

Evaluación de Impacto: RDD

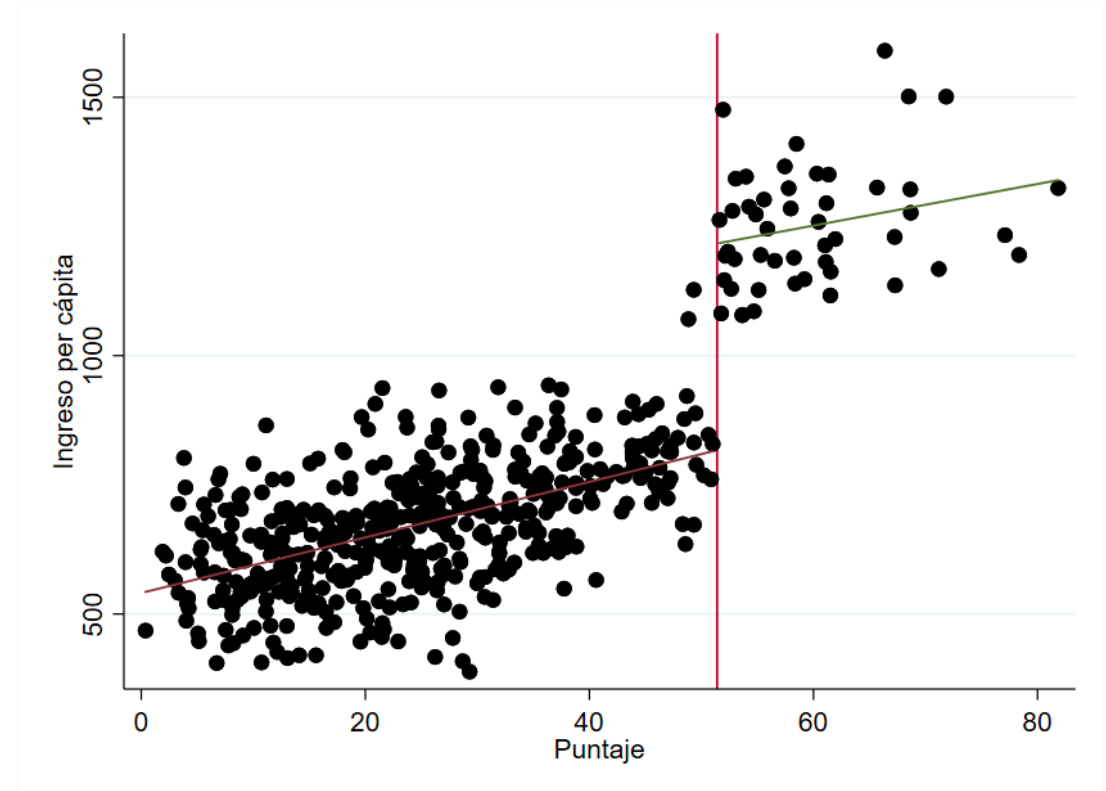
Francesco Bogliacino

¿Qué tienen en común todos estos casos?

- Entramos en la Nacional si alcanzamos C_0 en el puntaje de admisión;
- Nos multan si conducimos con C_0 de alcohol en la sangre;
- En Israel abren una nueva clase si se superan los 40 niños (regla de Maimónides);
- Ganas las elecciones si tienes el 50.1% de votos;
- Perdemos el estatus de estudiantes si bajamos el PAPA de 3;
- Recibimos Familias en Acción si no alcanzamos el C_0 del puntaje SISBEN
- ...

Let's work through an example

- El caso de la admisión a la Nacho
- ¿Qué está pasando?

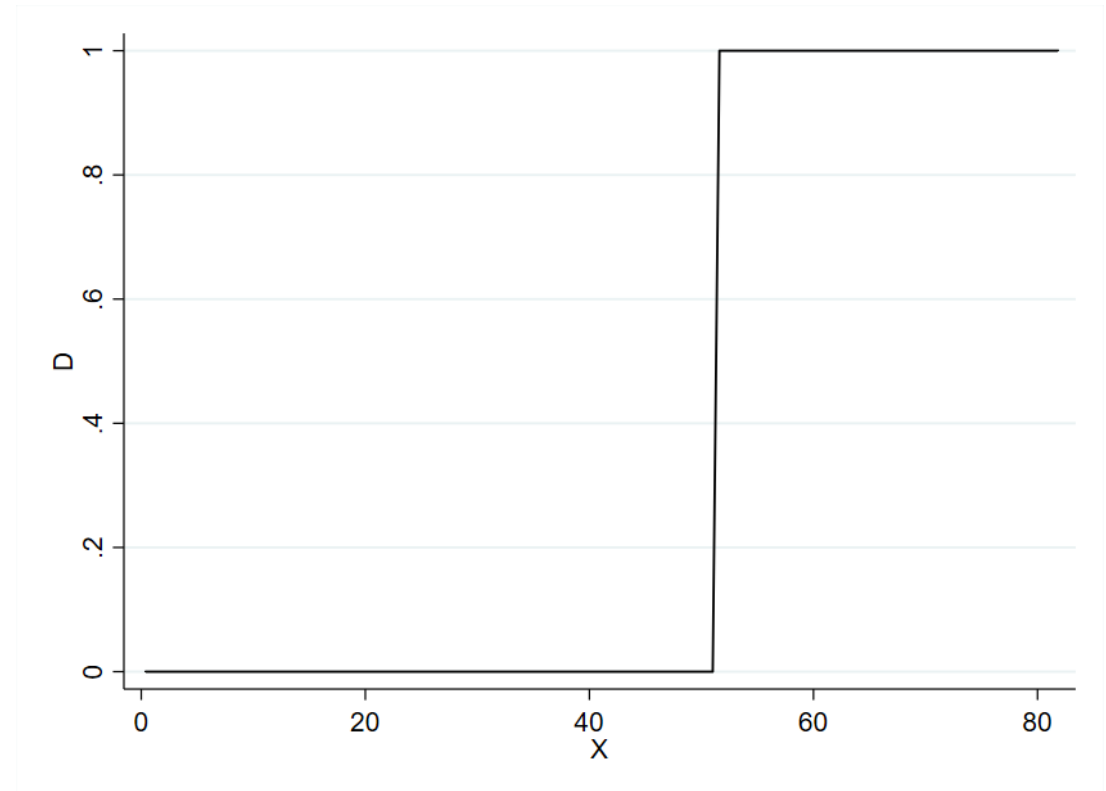


Sharp RDD

- Este es un caso (hipotético) de asignación determinística

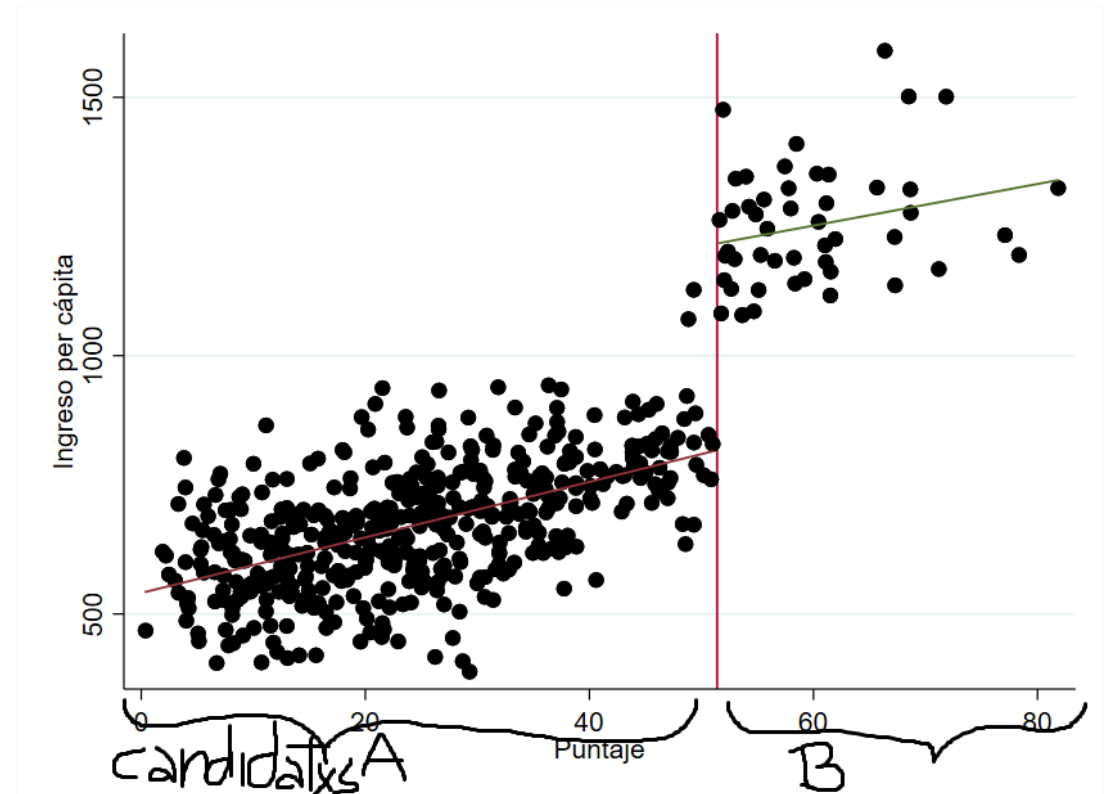
$$D_i = \begin{cases} 1 & \text{si } X \geq c_0 \\ 0 & \text{si } X < c_0 \end{cases}$$

- Tenemos una variable **continua** que llamamos *running variable* (X, el puntaje)
- Tenemos un umbral (**cutoff**), c_0
- Tenemos una regla de asignación que es determinística, es decir, conocido el puntaje conocemos el estatus de cada unidad



Pensamos en término de Potential Outcome

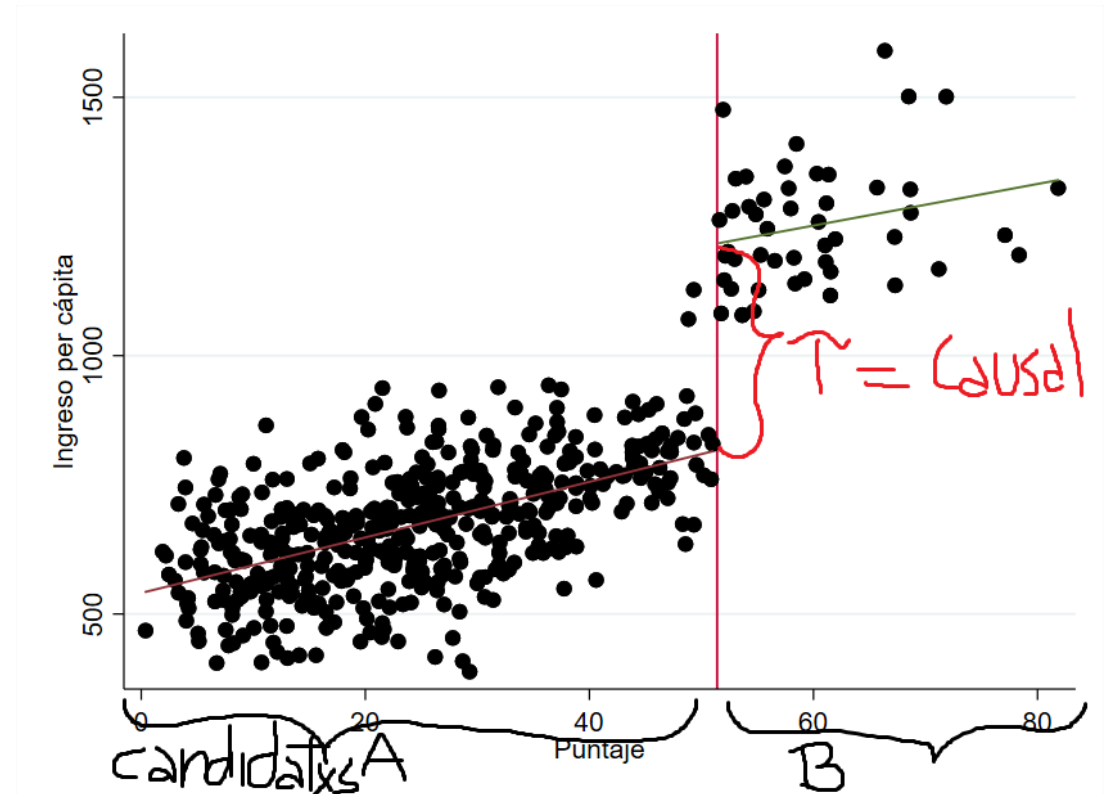
- Es obvio que A y B NO son comparables, serán presumiblemente diferentes en muchas dimensiones;
- Pero exactamente en el cutoff la única cosa que plausiblemente varía (*more on this later*) es la asignación a tratamiento



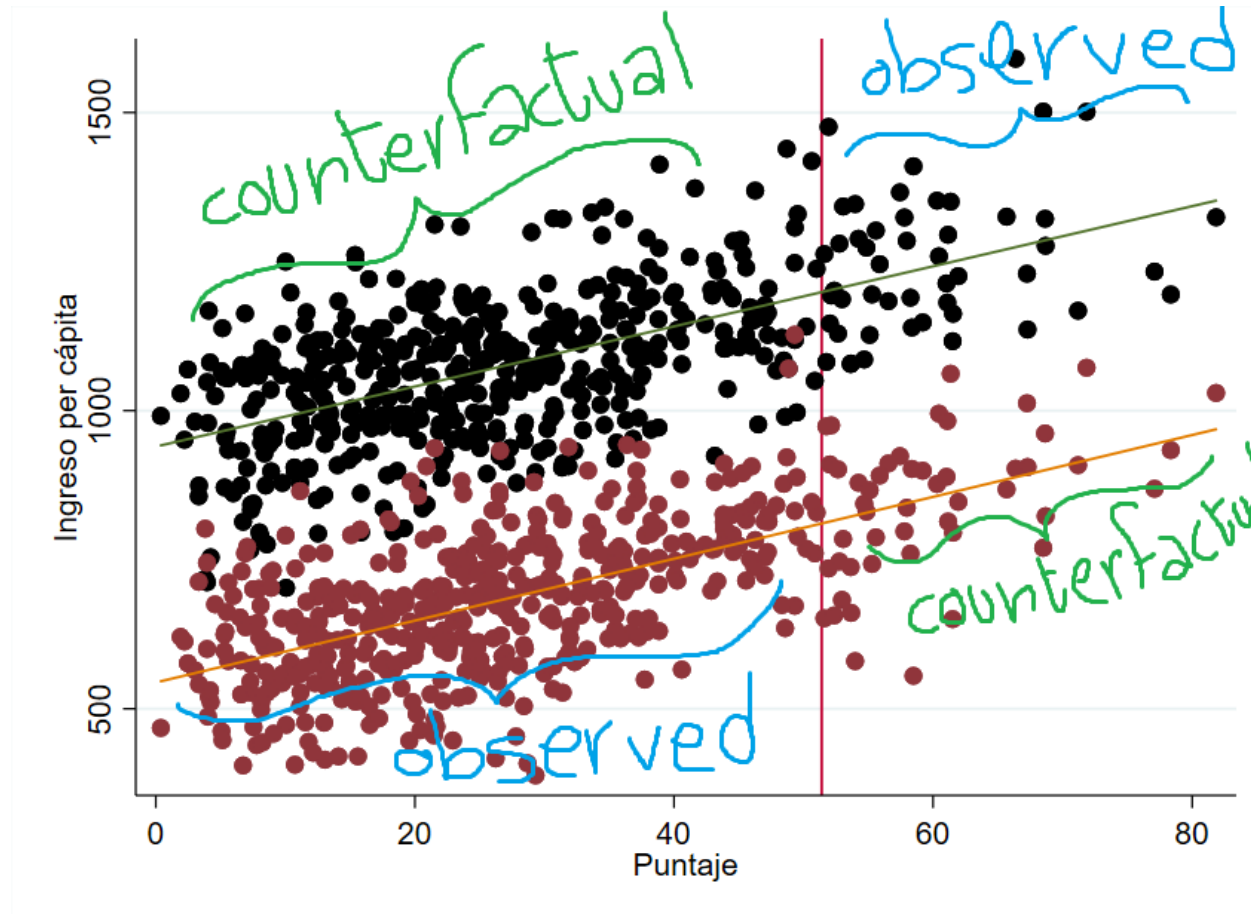
Pensamos en término de Potential Outcome

- El parametro causal es la discontinuidad alrededor del cutoff
- A la derecha son todos tratados, a la izquierda son todos no tratados

$$\begin{aligned} \tau &= \lim_{X \rightarrow c_0} E[Y_i^1 | X = c_0] \\ &- \lim_{X \rightarrow c_0} E[Y_i^0 | X = c_0] = \\ &\lim_{X \rightarrow c_0} E[Y_i | X = c_0] \\ &- \lim_{X \rightarrow c_0} E[Y_i | X = c_0] \end{aligned}$$



Pensamos en término de Potential Outcome



Pensamos en término de Potential Outcome

- Los contrafactuales son producto de extrapolación porque no tenemos un solo punto de traslape
- Esto es muy diferente de RCT, donde usábamos el observado para estimar el contrafactual, aquí usamos el observado para extrapolar el contrafactual

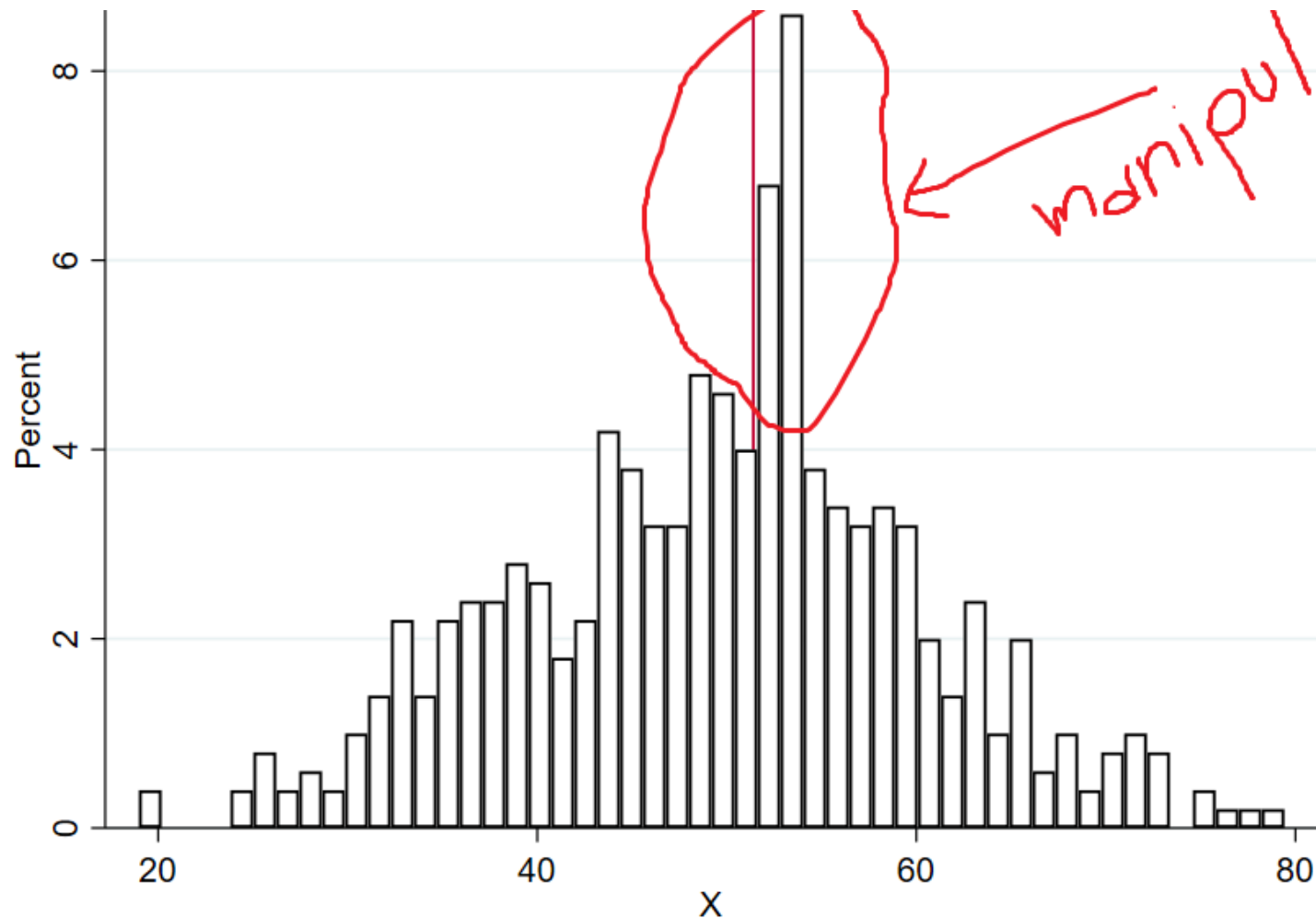
Cuál es la hipótesis de identificación en este caso?

- **Continuidad**

- La identificación se basa en que las CEF sean *smooth* alrededor del cutoff, si saltan por cualquier otra razón no puedo identificar el impacto del tratamiento porque tengo explicación plausible alternativa

$$E[Y_i^1 | X = c_0] \text{ y } E[Y_i^0 | X = c_0] \text{ son continuas en } X \text{ a } c_0$$

¿Qué pasaría si los participantes tienen **control** sobre asignación?

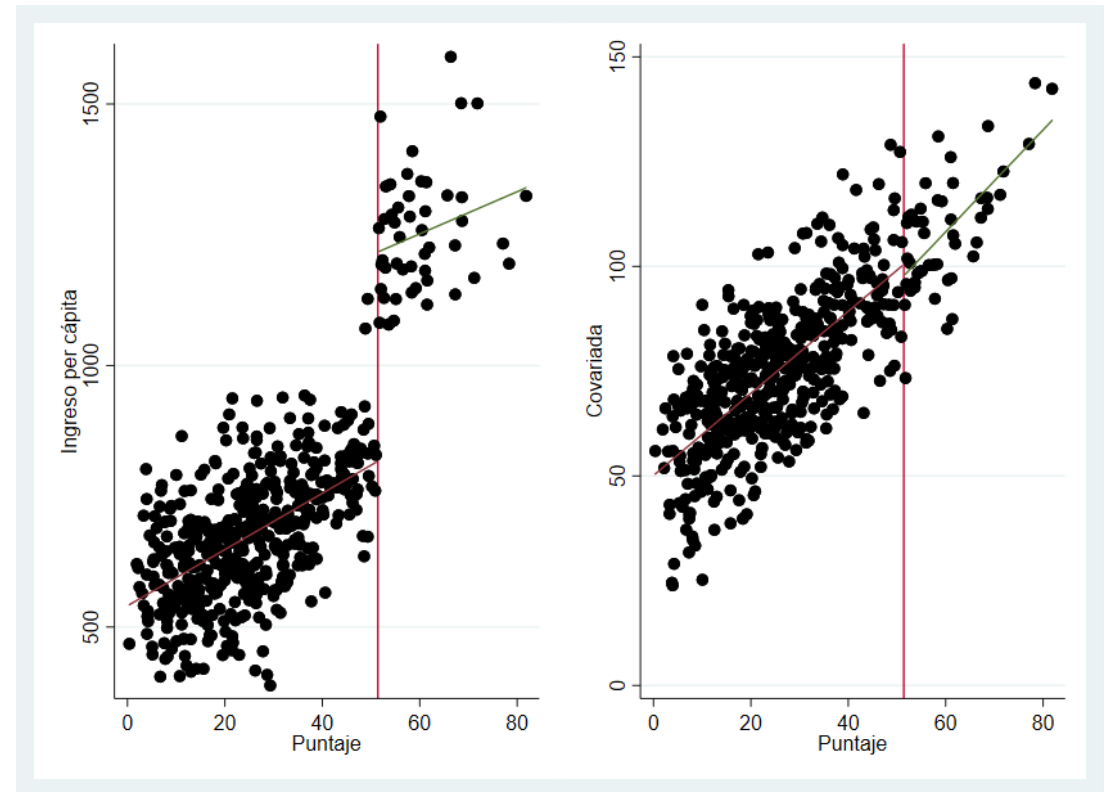
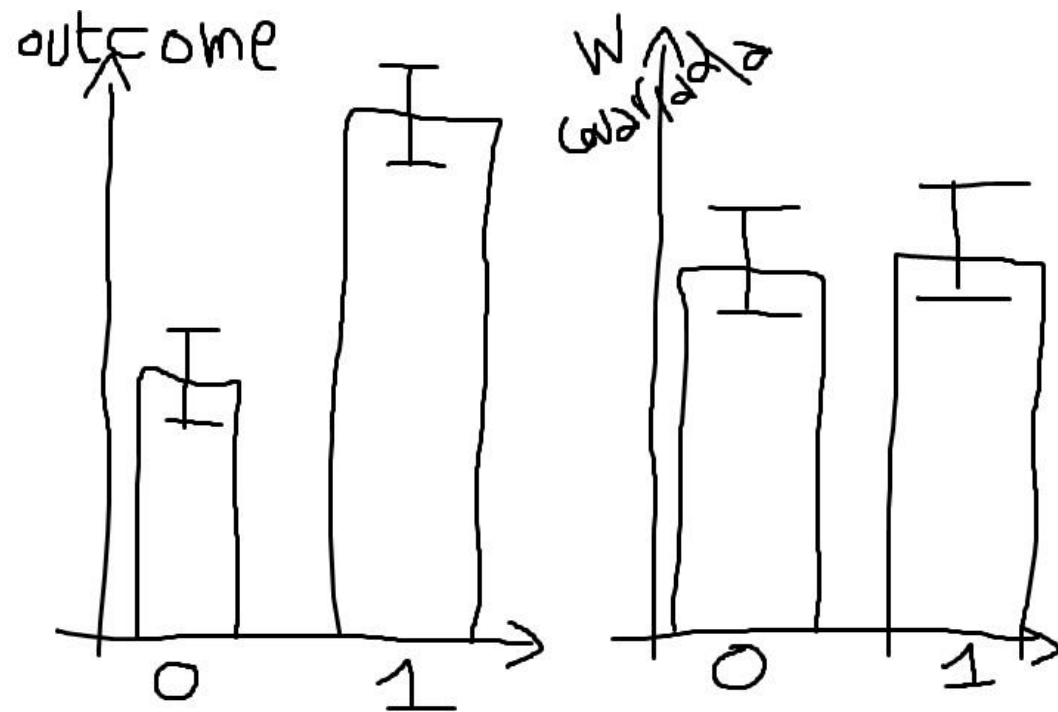


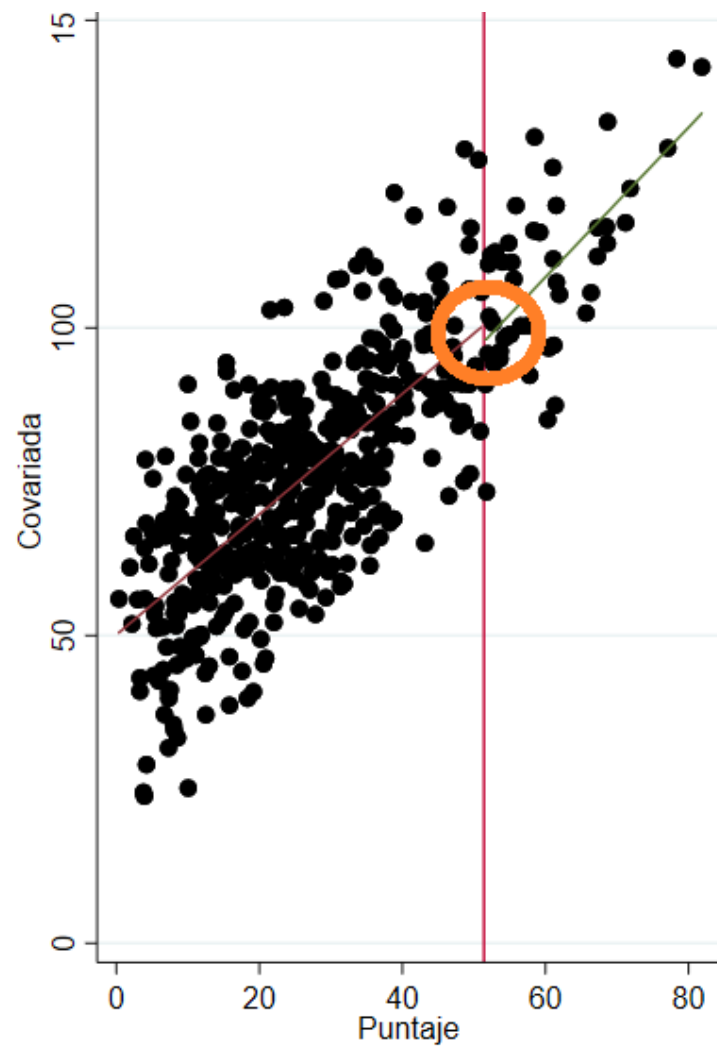
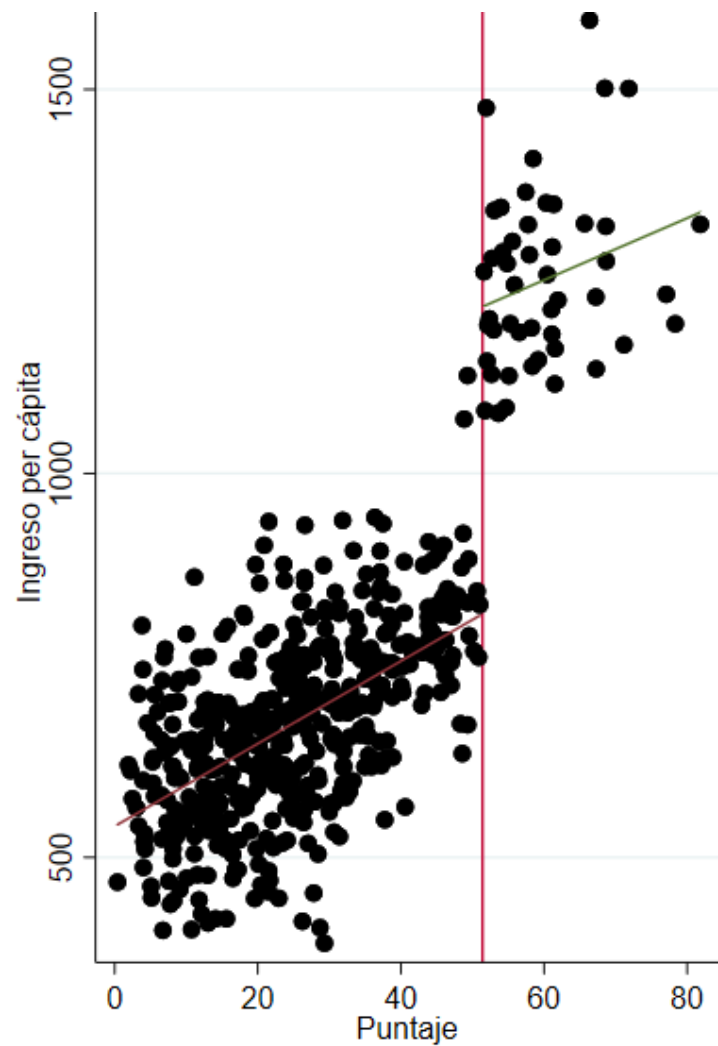
¿Qué pasaría
si los
participantes
tuvieran
control sobre
asignación?

Más sobre identificación

- El supuesto de identificación es la ausencia de “saltos” en las covariadas;
- Podemos hacer algo parecido al RCT:
 - RCT: testeo como diferencia de promedios y muestro que el supuesto de identificación es consistente con los datos mostrando las diferencia de promedios para covariadas
 - RDD: testeo como discontinuidad al cutoff y muestro que el supuesto es consistente con los datos mostrando la (ausencia) de discontinuidad para covariadas

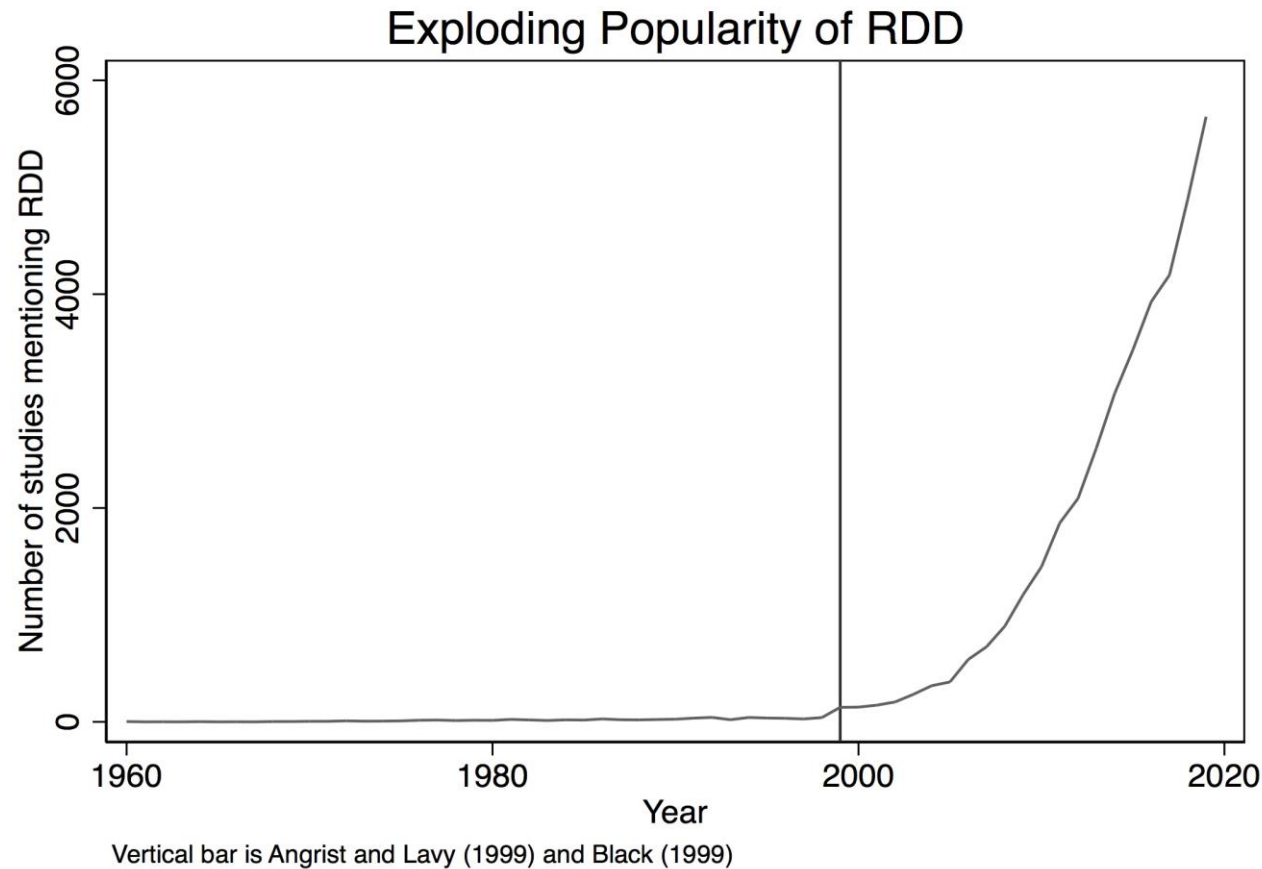
RCT vs RDD



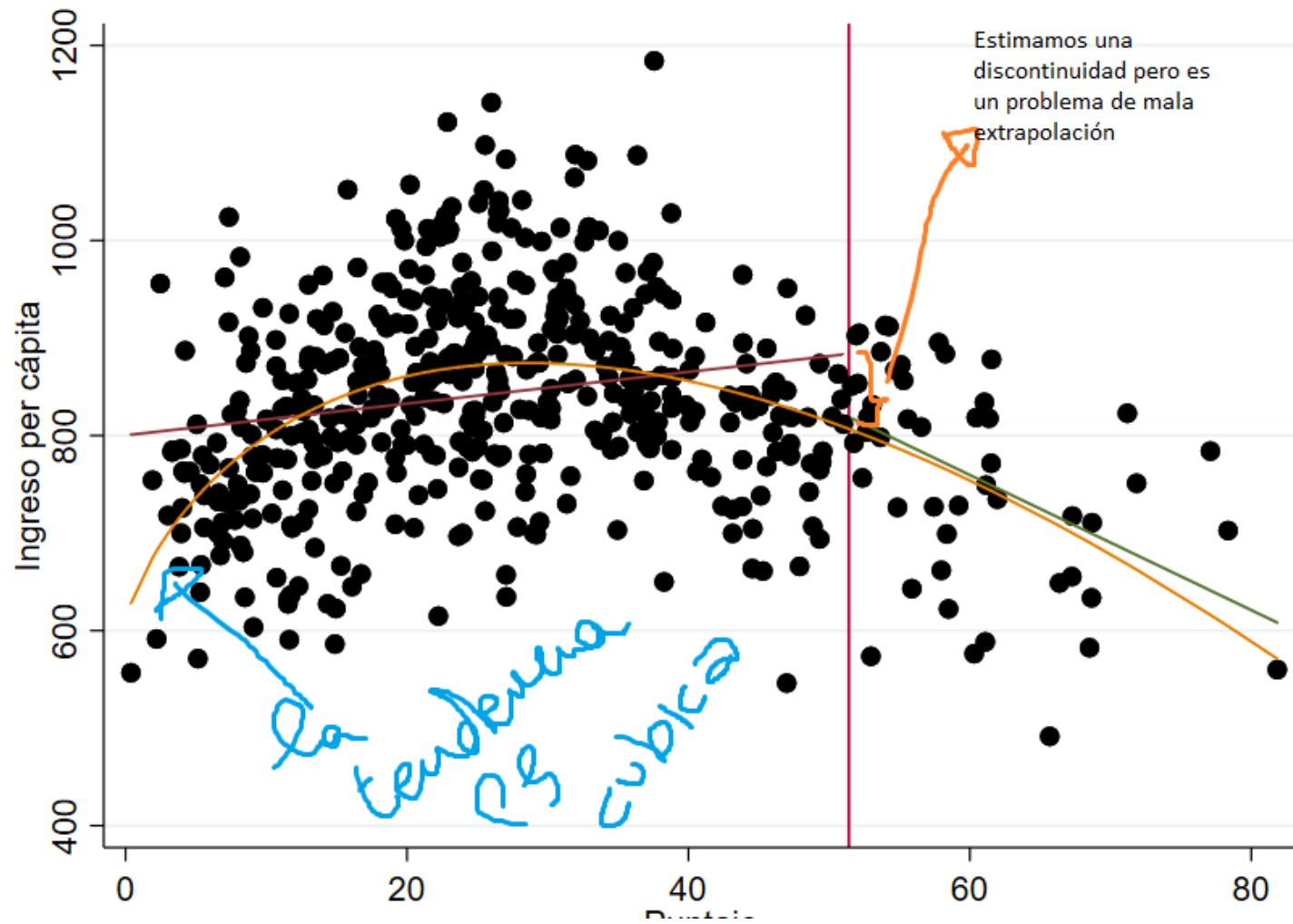


Obviamente hay que hacer una prueba para ver si la discontinuidad es diferente de cero

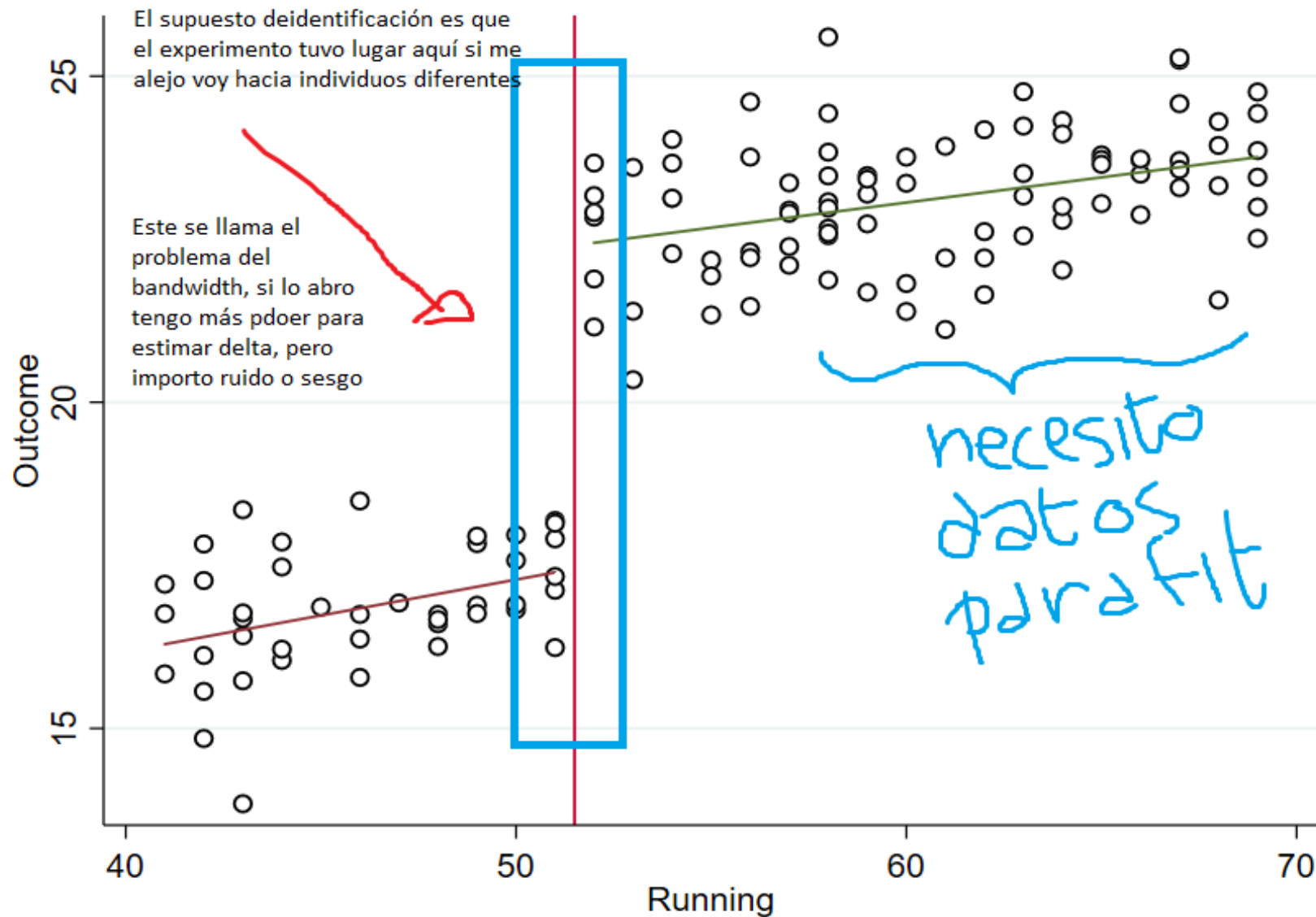
Su “paralelo” con RCT explica éxito con economistas



Source: Cunningham



Now,
extrapolation...



More on
extrapolation
(RDD es
intensive en
datos) ...



Un ejemplo



The Real Winner's Curse



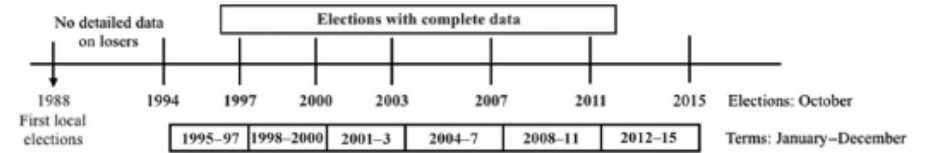
Leopoldo Fergusson	Universidad de los Andes
Pablo Querubin	New York University
Nelson A. Ruiz	University of Oxford
Juan F. Vargas	Universidad del Rosario

Contexto y motivación

- Sabemos que el cierre político puede llevar a formación de grupos armados. Eso implica que hay algo de “sustituibilidad” de los dos mecanismos;
- Pero hay otro mecanismo que ha sido menos explorado: abertura política genera respuesta violenta (Argentina, Chile, Haiti ...) si:
 - Elite tiene ventaja comparativa en manejo de la violencia;
 - El contexto institucional es debil
- Empíricamente es un challenge:
 - Intro de la democracia es endógena respeto a la dinámica de violencia;
 - Es difícil encontrar un contrafactual plausible

Empirical strategy

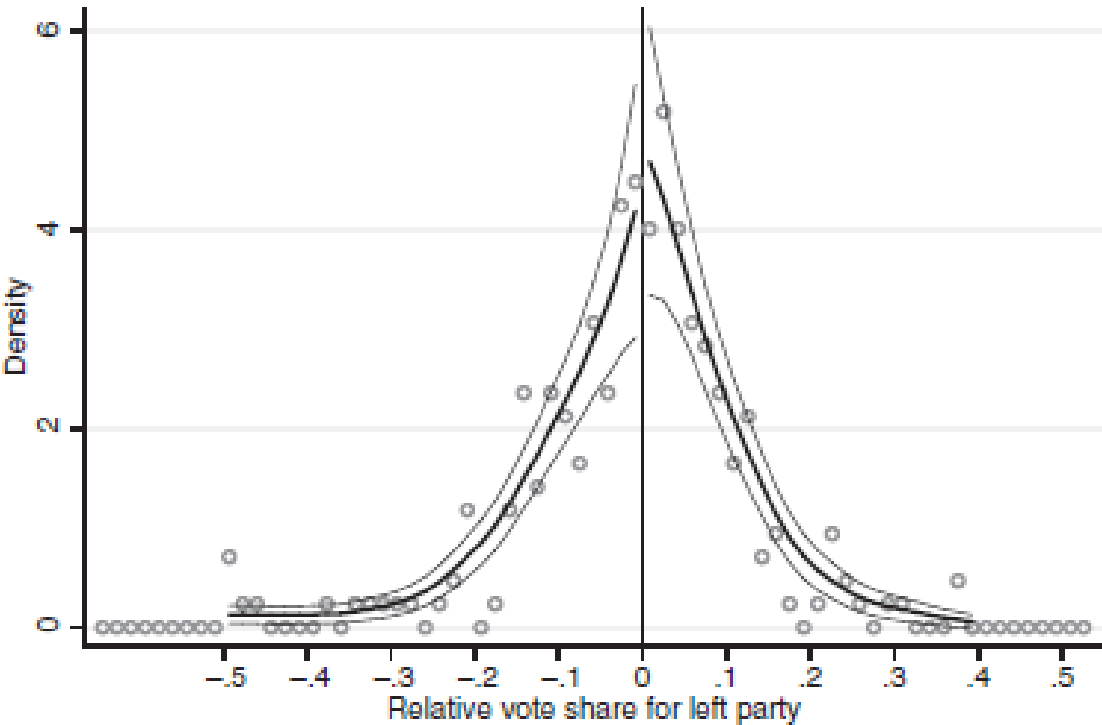
FIGURE 1 Election Years and Available Data



Source: Authors' own.

- Uso de la intro de elecciones en 1988-1991
- RDD para la victoria de gobiernos de izquierda:
 - Uso el margen de lograr la victoria por pocos votos, el “experimento” es sobre el margen entre quien ganó por poco y quien perdió por poco
- Respuesta paramilitar (ventaja comparativa en uso de violencia por parte de la elite)
- Datos:
 - Elecciones con categorización de los partidos según ideología
 - Base de datos de Restrepo et al (2003) sobre actos violentos con autor
 - Otros datos para estudiar mecanismos etc

FIGURE 2 McCrary Test: Sorting Around the Winning Threshold for the Left



Dependent variable	Mean	Standard Deviation	Left victory	Std. Error.	Obs	Bandwidth
Panel A. <i>Election year</i>						
Year elected	2002.701	5.447	1.257	2.426	167	.1
Panel B. <i>Geographic characteristics</i>						
Altitude, meters	1752.587	3469.152	69.265	671.765	94	.042
Average precipitation	93.010	18.665	1.505	8.379	152	.09
Distance to department capital, km	81.129	53.511	-1.778	23.432	152	.089
Distance to closest market place, km	145.999	91.279	6.226	41.881	129	.069
Andean region dummy	0.417	0.494	-.128	.203	142	.08
Pacific region dummy	0.398	0.490	.025	.165	156	.092
Eastern region dummy	0.098	0.298	-.065	.112	126	.066
Caribbean region dummy	0.087	0.282	.143	.105	166	.099
Panel C. <i>Socioeconomic characteristics</i>						
Vote % for left-wing presidential candidates, 1994	0.067	0.070	-.019	.019	111	.056
Vote % for conservative presidential candidates, 1994	0.422	0.209	-.006	.103	129	.069
La Violencia incidence (1948-1953)	0.146	0.353	-.146	.16	142	.079
Rurality index	0.654	0.238	-.035	.091	178	.114
Initial population, 1993	26328.799	33888.293	18839.103	17523.015	148	.083
Coca, 1994	0.075	0.264	.089	.093	120	.062
Literacy rate, 1993	85.452	8.783	-.664	3.646	150	.088
Panel D. <i>Land variables</i>						
Land GINI, based plot sizes	0.707	0.118	-.023	.05	81	.075
Land GINI, based on landowner holdings	0.722	0.103	-.013	.049	67	.06
Panel E. <i>Tax revenue</i>						
Tax income (per capita)	0.071	0.308	-.191	.156	206	.145
Non tax income (per capita)	0.015	0.029	-.015	.012	199	.136
Tax income from land taxes (per capita)	0.017	0.057	-.046	.035	196	.129
Tax income from commerce and industry (per capita)	0.027	0.188	-.092	.087	216	.163
Panel F. <i>Electoral variables</i>						
Number of candidates in election	3.949	1.968	.107	.647	135	.075
Number of parties in election	3.587	1.968	.455	.668	142	.079
Voter turnout	0.590	0.170	.014	.054	146	.088

Step 1: identificación

Step 2: Main Results

TABLE 1 Effect of Electing a Left-Wing Mayor on Paramilitary Violence

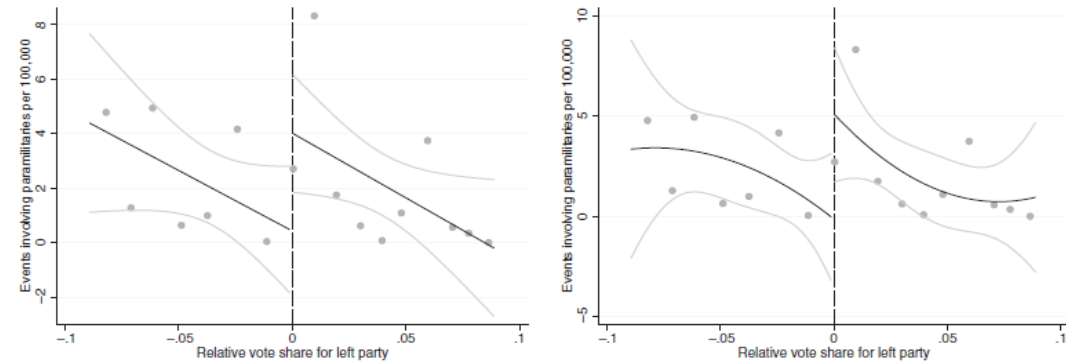
Dependent variable: Average yearly violent events involving paramilitaries (per 100,000 inhabitants) during term in office

	Linear Polynomials				Quadratic Polynomials			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Nonparametric estimates and robust standard errors of Calonico, Cattaneo, and Titiunik (2014)								
Left-wing mayor elected	4.963** (2.344)	5.455** (2.307)	6.719*** (2.445)	7.273*** (2.621)	5.665** (2.469)	5.644** (2.400)	6.623*** (2.522)	6.769** (2.689)
Observations	152	129	108	99	205	135	156	140
Bandwidth	0.090	0.070	0.054	0.048	0.144	0.076	0.093	0.078
Panel B: Parametric estimates								
Left-wing mayor elected	4.225** (2.113)	4.136 (2.231)	4.128 (2.175)	4.198 (2.269)	5.469** (2.515)	5.475** (2.533)	6.148** (2.657)	6.215** (2.845)
Observations	152	152	151	151	152	152	151	151
Bandwidth	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090
Controls	No	Geographic	Socioeconomic	All	No	Geographic	Socioeconomic	All

Notes: Standard errors are in parentheses.

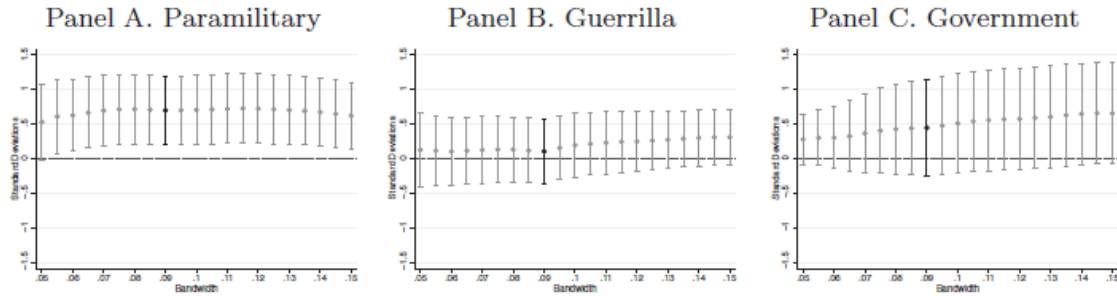
Significant at 5% level; *significant at 1% level.

FIGURE 4 Effect of Electing a Left-Leaning Mayor on the Average Yearly Violent Events Involving Paramilitaries



Notes: Observations displayed are within Calonico, Cattaneo, and Titiunik's (2014) optimal bandwidth for the linear, nonparametric, and no-controls case (0.09). Left: linear fit. Right: quadratic fit.

Violent events during term in office by group



Violent events in year before elections by group

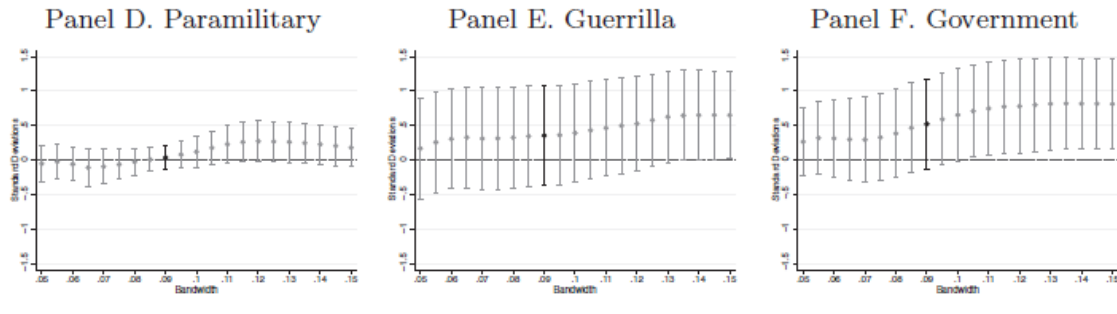


TABLE 2 Effect of Electing a Left-Wing Mayor on Paramilitary Violence Robustness to Alternative Measures of the Dependent Variable

Dependent variable:	Conflict Casualties		Dummy for Any Paramilitary Violence		Count of Violent Events	
	(1)	(2)	(3)	(4)	(5)	(6)
Left-wing mayor elected	13.967*** (5.093)	18.794** (7.299)	0.274** (0.118)	0.299 (0.165)	0.526*** (0.156)	0.750*** (0.212)
Observations	151	151	151	151	151	151
Bandwidth	0.090	0.090	0.090	0.090	0.090	0.090
Polynomial order	1	2	1	2	1	2
Controls	All	All	All	All	All	All
Estimation Method	OLS	OLS	Probit	Probit	Neg. Bin.	Neg. Bin.

Notes: Standard errors are in parentheses. Nonparametric estimates and robust standard errors of Calonico, Cattaneo, and Titiunik (2014), are used in columns 1–2.

Significant at 5% level; *significant at 1% level.

Step 2: Robustness

Sugiere que es por la ventaja comparativa en la violencia

TABLE 5 Heterogeneous Effects by Timing of AUC Demobilization

Dependent variable: Average yearly events involving paramilitary (per 100,000 inhabitants) during term in office		
	(1)	(2)
A Left-wing mayor elected	5.824** (2.582)	8.061** (3.122)
Post-AUC demobilization	1.845 (1.572)	1.870 (1.607)
B Post AUC demobilization × Left-wing mayor elected	−4.314** (1.794)	−4.456** (1.834)
Observations	151	151
R-squared	0.251	0.262
Bandwidth	0.090	0.090
A + B	1.51	3.605
Ho: A + B = 0		
F-statistic	.53	1.96
P-value	.47	.16
(Local) polynomial order	1	2
Controls	All	All

Notes: Standard errors are in parentheses. Parametric estimates are used.
Significant at 5% level; *significant at 1% level.

TABLE 4 Effect of Electing a Right-Wing Mayor on Violence

Dependent variable: Average yearly events per 100,000 inhabitants during term in office involving:				
	All Groups (1)	Paramilitary (2)	Guerrilla (3)	Government (4)
Right-wing mayor elected	1.220 (1.250)	0.485 (0.629)	0.154 (0.404)	−0.071 (0.549)
Observations	480	480	480	480
Bandwidth	0.090	0.090	0.090	0.090
Polynomial order	1	1	1	1
Controls	All	All	All	All

Notes: Standard errors are in parentheses. Nonparametric estimates and robust standard errors of Calonico, Cattaneo, and Titiunik (2014) are used in all regressions.
Significant at 5% level; *significant at 1% level.

Tampoco es consistente con que sean “nuevos” partidos

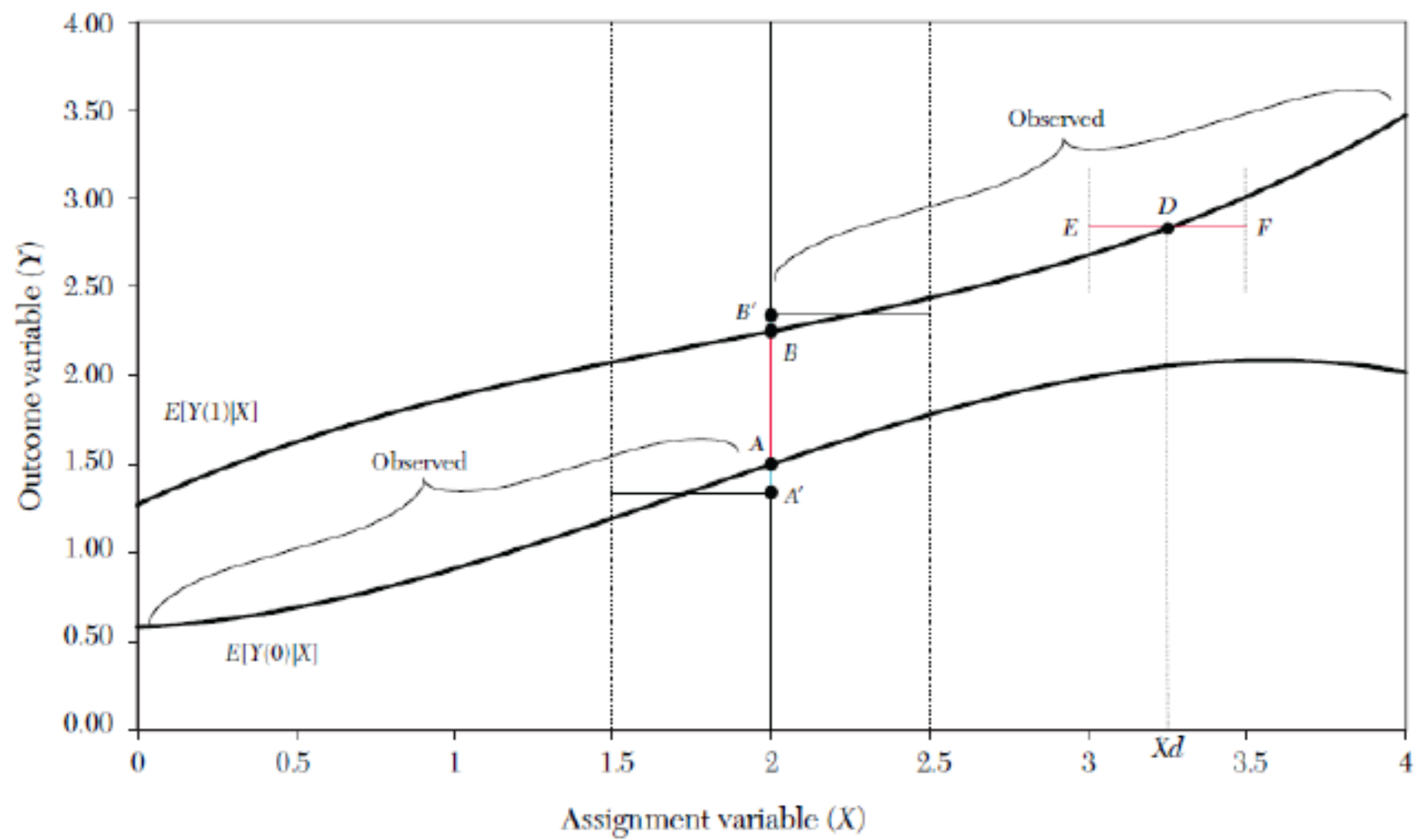
Step 4: Mecanismos



Ok let's get
practical

Estimación

- Descarguen rdrobust
- stata



Local linear non parametric regression

$$\sum_{i=1}^n (y_i - a - b(x_i - c_0))^2 K\left(\frac{x_i - c_0}{h}\right) 1(x_i > c_0)$$

h es el bandwidth

https://twitter.com/page_eco/status/958687180104245248

Kernel:

- Si es rectangular le da el mismo peso dentro de un bin
- Si es triangular le da más a los datos que están en el centro

Making the case for identification

- En la realidad, “defender” un RCT en seminario es smooth
- Para el RDD es necesario tener más argumentos
- Descarguen rdrobust tanto para Stata como para R

Violación de continuidad

- Primero los agentes tienen que tener interés en manipular (que no me multen por conducir bajo el efecto, entrar a la U, recibir subsidio):
 - Que el cutoff sea conocido con antelación;
 - Que se pueda hacer ajuste (retomar examen);
 - Que haya razones administrativas para condicionar los datos (redondeando las cifras)-> *heaping*

Como siempre para estudiar bien un tema necesito conocer a fondo los detalles institucionales

Manipulation: Heaping

- ¿Qué hacer en este caso?
- Eliminar los datos en exactamente 1500 (Donut Hole) y estimar en el entorno

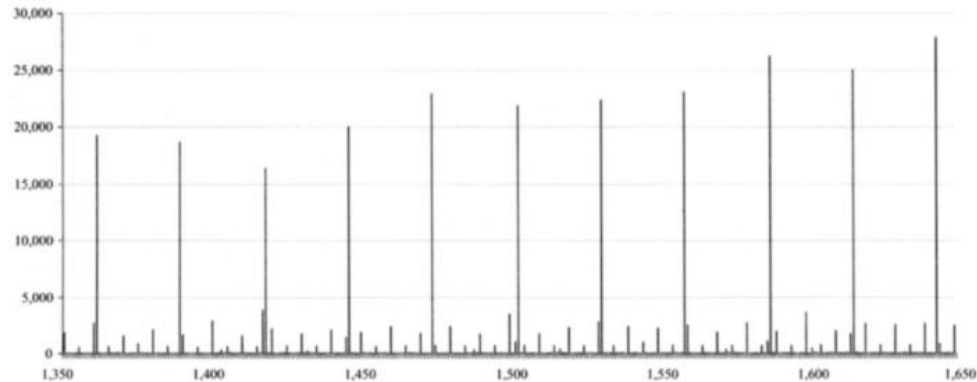


FIGURE I

Frequency of Births by Gram: Population of U.S. Births
between 1,350 and 1,650 g

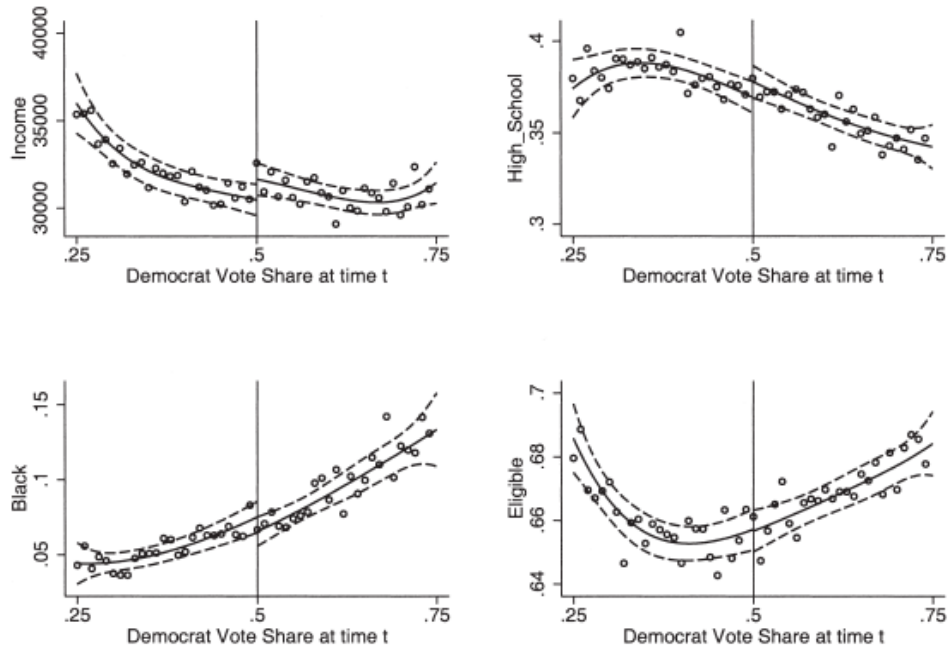
NCHS birth cohort linked birth/infant death files, 1983–1991 and 1995–2003,
as described in the text.

Fuente: Distribución por peso al nacer, Almond et al. 2010

Endogenous cutoff

- Es el caso donde el umbral está correlacionado con otras variables (ej 18 años-> voto,)

Balance test



Fuente: Lee Moretti and Butler (2004)

- Este paper es sobre victorias electorales (close elections)
- Hay tantos estudios de RDD sobre esto que prácticamente es un campo de estudio a parte

Fuzzy RDD

- TBC (después de variable instrumental)