

Economía Aplicada

Cuasi experimentos: Variables Instrumentales y Regresión con  
Discontinuidad

---

Departamento de Economía  
Universidad Carlos III de Madrid

# Evaluación de política con cuasi experimentos

- En un cuasi experimento o experimento natural existe alguna fuente de aleatoriedad que permite considerar a alguna variable como si fuese asignada de manera aleatoria. Consideramos dos tipos de cuasi experimentos:
- El caso de que el tratamiento ( $D$ ) pueda considerarse como si fuese asignado de manera aleatoria (quizás condicionando en ciertas variables adicionales  $X$ ).
- El caso de que una variable ( $Z$ ) que afecta al tratamiento ( $D$ ) se puede considerar como si fuese asignada de manera aleatoria (quizás condicionando en ciertas variables adicionales  $X$ ). En ese caso  $Z$  puede utilizarse como variable instrumental para  $D$ .
  - El artículo de Angrist es un ejemplo de este caso.

## Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records, Angrist, AER(1990)

- ¿Hacer el servicio militar tiene un efecto negativo en los ingresos?
- Que los veteranos tengan menores ingresos que los no veteranos no implica que ser veterano sea la causa de los menores ingresos.
- Esta comparación simple de los ingresos entre los dos grupos no es adecuada si ser veterano no es independiente de los ingresos potenciales.
- Comparar los ingresos controlando por un conjunto de características observables solamente tiene sentido si ser veterano es independiente de los ingresos potenciales una vez que tenemos en cuenta esas características.

## Efecto del estatus de veterano en los ingresos

Llamemos  $Y_i$  a los ingresos,  $D_i$  representa el estado de veterano en la era de Vietnam, y  $X_i$  un conjunto de controles:

¿Es correcto estimar la siguiente esperanza condicional para estimar el efecto de ser veterano de Vietnam en los ingresos?

$$E[Y_i | D_i, X_i] = \beta_0 + \alpha D_i + \gamma' X_i$$

- Probablemente existan diferencias no observables entre los hombres que eligen enrolarse en el servicio militar y los que no, y probablemente esas diferencias estén correlacionadas con los ingresos potenciales.
- Si  $D_i$  está correlacionada con variables no observadas que pertenecen a la ecuación, las estimaciones de MCO serán inconsistentes.
- Una posible solución es encontrar una variable instrumental válida.

# Una VI para el estatus de veterano

- Preocupaciones sobre la justicia de la política de conscripción en EEUU llevó a instituir una selección por sorteo en 1970.
- Este sorteo se realizó anualmente en 1970, 1971 y 1972. Se le asignó un número aleatorio (del 1 al 365) a cada fecha de nacimiento de las cohortes de 19 años, los hombres con números debajo de un cierto valor (fijado por el Ministerio de Defensa) eran llamados a servir en el ejército.
- El estatus de veterano no está completamente determinado por el sorteo: algunos se alistaban voluntariamente, otros evitaban el enrolamiento por razones de salud o de estudios. Pero el sorteo está altamente correlacionado con el estatus de veterano.

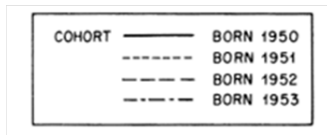
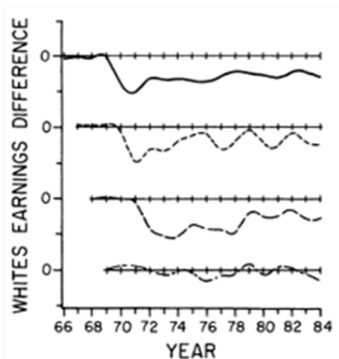
## Reclutamiento Aleatorio como instrumento 1/2

- Llamemos  $Z_i$  a la variable binaria de reclutamiento aleatorio.
- De modo de poder identificar el efecto causal de  $D_i$  en los ingresos es crucial que la única razón para que  $E(Y_i|Z_i)$  cambie con  $Z_i$  sea la variación en  $E(D_i|Z_i)$ : o sea que el reclutamiento aleatorio afecte a los ingresos sólo a través de su efecto en el estatus de veterano.
- Un chequeo simple es ver si existe asociación entre  $Z_i$  y alguna característica personal que no puede ser afectada por  $D_i$ . Otro chequeo es mirar si existe asociación entre  $Z_i$  y la variable de interés en muestras donde no hay relación entre  $D_i$  y  $Z_i$ .

## Reclutamiento Aleatorio como instrumento 2/2

- Angrist analiza los ingresos de 1969 (previos al sorteo de 1970) y encuentra que no hay efecto del reclutamiento aleatorio en los ingresos.
- También analiza la cohorte de hombres nacidos en 1953. Aunque hubo sorteo para esa cohorte, nadie fue finalmente seleccionado. Por lo tanto no hay correlación entre el número obtenido en el sorteo y el estatus de veterano para esa cohorte. Angrist encuentra que no existe una relación significativa entre los ingresos y el reclutamiento aleatorio para los nacidos en 1953.
- Ambos resultados apoyan el argumento de que el reclutamiento aleatorio afecta a los ingresos sólo a través del estatus de veterano.

## Diferencias en Ingresos por Reclutamiento Aleatorio - Un Gráfico



*Notes:* The figure plots the difference in FICA taxable earnings by draft-eligibility status for the four cohorts born 1950–53. Each tick on the vertical axis represents \$500 real (1978) dollars.

## Diferencias en Ingresos por Reclutamiento Aleatorio - Regresiones

| TABLE 1—DRAFT-ELIGIBILITY TREATMENT EFFECTS FOR EARNINGS |                       |                  |                  |                 |                        |      |      |      |
|----------------------------------------------------------|-----------------------|------------------|------------------|-----------------|------------------------|------|------|------|
| Whites                                                   |                       |                  |                  |                 |                        |      |      |      |
| Year                                                     | FICA Taxable Earnings |                  |                  |                 | Total W-2 Compensation |      |      |      |
|                                                          | 1950                  | 1951             | 1952             | 1953            | 1950                   | 1951 | 1952 | 1953 |
| 66                                                       | -21.8<br>(14.9)       |                  |                  |                 |                        |      |      |      |
| 67                                                       | -8.0<br>(18.2)        | 13.1<br>(16.4)   |                  |                 |                        |      |      |      |
| 68                                                       | -14.9<br>(24.2)       | 12.3<br>(19.5)   | -8.9<br>(19.2)   |                 |                        |      |      |      |
| 69                                                       | -2.0<br>(34.5)        | 18.7<br>(26.4)   | 11.4<br>(22.7)   | -4.0<br>(18.3)  |                        |      |      |      |
| 70                                                       | -233.8<br>(39.7)      | -44.8<br>(36.7)  | -5.0<br>(29.3)   | 32.9<br>(24.2)  |                        |      |      |      |
| 71                                                       | -325.9<br>(46.6)      | -298.2<br>(41.7) | -29.4<br>(40.2)  | 27.6<br>(30.3)  |                        |      |      |      |
| 72                                                       | -203.5<br>(55.4)      | -197.4<br>(51.1) | -261.6<br>(46.8) | 2.1<br>(42.9)   |                        |      |      |      |
| 73                                                       | -226.6<br>(67.8)      | -228.8<br>(61.6) | -357.7<br>(56.2) | -56.5<br>(54.8) |                        |      |      |      |
| 74                                                       | -243.0<br>(81.4)      | -155.4<br>(75.3) | -402.7<br>(68.3) | -15.0<br>(68.1) |                        |      |      |      |
| 75                                                       | -295.2<br>(94.4)      | -99.2<br>(89.7)  | -304.5<br>(85.0) | -28.3<br>(79.6) |                        |      |      |      |

## Estimador de Wald 1/2

- En un modelo con  $D$  endógena:  $Y_i = \beta_0 + \alpha D_i + \varepsilon_i$
- Con  $Z$  una VI válida, se puede escribir  $\alpha = \text{Cov}(Y_i, Z_i) / \text{Cov}(D_i, Z_i)$
- Si  $Z$  es una variable binaria, que toma el valor uno con probabilidad  $p$ , para cualquier variable  $W$  podemos escribir:

$$\begin{aligned}\text{Cov}(W_i, Z_i) &= E[W_i Z_i] - E[W_i]E[Z_i] \\ &= \{E[W_i | Z_i = 1] - E[W_i | Z_i = 0]\} p * (1 - p)\end{aligned}$$

- Por lo tanto:

$$\alpha = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(D_i, Z_i)} = \frac{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]}{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]}$$

## Estimador de Wald 2/2

- Si además  $D$  es binaria, como en el caso de representar al tratamiento:
  - $E[D_i|Z_i = 1]$  es la proporción de tratados ( $D = 1$ ) en el total de observaciones con  $Z = 1$ .
  - $E[D_i|Z_i = 0]$  es la proporción de tratados ( $D = 1$ ) en el total de observaciones con  $Z = 0$ .
  - El denominador captura el impacto del instrumento en la probabilidad de formar parte del grupo de tratamiento.
- El análogo muestral de  $\alpha$  se conoce como estimador de Wald.

Estimador de Wald (incluyendo controles  $X$ ):

$$\hat{\alpha}_W(X) = \frac{\bar{Y}(X, Z=1) - \bar{Y}(X, Z=0)}{\bar{P}_{D=1}(X, Z=1) - \bar{P}_{D=1}(X, Z=0)}$$

## Estimador de Wald en este caso

- Numerador:

$\bar{Y}(X, Z = 1)$ : ingresos medios de los sorteados ( $Z_i = 1$ ).

$\bar{Y}(X, Z = 0)$ : ingresos medios de los no sorteados ( $Z_i = 0$ ).

- Denominador:

$\bar{P}_{D=1}(X, Z = 1)$ : proporción de veteranos ( $D_i = 1$ ) en el grupo de sorteados ( $Z_i = 1$ ): de todos los elegidos en el sorteo, qué proporción efectivamente fue al ejército.

$\bar{P}_{D=1}(X, Z = 0)$ : proporción de veteranos en el grupo de no sorteados ( $Z_i = 0$ ): de los que no fueron elegidos en el sorteo, qué proporción igualmente fue al ejército.

# Resultados

**Table 2 IV estimates of the effects of military service on US white men born 1950**

| Earnings year | Earnings |                    | Veteran status |                    | Wald estimate of veteran effect |
|---------------|----------|--------------------|----------------|--------------------|---------------------------------|
|               | Mean     | Eligibility effect | Mean           | Eligibility effect |                                 |
|               | (1)      | (2)                | (3)            | (4)                | (5)                             |
| 1981          | 16,461   | - 435.8<br>(210.5) | 0.267          | 0.159<br>(.040)    | - 2,741<br>(1,324)              |
| 1970          | 2,758    | - 233.8<br>(39.7)  |                |                    | - 1,470<br>(250)                |
| 1969          | 2,299    | - 2.0<br>(34.5)    |                |                    |                                 |

Notes: Figures are in nominal US dollars. There are about 13,500 observations with earnings in each cohort. Standard errors are shown in parentheses.

Tabla tomada de Angrist y Pischke, Mostly Harmless Econometrics.

- Para los hombres nacidos en 1950, hay efectos negativos significativos de haber sido seleccionado en los ingresos en 1970, cuando estos hombres comenzaban el servicio militar y en 1981, diez años después.
- En contraste, no hay evidencia de una asociación entre haber sido elegido en el sorteo y los ingresos en 1969, el año que el sorteo para los hombres nacidos en 1950 se realizó pero nadie fue realmente seleccionado.

## Estimador de Wald

- Para pasar de los efectos del sorteo a los efectos de ser veterano se necesita el denominador del estimador de Wald, que es el efecto del sorteo en la probabilidad de ser veterano.
- Esta información se reporta en la columna (4), que muestra que los hombres sorteados tienen una probabilidad mayor de servir en el ejército en Vietnam que los no sorteados. La diferencia  $\bar{P}_{D=1}(X, Z=1) - \bar{P}_{D=1}(X, Z=0)$  es de 0.16 puntos.
- En 1981, mucho después de que estos veteranos dejaran el ejército, el estimador de Wald es aproximadamente 17% del ingreso medio.
- Los efectos son en términos porcentuales mayores en 1970, cuando los soldados afectados estaban todavía en el ejército.

## Regresión con discontinuidad: Introducción

- Otro método usado en cuasi experimentos es el llamado de regresión con discontinuidad (RD)
- RD es útil en casos en que:
  - una variable  $W$  afecta  $Y$ , el resultado de interés.
  - el tratamiento ( $D$ ) es una función discontinua de  $W$ . En concreto, participar en el tratamiento depende de que  $W$  supere un umbral  $w_0$ .
- Notar que  $W$  no es un instrumento válido ya que no satisface el supuesto de exogeneidad.
- RD explota que existe una discontinuidad en la relación entre  $D$  y  $W$  junto con continuidad en la relación entre  $Y$  y  $W$ .

## Ejemplo: El efecto de los cursos de verano en las notas

- Matsudaira en el Journal of Econometrics (2008) utiliza una política para alumnos de primaria y secundaria en EEUU por la cual los estudiantes cuya nota final no supera un mínimo deben asistir a un curso de verano.
- Una manera de estimar el efecto de estos cursos de verano en las notas del año siguiente sería comparar a estudiantes un poco por debajo del umbral (y que por tanto asistirán al curso) con los estudiantes un poco por encima del umbral (y que por tanto no asistirán al curso por muy poco) .
- En tanto el umbral no se utilice para tomar otras decisiones, parece razonable pensar que si se observa un salto en las notas del año siguiente alrededor del umbral, se deba al efecto del curso de verano.

## Idea de RD

- La idea de la regresión con discontinuidad es entonces estimar el efecto del tratamiento comparando individuos con  $W$  justo por debajo de un umbral  $w_0$  (que serán tratados) con individuos con  $W$  justo por encima de  $w_0$  (no tratados).
- Si el efecto directo de  $W$  en el resultado  $Y$  es continuo, el efecto del tratamiento debería aparecer como un salto en  $Y$  alrededor de  $w_0$ . La magnitud de este salto estima el efecto del tratamiento.
- Supuesto clave: individuos justo por encima o por debajo de  $w_0$  son comparables. Cierta variación aleatoria hace que alguien quede por encima de  $w_0$  o por debajo, generando diferencias en el tratamiento. Entonces, cualquier diferencia en  $Y$  en  $w_0$  se debe al tratamiento.

# Tipos de RD

Existen dos tipos de diseño para RD:

- Diseño brusco: todas las unidades por encima (o por debajo) del umbral  $w_0$  reciben tratamiento.
- Diseño difuso: superar el umbral  $w_0$  afecta la probabilidad de recibir tratamiento pero no es el único determinante.

## Diseño brusco para RD 1/2

- Se conoce como diseño brusco para RD cuando el tratamiento ( $D_i$ ) es una función determinística y discontinua de una variable observable

$W$ . Por ejemplo:

$$D_i = \begin{cases} 1 & \text{si } W_i < w_0 \\ 0 & \text{si } W_i \geq w_0, \end{cases}$$

donde  $w_0$  es un umbral conocido.

- Por ejemplo, todos los estudiantes con  $W < w_0$  asisten al curso de verano y ningún estudiante con  $W \geq w_0$  asiste.
- En este caso el salto en  $Y$  en  $w_0$  es el efecto medio del tratamiento para los individuos en el umbral. Bajo ciertos supuestos puede ser una buena aproximación al efecto para otros individuos.

## Diseño brusco para RD 2/2

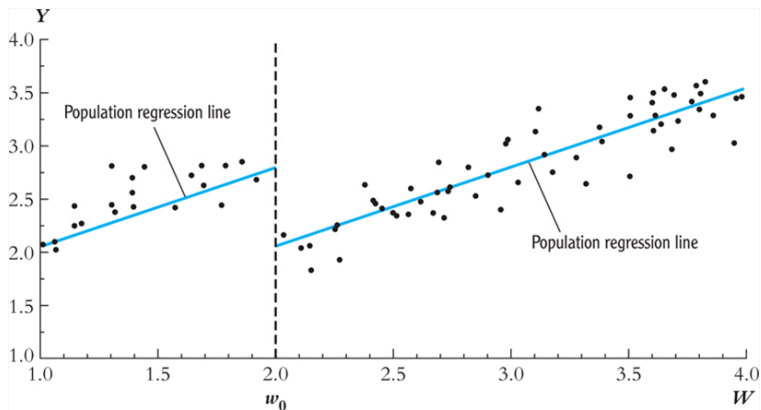
- Si el modelo de regresión es lineal en  $W$ , excepto por el salto debido al tratamiento, el efecto del tratamiento  $\beta_1$  puede estimarse por MCO:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 W_i + u_i$$

- Si cruzar el umbral afecta a  $Y_i$  solamente a través de  $D_i$ , entonces los estimadores MCO son consistentes.
- En un diseño brusco,  $D_i$  es una función determinística de  $W_i$ . El efecto causal de  $D_i$  se identifica al controlar por la relación entre  $W$  e  $Y$  (representada por  $W_i$ ).
- Es posible, y común, incluir  $f(W_i)$  en lugar de  $W_i$  en la regresión, con  $f(W_i)$  continua en el entorno de  $w_0$ . El objetivo de  $f(W_i)$  es capturar lo mejor posible la relación entre  $W$  e  $Y$ .

## Diseño brusco para RD - figura en caso lineal

- Todos los individuos con  $W$  debajo de  $w_0$  son tratados, el efecto del tratamiento es el salto o “discontinuidad”



De Stock y Watson, cap. 13.

## MCO en el ejemplo de los cursos de verano

- Consideremos el siguiente modelo:  $Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$ , donde  $D_i$  es una variable binaria que vale uno si  $i$  asiste al curso de verano.
- El problema si quisiéramos estimar el efecto de  $D_i$  utilizando MCO es que la asistencia es una variable endógena.
- De hecho, el autor muestra que los estudiantes que asisten a los cursos de verano tienen peores notas, es más probable que sean negros o hispanos, que tengan almuerzo gratis y que vivan en barrios más pobres. Además podemos pensar que existen diferencias no observables.
- Por lo tanto es de esperar que exista un sesgo negativo en las estimaciones MCO de una ecuación como la anterior.

## Solución con diseño brusco para RD 1/2

- Si todos los estudiantes con notas debajo del umbral asisten al curso de verano con probabilidad uno, estamos en un caso de diseño brusco.
- En ese caso se puede mostrar que el efecto de asistir al curso de verano se logra identificar comparando las notas del año posterior al curso de verano para estudiantes justo por debajo y justo por encima del umbral.
- Porque los estudiantes que obtuvieron notas justo por debajo del umbral son en promedio parecidos en todos los factores relevantes a los estudiantes que obtuvieron notas justo por encima del umbral.
- El supuesto que se necesita es que la esperanza condicional de todas las características que afectan a las notas son continuas alrededor del umbral.

## Solución con diseño brusco para RD 2/2

- Si los estudiantes que obtienen notas por debajo del umbral asisten al curso de verano deberíamos observar un salto en la asistencia en  $W = w_0$ .
- Y si el programa tiene un efecto en  $Y_i$ , deberíamos observar un salto  $W = w_0$  también para  $Y$ .
- En ese caso es posible estimar el efecto de asistir al curso de verano en una ecuación como:

$$Y_i = \beta_0 + \beta_1 D_i + f(W_i) + u_i,$$

donde  $f(W_i)$  es una función de  $W$ , por ejemplo un polinomio de orden  $p$ .

## Diseño difuso para RD

- En el caso del diseño llamado difuso para la regresión con discontinuidad, cruzar el umbral afecta la probabilidad de ser tratado.
- Por ejemplo, los estudiantes con notas debajo del umbral tienen mayor probabilidad de asistir al curso de verano, pero por otros motivos pueden ser eximidos. A su vez, estudiantes con notas encima del umbral es menos probable que asistan pero por otros motivos puede que deban hacerlo. Es el caso en el que las reglas que determinan la participación en el tratamiento no solamente dependen de  $W$ .
- Por lo tanto cruzar el umbral no obliga a los individuos a recibir tratamiento sino que otros factores entran en consideración.

## Solución con diseño difuso para RD

- Definiendo una variable binaria para cruzar el umbral:

$$Z_i = \begin{cases} 1 & \text{si } W_i < w_0 \\ 0 & \text{si } W_i \geq w_0, \end{cases}$$

- Cruzar el umbral afecta la probabilidad de recibir tratamiento:  
 $C(Z_i, D_i) \neq 0$ :  $Z_i$  es un instrumento relevante para  $D_i$ .
- Si cruzar el umbral no tiene efecto directo sobre  $Y_i$ , sino que solamente afecta a  $Y_i$  via afectar la probabilidad de tratamiento, entonces  $Z_i$  es un instrumento exógeno para  $D_i$ .
- El diseño difuso para RD implica estimar la siguiente ecuación:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 W_i + u_i,$$

usando  $Z_i$  como instrumento para  $D_i$ .

## Matsudaira (2008)

- Dado que las reglas para asistir al curso de verano dependen de otros factores además de la nota final (una asistencia mínima por ejemplo), el autor utiliza un diseño difuso para RD.
- Sea  $W_i$  el resultado del estudiante  $i$  en el examen de matemáticas en 2001. Entonces:

$$M_i = \begin{cases} 1 & \text{si } W_i < 0 \\ 0 & \text{si } W_i \geq 0, \end{cases}$$

$M_i$  indica que el estudiante es obligado a asistir al curso de verano (el umbral se normaliza a cero).

- El autor usa  $M_i$  (la variable que vale uno si no se pasa el umbral) como instrumento para asistir al curso de verano ( $D_i$ ) y un polinomio de orden tres para  $W$ .

# Discusión

- RD resulta atractivo:
  - La estructura de umbrales para asignar tratamiento existe en varios programas y se puede explotar la existencia de reglas claras.
  - Muy intuitivo: fácil de comunicar los resultados.
- Problemas potenciales:
  - La necesidad de datos: es importante contar con suficientes observaciones alrededor del umbral.
  - El mismo valor del umbral puede ser utilizado para otros objetivos.
  - La manipulación del umbral: los individuos pueden intentar modificar la variable de asignación para terminar justo debajo o encima del umbral dependiendo si desean ser parte o no del tratamiento.
  - La extrapolación a la población de interés requiere más supuestos.
  - La sensibilidad a la forma funcional: la relación entre la variable de asignación y el resultado debe capturarse correctamente.