RD and Fuzzy RD exploit discontinuities in treatment *level* as one moves across a cut point in the forcing variable ($E[D|X < x_o]$ to $E[D|X \geq x_o]$).
- Another type discontinuity is in the treatment *slope* around as one moves across the cut point—a "kink."

Kink design

Figure 2. Louisiana: Schedule of UI Weekly Benefit Amount, jan1979-Dec1983

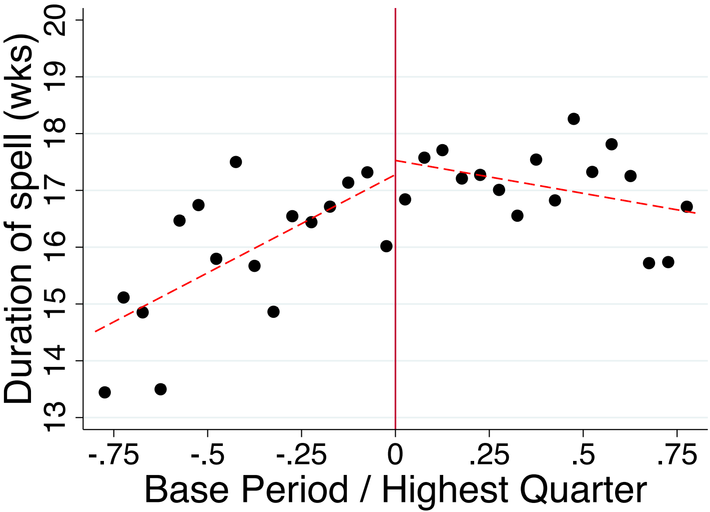


Source: Louisiana Revised Statutes RS 23:1592 and yearly *Significant Provisions of State Unemployment Insurance Laws* 1976 to 1984, Dpt of Labor, Employment & Training Administration.
Note: The graph shows the evolution of the schedule of the weekly benefit amount (WBA) in nominal terms as a deterministic and kinked function of the highest quarter of earnings in Louisiana. The schedule applies based on the date the UI claim was filed, so that a change in the maximum weekly benefit amount does not affect the weekly benefit amount of ongoing spells.

(Landais, 2014)

Kink design

B. Effect of potential duration



(Landais, 2014)

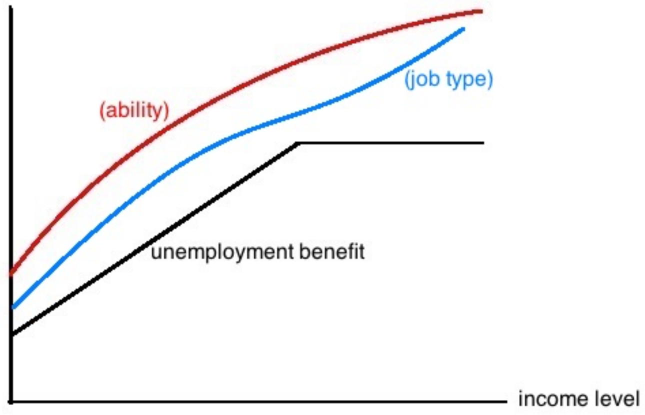
Kink design

- Identification of a causal effect requires smoothness in $\partial Y/\partial X$ and in confounders over X.
- E.g., following Nielsen et al. (2010) suppose

$$Y_i = \tau D_i + g(X_i) + \eta_i,$$

  where $g(\cdot)$ is smooth, and D_i is a deterministic function of X_i.
- Suppose there is a kink in D_i at $X_i = x_0$ but that $g(\cdot)$ and η_i vary smoothly wrt X_i at x_0.

Kink design



Kink design

- A kink in D_i:

$$d_1 \equiv \lim_{x \downarrow x_0} \frac{\partial D_i}{\partial x} \neq \lim_{x \uparrow x_0} \frac{\partial D_i}{\partial x} \equiv d_0.$$

  but smoothness in $g(\cdot)$ and η_i:

$$\lim_{x \uparrow x_0} \frac{\partial g(x)}{\partial x} = \lim_{x \downarrow x_0} \frac{\partial g(x)}{\partial x} \text{ and } \lim_{x \uparrow x_0} \frac{\partial \mathrm{E}\left[\eta_i | X_i = x\right]}{\partial x} = \lim_{x \downarrow x_0} \frac{\partial \mathrm{E}\left[\eta_i | X_i = x\right]}{\partial x}.$$

- Then,

$$\begin{aligned}
&\frac{\lim_{x \downarrow x_0} \frac{\partial \mathrm{E}[Y_i|X_i=x]}{\partial x} - \lim_{x \uparrow x_0} \frac{\partial \mathrm{E}[Y_i|X_i=x]}{\partial x}}{d_1 - d_0} \\
&= \frac{\lim_{x \downarrow x_0} \frac{\partial(\tau D_i + g(x) + \mathrm{E}[\eta_i|X_i=x])}{\partial x} - \lim_{x \uparrow x_0} \frac{\partial(\tau D_i + g(x) + \mathrm{E}[\eta_i|X_i=x])}{\partial x}}{d_1 - d_0} \\
&= \frac{\tau(d_1 - d_0)}{d_1 - d_0} \\
&= \tau.
\end{aligned}$$

Kink design

- If policy rule is deterministic, then you can get $d_1 - d_0$ from the rule. Otherwise, a local linear specification for i near x_0:

$$D_i = \gamma_0 + \gamma_1 X_i + \delta X_i I(X_i > x_0) + \varepsilon_i,$$

where $\hat{\delta}$ estimates $d_1 - d_0$.

- To estimate the kink in the outcome, again local linear specification:

$$Y_i = \lambda_0 + \lambda_1 X_i + \beta X_i I(X_i > x_0) + \nu_i,$$

- Regression kink estimator is $\hat{\beta}/(d_1 - d_0)$ if you know $d_1 - d_0$ or $\hat{\beta}/\hat{\delta}$ if you have to estimate.
- See Card et al. (2012) for generalizations.
- Implementation with `rdrobust`.

Multiway & Geographic RD

- Multiway RD looks at assignment to treatments based on multiple indices (Papay et al., 2011).
- Examples:
  - Block grants to communities based on multiple characteristics.
  - Geographic boundaries.

Multiway & Geographic RD

TABLE 1.1: VILLAGE BLOCK GRANT ALLOCATIONS

		Population		
		Large (>=700 persons)	Medium (300-699 persons)	Small (<299 persons)
Conflict Intensity	High	170,000,000 (US$19,000)[6]	150,000,000 (US$17,000)	120,000,000 (US$13,000)
	Medium	120,000,000 (US$13,000)	100,000,000 (US$11,000)	80,000,000 (US$8,900)
	Low	80,000,000 (US$8,900)	70,000,000 (US$7,800)	60,000,000 (US$6,700)

(Barron et al. 2009)

Multiway & Geographic RD

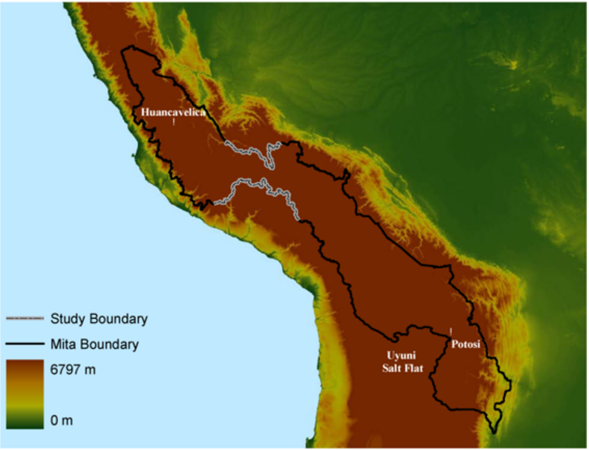


FIGURE 1.—The *mita* boundary is in black and the study boundary in light gray. Districts falling inside the contiguous area formed by the *mita* boundary contributed to the *mita*. Elevation is shown in the background.

(Dell 2011)

Multiway & Geographic RD

- Suppose 2 forcing variables each with one cut point.
- For forcing variable j, define,

$$D_{ji} = 1(X_{ji} > x_{0j}),$$

for $j = 1, 2$. This establishes $2^2 = 4$ treatment conditions:

Condition	D_{1i}	D_{2i}
A	1	0
B	1	1
C	0	0
D	0	1

Multiway & Geographic RD

TABLE 1.1: VILLAGE BLOCK GRANT ALLOCATIONS

		Population		
		Large (>=700 persons)	Medium (300-699 persons)	Small (<299 persons)
Conflict Intensity	High	170,000,000 (US$19,000)[6]	150,000,000 (US$17,000)	120,000,000 (US$13,000)
	Medium	120,000,000 (US$13,000)	100,000,000 (US$11,000)	80,000,000 (US$8,900)
	Low	80,000,000 (US$8,900)	70,000,000 (US$7,800)	60,000,000 (US$6,700)

Damages

Population

Multiway & Geographic RD

- The conditional means as one approaches x_{0j} from below and above are,

$$\mu_{-j} = \lim_{\Delta \to 0} \mathrm{E}\left[Y_i | x_{0j} - \Delta < X_{ji} < x_{0j}\right]$$

  and

$$\mu_{+j} = \lim_{\Delta \to 0} \mathrm{E}\left[Y_i | x_{0j} \leq X_{ji} < x_{0j} + \Delta\right]$$

- Expected value marginalizes over boundary interval (since effects can vary along the boundary).
- Effects based on contrasts between μ_{-1}, μ_{+1}, μ_{-2}, μ_{+2}.
- Requires modeling conditional means below and above the different cut points.

Multiway & Geographic RD

- Papay et al. use local linear approximation for 2-way RD:

$$
\begin{aligned}
Y_i = & \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{1i} D_{2i} + \beta_4 \tilde{X}_{1i} + \beta_5 \tilde{X}_{2i} \\
& + \beta_6 (\tilde{X}_{1i} \tilde{X}_{2i}) + \beta_7 (\tilde{X}_{1i} D_{1i}) + \beta_8 (\tilde{X}_{2i} D_{2i}) \\
& + \beta_9 (\tilde{X}_{1i} D_{2i}) + \beta_{10} (\tilde{X}_{2i} D_{1i}) + \beta_{11} (\tilde{X}_{1i} \tilde{X}_{2i} D_{1i}) + \beta_{12} (\tilde{X}_{1i} \tilde{X}_{2i} D_{2i}) \\
& + \beta_{13} (\tilde{X}_{1i} D_{1i} D_{2i}) + \beta_{14} (\tilde{X}_{2i} D_{1i} D_{2i}) + \beta_{15} (\tilde{X}_{1i} X_{2i} D_{1i} D_{2i}),
\end{aligned}
$$

  where i within a bandwidth of the cutting lines.
- Specifies 4 response surfaces.

Multiway & Geographic RD

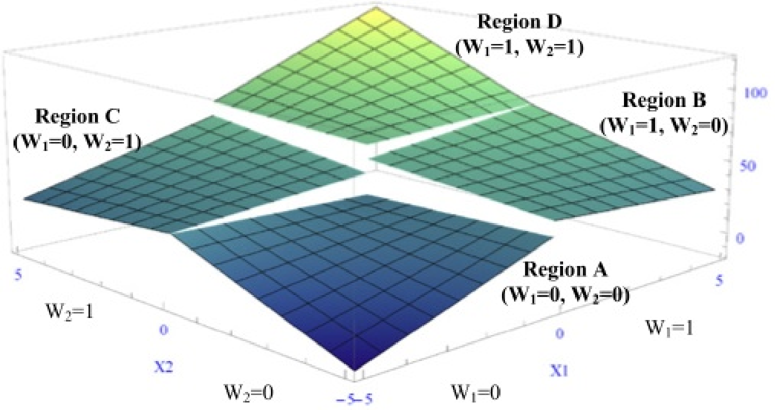


Fig. 1. Hypothetical population representation of the four surfaces (A, B, C, and D) from Eq. (1) defined by the two cut scores, $X_1 = 0$ and $X_2 = 0$.

(Papay et al. 2011)

Multiway & Geographic RD

- This regression estimates how causal effects differ over the edges ($\tilde{X}_{ji} D_{ki}$ terms).
- We can construct different kinds of average by applying different amounts of weight to different points along the edges.
- Papay et al. use cross-validation for bandwidth selection.

Multiway & Geographic RD

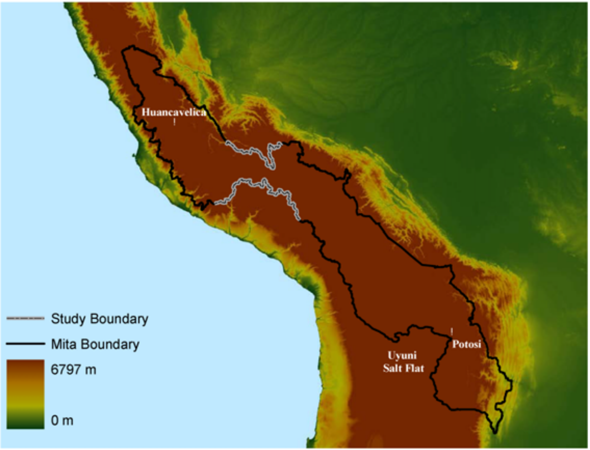


FIGURE 1.—The *mita* boundary is in black and the study boundary in light gray. Districts falling inside the contiguous area formed by the *mita* boundary contributed to the *mita*. Elevation is shown in the background.

(Dell 2011)

Multiway & Geographic RD

$$(1) \qquad c_{idb} = \alpha + \gamma mita_d + X'_{id}\beta + f(\text{geographic location}_d) + \boldsymbol{\phi}_b + \varepsilon_{idb},$$

where c_{idb} is the outcome variable of interest for observation i in district d along segment b of the *mita* boundary, and $mita_d$ is an indicator equal to 1 if district d contributed to the *mita* and equal to 0 otherwise; X_{id} is a vector of covariates that includes the mean area weighted elevation and slope for district d and (in regressions with equivalent household consumption on the left-hand side) demographic variables giving the number of infants, children, and adults in the household; $f(\text{geographic location}_d)$ is the RD polynomial, which controls for smooth functions of geographic location. Various forms will be explored. Finally, $\boldsymbol{\phi}_b$ is a set of boundary segment fixed effects that denote which of four equal length segments of the boundary is the closest to the observation's district capital.[10] To be conservative, all analysis excludes metropol-

(Dell 2011)

For $f(.)$:

- Polynomial in lat-lon coordinates and their interaction.
- Polynomial in distance to a key point.
- Polynomial in distance to nearest boundary point.

Multiway & Geographic RD

- Specification ignores effect heterogeneity along the border.
- Different ways to incorporate that into analysis.
  - Interactions with the segment fixed effects.
  - Estimates for specific boundary points, defining kernel weights in terms of distance to such points (Keele & Titiunik, 2015).
  - Weighting to account for variable population density along boundary (marginalizes over boundary).
- In geo-RD, sorting and spillover across boundary a concern.
  - "Donut RD" as a robustness check (Barreca et al. 2011, *QJE*).
  - Matching to address confounding due to sorting (Keele et al. 2015, *JRSS A*).

Multiway & Geographic RD

234 L. Keele, R. Titiunik and J. R. Zubizarreta

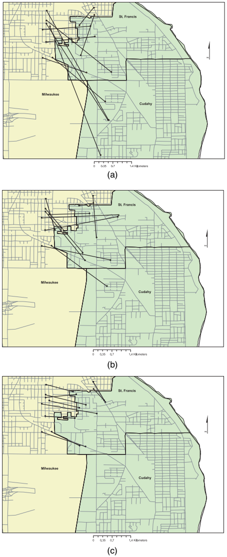


Fig. 3. 10 pairs of matches randomly sampled from legislative district exact match I: (a) design 1, covariates-only match; (b) design 2, distance-only match; (c) design 3, covariates and distance match

Matching to address confounding due to sorting across boundary.

Remarks

RD popular method for good reason:

Remarks

RD popular method for good reason:

- "Half-way house" between experiments and observational studies: assignment mechanism *known* though not *randomized*.

Remarks

RD popular method for good reason:

- "Half-way house" between experiments and observational studies: assignment mechanism *known* though not *randomized*.
- Second-best design when an experiment is infeasible.

Remarks

RD popular method for good reason:

- "Half-way house" between experiments and observational studies: assignment mechanism *known* though not *randomized*.
- Second-best design when an experiment is infeasible.
- Graphical underpinnings make it especially transparent.

Remarks

RD popular method for good reason:

- "Half-way house" between experiments and observational studies: assignment mechanism *known* though not *randomized*.
- Second-best design when an experiment is infeasible.
- Graphical underpinnings make it especially transparent.
- Green et al. (2009) test RD estimates against an experimental benchmark:
  - Local linear approach does well when the RD design is legit.

Remarks

There are pitfalls though:

- Researchers may be too cavalier in assuming identifying assumptions hold (see next slide).
- RD LATE may not answer your question — what counterfactual is being considered?

Remarks

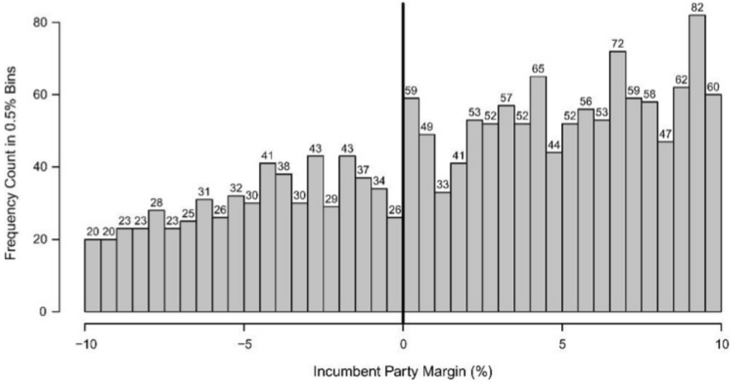
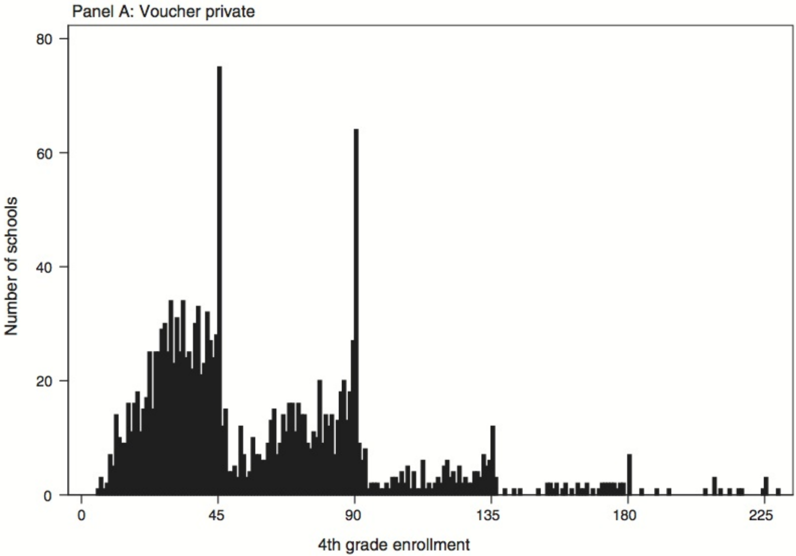


Fig. 1 Histogram of the incumbent party's margin in U.S. House elections (bin width = 0.5%).

(Caughy and Sekhon 2011)

Remarks



(Urquiola and Verhoogen 2009)

Remarks

Conditioning strategies may help:

- Match or work with within covariate strata for which identification holds.
- Use FE to account for subpopulation differences.
- Extension of local linear regression approach with kernel based methods for controlling for covariates (Froelich 2007).

Remarks

Table 1: p-values from placebo tests in Caughey and Sekhon (2011) with and without controlling for incumbency. These tests cover all those with a reported imbalance in Caughey and Sekhon (2011).

Dependent Variable	Original Specification	Including Dem Win $t-1$
Democratic Win $t-1$	.00	–
Democratic % Vote $t-1$	.10	.33
Democratic % Margin $t-1$	.03	.58
Incumbent D1 Nominate	.00	.60
Democratic Incumb in Race	.00	.58
Republican Incumb in Race	.00	.44
Democratic # Previous Terms	.08	.74
Republican # Previous Terms	.00	.10
Democratic Experience Adv	.00	.70
Republican Experience Adv	.00	.31
Partisan Swing	.00	.24
CQ Rating	.00	.47
Democratic Spending %	.01	.22
Democratic Donation %	.07	.53

NOTE: Cell entries are p-values for the variable *Democratic Win t* from linear regressions on the set of races in a 0.5 point window, with robust standard errors. In the column labeled Original Specification the only regressor is *Democratic Win t*. In the column labeled Including *Democratic Win t−1* the two regressors are *Democratic Win t* and *Democratic Win t−1*. For full variable definitions see Caughey and Sekhon (2011).

(Eggers et al. 2014)

Remarks

Nature of the LATE (what counterfactuals?):

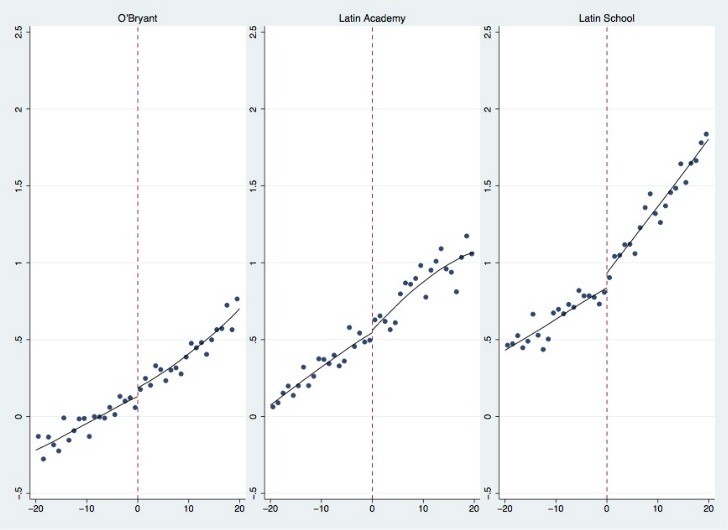


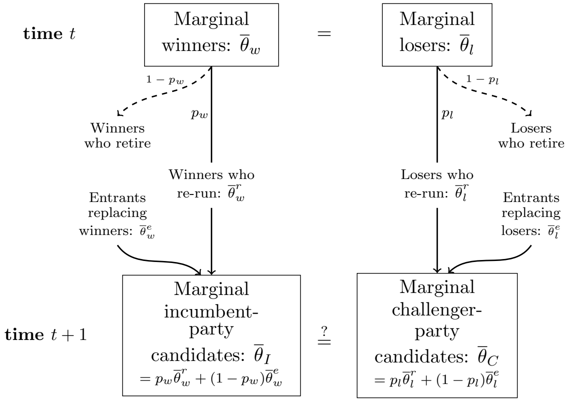
Figure 16. SAT Scores for 7th (2000-2005) and 9th (2001-2006) Grade Applicants in Boston

(Abdulkadiroglu, Angrist, and Pathak, 2011)

Remarks

Nature of the LATE (what counterfactuals?):

Figure 1: Channels for quality-based incumbency effects



NOTE: Marginal winners and marginal losers have the same average quality ($\overline{\theta}_w = \overline{\theta}_l$), but the candidates who compete in the next election for the incumbent party and the challenger party may differ in quality depending on the proportion of marginal winners and losers who run in the next election (p_w and p_l), the quality of these re-running candidates ($\overline{\theta}_w^r$ and $\overline{\theta}_l^r$), and the quality of entering candidates ($\overline{\theta}_w^e$ and $\overline{\theta}_l^e$).

(Eggers 2016)