

Identificación Efectos Causales: Una Introducción

Alvaro J. Riascos Villegas¹

¹Universidad de los Andes y Quantil

16 de junio de 2022

Introducción

- Supongamos que una variable de interés y depende de un conjunto de variables x (observables y no observables):
 $y = f(x)$.
- El efecto causal de una variable x_i en y se define como el efecto en y de mover x_i dejando todas las demás variables x_{-i} fijas (i.e., ceteris paribus).
- Para variables continuas es la derivada parcial y para discretas la diferencia finita.
- Como un modelo realista de la relación entre y y x es un modelo estadístico, nos concentramos en modelos de la forma $f(x) = E[y | x] + \mu$. Esto es sin pérdida de generalidad porque cualquier variable aleatoria y podemos escribirla como $y = E[y | x] + \mu$ (Teorema de Descomposición), donde $E[\mu | X] = 0$. En particular μ es ortogonal a X .

Introducción

- La ciencia y el arte para identificar y cuantificar un efecto causal es:
 - ① Tener una especificación adecuada de f .
 - ② Argumentar convincentemente que, por lo menos conceptualmente, es posible mover x_i manteniendo fijas las variables x_{-i} .
- Hay varias formas para hacerlo: usando modelos estructurales (i.e., modelos físicos, teóricos, etc.), experimentos controlados o, estrategias de identificación basadas en datos observados.

Modelo de regresión lineal

- Considere la siguiente especificación lineal (poblacional):

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K + \mu \quad (1)$$

- Típicamente no conocemos la verdadera forma funcional y solo observamos una muestra de datos (i.e., no conocemos el verdadero proceso generador de los datos).
- Suponiendo una muestra aleatoria de observaciones (i.e., i.i.d), una hipótesis clave para poder estimar β de forma consistente¹ es:

$$E(\mu) = 0 \quad (2)$$

$$\text{Cov}(x_j, \mu) = 0 \text{ hipótesis de exogeneidad.} \quad (3)$$

¹La consistencia de un estimador es lo mínimo que podemos pedir para que este sea de interés.

Modelo de regresión lineal

- Si \mathbf{x} tiene una constante, las condiciones anteriores son equivalentes a $E[\mathbf{x}'u] = 0$, condición de ortogonalidad.
- Ambas condiciones son una consecuencia de $E[\mu | \mathbf{x}] = 0$, hipótesis de esperanza condicional cero.
- La hipótesis de ortogonalidad es más débil que la hipótesis de independencia (e.g., $Cov(\mathbf{x}, \mathbf{x}^2) = 0, \mathbf{x} \sim U[-1, 1]$).

Formas de endogeneidad

- Tres formas:
 - 1 Variable omitida: el salario de una persona lo determina el nivel de educación observable y la habilidad (no observable). Otro caso común es la autoselección: las personas que reciben un tratamiento son elegibles de forma aleatoria pero el tratamiento es voluntario y puede depender de variables no observadas.
 - 2 Error de medición: medición de sobredosis de consumo de drogas con base en entrada a urgencias en IPS.
 - 3 Simultaneidad: demanda de un bien se determina de forma simultánea con oferta.

OLS

- Si se cumple la hipótesis de ortogonalidad y del rango completo entonces β queda **identificado**:

$$\beta = (E[x'x])^{-1}E[x'y] \quad (4)$$

- Dada una muestra aleatoria el estimador muestral de β es **consistente** (estimador OLS).
- Si se cumple la hipótesis ligeramente más fuerte de esperanza condicional cero, entonces el estimador OLS está identificado, es consistente y es **no sesgado**.

Variables instrumentales

- Suponga que x_K está correlacionado con μ .
- El método de VI consiste en encontrar una variable z_1 tal que:
 - 1 $Cov(z_1, \mu) = 0$.
 - 2 Si proyectamos x_K en $[1, x_1, \dots, x_{K-1}, z_1]$:

$$x_K = \delta_0 + \delta_1 x_1 + \dots + \theta_1 z_1 + r_K \quad (5)$$

$\theta_1 \neq 0$ (obsérvese que por construcción se cumple la condición de ortogonalidad de r_K y $[1, x_1, \dots, x_{K-1}, z_1]$).

- Esta hipótesis se puede enunciar como, una vez se controla por las demás variables exógenas, x_K está correlacionado con z_1 .
- En este caso decimos que z_1 es un candidato a VI de x_K .

Variables instrumentales

- La segunda ecuación se llama la ecuación en **forma reducida**.
- Con las hipótesis mencionadas se pueden identificar, los coeficientes β como el estimador de VI.
- La primera ecuación, ortogonalidad del instrumento, no es testeable. La forma reducida sí.

Ejemplos: Experimentos Aleatorios Controlados (RCT)

- El tratamiento es exógeno porque es aleatorio.
- En este caso no hay endogeneidad y se puede usar OLS, no es necesario usar VI.

$$y = \beta_0 + \beta_1 d + \mu \quad (6)$$

- En la evaluación de un programa los elegibles se determinan de forma aleatoria.
- Los tratados se autoseleccionan.
- La variable de elegibilidad puede usarse como un instrumento de la variable participación en el programa.

Ejemplos: Salario y educación

- Regresar salarios en experiencia y educación.
- Educación puede no ser exógena porque el error puede incluir variables omitidas no observables: habilidad intelectual, características familiares o variables potencialmente observables: educación de la madre.
- Si usamos la educación de la madre como instrumento de educación probablemente se cumple la condición de la ecuación en forma reducida.
- Sin embargo, la primera queda en duda: características familiares (i.e., educación del padre) pueden correlacionarse con la educación de la madre (i.e., no se cumpliría la hipótesis de exogeneidad del instrumento).
- Alternativamente, el último dígito de la cédula de la persona cumpliría exogeneidad pero no estaría correlacionado con la educación de la persona.

Ejemplos: Salario y educación (Angrist y Krueger (1991))

- Solución de Angrist y Krueger (1991). Considere como instrumento si la persona nació en el primer trimestre del año.
- Exogeneidad es creíble, correlación en forma reducida: la ley obliga por lo menos a algunas personas a matricularse al colegio, los que nacen en el primer trimestre estudian más.
- Problema: que el efecto que en realidad se este midiendo sea solo el de aquellos que fueron obligados a ir a la escuela y por lo tanto tener más tiempo de educación.

- En ocasiones tenemos una variación exógena (aleatoria) en el contexto económico que se correlaciona con la variable endógena: anuncios en TV de desmovilización de las FARC y el clima en el país.

Ejemplos: Experimentos naturales: Salario y participación Vietnam

- Angrist - Krueger (1991).
- Estudian el efecto de participar en la guerra de Vietnam en los ingresos de los hombres.
- Participar puede no ser exógena a características no observadas.
- Usan como instrumento el número de la balota de prestación del servicio: entre menor el valor más probable ser llamado a prestar servicio.
- Pueden persistir problemas de endogeneidad: si no se controla por educación en la ecuación de ingresos, hombres con balota con número bajo saben que su probabilidad aumenta y eligen tener mayor educación buscando una oportunidad de no prestar servicio.

Ejemplos: No Experimentales (Evans y Schab (1995))

- Estudian el efecto de asistir a colegios católicos en diferentes resultados de los estudiantes.
- Instrumento de atender a un colegio católico: ser católico o no.
- En ocasiones se desea estimar el efecto de una variable endógena y se usa como instrumento la variación en precios de las regiones o impuestos como instrumentos.
- Ejemplo: Si se desea estimar el efecto del consumo de alcohol en el rendimiento universitario se puede usar el precio del alcohol como instrumento del consumo.

Ejemplos: Sesgo de usar OLS

- Usar IV no siempre disminuye el efecto de la variable de interés.
- Card (1995) usa como instrumento si una persona nació cerca a una universidad como instrumento de años de escolaridad.
- Al estimar el efecto de los retornos a la educación los resultados pasan de 7.5 % (OLS) a 13.2 % (IV).

Variables Instrumentales y Doble Selección

- Una tentación en el modelo de VI es incluir la mayor cantidad de controles observados.
- La especificación lineal es una hipótesis. Introducir interacciones, polinomios incrementa las variables considerablemente.
- Podemos terminar con $K > n$ (i.e., por ejemplo si queremos estimar una demanda por SKU y controlamos por los precios de todos los SKUs).
- La técnicas de reducción de dimensionalidad como LASSO pueden ayudarnos.

- Esta estrategia consiste en utilizar una técnica de reducción de dimension en la primera etapa.
- La segunda etapa se estima de la forma tradicional.

- Sustituir la ecuación en forma reducida en la ecuación estructural.
- El nuevo sistema de ecuaciones es un nuevo sistema en forma reducida.
- Hacer reducción de dimensión en cada ecuación para obtener las variables: x^r y z^r respectivamente.
- Regresar y en x^K, x^r, z^r .

Ejemplos: Leyes de Extinción de Dominio y Precio de Propiedades

- Regresar precio vivienda en la decisión de leyes de extinción de dominio y controles ($n=183$).
- Se usa como instrumento características de los jueces. Como estos se asignan aleatoriamente se cumple la restricción de ortogonalidad.
- Se prueba la segunda hipótesis: correlación de características con decisión.
- Se incluyen interacciones, polinomios cúbicos (147 instrumentos).
- Efectos estimados de una decisión adicional de protección de derechos de propiedad privada aumenta el precio de las viviendas entre 2 – 11 %.

- Donohue y Levitt (2001) estudian los efectos de legalización del aborto en el crimen veinte años después (n=600).
- Incluyen 248 instrumentos

Ejemplos: Legalización aborto y crimen

Table 1

Effect of Abortion on Crime

<i>Estimator</i>	<i>Type of crime</i>					
	<i>Violent</i>		<i>Property</i>		<i>Murder</i>	
	<i>Effect</i>	<i>Std. error</i>	<i>Effect</i>	<i>Std. error</i>	<i>Effect</i>	<i>Std. error</i>
First-difference	-.157	.034	-.106	.021	-.218	.068
All controls	.071	.284	-.161	.106	-1.327	.932
Double selection	-.171	.117	-.061	.057	-.189	.177

Notes: This table reports results from estimating the effect of abortion on violent crime, property crime, and murder. The row labeled “First-difference” gives baseline first-difference estimates using the controls from Donohue and Levitt (2001). The row labeled “All controls” includes a broad set of controls meant to allow flexible trends that vary with state-level characteristics. The row labeled “Double selection” reports results based on the double selection method outlined in this paper and selecting among the variables used in the “All controls” results.

- Acemoglu, Johnson y Robinson (2001).
- Utilizan la mortalidad de los primeros colonizadores europeos como instrumento de la calidad de las instituciones.
- Controlan por variables geográficas ($n=64$).

Table 2

Effect of Institutions on Output

	<i>Latitude</i>	<i>All controls</i>	<i>Double selection</i>
First stage	-0.5372 (0.1545)	-0.2182 (0.2011)	-0.5429 (0.1719)
Second stage	0.9692 (0.2128)	0.9891 (0.8005)	0.7710 (0.1971)

Notes: In an exercise that follows the work of Acemoglu, Johnson, and Robinson (2001), this table reports results from estimating the effect of institutions, using settler mortality as an instrument. The row “First Stage” gives the first-stage estimate of the coefficient on settler mortality obtained by regressing “*ProtectionfromExpropriation_i*” on “*SettlerMortality_i*” and the set of control variables indicated in the column heading. The row “Second stage” gives the estimate of the structural effect of institutions on log(GDP per capita) using “*SettlerMortality_i*” as the instrument and controlling for variables as indicated in the column heading (see text for details). Each column reports the results for different sets of control variables. The column “Latitude” controls linearly for distance from the equator. The column “All controls” includes 16 controls defined in the main text and in footnote 9, and the column “Double selection” uses the union of the set of controls selected by LASSO for predicting GDP per capita, for predicting institutions, and for predicting settler mortality. Standard errors are in parentheses.