Semana 5. Regresión Discontinua Nítida

Equipo Econometría Avanzada

Universidad de los Andes

9 de septiembre de 2022

Contenido

Contexto

Meyersson (2014) utiliza datos de Turquía en 1994 para estudiar si el control político por parte de un partido islámico tiene efectos sobre el acceso a la educación de las mujeres.

- Múltiples candidatos del partido Islámico ganaron las elecciones locales de 1994 en Turquía.
- El alcalde elegido es aquel candidato que obtiene la mayoría de votos.
- El **margen de votos** del candidato del partido islámico es la diferencia entre los votos dicho candidato y el (*otro*) candidato con mayor número de votos.
- Si el **margen de votos es positivo** el candidato del partido islámico fue elegido como alcalde.

Pregunta de investigación

¿Cuál es el efecto del control político por parte de un partido islámico (D_i) sobre el acceso a la educación de las mujeres (y_i)?

$$y_i = \rho D_i + f(x_i) + u_i$$

donde, para cada municipio i,

- y_i: proporción de mujeres entre 15 y 20 años con educación completa.
- x_i: margen de votos del partido islámico.
- $f(\cdot)$: función suave de x_i
- $D_i = \mathbb{1}[x_i \geq 0]$.

Regresión discontinua nítida

El tratamiento *es una función* **discontinua** *y* **determinísica** de una variable x_i (variable de focalización, *running variable* o *forcing variable*).

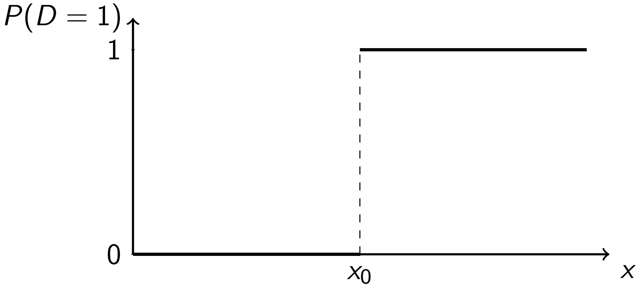
Por ejemplo,

$$D_i = \begin{cases} 1 & \text{si } x_i \geq x_0 \\ 0 & \text{si } x_i < x_0 \end{cases} \tag{1}$$

donde x_0 es el corte que determina el tratamiento.

- También puede ser el caso de que $D_i = \mathbb{1}[x_i \leq x_0]$.
- No hay **ningún** valor de x_i en el que se observen tratados y no tratados al mismo tiempo.

Es decir, en RDN hay un **salto determinístico en la probabilidad de tratamiento:** La probabilidad de ser tratado salta de 0 a 1 (o de 1 a 0) cuando se pasa de un lado a otro del umbral.



Ya veremos qué pasa cuando el salto no es determinístico.

La idea

En términos poblacionales, el efecto de interés está dado por

$$ATE(X_i) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i]$$

No obstante, el problema fundamental de la identificación causal es que no observamos $Y_i(1)$ y $Y_i(0)$ al mismo tiempo para ningún individuo.

En el caso particular de RD, en términos poblacionales

$$\mathbb{E}[Y_i|X_i] = \begin{cases} \mathbb{E}[Y_i(1)|X_i] & \text{si} \quad X_i \geq x_0 \\ \mathbb{E}[Y_i(0)|X_i] & \text{si} \quad X_i < x_0 \end{cases} \tag{2}$$

Es decir, para ningún valor de $X_i = x_i$ observaremos tanto $\mathbb{E}[Y_i(1)|X_i = x_i]$ como $\mathbb{E}[Y_i(0)|X_i = x_i]$.

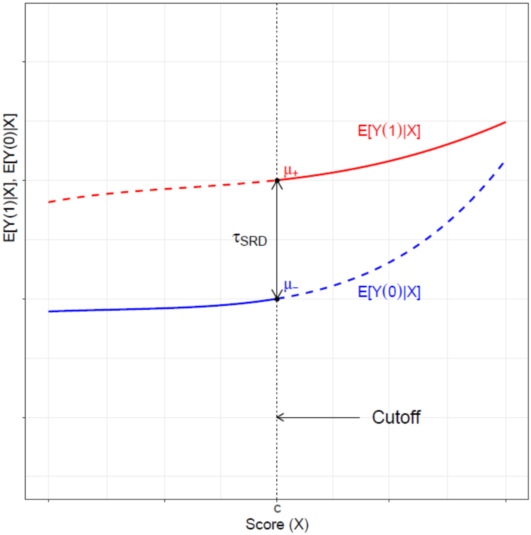


Figure 2: RD Treatment Effect in Sharp RD Design

- Gráficamente, el efecto causal $ATE(x_i)$ es la diferencia entre la curva roja y la curva azul para un valor de $X_i = x_i$. Esto es imposible de calcular en la práctica, pues en la región que observamos la curva roja, no observamos la azul, y viceversa. Sin embargo, algo especial pasa en $X_i = x_0$: casi vemos ambas curvas $\mathbb{E}[Y_i(1)|X_i]$ y $\mathbb{E}[Y_i(0)|X_i]$ simultáneamente.
- Así, si encontráramos la manera de calcular ambas funciones en $X_i = x_0$ (μ^+ y μ^- en la figura), podríamos estimar el efecto causal ($ATE(X_i)$) en dicho punto. La clave para lograrlo es la **continuidad local**.

Continuidad local

Supuesto: Las funciones de los resultados potenciales promedio condicionales en la variable de focalización, $E[Y_i(1)|X_i = x]$ y $E[Y_i(0)|X_i = x]$, son continuas en el punto $x = x_0$.

Bajo este supuesto podemos identificar $ATE(x_0)$:

$$\begin{aligned}
\tau_{RDN} &= \lim_{x\downarrow x_0} \mathbb{E}[Y_i|X_i = x] - \lim_{x\uparrow x_0} \mathbb{E}[Y_i|X_i = x] \\
&= \lim_{x\downarrow x_0} \mathbb{E}[Y_i(1)|X_i = x] - \lim_{x\uparrow x_0} \mathbb{E}[Y_i(0)|X_i = x] \\
&= \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x_0]
\end{aligned}$$

El anterior es el efecto causal del tratamiento *para aquellos individuos tales que* $X_i = x_0$. Esto se conoce como un **Local Average Treatment Effect** (LATE).

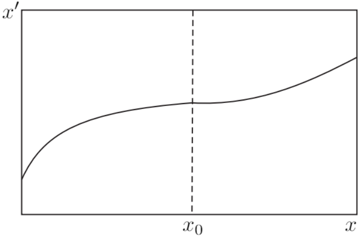
Supuesto de identificación

Si las funciones de resultados potenciales condicionales en X_i son continuas en x_0, debe ser que:

1. Las funciones del valor esperado condicional de las covariables también son continuas en el punto x_0.
    - Los demás determinantes de la variable de resultado no cambian abruptamente en x_0.
2. La densidad de individuos a un lado o el otro del corte es aproximadamente igual.
    - Los individuos no pueden manipular el puntaje de su variable de focalización para quedar a un lado o el otro del corte.
    - Saltos en la densidad de individuos de un lado al otro hace que los resultados sean menos creíbles.

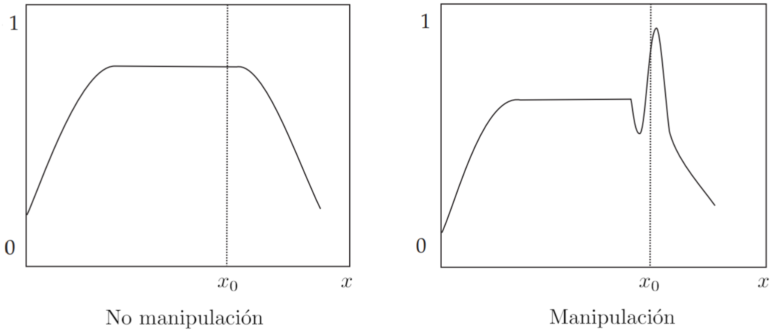
¡Estas condiciones pueden ser verificadas empíricamente!

Continuidad local de otros confounders: Las variables observadas no deben presentar ningún salto en el punto de corte. Esto sugiere que las funciones de resultados potenciales condicionales en X_i son efectivamente continuas.



Continuidad local   No continuidad local

No manipulación: No hay discontinuidad la función de densidad de X_i en x_0. Esto sugiere que las unidades justo debajo del corte son comparables a las unidades justo después.

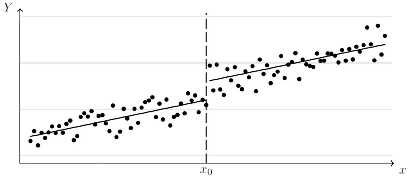


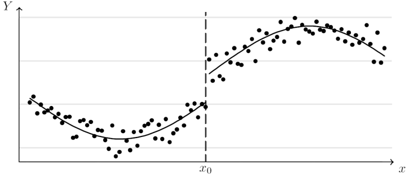
- Ejemplo de manipulación del puntaje SISBEN en Colombia: Camacho y Conover (2011) pág. 43.

Estimación paramétrica: regresión lineal

$$Y_i = f(X_i - x_0) + \rho D_i + u_i \tag{3}$$

- Escenario lineal
- Escenario no lineal



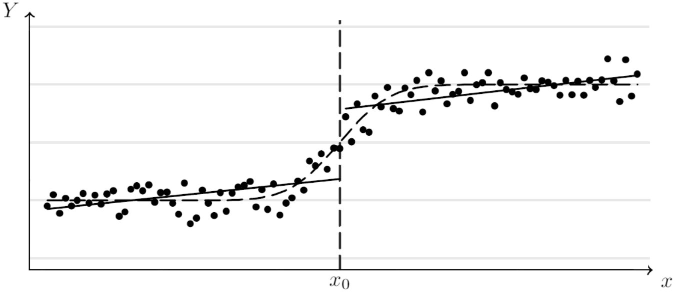


$f(X_i) = \alpha_1 X_i$ ó
$f(X_i) = \alpha_1 X_i + \delta_1 X_i D_i$
$f(X_i) = \alpha_1 X_i + \alpha_2 X_i^2$ ó

$f(X_i) = \alpha_1 X_i + \alpha_2 X_i^2 + \delta_1 X_i D_i + \delta_2 X_i^2 D_i$

Importancia de la forma funcional

La mala especificación de $f(.)$ puede llevar a conclusiones erradas:



- Si hay dos programas que se asignan con el mismo punto de corte, ¿tendría esto implicaciones en la identificación del efecto de uno de los programas?

Estimación no paramétrica: polinomios locales

- Realmente el objetivo de la regresión discontinua es hallar el valor de la discontinuidad, no la forma funcional del valor esperado de Y condicional en X. Esto es, queremos aproximar:

$$\tau = \lim_{\eta \downarrow 0} \mathbb{E}[Y|X_i = x_0 + \eta] - \lim_{\eta \uparrow 0} \mathbb{E}[Y|X_i = x_0 + \eta]$$

- La estimación por MCO busca una aproximación global del valor esperado condicional. Esto supone dos problemas:
  - Las aproximaciones globales son malas para ajustar los extremos del intervalo donde se aproxima (fenómeno de Runge).
  - La regresión pesa a todas las observaciones por igual, aun cuando nos interesan más las observaciones cerca al umbral.
- Por esto la literatura ha convergido en que la estimación más adecuada es a través de polinomios locales. Esta metodología puede usarse en Stata o en R usando el comando *rdrobust* desarrollado por Calonico, Cattaneo, Titiunik (2014).

¿Cómo funcionan los polinomios locales?

Para realizar una estimación de un polinomio local de grado $p \geq 0$, realizamos un esquema de mínimos cuadrados pesados por un "kernel" $K(\cdot)$ en un ancho de banda $h > 0$.

Entonces a la derecha del umbral resolvemos:

$$\min_{\hat{\mu}_+, \hat{\mu}_{+,j}} \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \left[y_i - \hat{\mu}_+ - \sum_{j=1}^{p} \hat{\mu}_{+,j}(X_i - x_0)^j\right]^2 \mathbb{1}[X_i \in (x_0, x_0 + h)]$$

y a la izquierda:

$$\min_{\hat{\mu}_-, \hat{\mu}_{-,j}} \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \left[y_i - \hat{\mu}_- - \sum_{j=1}^{p} \hat{\mu}_{-,j}(X_i - x_0)^j\right]^2 \mathbb{1}[X_i \in (x_0 - h, x_0)]$$

Finalmente:

$$\hat{\tau}_{RDD} = \hat{\mu}_+ - \hat{\mu}_-; \qquad (\hat{\tau}_{RDD} - \tau) \xrightarrow{d} \mathcal{N}(\mathcal{B}, \Sigma)$$

Podemos estimar el sesgo $\hat{\mathcal{B}}$, y corregir por él: $\hat{\tau}_{\text{Bias-corrected}} = \hat{\tau}_{RDD} - \hat{\mathcal{B}}$.

¿Cómo elegir los parámetros?

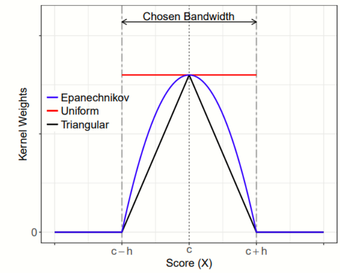
- **Grado del polinomio (p):** Gelman & Imbens (2016) sugieren usar $p = 1, 2$. (Mejora eficiencia del estimador y la cobertura de los intervalos de confianza)

- **Ancho de Banda (h):** Equilibrar el *trade-off* sesgo-varianza. La manera más famosa de hacerlo es minimizar el error cuadrático medio (ECM):

$$\text{ECM} = \text{Sesgo} + \text{Varianza}$$

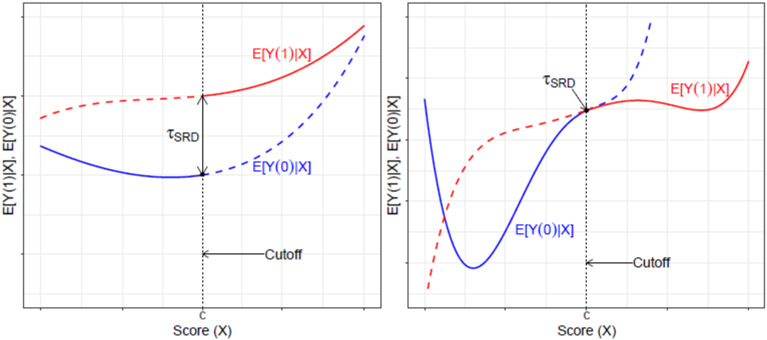
  Un chequeo adicional es que el efecto estimado no sea altamente sensible a h.

- **Kernel ($K(\cdot)$):** Cattaneo, Idrobo, Titiunik (2019) argumentan que usar un kernel triangular en conjunto con un ancho de banda elegido por minimización del ECM resulta en estimadores con propiedades óptimas.



Tipos de Kerneles. Cattaneo, Idrobo, Titiunik (2019)

Localidad del efecto



(a) Mild Heterogeneity

(b) Severe Heterogeneity

Figure 4: Local Nature of the RD Effect

El efecto al final es local porque no sabemos cómo sería el salto en otros puntos distintos a x_0.

Simulador de RDD

- Simulador de RDD para que jueguen.

Conclusión

- En RDN, calculamos el efecto local de una política que asigna un tratamiento mediante una regla particular: una unidad es tratada *si, y solo si*, un atributo suyo (modelado como una variable continua) es mayor a un valor exógeno llamado umbral.
- La ecuación (3) se estima por MCO o por polinomios locales en una vecindad cercana al umbral.
- En la elección del ancho de banda hay un *trade-off* entre sesgo y varianza.
  - Este sesgo se da por el fenómeno de Runge.
- Actualmente, RDN es una de las metodologías más robustas pues podemos validar bastante bien el supuesto de identificación.
- El enfoque tradicional de RDN que revisamos hoy está basado en la continuidad de $E[Y_i(d)|X_i = x]$ para $d \in \{0, 1\}$ en el umbral.
  - Hay un enfoque alternativo que se basa en la aleatorización local que existe de estar a justo un lado o el otro del corte.
  - Los supuestos asociados a este enfoque son distintos.
  - Ver Cattaneo, Titiunik y Vazquez-Bare (2017) para una discusión detallada.

¡Gracias!