

# 5 – Regresión Discontinua (RD)

Evaluación de Impacto  
Rachid Laajaj



Universidad de  
los Andes

Facultad  
de Economía

Educación  
**Continua**  
Vicerrectoría Académica

Universidad de los Andes | Vigilada Mineducación  
Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964.  
Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia



# El Método de Regresión Discontinua

El método de Regresión Discontinua es un método que aprovecha una discontinuidad arbitraria en una regla de asignación a un tratamiento.

Ejemplos:

- Solo los hogares con Sisbén debajo de un umbral se benefician de un programa social
- Solo los municipios de más de X habitantes son afectados por una reforma
- El candidato con más votos gana una elección

Típicamente proviene de una restricción/decisión administrativa. Un índice de elegibilidad y un umbral estricto se usan para determinar la elegibilidad a un tratamiento.

En este caso las unidades justo abajo y encima del umbral deben ser muy parecidas y ofrecen un buen contrafactual.

Difieren un poco en el índice de elegibilidad, pero se puede controlar por eso.

# ¿Cuándo se puede aplicar el Método de Regresión Discontinua?

- Existe un índice de elegibilidad observable que se usó como criterio de participación en una intervención.
- Este índice debe ser continuo o “fluido” (no puede ser categórico con un numero pequeño de categorías)
- Se aplica un umbral predefinido: las unidades por encima (o por debajo) del umbral son elegibles a recibir el tratamiento.
- El umbral solamente determina el tratamiento de interés: no hay otros tratamientos que se aplican en este mismo umbral.
- El umbral del índice de elegibilidad no es manipulable
- Requiere datos sobre el índice de elegibilidad, el tratamiento, el resultado de interés y preferiblemente de controles adicionales predefinidos.

# La Regresión Discontinua: ¡Un Salto es Sospechoso!

Suponga que la asignación al grupo de tratados  $D_i$  ocurre solo para las unidades  $i$  con un índice de elegibilidad mayor al umbral  $\bar{x}$ :

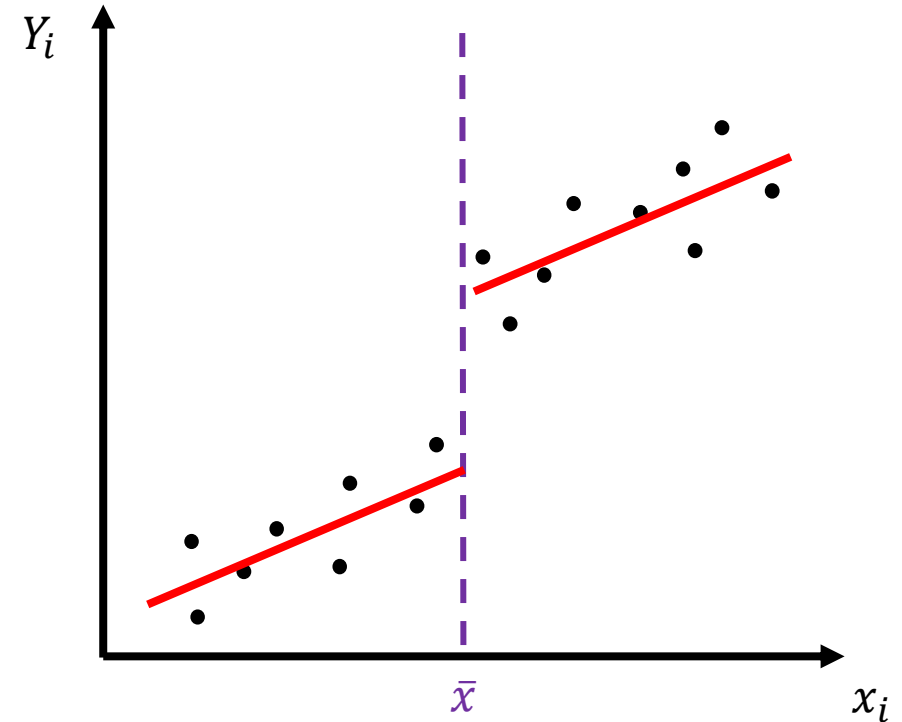
$$D_i = \begin{cases} 1 & \text{si } x_i \geq \bar{x} \\ 0 & \text{si } x_i < \bar{x} \end{cases}$$

$x_i$  es el índice de elegibilidad o en inglés, la *running variable* o *forcing variable*

$Y_i$  es el resultado de interés.

Si observamos un salto en  $Y_i$  justo al nivel de  $\bar{x}$ , entonces es probable que este salto este causado por la elegibilidad al tratamiento.

¿Qué más podría explicar una diferencia importante en  $Y_i$  entre las unidades  $\bar{x} - \varepsilon$  y  $\bar{x} + \varepsilon$ ?  
( $\varepsilon$  es un valor muy pequeño)



# El Método de Estimación Lineal de la Regresión Discontinua

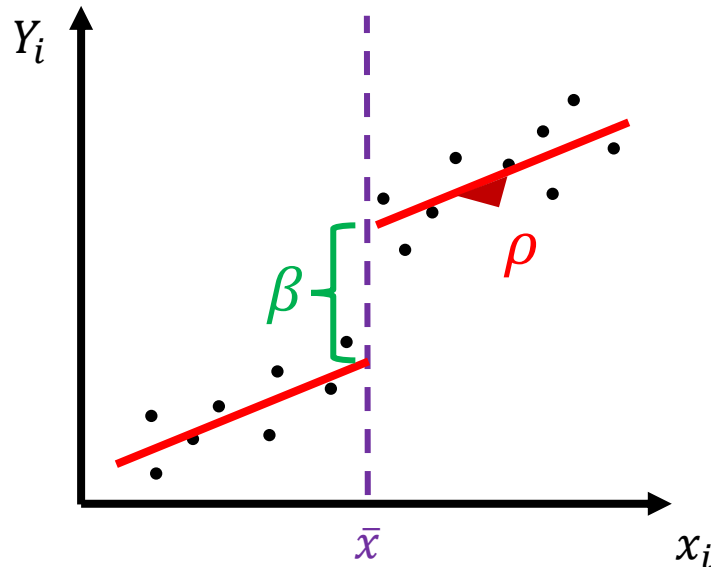
Debemos controlar por el índice de elegibilidad ya que sabemos que existe esta diferencia.

Primero suponemos que el efecto del índice de elegibilidad sobre  $Y_i$  es lineal. Entonces puedo estimar  $\rho$ , el efecto del tratamiento:

$$Y_i = \alpha + \beta D_i + \rho x_i + \eta_i \quad (1)$$

$\rho$  controla por la diferencia entre los tratados y no tratados en  $x_i$

$\beta$  es el impacto de ser elegible para el tratamiento sobre  $Y_i$



# Estimación no Lineal de la Regresión Discontinua

Pero la relación entre  $x_i$  y  $Y_i$  no siempre es lineal

Si no se captura bien la relación  $E[Y_{oi}|x_i]$ , se puede generar un sesgo en la estimación del salto que ocurre en  $\bar{x}$  cuando alrededor del umbral.

El modelo más general de RD es:

$$Y_i = \alpha + f(x_i) + \beta D_i + \eta_i \quad (2)$$

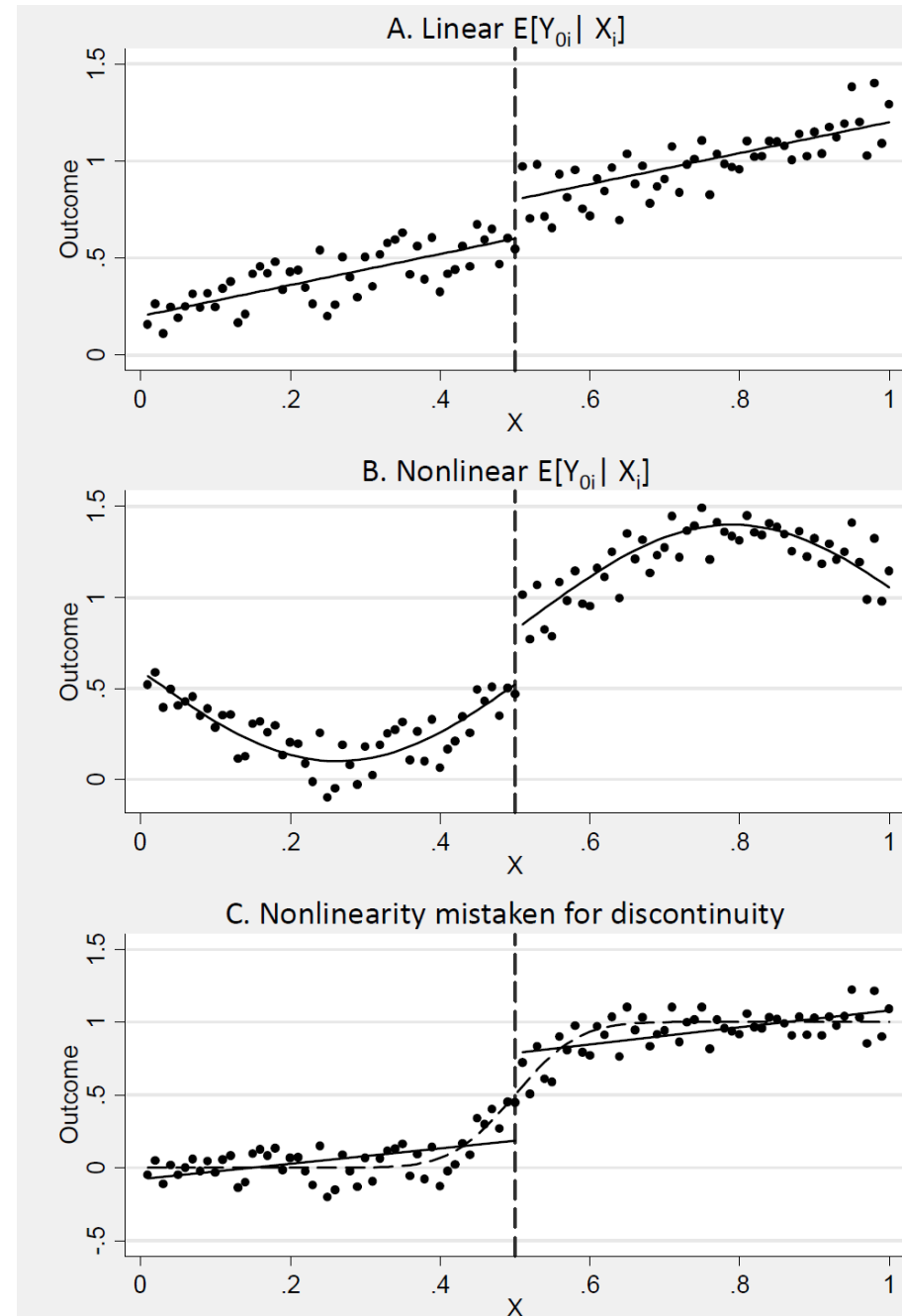
$f(x_i)$  es la forma funcional que se usa para capturar la relación  $E[Y_{oi}|x_i]$

$\beta$  es el impacto de ser elegible para el tratamiento, sobre  $Y_i$

En  $f(x_i)$  se pueden usar funciones polinomiales, diferentes de cada lado.

La literatura recomienda una regresión lineal local (lineal, pero no paramétrica, ponderada).

También es posible incluir otros controles en la regresión





# Regresión Discontinua “Nítida” o “Difusa”

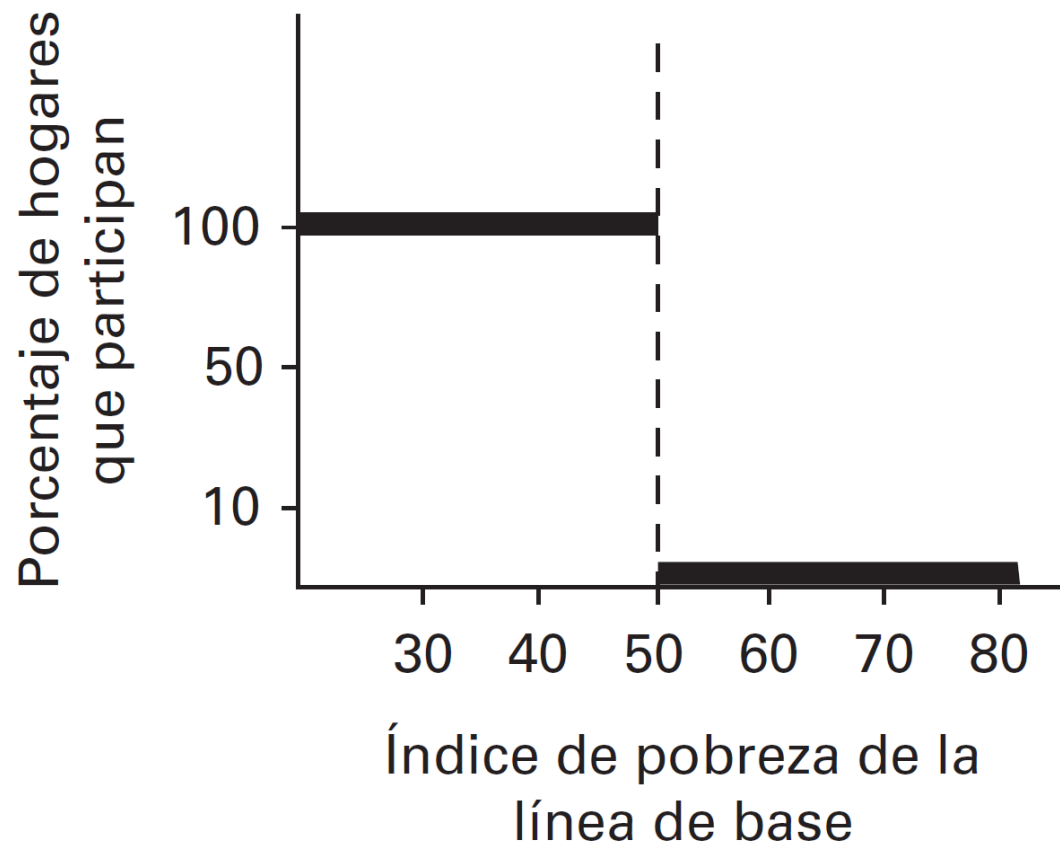
- En una Regresión Discontinuada Nítida, el % de tratados pasa del 0% al 100% entre un lado del umbral de elegibilidad y el otro.
- Mientras que en una Regresión Discontinua “Difusa” (*Fuzzy Regression Discontinuity* en inglés), el cambio en la probabilidad de estar tratado debe ser significativo (y suficientemente alto para el poder de la estimación), pero no es del 100%.

Si  $D_i = \begin{cases} 1 & \text{si } x_i > \bar{x} \\ 0 & \text{si } x_i < \bar{x} \end{cases}$  es la variable binaria de asignación al tratamiento, y  $T_i$  es el hecho de realmente beneficiarse del tratamiento. En una regresión nítida,  $D_i = T_i$  pero en una regresión difusa, la correlación no es perfecta.

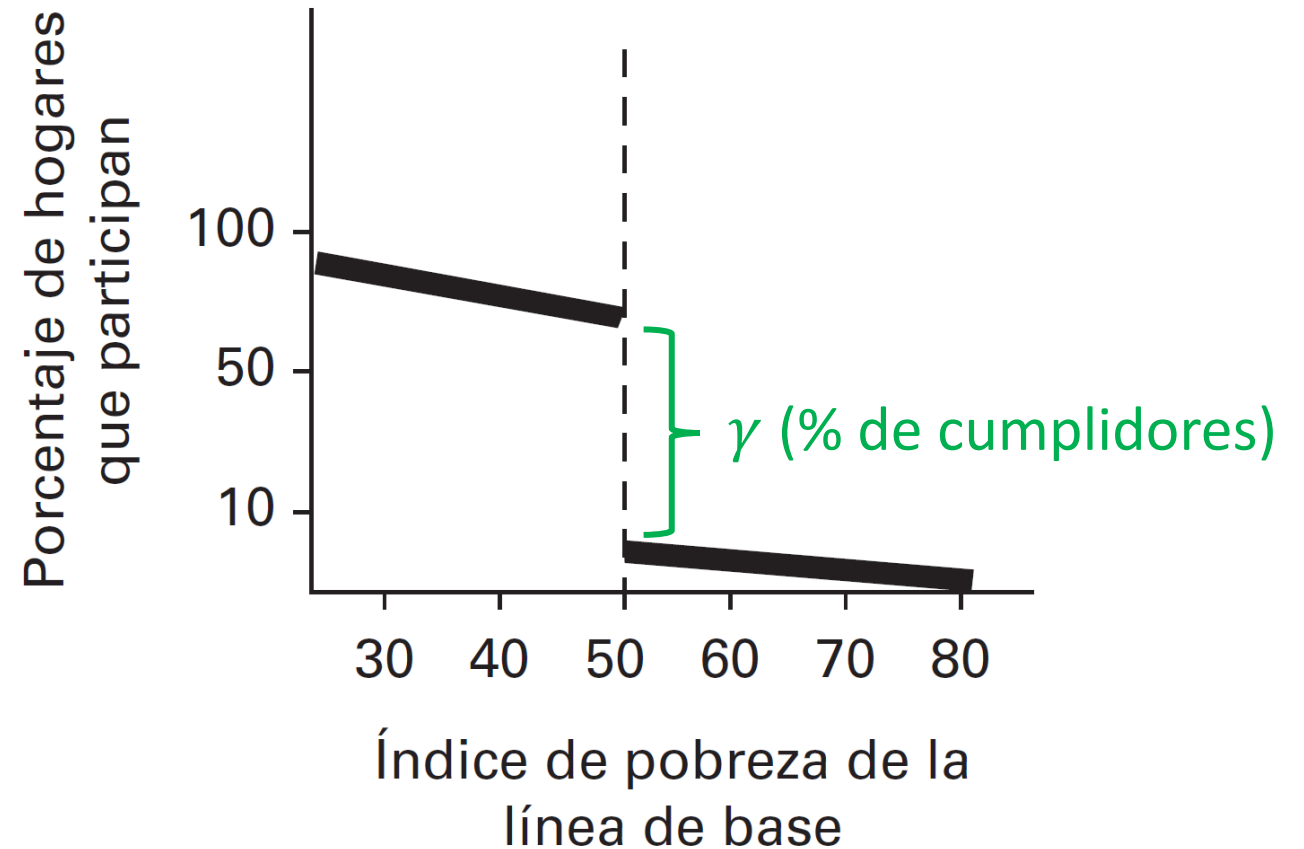
Por ejemplo, en un programa social de transferencias condicionales un hogar puede cumplir con el criterio de Sisbén para ser elegible, pero no implica que necesariamente se beneficie del programa.

# Regresión Discontinua “Nítida” o “Difusa” (Grafica)

a. DRD nítido  
(pleno cumplimiento)



b. DRD difuso  
(cumplimiento incompleto)





# El Método de Regresión Discontinua Difusa con Variable Instrumental

La Regresión Discontinua difusa permite una estimación del efecto del tratamiento  $T_i$  a través del método de Variable Instrumental:  $D_i \rightarrow \hat{T}_i \rightarrow Y_i$

En la primera etapa, se instrumenta  $\hat{T}_i$  por  $D_i$  y otros controles (incluyendo  $f(x)$ ), por ejemplo, si es un polinomio de primer orden, permitiendo coeficientes diferentes de cada lado (con  $\alpha_1$  y  $\alpha_2$ ):

$$T_i = \alpha_0 + \gamma D_i + \alpha_1 x_i + \alpha_2 x_i * D_i + \epsilon_i \quad (1)$$

$\gamma$  indica el salto en la probabilidad de estar tratado (en la discontinuidad).

También  $\gamma$  indica la proporción de cumplidores: los que están tratados solamente si son asignados al grupo de los tratados.

En la segunda etapa se estima el efecto de  $\hat{T}_i$  sobre  $Y_i$ :

$$Y_i = \alpha_0 + \tau \hat{T}_i + \rho_1 x_i + \rho_2 x_i * D_i + \epsilon_i \quad (2)$$

En ciertos casos se puede correr la forma reducida para aprender sobre el efecto de la elegibilidad.

## Ejemplos prácticos (1)

- Gagliarducci, S. and Nannicini, T., 2013. *Do better paid politicians perform better? Disentangling incentives from selection*. Journal of the European Economic Association, 11(2), pp.369-398.
- El índice de elegibilidad es el tamaño de la población de los municipios.
- En Italia, se aplica un aumento importante del salario del alcalde para municipios de más de 5,000 habitantes.
- Primero, se observa qué tan diferentes son los alcaldes cuando el salario es mayor: aumenta el nivel de educación, etc.
- Segundo, se observa una mejoría en varias medidas de eficiencia institucional.

## Ejemplos prácticos (2)

Hoekstra, Mark. 2009. *The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach*. Review of Economics and Statistics 91 (4): 717–24.

El índice de elegibilidad es el SAT score (parecido a Saber 11 en Colombia)

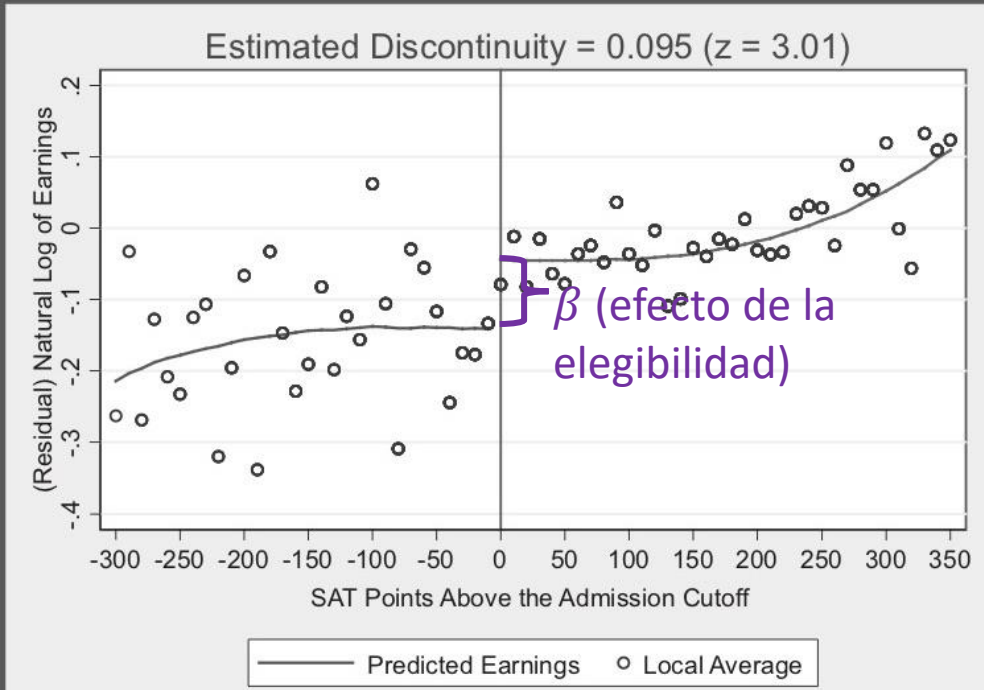
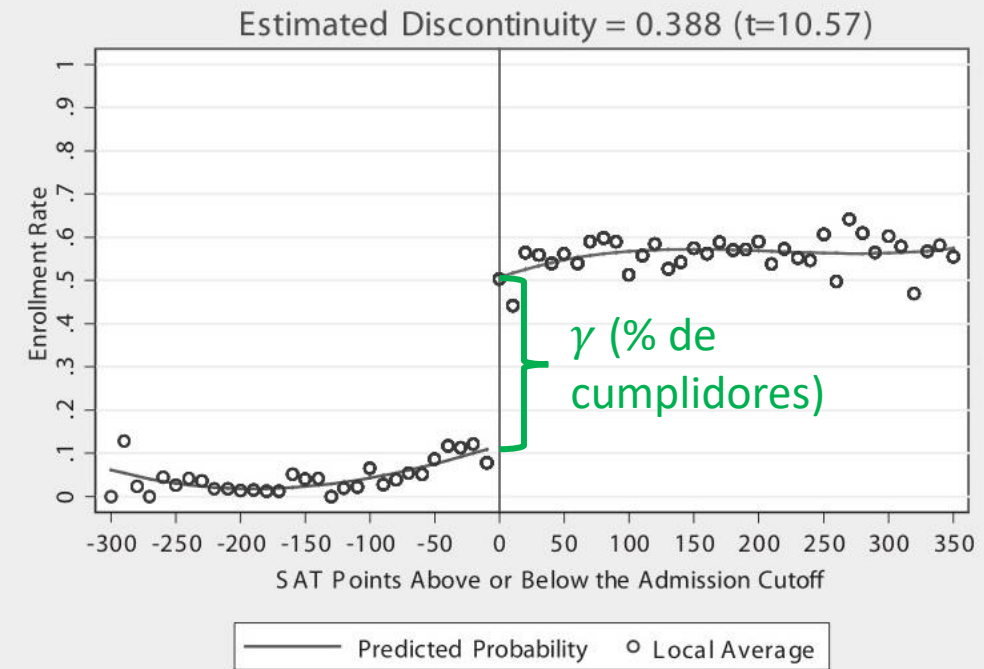
El umbral es el puntaje en el SAT que cada universidad usa como criterio de admisión.

$D_i = 1$  si obtuvo el SAT mínimo para admisión

$T_i = 1$  si el estudiante está admitido en el “Flagship State University”

$Y_i$  es el ingreso del estudiante varios años después

➤ ¿Cómo obtener  $\tau$  a partir de  $\gamma$  y  $\beta$  ?



# La Regresión Discontinua y la Estimación de un Efecto Local

- La Regresión Discontinua estima un Efecto Local del Tratamiento (LATE).
- 1) Nos informa sobre el efecto para un rango específico del índice de elegibilidad (muy cerca del umbral)
- En el último ejemplo, los estudiantes admitidos cerca del umbral son los que tienen un SAT apenas suficiente para estar admitidos: son los “menos buenos” dentro de los admitidos.
- No nos informa sobre el efecto que tendría una admisión a una universidad para los estudiantes más abajo o más arriba en la distribución de notas de SAT.
- 2) En una RD difusa, como cualquier IV, el método estima el efecto sobre los cumplidores.
- En el último ejemplo, los estudiantes que entran en el *Flagship State University* solo si tienen el SAT suficiente, pueden ser diferente de los siempre tratados o nunca tratados
- ¿A que corresponden los siempre tratados o nunca tratados y por qué podrían tener un impacto diferente sobre sus ganancias?

# Los Supuestos de Identificación de la Regresión Discontinua

El supuesto de identificación se puede describir de la manera siguiente:

$$\underbrace{\lim_{x_i \uparrow \bar{x}} (Y_{0i} | D_i = 0, \text{Controles})}_{\text{resultado de no asignados a tratamiento en } \bar{x}} = \underbrace{\lim_{x_i \downarrow \bar{x}} (Y_{0i} | D_i = 1, \text{Controles})}_{\text{contrafactual de los asignados a tratamiento en } \bar{x}}$$

Se supone que los no asignados son un buen contrafactual cuando  $x_i = \bar{x}$

El segundo supuesto esencial es que la forma funcional sea adecuada para estimar  $E[Y_i | x_i = \bar{x}]$  de ambos lados del umbral.

# Robustez y Verificación de los Supuestos de la Regresión Discontinua

## Verificación de las condiciones para la aplicación de la RD

- Ausencia de otros Programas que comparten el mismo índice y umbral de elegibilidad
- Aparte de su tratamiento, ¿existe algún otro programa o cualquier cosa que afecte solamente a los que están a un lado del umbral?

## Verificación de los supuestos de la RD:

- No manipulación (no agrupamiento en la distribución del índice de elegibilidad)
- Continuidad en variables predeterminadas, alrededor del umbral

## Pruebas Típicas de Robustez:

- Robustez a cambios en el ancho de banda aplicado
- Robustez a cambios en la forma funcional



# Verificación de los Supuestos de la Regresión Discontinua:

## 1) Ausencia de otros Programas con mismo Umbral

### Ausencia de otros Programas con mismo Umbral

Ejemplo: Laajaj, R., Moya, A., & Sánchez, F. (2022). Equality of opportunity and human capital accumulation: Motivational effect of a nationwide scholarship in Colombia. *Journal of Development Economics*, 154, 102754.

Other programs that share the same sisben eligibility cutoff.

Program	Benefits	<i>Sisbén</i> Cut-off			Starting year
		14 Cities	Other Urban	Rural	
SPP	Merit and need-base college scholarship	57.21	56.32	40.75	2014
ICBF Primera Infancia	Education and nutrition of young children (0–5 years old)	57.21†	56.32†	40.75†	2007
Vivienda Rural	Rural housing building or improvement	NA	56.32†	40.75†	2013
BEPS	Savings program for the elderly without a pension	57.21†	56.32†	40.75†	2015
Access - Icetex	Long-term credit for tertiary education (tuitions or life expenditures) with subsidized interest rate	57.21†	56.22	40.75†	2013
Atención Humanitaria	Financial aid, food, housing and health services for victims of the armed conflict	57.21†	56.32†	56.32	2011

Como alternativa, usamos el método de Diferencias-en-Discontinuidades

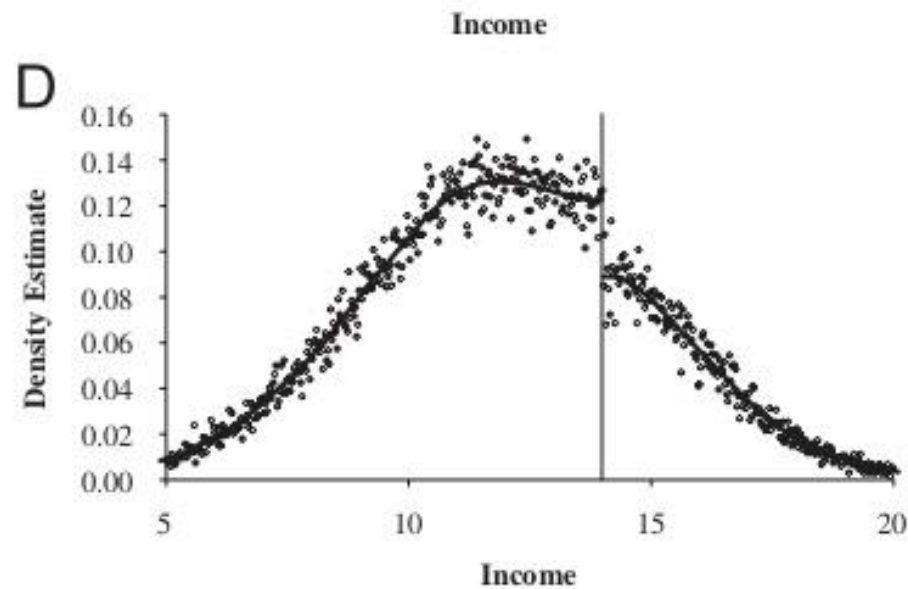
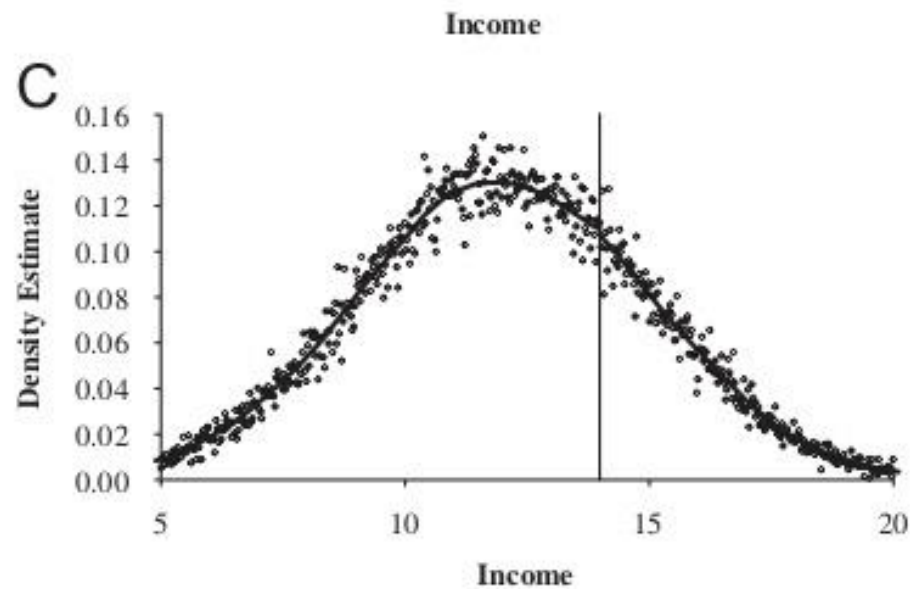
# Las Pruebas de Verificación de los Supuestos de la Regresión Discontinua: 1) La no manipulación

Existe evidencia de manipulación cuando se observa un agrupamiento en la distribución del índice de elegibilidad alrededor del umbral.

Por ejemplo, en un programa de ayuda social. Si los individuos con más conexiones y que parecen no ser elegibles (por muy poco) consiguen que se modifique su índice de elegibilidad.

La prueba más común es la de McCrary (2008)

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Design Test." *Journal of Econometrics* 142: 698–714.



En el caso del Sisbén, muchos contestan para parecer más pobres. ¿Afecta esto la posibilidad de usarlo como índice de elegibilidad para una RD?

# Las Pruebas de Verificación de la RD:

## 2) La continuidad en las variables predeterminadas

- Se puede correr la estimación (2), pero usando como variables explicadas o dependientes las variables predeterminadas (que típicamente sirven de control en la regresión).
- En ciertos casos las variables incluyen  $Y_{t-1}$  (cuando está disponible) o variables como educación de los papás, género, edad, etc.
- Se espera que no cambien de manera significativa ya que las unidades justo arriba y abajo del umbral deberían ser muy similares (“as good as random”) una vez se controla por  $x_i$

# Pruebas Típicas de Robustez (1): Robustez a Cambios en la Forma Funcional

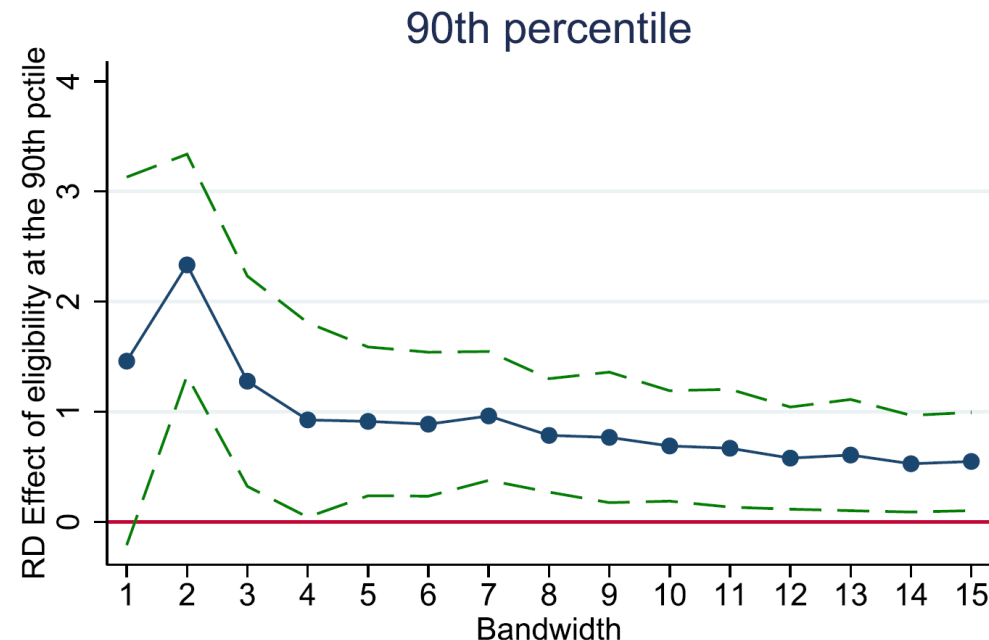
- La decisión de la forma funcional es clave: debe controlar por variación en  $x_i$  de manera suficientemente flexible, sin ser demasiado sensibles a unas pocas observaciones ni a observaciones lejos del umbral.
- Imbens y Lemieux (2008)* recomiendan una estimación local lineal con coeficientes diferentes de cada lado, es decir polinomial 1, pero con Kernel (ponderaciones decrecientes cuando se aleja del umbral).
- Pero en todo caso, el resultado debe ser robusto al uso de otras formas funcionales como otros polinomios.
- Es común incluir estas pruebas de robustez en un apéndice.

## Pruebas Típicas de Robustez (2): Robustez a Cambios en el Ancho de Banda

- Típicamente cuando se estima la ecuación (2), se restringen los datos a un ancho de banda (un rango dentro de los valores del índice de elegibilidad alrededor del umbral).
- Existen métodos, como *Calonico et al. (2014)*, que proponen un ancho de banda óptimo aplicando un arbitraje entre sesgo (si es muy grande) y ruido (si es muy pequeño).
- Pero el resultado debe ser robusto a cambios en el ancho de banda elegido

- Ejemplo de gráfica, replicando la estimación principal con varios anchos de banda:

*Laajaj et al. (2021)*

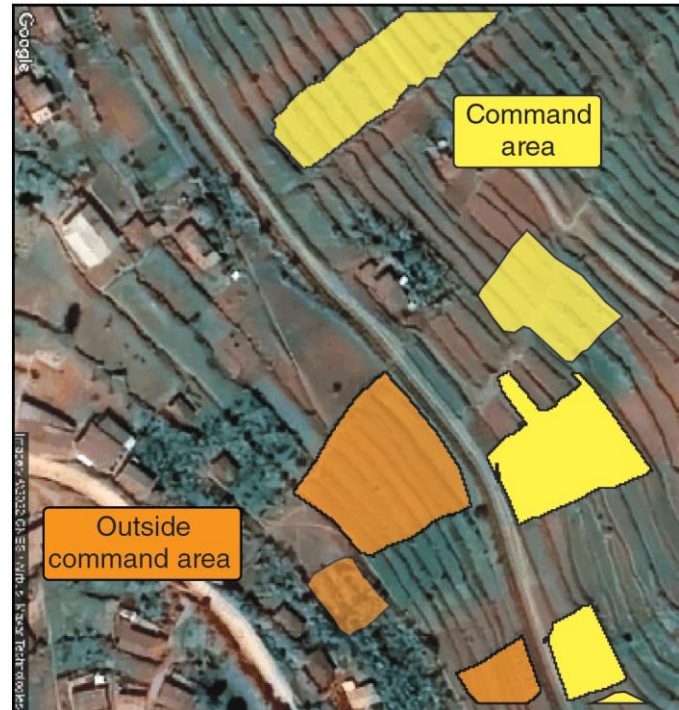


# Ejemplos de RD's comunes

- Elecciones y votos como índice de elegibilidad
- Índice social, como el Sisbén, determina la elegibilidad a un programa\*
- El tamaño de un municipio (u otra regla de participación como % de pobreza...)
- Calificación de una prueba
- Ranking para una admisión
- Coordenadas/fronteras geográficas\*

\* Mucho cuidado con otros cambios que pueden ocurrir en el mismo umbral

Panel A. Plots inside and outside command area



Panel B. Sharp increase in irrigation at boundary

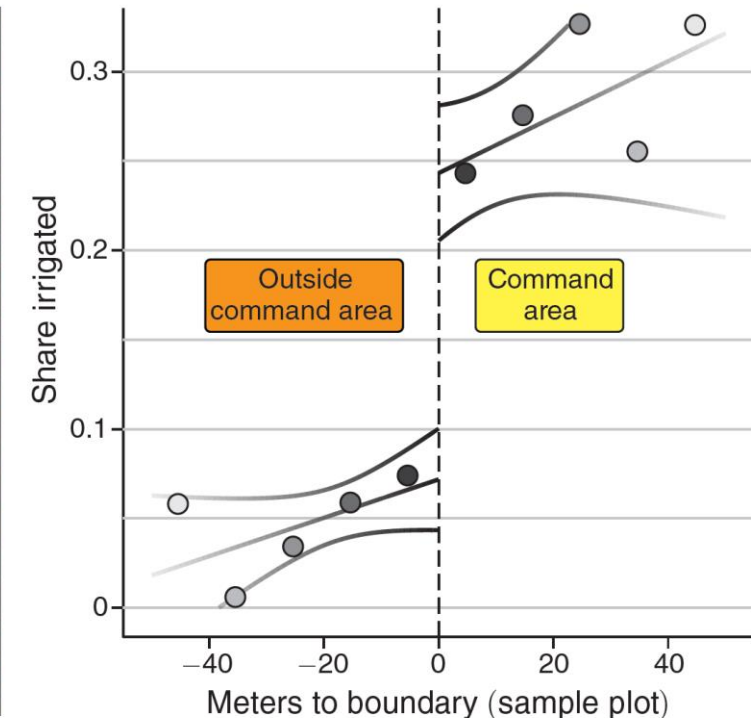


FIGURE 2. ESTIMATING THE IMPACT OF IRRIGATION EXPLOITING SPATIAL DISCONTINUITY IN ACCESS



# ¡Gracias!