

Inferencia, Causalidad y Políticas Públicas

ECO-60116

Week 05: Regresión discontinua II

Eduard F. Martinez Gonzalez, Ph.D.

Departamento de Economía, Universidad Icesi

September 26, 2025

RECAP: Regresión Discontinua (Intuición)

- En RD, usamos la “*arbitrariedad*” de ciertas reglas de asignación para estimar *efectos causales*.
- La *asignación a tratamiento* depende únicamente de una variable *continua* z_i observada que determina la *elegibilidad* a un tratamiento en función de un *umbral* de elegibilidad (conocido) z_0 .

RECAP: Regresión Discontinua (Intuición)

- En RD, usamos la “*arbitrariedad*” de ciertas reglas de asignación para estimar *efectos causales*.
- La *asignación a tratamiento* depende únicamente de una variable *continua* z_i observada que determina la *elegibilidad* a un tratamiento en función de un *umbral* de elegibilidad (conocido) z_0 .
- Supuestos básicos:
 - 1 No es posible *manipular* (perfectamente) la *regla de asignación*.
 - 2 Los resultados potenciales son continuos en la vecindad del *umbral de elegibilidad*:

$$\lim_{h \rightarrow 0} \mathbb{E}[Y_i^j | z_i = z_0 + \eta] = \lim_{\eta \rightarrow 0} \mathbb{E}[Y_i^j | z_i = z_0 - \eta] \quad \text{para } j \in \{1, 0\} \quad (1)$$

RECAP: Regresión Discontinua (Intuición)

- En RD, usamos la “*arbitrariedad*” de ciertas reglas de asignación para estimar *efectos causales*.
- La *asignación a tratamiento* depende únicamente de una variable *continua* z_i observada que determina la *elegibilidad* a un tratamiento en función de un *umbral* de elegibilidad (conocido) z_0 .
- Supuestos básicos:
 - 1 No es posible *manipular* (perfectamente) la *regla de asignación*.
 - 2 Los resultados potenciales son continuos en la vecindad del *umbral de elegibilidad*:

$$\lim_{h \rightarrow 0} \mathbb{E}[Y_i^j | z_i = z_0 + \eta] = \lim_{\eta \rightarrow 0} \mathbb{E}[Y_i^j | z_i = z_0 - \eta] \quad \text{para } j \in \{1, 0\} \quad (1)$$

- En **regresión discontinua nítida**, hacemos un supuesto adicional:
 - Todos los elegibles reciben el tratamiento, y los no elegibles no reciben el tratamiento (“*perfect compliance*”).

Roadmap

1 RECAP: Regresión Discontinua Nitida (RDN)

- Especificación de $f()$
- Elegir la vecindad de Z_0

2 Regresión Discontinua Borrosa (RDB)

3 Aplicación en R

- Building Rural Entrepreneurial Capacities Programme: Trust and Opportunity

¿Por qué importa modelar $f(z)$?

Modelo de partida: $y_i = \beta_0 + f(z_i) + \tau D_i + \epsilon_i$, $D_i = \mathbf{1}\{z_i \geq z_0\}$

Contraste ingenuo cerca del umbral (con $\eta > 0$ pequeño):

$$\Delta Y(\eta) \equiv E[y_i \mid z_0 + \eta] - E[y_i \mid z_0 - \eta] = \tau + \underbrace{f(z_0 + \eta) - f(z_0 - \eta)}_{\text{sesgo por tendencia en } z}$$

¿Por qué importa modelar $f(z)$?

Modelo de partida: $y_i = \beta_0 + f(z_i) + \tau D_i + \epsilon_i$, $D_i = \mathbf{1}\{z_i \geq z_0\}$

Contraste ingenuo cerca del umbral (con $\eta > 0$ pequeño):

$$\Delta Y(\eta) \equiv E[y_i \mid z_0 + \eta] - E[y_i \mid z_0 - \eta] = \tau + \underbrace{f(z_0 + \eta) - f(z_0 - \eta)}_{\text{sesgo por tendencia en } z}$$

Supuesto clave de identificación: Si $E[y_i^0 \mid z]$ y $E[y_i^1 \mid z]$ son *continuas* en z_0 , entonces $f(z)$ es continua en z_0 y

$$\lim_{\eta \downarrow 0} \{f(z_0 + \eta) - f(z_0 - \eta)\} = 0 \quad \Rightarrow \quad \tau = \lim_{z \downarrow z_0} E[Y \mid z] - \lim_{z \uparrow z_0} E[Y \mid z].$$

¿Por qué importa modelar $f(z)$?

Modelo de partida: $y_i = \beta_0 + f(z_i) + \tau D_i + \epsilon_i$, $D_i = \mathbf{1}\{z_i \geq z_0\}$

Contraste ingenuo cerca del umbral (con $\eta > 0$ pequeño):

$$\Delta Y(\eta) \equiv E[y_i | z_0 + \eta] - E[y_i | z_0 - \eta] = \tau + \underbrace{f(z_0 + \eta) - f(z_0 - \eta)}_{\text{sesgo por tendencia en } z}$$

Supuesto clave de identificación: Si $E[y_i^0 | z]$ y $E[y_i^1 | z]$ son *continuas* en z_0 , entonces $f(z)$ es continua en z_0 y

$$\lim_{\eta \downarrow 0} \{f(z_0 + \eta) - f(z_0 - \eta)\} = 0 \quad \Rightarrow \quad \tau = \lim_{z \downarrow z_0} E[Y | z] - \lim_{z \uparrow z_0} E[Y | z].$$

Implicación empírica: Para “quitarse el sesgo de encima” hay que *capturar bien la tendencia suave $f(z)$ en la vecindad del corte*:

- Ajustes *locales* a ambos lados (p. ej., lineal local) en lugar de polinomios globales rígidos.
- Elección de vecindad (h) y forma funcional que hagan $f(z)$ “bien aproximada” justo alrededor de z_0 .

¿Por qué importa modelar $f(z)$?

Modelo de partida: $y_i = \beta_0 + f(z_i) + \tau D_i + \epsilon_i$, $D_i = \mathbf{1}\{z_i \geq z_0\}$

Contraste ingenuo cerca del umbral (con $\eta > 0$ pequeño):

$$\Delta Y(\eta) \equiv E[y_i | z_0 + \eta] - E[y_i | z_0 - \eta] = \tau + \underbrace{f(z_0 + \eta) - f(z_0 - \eta)}_{\text{sesgo por tendencia en } z}$$

Supuesto clave de identificación: Si $E[y_i^0 | z]$ y $E[y_i^1 | z]$ son *continuas* en z_0 , entonces $f(z)$ es continua en z_0 y

$$\lim_{\eta \downarrow 0} \{f(z_0 + \eta) - f(z_0 - \eta)\} = 0 \quad \Rightarrow \quad \tau = \lim_{z \downarrow z_0} E[Y | z] - \lim_{z \uparrow z_0} E[Y | z].$$

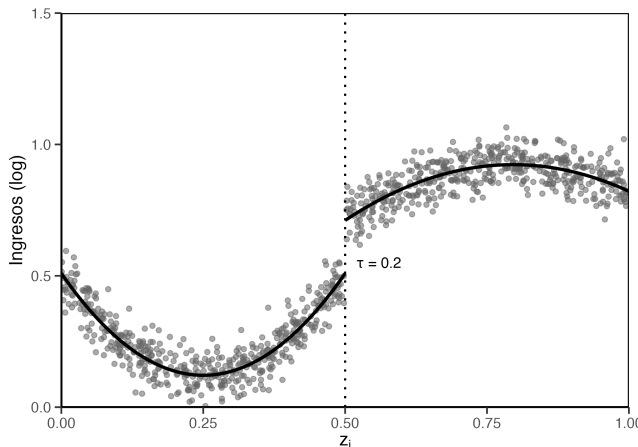
Implicación empírica: Para “quitarse el sesgo de encima” hay que *capturar bien la tendencia suave $f(z)$ en la vecindad del corte*:

- Ajustes *locales* a ambos lados (p. ej., lineal local) en lugar de polinomios globales rígidos.
- Elección de vecindad (h) y forma funcional que hagan $f(z)$ “bien aproximada” justo alrededor de z_0 .

Takeaway: La RD identifica el **salto causal** sólo si el **componente suave** $f(z)$ está bien especificado cerca de z_0 ; si no, la diferencia absorbe tendencia y queda sesgada.

Cuando $f(z)$ está bien especificada (polinomio por lado)

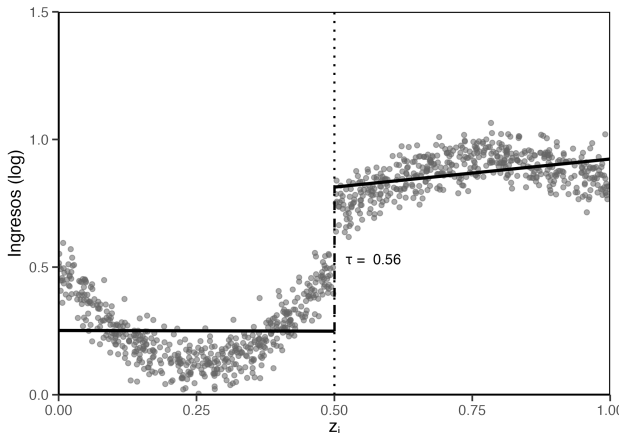
Ecuación: $\log(\text{Ingresos}) = \beta_0 + f(z_i) + \tau D_i + \epsilon_i$



El estimador MCO de τ es un estimador del salto en la variable de resultado que se genera en z_0 .

Cuando $f(z)$ está mal especificada

Ecuación: $\log(\text{Ingresos}) = \beta_0 + \beta_1 z_i + \tau D_i + \epsilon_i$



- Si $f(z)$ se modela mal, el MCO **atribuye a τ** parte de la *curvatura* de $E[Y | z]$ (sesgo de forma funcional).
- La “discontinuidad” estimada **no es el salto causal** en z_0 , sino un artefacto del ajuste global.

Polinomios globales

Idea: aproximar la tendencia $f(z)$ en *todo* el soporte con un polinomio de grado P .

$$\underbrace{f(z_i)}_{\text{tendencia}} = \sum_{p=1}^P \beta_p z_i^p$$

Entonces,

$$y_i = \beta_0 + \sum_{p=1}^P \beta_p z_i^p + \tau D_i + \varepsilon_i$$

- **Pros:** fácil de estimar (MCO), más flexibilidad al subir P .
- **Contras:** sensibilidad a P , oscilaciones y *peso excesivo* de datos lejanos al umbral \Rightarrow sesgo.
- **Recomendación:** como máximo grados **2–3** y, mejor aún, **polinomios locales por lado** con ventana (h) adecuada.

Referencia: Gelman & Imbens (2019), “Why High-Order Polynomials Should Not Be Used in RD”.

Roadmap

1 RECAP: Regresión Discontinua Nitida (RDN)

- Especificación de $f()$
- Elegir la vecindad de Z_0

2 Regresión Discontinua Borrosa (RDB)

3 Aplicación en R

- Building Rural Entrepreneurial Capacities Programme: Trust and Opportunity

Estimación en la vecindad del umbral z_0

- Usamos sólo $|z_i - z_0| \leq h$: cerca del corte, $f(z)$ es suave.
- La comparación izquierda–derecha en esa vecindad aproxima el *salto causal*.
- Justo bajo el umbral \Rightarrow buen contrafactual para justo sobre el umbral.

$$\hat{\tau}(h) = \bar{Y}_R(h) - \bar{Y}_L(h), \quad z \in [z_0 - h, z_0 + h]$$

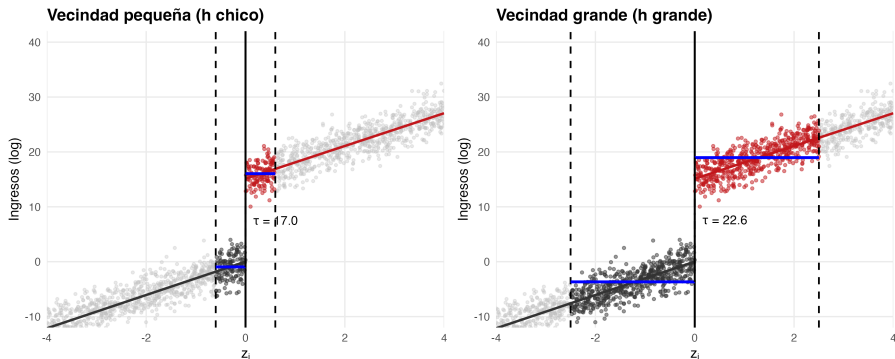
El problema es que a medida que h se acerca a cero, vamos perdiendo observaciones \rightarrow **menor precisión** del estimador.

Estimación para un ancho de banda h :

$$\hat{\tau}(h) = \underbrace{\left(\frac{1}{n_R(h)} \sum_{i: 0 \leq z_i - z_0 \leq h} Y_i \right)}_{\bar{Y}_R(h)} - \underbrace{\left(\frac{1}{n_L(h)} \sum_{i: -h \leq z_i - z_0 < 0} Y_i \right)}_{\bar{Y}_L(h)}$$

donde usamos sólo $z_i \in [z_0 - h, z_0 + h]$.

Sensibilidad del estimador a diferentes h



Nota: h controla el compromiso **sesgo-varianza**: h pequeño \rightarrow menos sesgo y más varianza; h grande \rightarrow más sesgo y menos varianza. (Puede ponderarse por distancia con kernels.)

Regresión local: combinar vecindad + forma funcional

En la práctica, usamos una combinación de los dos métodos: restringimos las observaciones a una vecindad de z_0 y usamos alguna aproximación de $f()$.

Tres pasos para la implementación:

- 1 Seleccionar el orden del polinomio (p) y una función de pesos o función de Kernel $K(.)$
- 2 Seleccionar un ancho de banda (h).
- 3 Estimar los parámetros minimizando el cuadrado de los errores ponderados y reportar $\hat{\tau} = \hat{\alpha}^+ - \hat{\alpha}^-$

Para seleccionar el ancho de banda se pueden usar métodos alternativos para la inferencia estadística (Calonico et al., 2014).

Roadmap

1 RECAP: Regresión Discontinua Nitida (RDN)

- Especificación de $f()$
- Elegir la vecindad de Z_0

2 Regresión Discontinua Borrosa (RDB)

3 Aplicación en R

- Building Rural Entrepreneurial Capacities Programme: Trust and Opportunity

Regresión Discontinua Borrosa (RDB)

- En **regresión discontinua borrosa** relajamos el supuesto de “*perfect compliance*”: hay una relación probabilística entre participación y z_i .
- La **probabilidad** de estar en el grupo de tratamiento o de control cambia de forma **discontinua** en el umbral. Para un $\eta > 0$,

$$\lim_{\eta \rightarrow 0} \Pr[D_i = 1 | z_i = z_0 - \eta] \neq \lim_{\eta \rightarrow 0} \Pr[D_i = 1 | z_i = z_0 + \eta]$$

- De forma general:

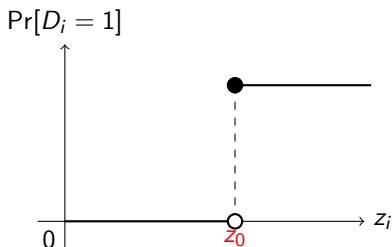
$$\Pr[D_i = 1 | z_i] = \begin{cases} g_1(z_i) & \text{si } z_i \geq z_0 \\ g_0(z_i) & \text{si } z_i < z_0 \end{cases}$$

donde $g_1(z_0) \neq g_0(z_0)$.

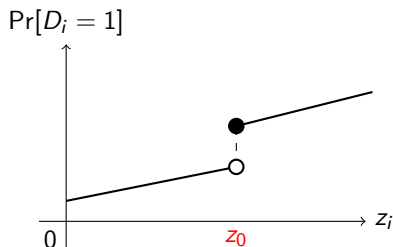
- Podemos ser agnósticos sobre las formas funcionales de g_0 y g_1 , pero la probabilidad debe “saltar” en el umbral z_0 .

Discontinuidad de la probabilidad de ser tratado

La **asignación al tratamiento** ya no es determinística: otros factores (observados y no observados) pueden afectar la probabilidad de estar en el grupo de tratamiento o de control.



(a) RDN



(b) RDB

Relajando los supuestos de RDN

La probabilidad de ser tratado “salta” exactamente en el punto en el que la variable dummy z_i^* pasa de cero a uno.

- z_i^* está definida por un umbral arbitrario y puede ser usada como una **variable instrumental, condicional en z_i** : afecta la probabilidad de ser tratado, pero es independiente de los resultados potenciales.
- Hacemos un supuesto de **monotonicidad local**: el efecto del instrumento (estar a un lado u otro de umbral) sobre la probabilidad de ser tratado va en la misma dirección para todas las unidades.
- Es local en la medida que el instrumento sólo se “activa” en la vecindad de z_0 .

Identificación (I)

- Ecuación estructural:

$$y_i = \beta_0 + f(z_i) + \tau D_i + \epsilon_i \quad (1)$$

- Primera etapa:

$$D_i = \gamma_0 + g(z_i) + \pi z_i^* + \eta_i \quad (2)$$

donde

- ▶ D_i : tratamiento **recibido** (endógeno).
 - ▶ z_i^* : indicador de **elegibilidad** (instrumento exógeno).
 - ▶ $f(\cdot)$ y $g(\cdot)$: funciones suaves de z_i (misma base por lado y misma vecindad).
- La magnitud del salto en la probabilidad en el umbral de elegibilidad, capturado por π , nos indica la “fuerza” de la primera etapa.
 - Si la forma funcional de $g(z_i)$ y $f(z_i)$ coinciden, tenemos un estimador de MC2E.

Identificación (II)

- Suponga que ambas siguen un polinomio del mismo grado

$$f(z_i) = \sum_{p=1}^P \beta_p z_i^p, \quad g(z_i) = \sum_{p=1}^P \gamma_p z_i^p. \quad (3)$$

- Podemos reemplazar D_i en la Ecuación (1) usando (2) y escribir la forma reducida:

$$y_i = \alpha_0 + h(z_i) + \rho z_i^* + \zeta_i, \quad (4)$$

donde

- ▶ $\alpha_0 \equiv \beta_0 + \gamma_0 \tau$
- ▶ $h(z_i) \equiv \sum_{p=1}^P (\beta_p + \gamma_p \tau) z_i^p$
- ▶ $\zeta_i \equiv \epsilon_i + \eta_i \tau$
- ▶ $\rho \equiv \tau \pi$.

- En ese caso, el efecto causal también se puede recuperar via:

$$\tau = \frac{\rho}{\pi} \quad (5)$$

Interpretación del Efecto

- El tratamiento no afecta a todos los individuos (*"imperfect compliance"*), así que tenemos que **ajustar** la magnitud del salto en y_i por el cambio en la probabilidad de ser tratado. Similar al *estimador de Wald*.
- Sin **ajustar**, el estimador tiene una interpretación *ITT*: efecto de ser *elegible* sobre y_i .
- El efecto estimado por RDB es un *LATE*: efecto sobre aquellos individuos cuya condición de tratamiento puede ser afectada por z_i^* .
- **No obstante:**
 - ▶ La identificación sigue dependiendo de la capacidad de *separar* la discontinuidad que se genera en y_i cuando z_i^* pasa de cero a uno, de las funciones continuas $f(z_i)$ y $g(z_i) \implies$ es importante especificar bien la forma funcional que usamos.
 - ▶ Igual que RDN, a medida que η se acerca a cero, vamos perdiendo observaciones y el estimador se hace más impreciso.
 - ▶ Dado el supuesto de continuidad en la vecindad de z_0 , aplican los mismos métodos para verificar supuestos que en RDN.

Roadmap

1 RECAP: Regresión Discontinua Nitida (RDN)

- Especificación de $f()$
- Elegir la vecindad de Z_0

2 Regresión Discontinua Borrosa (RDB)

3 Aplicación en R

- Building Rural Entrepreneurial Capacities Programme: Trust and Opportunity

Laajaj et al. (2023): RD difusa en un programa rural

- **Contexto:** programa de desarrollo rural con cofinanciación y asistencia técnica para planes de negocio de organizaciones, desplegado en 203 municipios. *Mecanismo de selección por ranking y cupos municipales.*
- **Regla de clase (simplificación a individuos):**
 - ▶ Cada municipio ordena a las personas por *puntaje* (idoneidad propuesta, vulnerabilidad, etc.).
 - ▶ Cupo fijo: **subsidios de COP 50 millones** a las propuestas con un puntaje superior a 75 puntos.
- **Diseño (FRD):** $\Pr(D=1 \mid z)$ salta en c_m pero < 1 (*take-up* imperfecto).
- **Variables clave:** $Z_{im}^* = \mathbf{1}\{z_{im} \geq c_m\}$ (elegible, instrumento) ; D_{im} (recibe subsidio) ; Y_{im} (resultado).
- **Lectura de resultados:** *LATE* en el umbral para *compliers* municipales (quienes toman el subsidio cuando Z^* cambia de 0 a 1).

Replication

- **Script R (descargable):** pipeline paso a paso para clase (`rd_fuzzy_class.R`). [Descargar script](#)