Clase 2: Regression Discontinuity Design

Tomás Rau Binder
QLab

Julio 2023

Contenidos

Introducción

- La clase pasada vimos Los 2 tipos de diseños de RDD y discutimos estimación.
- Ahora discutiremos algunos tópicos:
  - Ancho de banda en LLR
  - Test de McCrary (2008) y Cattaneo, Janson y Ma (2018) para manipulación de variable de asignación.
  - Ejemplo STATA, Almond et al. (2010).

Ancho de Banda en LLR

Recuerde que para un SHARP design tenemos que estimar $\lim_{x \to x_0^+} E(Y_i | X_i = x)$

$$(\hat{a}, \hat{b}) = argmin_{a,b} \sum_{i=1}^{n} (Y_i - a - b(X_i - x_0))^2 \mathrm{K}\left(\frac{X_i - x_0}{h}\right) \mathrm{I}_{(X_i > x_0)}$$

y $\lim_{X \to x_0^-} E(Y_i | X_i = x)$

$$(\hat{a}, \hat{b}) = argmin_{a,b} \sum_{i=1}^{n} (Y_i - a - b(X_i - x_0))^2 \mathrm{K}\left(\frac{X_i - x_0}{h}\right) \mathrm{I}_{(X_i \leq x_0)}$$

Es importante determinar el ancho de banda h óptimo.

Ancho de Banda en LLR

- La elección del Kernel no es tan relevante como la elección del ancho de banda (h). El más utilizado en estos casos es el triangular.
- La elección de h es muy importante debido al clásico trade-off sesgo-varianza de los métodos no paramétricos
- Un mayor h incorpora más datos en la estimación local lo que mejora la precisión (menor varianza) pero a su vez aumenta el sesgo porque incorpora observaciones más lejanas y por tanto más distintas
- Imbens y Kalyanaraman (2012) proponen un método plug-in para obtener un h óptimo:

$$h^*_{IK} = C_{IK} \cdot n^{-1/5}$$

donde C_{IK} es un término que depende del Kernel utilizado, de la función de densidad $f(x)$, de las varianzas condicionales de $Y|X$ y de las segundas derivadas a cada lado del umbral de la esperanza condicional de $Y|X$.

- Este ancho de banda "óptimo" de IK minimiza un criterio de AMSE(h), *Average Mean Square Error*

Ancho de Banda en LLR

- Sin embargo, de acuerdo a varios autores $\hat{h}_{IK}$ es típicamente muy grande, incluyendo observaciones lejanas al umbral (aumentando el sesgo)
- Calonico, Cattaneo y Titiunik (2014) refinan la idea de IK minimizando

$$MSE(h) = E[((\hat{\mu}_-(h) - \mu_-) - (\hat{\mu}_+(h) - \mu_+))^2]$$

donde $\hat{\mu}_-(h) = \lim_{X \to x_0^-} \hat{E}(Y_i|X_i = x, h)$ y $\mu_- = \lim_{X \to x_0^-} E[Y_i|X_i = x]$. Lo mismo para el límite por la derecha. Obteniendo:

$$h^*_{CCT} = C_{CCT} \cdot n^{-1/5}$$

en este caso C_{CCT} depende del Kernel utilizado y de las segundas derivadas a cada lado del umbral de la esperanza condicional de $Y|X$. Además el algoritmo propuesto por CCT utiliza anchos de banda piloto óptimos en el sentido MSE.

Ancho de Banda en LLR

- Luego, implementan el procedimiento de optimización descrito en Stata (*rdrobust*) con distintas opciones, como por ejemplo:
  - Un h común igual para cada lado del umbral para el estimador del LATE (*mserd*). Default de STATA.
  - Dos h, uno para cada lado del umbral para la estimación del LATE (*msetwo*)
- Para el *Fuzzy* design realizamos el mismo método para los términos del denominador.

Test de McCrary (2008) y Cattaneo, Janson y Ma (2018)

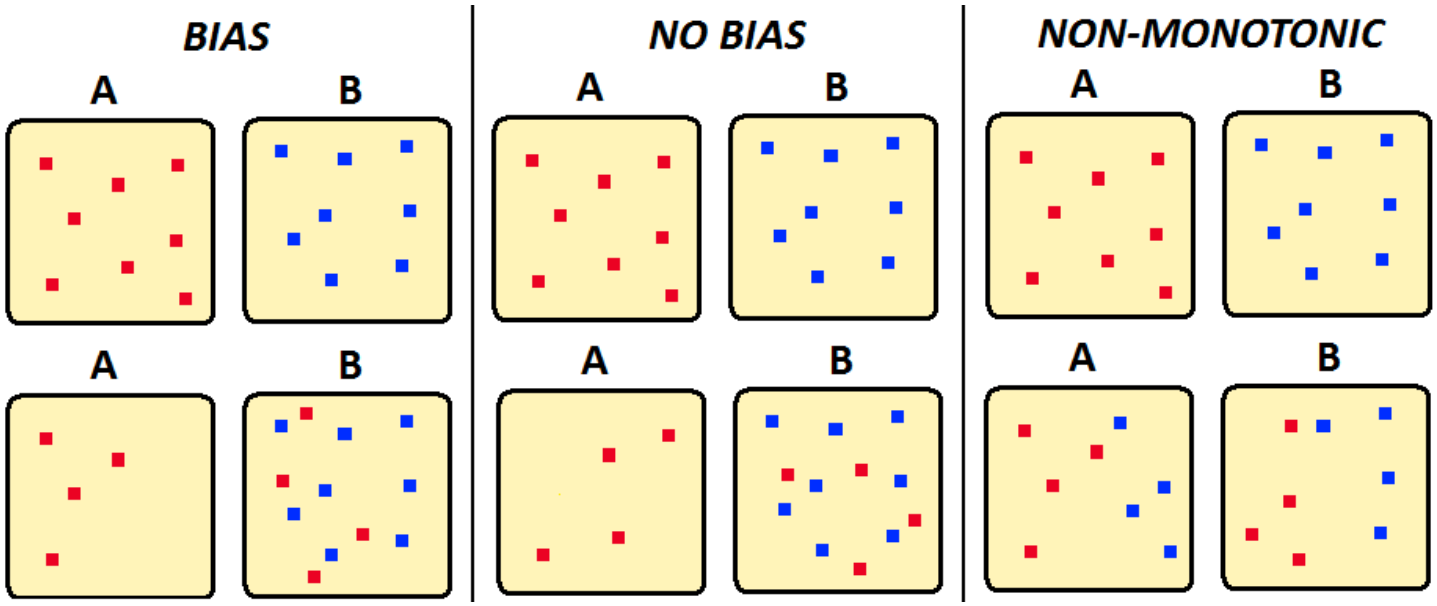
- McCrary (2008) argumenta que el supuesto de continuidad del valor esperado potencial puede verse comprometido si los individuos son capaces de **manipular** el valor de X para recibir tratamiento o para evitarlo.
- La intuición está en que los individuos que pueden cambiar su valor de X serán distintos a los que no y por tanto los grupos a cada lado del umbral dejarían de ser comparables.
- Ejemplos:
  1. Puntaje en una prueba de selección. Los individuos pueden tomar la prueba de nuevo y lograr cambiar su puntaje
  2. Votación en el congreso. Los votos pueden ser vendidos por favores políticos.
  3. Indices de Vulnerabilidad. Las familias subdeclaran ingresos o esconden activos para tener un menor puntaje y mayor prioridad en programas sociales.

Test de McCrary (2008)

- McCrary propone entonces determinar si la variable de asignación está siendo manipulada o no analizando la distribución de X.
- Así, una discontinuidad en la distribución de X en torno a x_0 podría considerarse como evidencia de manipulación e invalidar por tanto el diseño RD.
- Este es un test indirecto: una Densidad Continua de la variable de asignación no es condición necesaria ni suficiente para la identificación
- El test es informativo sólo con manipulación monotónica

Regression Discontinuity

Ejemplos de distintos tipos de manipulación:



Implementación del Test

En dos etapas:

1. Estimar un histograma suavizado, esto es, usando bines que muestren un histograma "suave" (requiere elección del **binsize** en el caso de McCrary. Para Cattaneo et al. (2018) se calcula óptimamente.)
2. Estimar una regresión local (local linear regression) sobre el histograma suavizado (requiere elección del ancho de banda o **bandwidth**)
3. La idea es ver si hay una discontinuidad en el histograma, justo antes y después del corte: $lim_{r \to c^+} f(r)$ y $lim_{r \to c^-} f(r)$

Parámetro de interés:

$$\theta = ln\ lim_{r \to c^+} f(r) - ln\ lim_{r \to c^-} f(r) \equiv ln\ f^+ - ln\ f^- \qquad (1)$$

La estimación usa 2 regresiones, una a cada lado de c, para estimar

$$\hat{\theta} \equiv ln\ \hat{f}^+ - ln\ \hat{f}^-$$

.

Elección de bin size y bandwidth

- La estimation of $\hat{\theta}$ NO requiere una elección cuidadosa de **binsize**
- La estimation de $\hat{\theta}$ SÍ requiere una elección cuidadosa de **bandwidth**

Metodos para elegir el *bandwidth*:

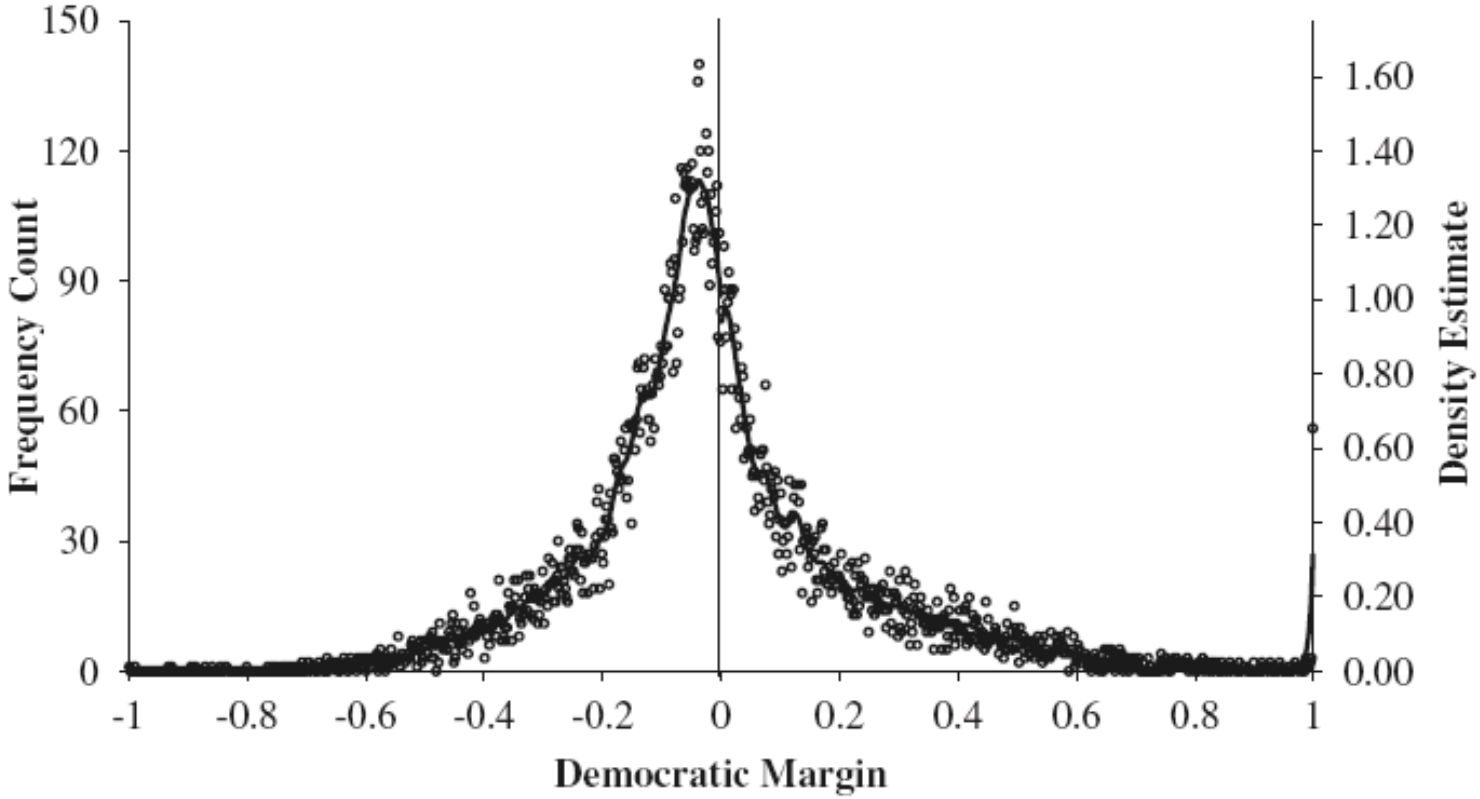
- Inspección visual usando ambas etapas
- Validación cruzada (Stone 1974) o Plug-in (Cheng 1997)
- Cattaneo, Janson y Ma (2018)

La versión alternativa a McCrary de Cattaneo, Janson y Ma (2018) usa un estimador distinto para la densidad y tiene mejor poder, bajo ciertas condiciones.

Ejemplo 1: Lee (2001)

Haber ganado las elecciones (House of representatives) hace más fácil ganar la siguiente.

La manipulación NO es esperada porque la coordinación de los votantes es muy difícil.



Ejemplo 1: Similar data que Lee (2001)

Usando Cattaneo, Janson y Ma (2018), se encuentra algo similar: haber ganado las elecciones (House of representatives) hace más fácil ganar la siguiente.

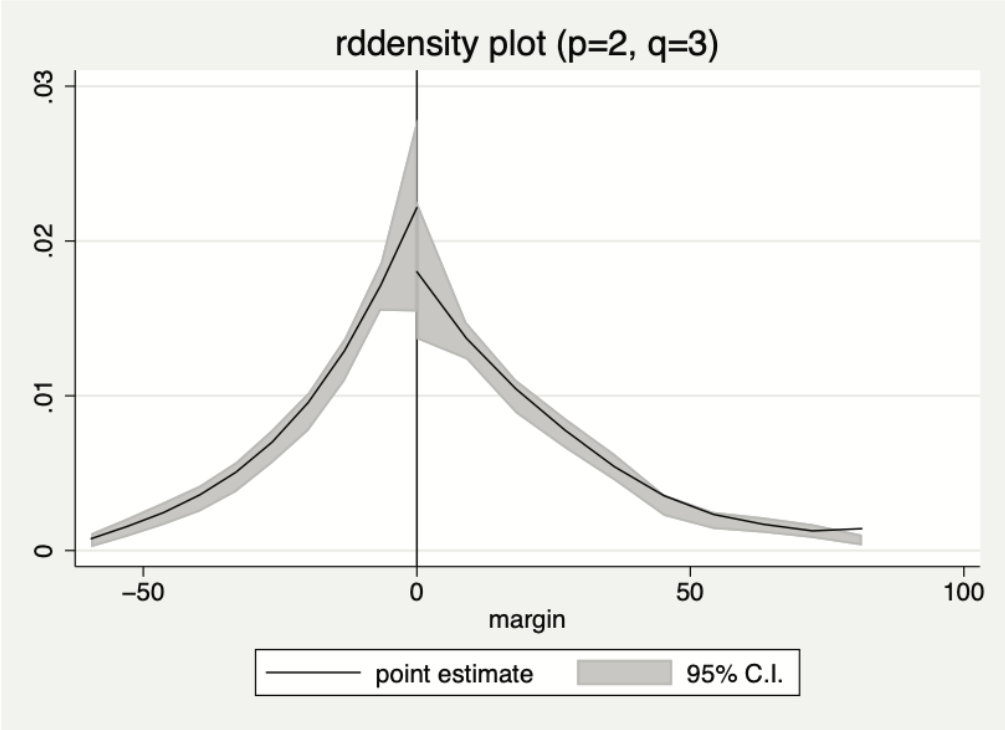
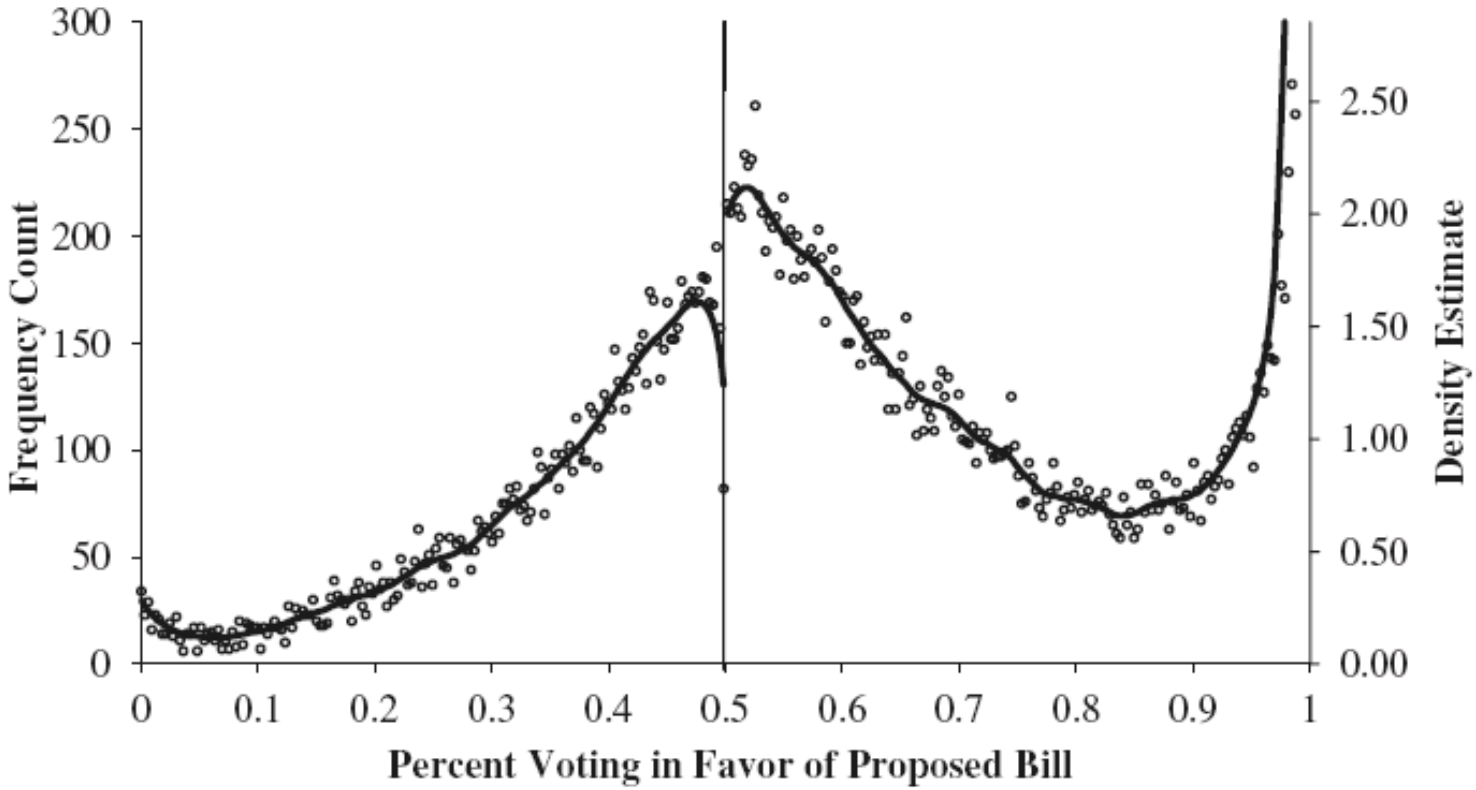


Figure 1. Manipulation test plot (default options)

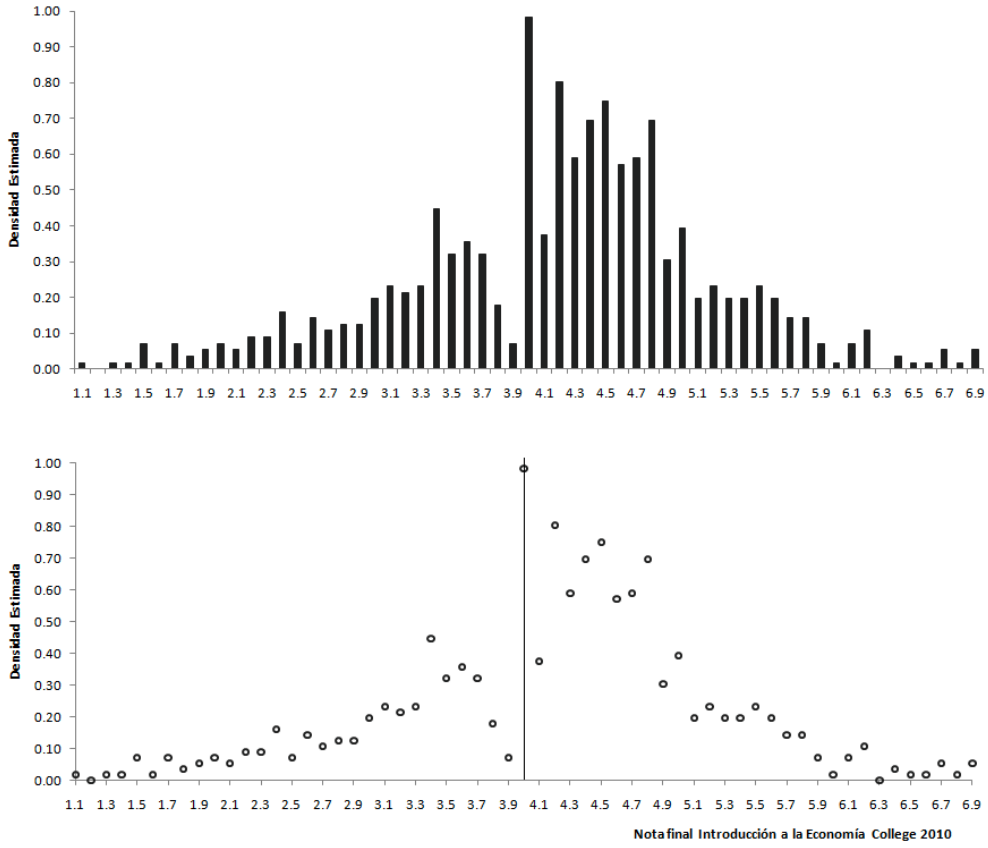
Ejemplo 2: Lee (2011)

Roll call vote': Debido a la naturaleza de juego repetido y dado que los votos son públicos, la coordinación de los representantes es esperable.

Cuando la votación es estrecha, los votos pueden ser "vendidos" a cambio de favores políticos.



Ejemplo 3: Notas de Economía, College UC (2011)



Nota aprobatoria en Chile es 4,0 (escala de 1,0 a 7,0).

Aplicación: Test de McCrary 2008

Test de McCrary 2008

- Veremos su implementación en STATA
- El comando en STATA se llama **DCdensity.ado** y puede ser descargado en la página:
- http://eml.berkeley.edu//~jmccrary/DCdensity/
- Veamos un ejemplo de "manipulación" con datos Chilenos.

Toma de Razón en Dirección de Vialidad

- La Toma de Razón (TdR) es un tipo de auditoría ex-ante que realiza la CGR
- Por medio de esta, la CGR visa la legalidad del acto administrativo que se realizará
- El año 2008 la Resolución 1600 dejó exentas de TdR a las obras menores a UTM 10.000
- Con Eduardo Engel y Andrea Repetto intentamos medir el impacto de la TdR en la calidad de las obras,
- pero nos encontramos con este problema:

Toma de Razón en Dirección de Vialidad

DCdensity monto, breakpoint(10000) generate(Xj Yj r0 fhat se_fhat)

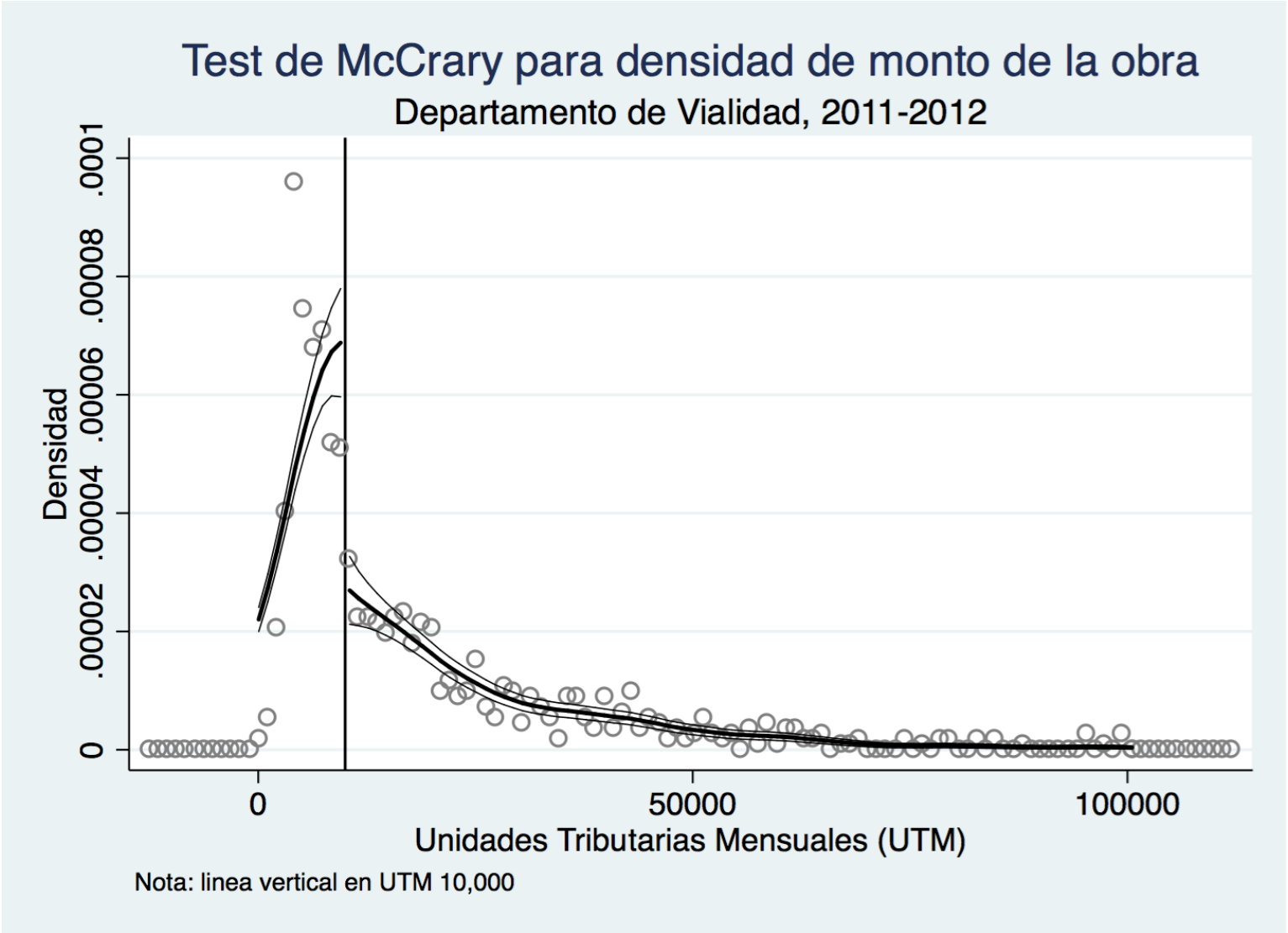
```
Using default bin size calculation, bin size = 1046.69236
Using default bandwidth calculation, bandwidth = 11594.5641

Discontinuity estimate (log difference in height): -.923294226
                                                   (.140409799)
Performing LLR smoothing.
97 iterations will be performed
.........
```

Donde **monto** es el monto de la obra, **breakpoint(10000)** es el umbral de corte y **generate(Xj Yj r0 fhat se_fhat)** son variables que genera el comando (las veremos en seguida).

Luego, el $t = -0{,}92/0{,}14 \simeq -6{,}6$. Para un $\alpha = 0{,}01$, el crítico es 2.576, $|-6,6| > 2{,}576$, se rechaza la nula de ausencia de discontinuidad al 1 %.

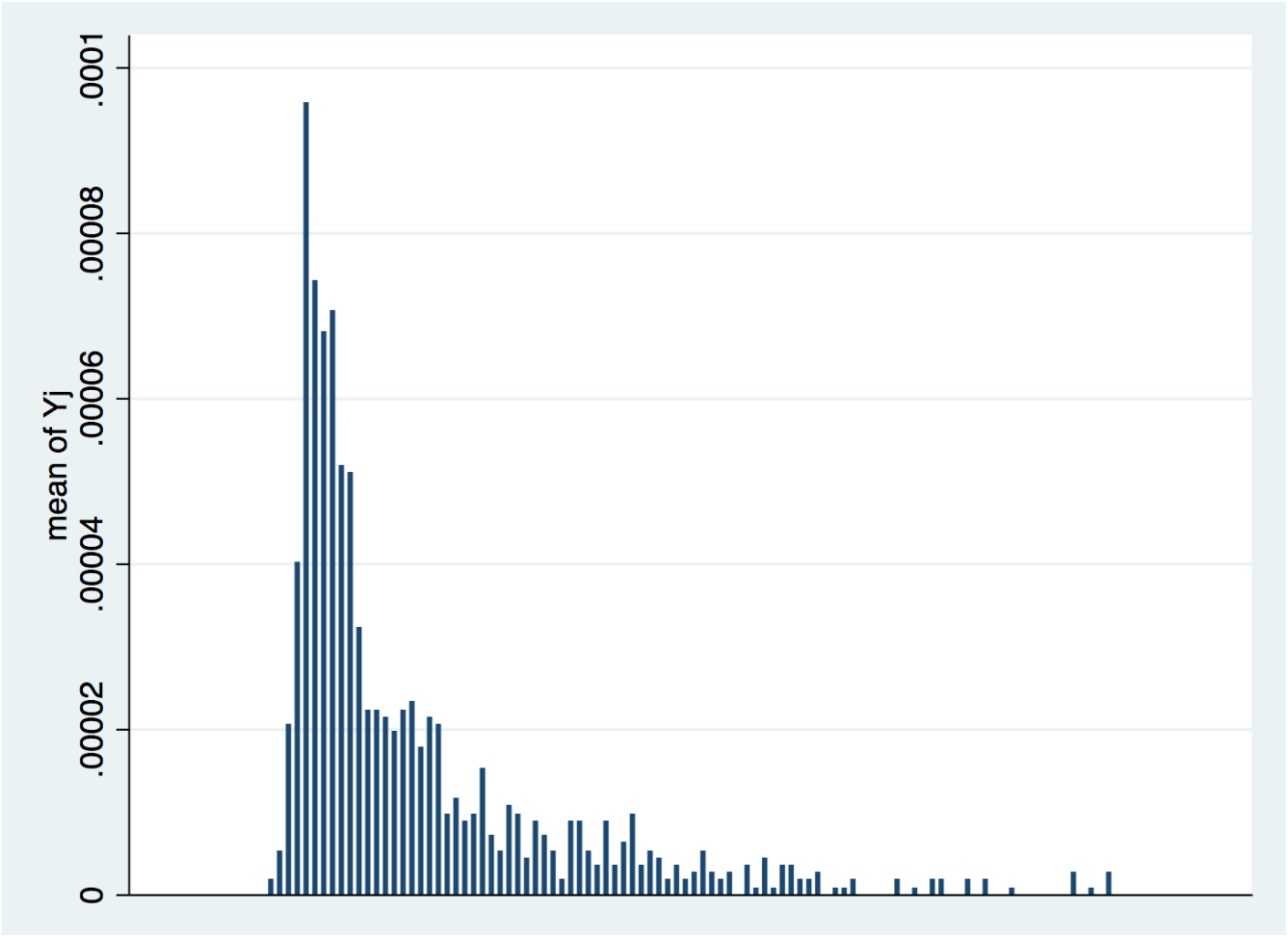
Toma de Razón en Dirección de Vialidad



Toma de Razón en Dirección de Vialidad

De las variables generadas Ud. puede replicar el histograma que genera el comando usando *Yj* y *Xj* como frecuencias y bines:

graph bar Yj, over(Xj,label(nolabel))



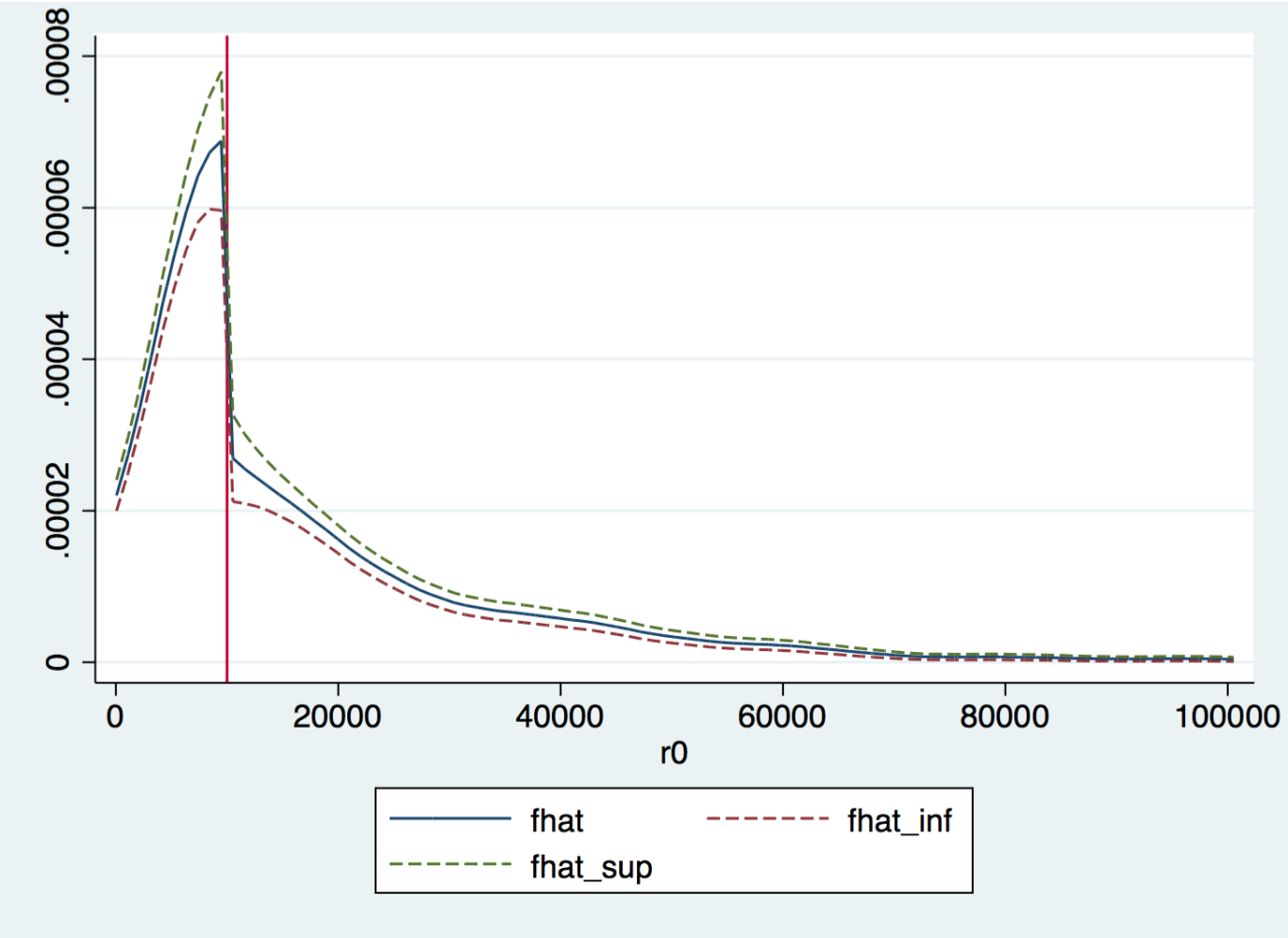
Toma de Razón en Dirección de Vialidad

También podemos agregarle al gráfico anterior intervalos de confianza:

*gen fhat_inf= fhat-1.96*se_fhat*

*gen fhat_sup= fhat+1.96*se_fhat*

line fhat fhat_inf fhat_sup r0, xline(10000) lpattern(solid dash dash)



Toma de Razón en Dirección de Vialidad

- Con una densidad así, es difícil perseverar con un RD
- En este caso, el *bunching* a la izquierda se justifica para evitar la TdR de CGR
- Pero no es la única explicación. Puede haber cierto fraccionamiento de proyectos para explicar tanta acumulación a la izquierda
- Hay evidencia actual de existencia de fraccionamiento en Fonasa

¿Qué hacer si hay manipulación?

Gerard, Rokkanen y Rothe (2020)

- Gerard et al (2020): "Bounds on Treatment Effects in Regression Discontinuity Designs with a Manipulated Running Variable "
- Proponen una forma de identificar parcialmente el efecto tratamiento bajo un escenario de manipulación
- Identifican cotas (inferiores y superiores) para el efecto de tratamiento y únicamente para individuos que no manipulan su variable de asignación
- El método propuesto consiste en:
  1. Estimar la proporción de individuos τ que manipulan; a partir del salto en la función de densidad de la variable de asignación X en el umbral x_0. Ejemplo un 10 %

Gerard, Rokkanen y Rothe (2020)

- El método propuesto consiste en:
  2. Ordenar la variable de resultado y eliminar el $\tau\,\%$ inferior de los datos para así obtener el bound superior de la esperanza condicional del outcome. Repetir el análisis eliminando el $\tau\,\%$ superior para obtener un bound inferior
  3. Con ambos bounds de $E(Y|X)$ a cada lado del umbral, derivar un intervalo para el *ATE at the cutoff for non-manipulators*.
- Como aplicación, evalúan el efecto que tiene el seguro de desempleo sobre la duración del desempleo en Brasil; presentando evidencia de manipulación del tiempo de empleo previo (la variable de asignación)

Chequeos de Robustez y más

Chequeos de Robustez

- Como sabemos, el supuesto de continuidad de los valores esperados potenciales implica que los individuos son similares a ambos lados del umbral. Por esta razón otro test indirecto de la validez de un diseño RD consiste en comparar características observables Z en torno a x_0 para determinar si en promedio son iguales.
- Formalmente, se evalúa si $E(Z|X = x)$ es continua en x_0 lo que se implementa a través de un diseño Sharp-RD utilizando el covariate Z como variable dependiente.
- De encontrarse una discontinuidad, podríamos decir que los individuos son distintos en esa dimensión y por tanto cuestionar la validez del diseño RD.

Ejemplo STATA: Almond et al. 2010

Ejemplo STATA

- Almond, Doyle, Kowalski & Williams (2010) "Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns"
- Los autores se preguntan si gastar más dinero en recién nacidos con peso bajo es rentable en términos de reducción de la tasa de mortalidad
- Para estimar los retornos del cuidado adicional y comparar luego con el gasto incurrido, Almond et al utilizan un diseño RDD (Sharp) en el que la variable de asignación es el peso al nacer
- El umbral es de 1.500 gramos, pues si un individuo nace con menos de 1.500 gramos entonces recibe cuidados adicionales de salud
- El principal outcome utilizado es si el niño murió al año siguiente de su nacimiento (binaria)

Ejemplo STATA

- Los autores utilizan cerca de 66 millones de nacimientos en USA entre 1983 y 2002. Para intentar replicar sus principales resultados, restringimos la muestra únicamente a aquellos nacimientos con peso entre 1.415 y 1.585 gramos
- Primero, podemos ver la representación gráfica con un polinomio usando el comando **rdplot**:
- Para estimar el retorno de los cuidados adicionales, utilizan dos modelos:
  1. Modelo paramétrico, en el que estiman una regresión lineal a cada lado del umbral (con distintas pendientes); controlando además por características individuales Z y por año y estado de nacimiento:

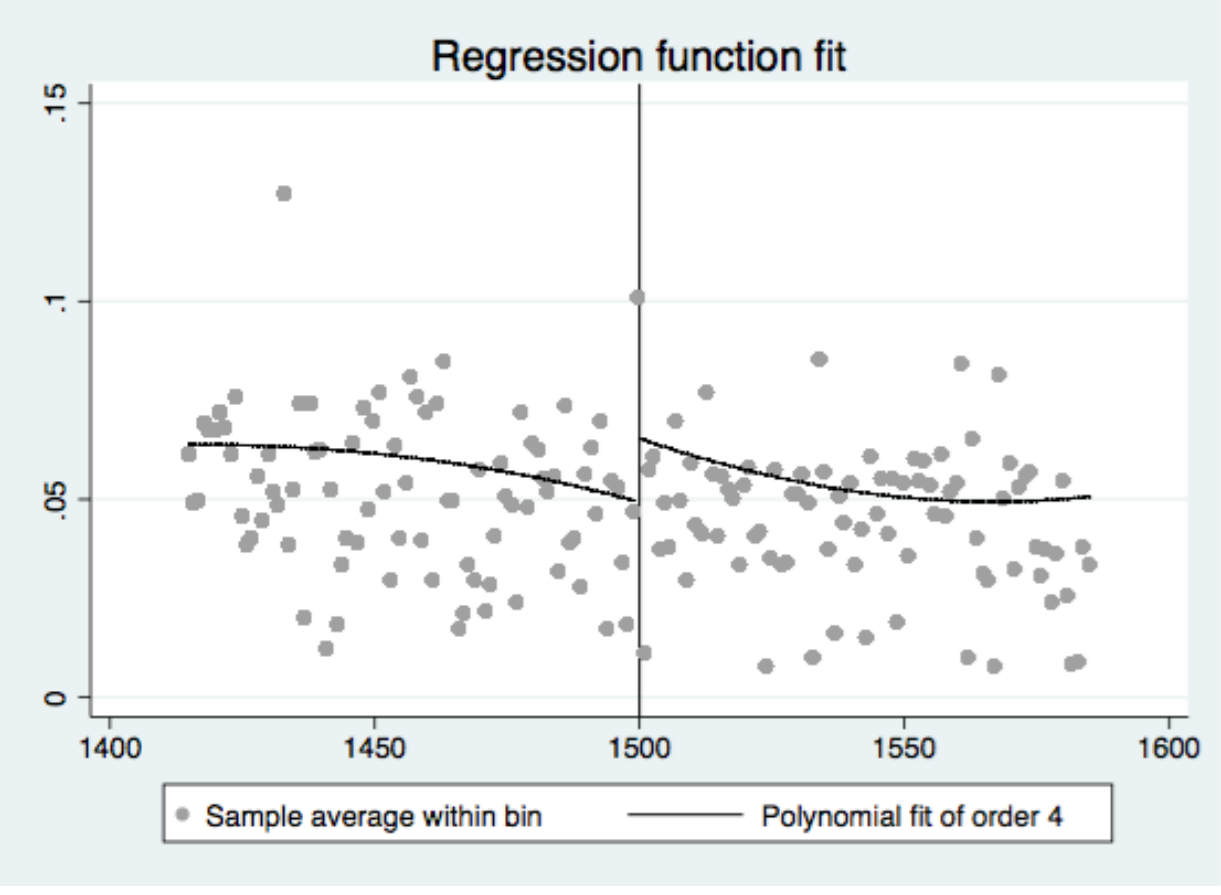
$$Y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 D_i(X_i - 1500) + \alpha_3(1 - D_i)(X_i - 1500) + \alpha_t + \alpha_s + \delta Z_i' + \epsilon_i$$

  2. Modelo no paramétrico, para el que utilizan una *Local Linear Regression*

Ejemplo STATA

1. Análisis gráfico, usando el comando *rdplot*

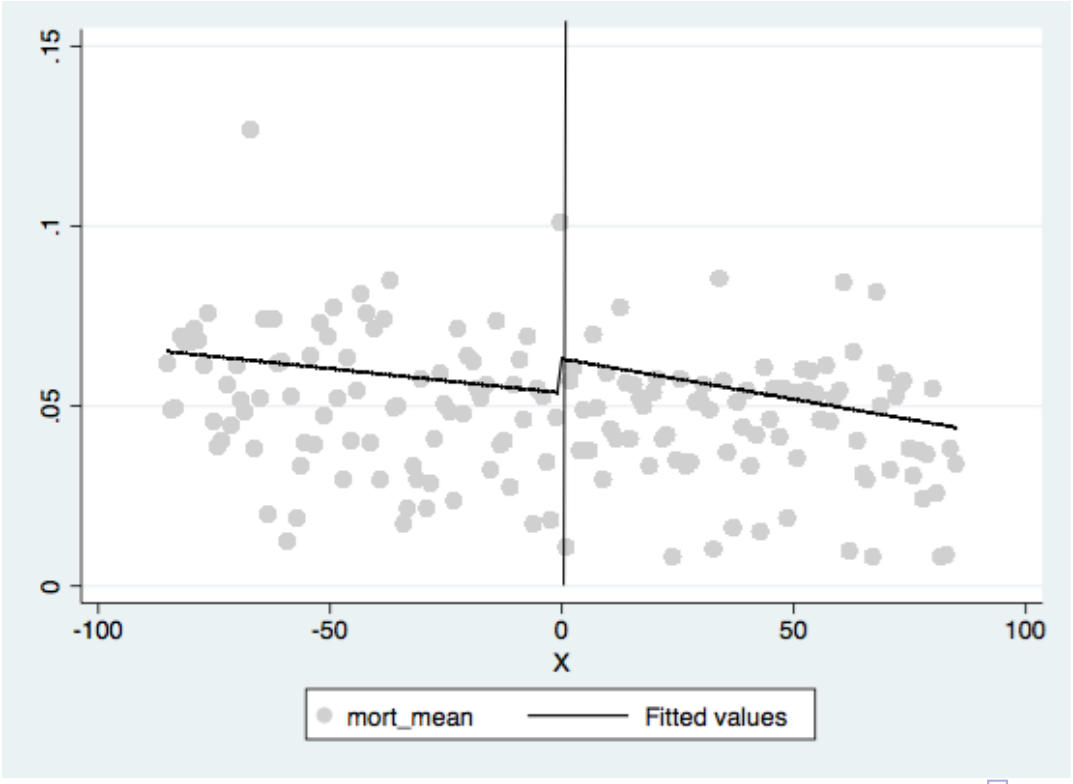
rdplot morta weight, c(1500)



Ejemplo STATA

1. Análisis gráfico, también podemos construir manualmente un gráfico a partir de una estimación lineal

gen $X = weight - 1500$
gen $D1 = X < 0$
gen D2=1-D1
reg morta D1 D1∗X D2∗X
predict y
bysort X: egen mort_mean = mean(morta)
graph twoway (scatter mort_mean X) (line y X)



Ejemplo STATA

2. Estimación, del modelo paramétrico

global covariates prenatal outside first orderm mage* meducm fage male ges* ...*

reg morta D1 D1∗X D2∗X $covariates i.year, vce(robust)

morta	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
D1	-.0071322	.0021539	-3.31	0.001	-.0113537	-.0029106
D1X	-.0118422	.0032129	-3.69	0.000	-.0181393	-.005545
D2X	-.0189199	.0028539	-6.63	0.000	-.0245134	-.0133264
prenatalm	.0091718	.002184	4.20	0.000	.0048912	.0134523
prenatal1	.0070673	.001454	4.86	0.000	.0042174	.0099171
prenatal7	.0001647	.001273	0.13	0.897	-.0023304	.0026597
prenatal11	0	(omitted)				
outside	.0040577	.0025618	1.58	0.113	-.0009633	.0090787
first	-.0200573	.0011479	-17.47	0.000	-.0223072	-.0178074
orderm	.0052086	.0072319	0.72	0.471	-.0089657	.0193829
mage1	.0016762	.0063898	0.26	0.793	-.0108477	.0142001
mage2	-.005181	.0046976	-1.10	0.270	-.0143883	.0040262
mage3	-.0033659	.0045681	-0.74	0.461	-.0123194	.0055876
mage4	-.0114303	.0044738	-2.55	0.011	-.0201988	-.0026618
mage5	-.0115175	.0044162	-2.61	0.009	-.0201732	-.0028619
mage6	-.0108375	.0045052	-2.41	0.016	-.0196676	-.0020073
mage7	0	(omitted)				

—more—

El acceso a cuidados adicionales reduce la mortalidad en 0.71pp.

Ejemplo STATA

2. Estimación, del modelo no-paramétrico usando el *bandwidth* de los autores

rdrobust morta weight, c(1500) h(85)

Sharp RD estimates using local polynomial regression.

Cutoff c = 1500	Left of c	Right of c
Number of obs	95224	106852
Eff. Number of obs	94572	106074
Order loc. poly. (p)	1	1
Order bias (q)	2	2
BW loc. poly. (h)	85.000	85.000
BW bias (b)	85.000	85.000
rho (h/b)	1.000	1.000

Number of obs	=	202076	
BW type	=	Manual	
Kernel	=	Triangular	
VCE method	=	NN	

h=85, determinado arbitrariamente

Outcome: morta. Running variable: weight.

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	.01202	.00247	4.8729	0.000	.007183	.01685
Robust	-	-	4.8868	0.000	.012421	.029057

El acceso a cuidados adicionales reduce la mortalidad en 1.2pp (es grande si la tasa de mortalidad para aquellos en torno a 1500 grs es 5.5 %)

Nota: El comando considera la diferencia "derecha-izquierda", luego el signo positivo significa que aquellos a la derecha del umbral, tienen un tasa de mortalidad más alta (e

Ejemplo STATA

2. Estimación, del modelo no-paramétrico usando el *bandwidth* automático del comando
rdrobust morta weight, c(1500)

Sharp RD estimates using local polynomial regression.

Cutoff c = 1500	Left of c	Right of c
Number of obs	95224	106852
Eff. Number of obs	3585	29910
Order loc. poly. (p)	1	1
Order bias (q)	2	2
BW loc. poly. (h)	11.299	11.299
BW bias (b)	22.509	22.509
rho (h/b)	0.502	0.502

Number of obs	=	202076
BW type	=	mserd
Kernel	=	Triangular
VCE method	=	NN

h óptimo según CCT, 2014

one "mean square error" bandwidth for RD

Outcome: morta. Running variable: weight.

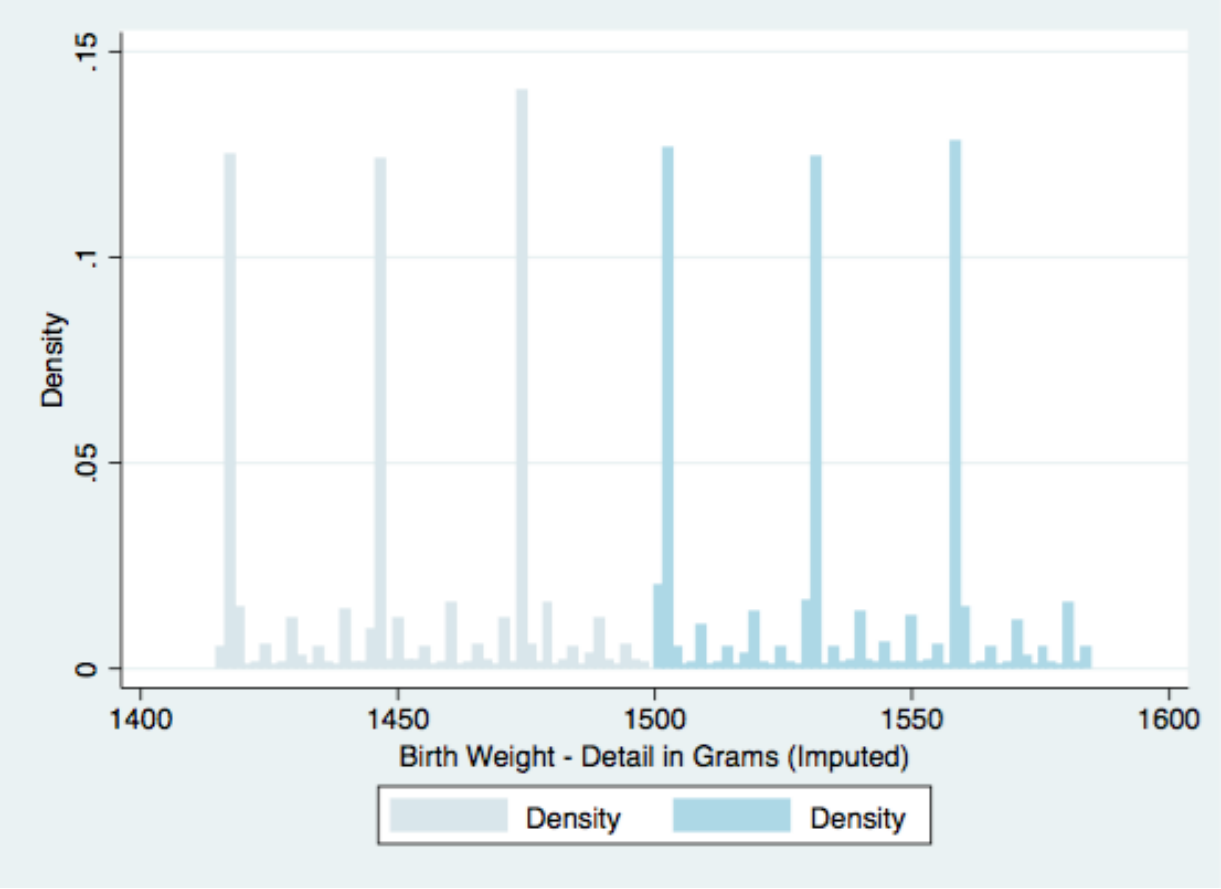
Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	.05372	.01076	4.9940	0.000	.032634	.074797
Robust	-	-	4.5360	0.000	.032735	.082547

El acceso a cuidados adicionales reduce la mortalidad en 5.4pp. Es poco creible...

Ejemplo STATA

3. McCrary (2008), Histograma del peso al nacer

twoway (hist weight if weight < 1500) (hist weight if weight >=1500)

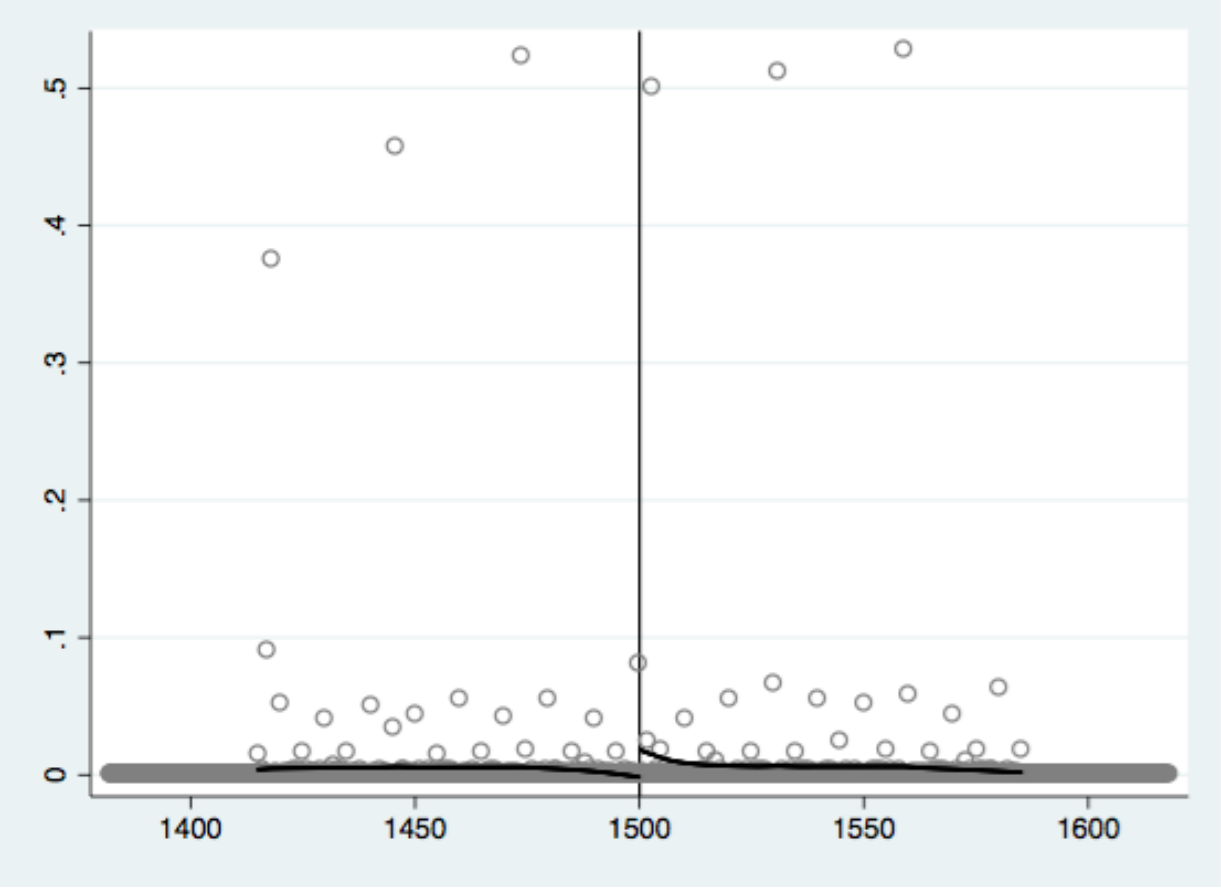


El Histograma es discontinuo en 1500. Los autores argumentan que esta discontinuidad se debe al redondeo y que sucede en 1500. En caso de ser manipulación debieramos observar acumulación en valores menores a 1500.

Ejemplo STATA

3. McCrary (2008), test de McCrary

DCdensity weight, breakpoint(1500) generate(Xj Yj r0 fhat se_fhat)



El test de McCrary también rechaza la hipótesis de continuidad. ¿Será válida la justificación de los autores sobre la no-manipulación en favor del redondeo?

Ejemplo STATA

4. Otros tests de Robustez, diseño RD sobre covariates

- Dummy de educación superior para la madre:
  rdrobust meduc2 weight, c(1500)

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	-.04696	.01964	-2.3907	0.017	-.085452	-.008459
Robust	-	-	-2.3279	0.020	-.096684	-.008297

- Dummy de raza blanca para la madre:
  rdrobust white weight, c(1500)

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	-.31495	.0391	-8.0549	0.000	-.39158	-.238312
Robust	-	-	-7.9334	0.000	-.402034	-.24274

Ejemplo STATA

- Parece ser que los recién nacidos en torno a 1.500 gramos no son tan parecidos.
- En particular, quienes nacen con menos de 1.500 gramos tienen mayor probabilidad de tener una madre con educación superior y de raza blanca.
- ¿Podrá ser esa una explicación de la menor tasa de mortalidad?
- Veamos otros chequeos de robustez

Ejemplo STATA

4. Otros tests de Robustez, diseño RD con distinto umbral

- Peso de corte en 1.490 gramos:
  rdrobust morta weight, c(1490)

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	.00998	.00232	4.2934	0.000	.005422	.01453
Robust	-	-	4.2367	0.000	.008819	.024002

- Peso de corte en 1.510 gramos:
  rdrobust morta weight, c(1510)

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	-.00789	.00241	-3.2780	0.001	-.012604	-.003171
Robust	-	-	-3.1516	0.002	-.020109	-.004688

Ejemplo STATA

- En ambos casos encontramos un efecto distinto de cero
- ¿Será que los autores realmente están identificando el retorno de los cuidados adicionales que se entregan a quienes nacen con un peso inferior a 1500?
- El artículo ha sido criticado por otros autores (Barreca et al. 2011). Ellos remueven los niños con peso exactamente igual a 1500grs y muestran que las estimaciones se reducen
- Cuando quitan aquellos a +- 3 gramos de 1500 los efectos desaparecen.
- Esta práctica de remover observaciones no está fundamentada, pero algunos lo hacen y le llaman *Donut-hole RD*
- Más que un método, algunos autores lo plantean como un chequeo de robustez adicional.

Extensiones

- Gerard et al (2020): "Bounds on Treatment Effects in Regression Discontinuity Designs with a Manipulated Running Variable ". Encontrar cotas cuando hay manipulación.
- Away of the cutoff: Si queremos alejarnos del umbral. *Angrist y Rokkanen (2015)*
- Regression Kink Design: En lugar de un salto, nos interesa un cambio en la pendiente. *Card, Lee, Pei y Weber (2012)*
- Multi cutoff RD: Cuando la selección al tratamiento se da a distintos niveles de X. *Cattaneo, Keele, Titiunik y Vazquez-Bare (2015)*
- RD with covariates: El beneficio de incluir covariates en la estimación. *Calonico, Cattaneo, Farrell y Titiunik (2016)*