

Repaso Regresión Discontinua e Introducción a Diferencias en Discontinuidades

Econometría Aplicada y Ciencia de Datos - CIDE Fall 2026

Carlos Brito

⁰El material de esta presentación se basa en los siguientes materiales: Huntington-Klein, N. (2022). *The Effect, An Introduction to Research Design and Causality*. Chapman & Hall; Cattaneo, M., Idrobo, N. & Titiunik, R. (2020). *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press; Grembi, V., Nannicini, T. & Troiano, U. (2016). Do Fiscal Rules Matter? *American Economic Journal: Applied Economics*.

Repaso: regresión discontinua

Regresión discontinua (RD)

- El tratamiento se asigna según una regla: una **variable de asignación** (*running variable*) X cruza un **umbral** c .
 - Becas si la calificación del examen $\geq c$; programa social si el puntaje de pobreza $\leq c$; reglas que cambian según la población del municipio.

Regresión discontinua (RD)

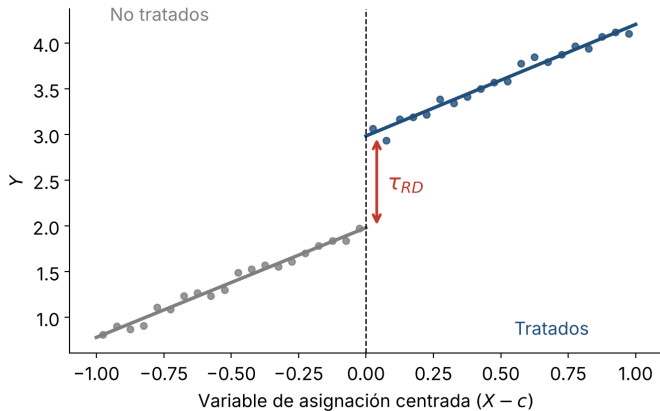
- El tratamiento se asigna según una regla: una **variable de asignación** (*running variable*) X cruza un **umbral** c .
 - Becas si la calificación del examen $\geq c$; programa social si el puntaje de pobreza $\leq c$; reglas que cambian según la población del municipio.
- **Idea:** justo alrededor del umbral, las unidades son prácticamente iguales; lo único que cambia es si reciben el tratamiento.
 - ↪ Las unidades justo debajo del umbral son el contrafactual de las que están justo arriba.

Regresión discontinua (RD)

- El tratamiento se asigna según una regla: una **variable de asignación** (*running variable*) X cruza un **umbral** c .
 - Becas si la calificación del examen $\geq c$; programa social si el puntaje de pobreza $\leq c$; reglas que cambian según la población del municipio.
- **Idea:** justo alrededor del umbral, las unidades son prácticamente iguales; lo único que cambia es si reciben el tratamiento.
 - $\hookrightarrow$ Las unidades justo debajo del umbral son el contrafactual de las que están justo arriba.
- El efecto se estima como el “salto” en Y en el umbral:

$$\tau_{RD} = \lim_{x \downarrow c} E[Y \mid X = x] - \lim_{x \uparrow c} E[Y \mid X = x]$$

RD gráficamente



- Ajustamos una línea a cada lado del umbral; la distancia entre ambas en $X = c$ es el efecto del tratamiento.

Supuesto clave: continuidad

- Los resultados potenciales $E[Y(0) \mid X = x]$ y $E[Y(1) \mid X = x]$ son **continuos** en $x = c$.
 - Si no hubiera tratamiento, no habría salto en Y en el umbral.
 - Todo lo demás que afecta a Y cambia de forma suave alrededor de c .

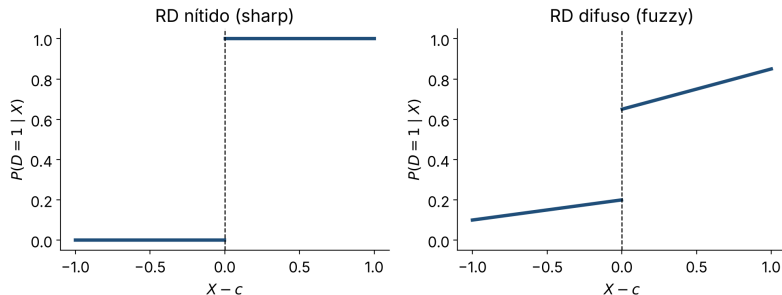
Supuesto clave: continuidad

- Los resultados potenciales $E[Y(0) | X = x]$ y $E[Y(1) | X = x]$ son **continuos** en $x = c$.
 - Si no hubiera tratamiento, no habría salto en Y en el umbral.
 - Todo lo demás que afecta a Y cambia de forma suave alrededor de c .
- ¿Qué puede romper este supuesto?
 - 1 **Manipulación:** las unidades pueden “acomodarse” de un lado del umbral (e.g. reportar menos ingreso para calificar).
 - 2 **Otros tratamientos en el mismo umbral:** algo más cambia exactamente en c . (Esto motiva diff-in-disc.)

Supuesto clave: continuidad

- Los resultados potenciales $E[Y(0) | X = x]$ y $E[Y(1) | X = x]$ son **continuos** en $x = c$.
 - Si no hubiera tratamiento, no habría salto en Y en el umbral.
 - Todo lo demás que afecta a Y cambia de forma suave alrededor de c .
- ¿Qué puede romper este supuesto?
 - 1 **Manipulación:** las unidades pueden “acomodarse” de un lado del umbral (e.g. reportar menos ingreso para calificar).
 - 2 **Otros tratamientos en el mismo umbral:** algo más cambia exactamente en c . (Esto motiva diff-in-disc.)
- El efecto estimado es **local**: es el efecto para las unidades cerca del umbral (LATE en $X = c$).

RD nítido vs. RD difuso



- **Nítido:** cruzar el umbral determina el tratamiento por completo.
- **Difuso:** cruzar el umbral solo aumenta la probabilidad de tratamiento $\Rightarrow$ se estima como variables instrumentales (el umbral es el instrumento):

$$\tau_{FRD} = \frac{\text{salto en } E[Y | X]}{\text{salto en } P(D = 1 | X)}$$

- Regresión lineal local dentro de una ventana (*bandwidth*) h alrededor del umbral:

$$Y_i = \beta_0 + \tau D_i + \beta_1(X_i - c) + \beta_2 D_i(X_i - c) + \varepsilon_i, \quad |X_i - c| \leq h$$

donde $D_i = 1[X_i \geq c]$.

- Regresión lineal local dentro de una ventana (*bandwidth*) h alrededor del umbral:

$$Y_i = \beta_0 + \tau D_i + \beta_1(X_i - c) + \beta_2 D_i(X_i - c) + \varepsilon_i, \quad |X_i - c| \leq h$$

donde $D_i = 1[X_i \geq c]$.

- τ es el estimador de RD. La interacción permite una pendiente distinta a cada lado.

- Regresión lineal local dentro de una ventana (*bandwidth*) h alrededor del umbral:

$$Y_i = \beta_0 + \tau D_i + \beta_1(X_i - c) + \beta_2 D_i(X_i - c) + \varepsilon_i, \quad |X_i - c| \leq h$$

donde $D_i = 1[X_i \geq c]$.

- τ es el estimador de RD. La interacción permite una pendiente distinta a cada lado.
- Decisiones del investigador:
 - **Ancho de banda** h : más chico $\Rightarrow$ menos sesgo, más varianza. Usar selección óptima (e.g. `rdrobust`).
 - **Forma funcional**: preferir polinomios de grado bajo (lineal o cuadrático).
 - **Kernel**: dar más peso a observaciones cerca del umbral (triangular).

Práctica moderna en RD: `rdrobust` y familia

- **No usar polinomios globales de grado alto** (Gelman & Imbens, 2019): dan pesos extraños a observaciones lejanas al umbral. Preferir lineal local (o cuadrático).

Práctica moderna en RD: rdrobust y familia

- **No usar polinomios globales de grado alto** (Gelman & Imbens, 2019): dan pesos extraños a observaciones lejanas al umbral. Preferir lineal local (o cuadrático).
- **Ancho de banda óptimo y basado en datos:** rdrobust elige h que minimiza el error cuadrático medio (MSE).

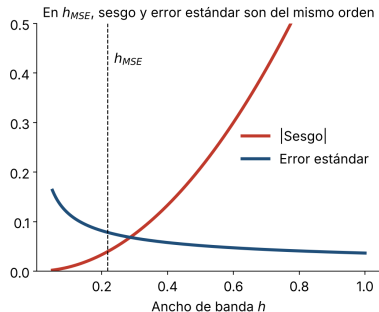
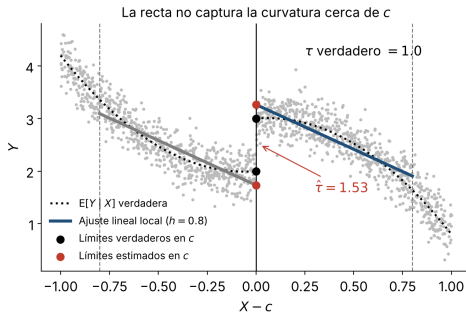
Práctica moderna en RD: rdrobust y familia

- **No usar polinomios globales de grado alto** (Gelman & Imbens, 2019): dan pesos extraños a observaciones lejanas al umbral. Preferir lineal local (o cuadrático).
- **Ancho de banda óptimo y basado en datos:** rdrobust elige h que minimiza el error cuadrático medio (MSE).
- **Inferencia robusta con corrección de sesgo** (Calonico, Cattaneo & Titiunik, 2014):
 - El h óptimo acepta algo de sesgo a cambio de menos varianza. Pero entonces $\hat{\tau}$ no está centrado en τ , y el intervalo $\hat{\tau} \pm 1.96 \hat{se}$ queda desplazado: **dice 95% pero cubre menos** (e.g. si sesgo = se, cubre $\approx 83\%$).
 - Para inferencia (IC y p-valores), usar la fila **Robust** de la salida de rdrobust.

Práctica moderna en RD: rdrobust y familia

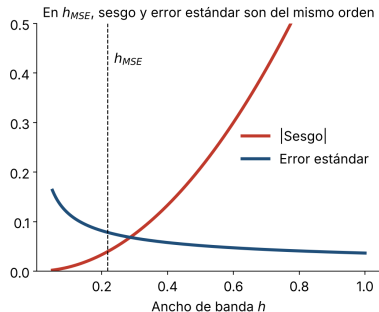
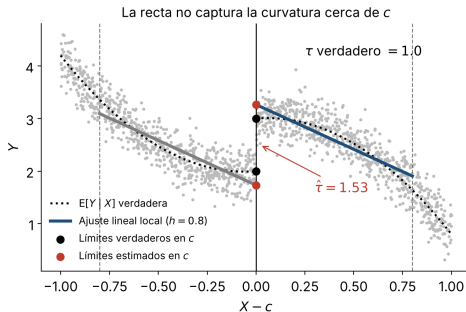
- **No usar polinomios globales de grado alto** (Gelman & Imbens, 2019): dan pesos extraños a observaciones lejanas al umbral. Preferir lineal local (o cuadrático).
- **Ancho de banda óptimo y basado en datos:** rdrobust elige h que minimiza el error cuadrático medio (MSE).
- **Inferencia robusta con corrección de sesgo** (Calonico, Cattaneo & Titiunik, 2014):
 - El h óptimo acepta algo de sesgo a cambio de menos varianza. Pero entonces $\hat{\tau}$ no está centrado en τ , y el intervalo $\hat{\tau} \pm 1.96 \hat{se}$ queda desplazado: **dice 95% pero cubre menos** (e.g. si sesgo = se, cubre $\approx 83\%$).
 - Para inferencia (IC y p-valores), usar la fila **Robust** de la salida de rdrobust.
- **Paquetes complementarios** (Stata, R y Python):
 - rdplot: gráfica con número de bins elegido con datos.
 - rddensity: prueba de manipulación (Cattaneo, Jansson & Ma, 2020).
 - rdlocrand (aleatorización local, útil con pocas obs. o X discreta) y rdmulti (múltiples umbrales).

¿Qué sesgo corrige rdrobust?



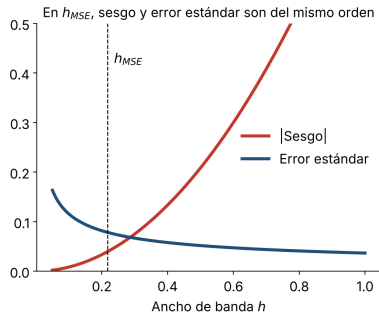
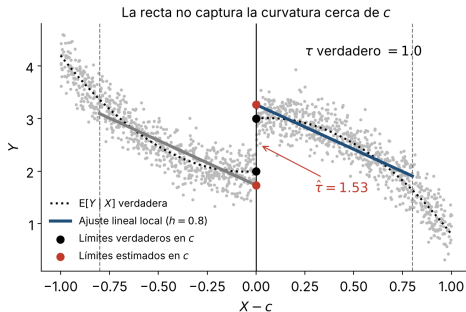
- **Sesgo de aproximación:** la recta local no captura la curvatura de $E[Y | X]$; el sesgo crece con h

¿Qué sesgo corrige rdrobust?



- **Sesgo de aproximación:** la recta local no captura la curvatura de $E[Y | X]$; el sesgo crece con h
- Con h_{MSE} el sesgo es del mismo orden que el error estándar $\Rightarrow \hat{\tau} \pm 1.96 \hat{s\hat{e}}$ cubre menos del 95%.

¿Qué sesgo corrige rdrobust?



- **Sesgo de aproximación:** la recta local no captura la curvatura de $E[Y | X]$; el sesgo crece con h
- Con h_{MSE} el sesgo es del mismo orden que el error estándar $\Rightarrow \hat{\tau} \pm 1.96 \hat{s\hat{e}}$ cubre menos del 95%.
- **CCT (2014):** (1) estimar el sesgo con un polinomio local cuadrático y restarlo; (2) ajustar el error estándar por el ruido de esa estimación.

- **Densidad de la variable de asignación** (McCrary; Cattaneo, Jansson & Ma):
 - Si hay un “amontonamiento” de observaciones justo de un lado del umbral, hay evidencia de manipulación.

- **Densidad de la variable de asignación** (McCrary; Cattaneo, Jansson & Ma):
 - Si hay un “amontonamiento” de observaciones justo de un lado del umbral, hay evidencia de manipulación.
- **Balance en covariables predeterminadas:**
 - Variables que no pueden ser afectadas por el tratamiento (edad, sexo, resultados pasados) no deberían saltar en c .

- **Densidad de la variable de asignación** (McCrary; Cattaneo, Jansson & Ma):
 - Si hay un “amontonamiento” de observaciones justo de un lado del umbral, hay evidencia de manipulación.
- **Balance en covariables predeterminadas:**
 - Variables que no pueden ser afectadas por el tratamiento (edad, sexo, resultados pasados) no deberían saltar en c .
- **Umbrales placebo:** no debería haber saltos en valores de X donde no cambia nada.

- **Densidad de la variable de asignación** (McCrary; Cattaneo, Jansson & Ma):
 - Si hay un “amontonamiento” de observaciones justo de un lado del umbral, hay evidencia de manipulación.
- **Balance en covariables predeterminadas:**
 - Variables que no pueden ser afectadas por el tratamiento (edad, sexo, resultados pasados) no deberían saltar en c .
- **Umbrales placebo:** no debería haber saltos en valores de X donde no cambia nada.
- **Sensibilidad** al ancho de banda y a la forma funcional.

Diferencias en discontinuidades

El problema: tratamientos compuestos

- Muchas veces, **varias políticas usan el mismo umbral.**

El problema: tratamientos compuestos

- Muchas veces, **varias políticas usan el mismo umbral**.
- *Ejemplo:* en Italia, el umbral de 5,000 habitantes determina:
 - Si el municipio está sujeto a reglas fiscales (Pacto de Estabilidad Interna, desde 2001).
 - El salario del alcalde, el tamaño del concejo municipal, el sistema electoral, ...

El problema: tratamientos compuestos

- Muchas veces, **varias políticas usan el mismo umbral**.
- *Ejemplo:* en Italia, el umbral de 5,000 habitantes determina:
 - Si el municipio está sujeto a reglas fiscales (Pacto de Estabilidad Interna, desde 2001).
 - El salario del alcalde, el tamaño del concejo municipal, el sistema electoral, ...
- Un RD en 5,000 captura el efecto de **todas** las políticas a la vez:

$$\tau_{RD} = \tau_{\text{reglas fiscales}} + \tau_{\text{otras políticas}}$$

↪ **Problema:** se viola el supuesto de continuidad para aislar el efecto de *una* política.

El problema: tratamientos compuestos

- Muchas veces, **varias políticas usan el mismo umbral**.
- *Ejemplo*: en Italia, el umbral de 5,000 habitantes determina:
 - Si el municipio está sujeto a reglas fiscales (Pacto de Estabilidad Interna, desde 2001).
 - El salario del alcalde, el tamaño del concejo municipal, el sistema electoral, ...
- Un RD en 5,000 captura el efecto de **todas** las políticas a la vez:

$$\tau_{RD} = \tau_{\text{reglas fiscales}} + \tau_{\text{otras políticas}}$$

↪ **Problema**: se viola el supuesto de continuidad para aislar el efecto de *una* política.

- Es el mismo problema que vimos al comparar antes y después: *otras cosas cambian al mismo tiempo*.

La idea de diff-in-disc

- Si la política de interés **entra en vigor en un momento del tiempo**, y las otras políticas ya estaban ahí, podemos combinar RD con DiD.

La idea de diff-in-disc

- Si la política de interés **entra en vigor en un momento del tiempo**, y las otras políticas ya estaban ahí, podemos combinar RD con DiD.
- **Antes** de la política, la discontinuidad en el umbral solo refleja a las políticas confusoras:

$$\tau_{pre} = \tau_{\text{otras políticas}}$$

- **Después** de la política, la discontinuidad refleja ambas:

$$\tau_{post} = \tau_{\text{política}} + \tau_{\text{otras políticas}}$$

La idea de diff-in-disc

- Si la política de interés **entra en vigor en un momento del tiempo**, y las otras políticas ya estaban ahí, podemos combinar RD con DiD.
- **Antes** de la política, la discontinuidad en el umbral solo refleja a las políticas confusoras:

$$\tau_{pre} = \tau_{\text{otras políticas}}$$

- **Después** de la política, la discontinuidad refleja ambas:

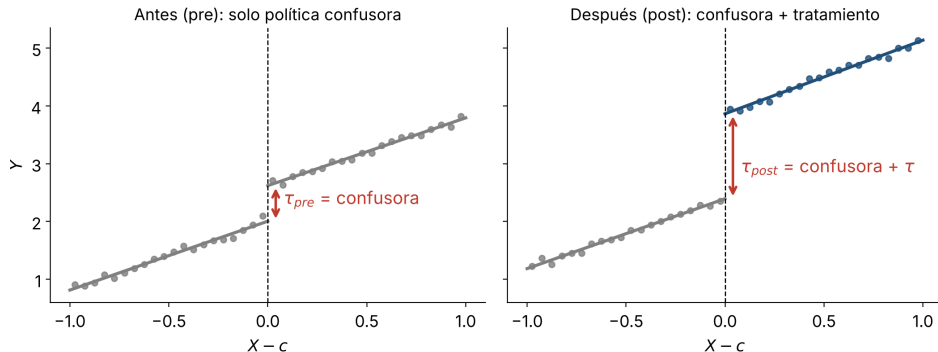
$$\tau_{post} = \tau_{\text{política}} + \tau_{\text{otras políticas}}$$

- Diferencias en discontinuidades:

$$\tau_{DiDisc} = \tau_{post} - \tau_{pre} = \tau_{\text{política}}$$

- La lógica es idéntica a DiD: el “salto” de antes funciona como el grupo de control.

Diff-in-disc gráficamente



- El salto *pre* mide el efecto de la política confusora. El salto *post* mide confusora + tratamiento.
- La diferencia entre ambos saltos es el efecto de la política de interés.

	DiD	RD	Diff-in-disc
Comparación entre grupos	Tratados vs. control	Arriba vs. abajo de c	Arriba vs. abajo de c
Comparación en el tiempo	Antes vs. después	—	Antes vs. después
Supuesto clave	Tendencias paralelas	Continuidad en c	Salto confusor constante
Efecto estimado	ATT	Local en c	Local en c

- 1 **Continuidad** de los resultados potenciales en el umbral, *dentro de cada periodo*, excepto por las políticas que cambian en c .

- ① **Continuidad** de los resultados potenciales en el umbral, *dentro de cada periodo*, excepto por las políticas que cambian en c .
- ② **El efecto de las políticas confusoras en el umbral no cambia en el tiempo.**
 - Análogo a tendencias paralelas: el salto *pre* es un buen contrafactual del salto que habría habido *post* sin la política.

- ① **Continuidad** de los resultados potenciales en el umbral, *dentro de cada periodo*, excepto por las políticas que cambian en c .
- ② **El efecto de las políticas confusoras en el umbral no cambia en el tiempo.**
 - Análogo a tendencias paralelas: el salto *pre* es un buen contrafactual del salto que habría habido *post* sin la política.
- ③ **No hay interacción** entre la política de interés y las confusoras (o, si la hay, el efecto se interpreta como el efecto *dado* que las otras políticas están vigentes).

Supuestos de identificación

- ❶ **Continuidad** de los resultados potenciales en el umbral, *dentro de cada periodo*, excepto por las políticas que cambian en c .
- ❷ **El efecto de las políticas confusoras en el umbral no cambia en el tiempo.**
 - Análogo a tendencias paralelas: el salto *pre* es un buen contrafactual del salto que habría habido *post* sin la política.
- ❸ **No hay interacción** entre la política de interés y las confusoras (o, si la hay, el efecto se interpreta como el efecto *dado* que las otras políticas están vigentes).
- ❹ **No hay manipulación** de la variable de asignación como respuesta a la nueva política.
 - Revisar la densidad de X antes y después.

- Interactuamos la regresión de RD con una dummy $Post_t$:

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 (X_i - c) + \beta_3 D_i (X_i - c) \\ + Post_t [\alpha_0 + \alpha_1 D_i + \alpha_2 (X_i - c) + \alpha_3 D_i (X_i - c)] + \varepsilon_{it}, \quad |X_i - c| \leq h$$

- Interactuamos la regresión de RD con una dummy $Post_t$:

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 (X_i - c) + \beta_3 D_i (X_i - c) \\ + Post_t [\alpha_0 + \alpha_1 D_i + \alpha_2 (X_i - c) + \alpha_3 D_i (X_i - c)] + \varepsilon_{it}, \quad |X_i - c| \leq h$$

- $D_i = 1[X_i \geq c]$; $Post_t = 1$ en los periodos después de que entra la política.
- β_1 : salto en el umbral **antes** de la política (efecto de las políticas confusoras).
- α_1 **es el estimador de diferencias en discontinuidades**: cambio en el salto entre antes y después.

- Interactuamos la regresión de RD con una dummy $Post_t$:

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 (X_i - c) + \beta_3 D_i (X_i - c) \\ + Post_t [\alpha_0 + \alpha_1 D_i + \alpha_2 (X_i - c) + \alpha_3 D_i (X_i - c)] + \varepsilon_{it}, \quad |X_i - c| \leq h$$

- $D_i = 1[X_i \geq c]$; $Post_t = 1$ en los periodos después de que entra la política.
- β_1 : salto en el umbral **antes** de la política (efecto de las políticas confusoras).
- α_1 **es el estimador de diferencias en discontinuidades**: cambio en el salto entre antes y después.
- Igual que en RD: elegir ancho de banda, kernel y polinomio; agrupar errores estándar por unidad (municipio).

¿Cómo evaluar los supuestos?

- **Saltos estables antes de la política:** con varios periodos *pre*, estimar el salto año por año.
 - Si el salto confusor era constante antes, es más creíble que lo hubiera seguido siendo.
 - Análogo a revisar tendencias paralelas en un estudio de eventos.
- **Diff-in-disc placebo:** usar una fecha falsa de inicio de la política o un umbral donde no cambió nada.
- **Balance en covariables:** diff-in-disc sobre variables predeterminadas debería ser cero.
- **Densidad:** la distribución de X alrededor de c no debería cambiar después de la política.

Ejemplo

Italia: reglas fiscales en municipios

Grembi, Nannicini & Troiano (2016)

- En 2001, Italia relajó las reglas fiscales del Pacto de Estabilidad Interna para los municipios con **menos de 5,000 habitantes**.
- Pero 5,000 habitantes también es el umbral de otras políticas municipales (salarios de alcaldes, tamaño del gobierno).

Italia: reglas fiscales en municipios

Grembi, Nannicini & Troiano (2016)

- En 2001, Italia relajó las reglas fiscales del Pacto de Estabilidad Interna para los municipios con **menos de 5,000 habitantes**.
- Pero 5,000 habitantes también es el umbral de otras políticas municipales (salarios de alcaldes, tamaño del gobierno).
- **Estrategia:** comparar el salto en 5,000 antes y después del cambio de 2001 en en las reglas fiscales.

Italia: reglas fiscales en municipios

Grembi, Nannicini & Troiano (2016)

- En 2001, Italia relajó las reglas fiscales del Pacto de Estabilidad Interna para los municipios con **menos de 5,000 habitantes**.
- Pero 5,000 habitantes también es el umbral de otras políticas municipales (salarios de alcaldes, tamaño del gobierno).
- **Estrategia:** comparar el salto en 5,000 antes y después del cambio de 2001 en en las reglas fiscales.
- **Resultado:** relajar las reglas fiscales aumentó el déficit y redujo los impuestos de los municipios pequeños.
 - Las reglas fiscales sí importan para la disciplina fiscal.

Grembi, V., Nannicini, T. & Troiano, U. (2016). Do Fiscal Rules Matter? *American Economic Journal: Applied Economics*, 8(3), 1–30.