Variables Instrumentales

2025-05-28

Introducción

Introducción

Como vimos, los experimentos son el estándar de oro en la inferencia causal.

Podemos intentar pensar en factores de confusión y controlarlos usando regresión, emparejamiento, etc.

Pero, como vimos, puede ser difícil saber si tenemos un buen grupo de comparación y si controlamos todos los factores relevantes.

Introducción

Reconociendo este hecho, la microeconomía aplicada se ha centrado más en métodos "cuasi-experimentales".

Los métodos "cuasi-experimentales" son métodos que intentan encontrar variación natural que "simula" un experimento.

Puede ser en forma de "experimentos naturales": situaciones que son plausiblemente similares a un experimento real.

O podemos ser capaces de aislar alguna variación que creemos que es "aleatoria" en algún sentido.

Introducción

Hoy estudiaremos uno de estos métodos, las **variables instrumentales**.

Primero, revisaremos lo que vimos sobre cómo lidiar con el cumplimiento imperfecto en los experimentos.

Cumplimiento Imperfecto

Cumplimiento Imperfecto

Pensemos en un experimento que implica el envío de cartas informativas que animan a los padres a vacunar a sus hijos.

De una muestra, un grupo recibe una carta que fomenta la participación ($Z = 1$).

Otros están en el grupo de control y no reciben una carta ($Z = 0$).

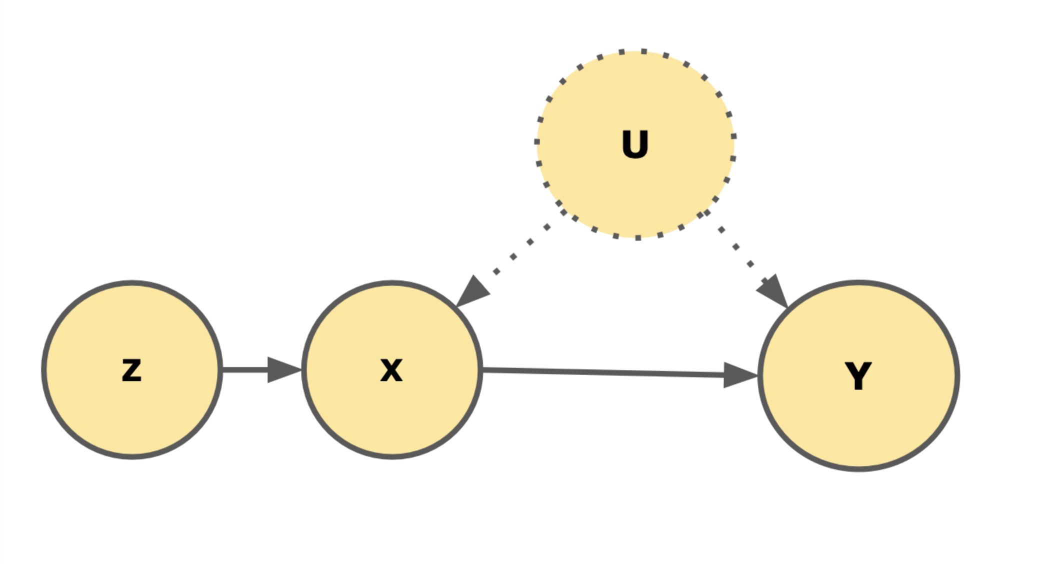
Nos interesa saber si los niños contraen una determinada enfermedad ($Y = 1$) o no ($Y = 0$).

Cumplimiento Imperfecto

Lo que queremos saber es cuánto afecta la administración de la vacuna (X) a la probabilidad de contraer la enfermedad.

La asignación al tratamiento Z es aleatoria, pero el tratamiento real X no lo es. El fomento aumenta la aceptación, pero algunas personas se vacunarán sin fomento, y otras no lo harán incluso con fomento.

Cumplimiento Imperfecto



Cumplimiento Imperfecto

Llamamos ITT (Intent-to-Treat) al efecto de recibir una carta sobre el resultado $Z \rightarrow Y$

El ITT es fácil de obtener: debido a que Z es aleatorio, no hay variables de confusión. No es necesario controlar nada. Simplemente comparar promedios.

$$ITT = E[Y|Z = 1] - E[Y|Z = 0]$$

Cumplimiento Imperfecto

También podemos encontrar fácilmente el efecto del fomento en la aceptación ($Z \rightarrow X$).

De nuevo, Z es aleatorio, así que solo hay que comparar promedios:

$$E[X|Z=1] - E[X|Z=0]$$

Este efecto se denomina típicamente la **Primera Etapa**.

Cumplimiento Imperfecto

Con un poco más de trabajo, podemos obtener el **LATE**, que es el efecto de recibir realmente el tratamiento **para los cumplidores**:

$$LATE = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[X|Z=1] - E[X|Z=0]} = \frac{ITT}{PrimeraEtapa}$$

En otras palabras, el LATE es el efecto de la asignación sobre el resultado dividido por el efecto de la asignación sobre el tratamiento.

Cumplimiento Imperfecto

Acuerdate que el ITT es igual a un promedio del efecto sobre los compliers, y 0 (el efecto sobre Tomadores constantes y Rechazadores constantes).

$$ITT = Pr(cumplidor) \cdot E[\delta | cumplidor] + Pr(tomador) \cdot 0 + Pr(rechazador) \cdot 0$$

Cuando dividimos por la Primera etapa, estamos simplemente eliminando el $Pr(cumplidor)$:

$$\frac{ITT}{Pr(cumplidor)} = E[\delta | cumplidor] = LATE$$

Cumplimiento Imperfecto

En cierto modo, esto tiene mucho sentido:

- Fomentar la vacunación solo puede afectar a la enfermedad si la gente se vacuna.
- Imagina que las cartas de fomento aumentaron la vacunación en 50 p.p.
- Así que si el fomento disminuyó la enfermedad en 5 p.p., eso significa que las personas que se vacunaron deben haber tenido un 10 p.p. menos de probabilidad de enfermedad.

Cumplimiento Imperfecto

Los supuestos clave que hicimos son:

1. La asignación es aleatoria, es decir, no es causada por ninguna otra variable que afecte a X o Y.

- Exogeneidad

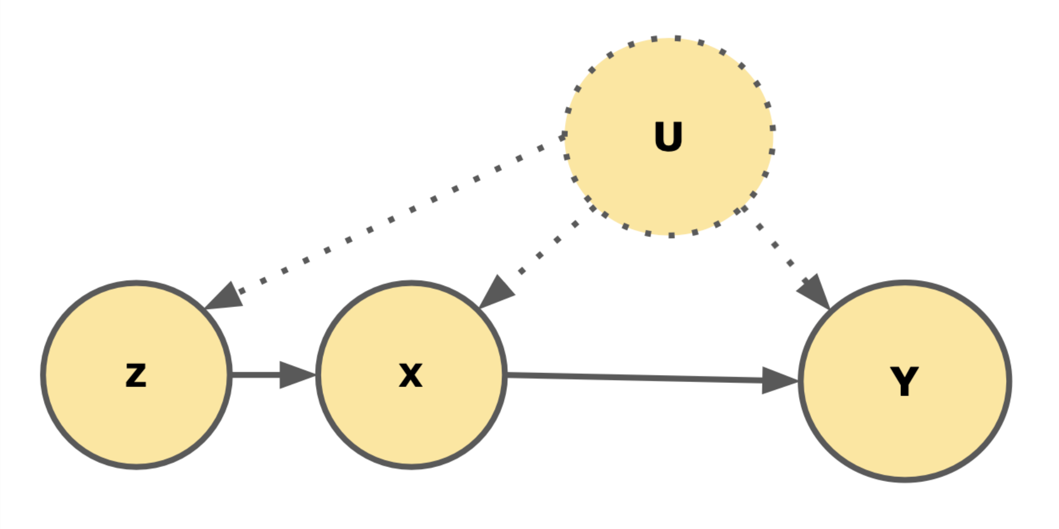
2. La asignación no puede afectar a Y directamente.

- Exclusión

3. No hay *defiers*. Es decir, individuos que son tratados cuando asignados al control y no-tratados cuando asignados al tratamiento.

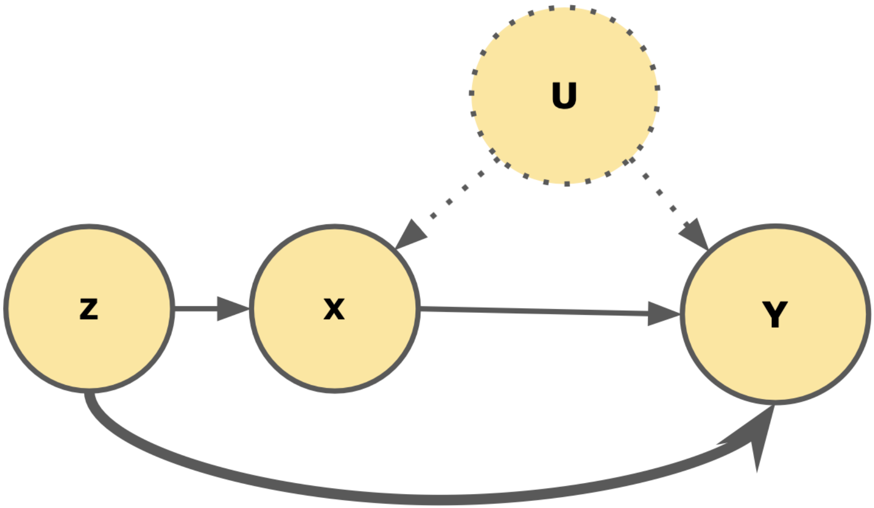
Cumplimiento Imperfecto

Violación de Exogeneidad



Cumplimiento Imperfecto

Violación de Exclusión



Cumplimiento Imperfecto

Pregunta: En el experimento de vacunación, si las cartas de fomento aumentaron la vacunación en 20 p.p. y disminuyeron la enfermedad en 4 p.p., ¿cuál es el efecto de tratamiento promedio local (LATE) estimado de la vacunación sobre la enfermedad?

Cumplimiento Imperfecto

Pregunta: En el experimento de vacunación, si las cartas de fomento aumentaron la vacunación en 20 p.p. y disminuyeron la enfermedad en 4 p.p., ¿cuál es el efecto de tratamiento promedio local (LATE) estimado de la vacunación sobre la enfermedad?

Respuesta: -0.04 / 0.2 = -0.2, o 20 p.p. En otras palabras, vacunarse reduce la probabilidad de enfermedad en 20 p.p. (para las personas sensibles al fomento).

Cumplimiento Imperfecto

En realidad, este es un caso particular de **Variables Instrumentales**.

Llamamos a Z el **instrumento**, y a X la **variable endógena** o instrumentada.

Ejercicio 1

Un experimento evalúa el efecto de un programa de tutoría en las calificaciones. Los estudiantes son asignados aleatoriamente a recibir una invitación al programa (Z=1) o no (Z=0).

Datos:

- Calificación promedio en grupo control (Z=0): 75
- Calificación promedio en grupo tratamiento (Z=1): 78
- Participación en tutoría en grupo control: 20%
- Participación en tutoría en grupo tratamiento: 60%

Preguntas:

- ¿Cuál es el ITT?

Ejercicio 1

Un experimento evalúa el efecto de un programa de tutoría en las calificaciones. Los estudiantes son asignados aleatoriamente a recibir una invitación al programa (Z=1) o no (Z=0).

Datos:

- Calificación promedio en grupo control (Z=0): 75
- Calificación promedio en grupo tratamiento (Z=1): 78
- Participación en tutoría en grupo control: 20%
- Participación en tutoría en grupo tratamiento: 60%

Preguntas:

- ¿Cuál es efecto en la Primera Etapa?

Ejercicio 1

Un experimento evalúa el efecto de un programa de tutoría en las calificaciones. Los estudiantes son asignados aleatoriamente a recibir una invitación al programa (Z=1) o no (Z=0).

Datos:

- Calificación promedio en grupo control (Z=0): 75
- Calificación promedio en grupo tratamiento (Z=1): 78
- Participación en tutoría en grupo control: 20%
- Participación en tutoría en grupo tratamiento: 60%

Preguntas:

- ¿Cuál es el LATE?

Cumplimiento Imperfecto

Recapitulación

Hemos revisado cómo las variables instrumentales pueden ayudarnos a estimar el efecto causal de un tratamiento (X) cuando la asignación al tratamiento (Z) es aleatoria pero el cumplimiento es imperfecto.

Idea clave: usar la asignación aleatoria como un 'instrumento' y ajustar por el hecho de que no todos siguen la asignación.

Esto nos da el Efecto de Tratamiento Promedio Local (LATE), que es el efecto sobre los *cumplidores*: aquellos cuyo estado de tratamiento está influenciado por el instrumento.

Tres supuestos fundamentales:

1. Exogeneidad,
2. Exclusión,
3. Monotonicidad.

Cumplimiento Imperfecto

Ahora vamos ver como ese mismo tipo de analisis puede ayudar en un contexto diferente.

La herramienta va ser la misma, pero la interpretación cambia.

Análisis de mediación

Análisis de mediación

Pensemos en un experimento diferente.

Estamos analizando un experimento que otorgó un valor monetario mensual a familias pobres.

No hay problema de cumplimiento: todos los asignados al tratamiento reciben el dinero, y nadie en el grupo de control.

Nuestro principal interés es la salud de los niños. Esperamos que el dinero extra ayude a lograr una mejor nutrición y, por lo tanto, una mejor salud.

También recopilamos datos sobre la asistencia escolar y las calificaciones.

Análisis de mediación

Nuestra medida promedio de salud es un índice compuesto basado en el peso, la altura, los niveles de hemoglobina, los niveles de vitamina A y la presencia de parásitos en los niños.

Los niños del grupo de control tuvieron una puntuación de salud de 6.5 sobre 10. El grupo de tratamiento tuvo una puntuación de salud de 7.0.

Por lo tanto, concluimos que el tratamiento aumenta la salud en 0.5 en esta escala.

Análisis de mediación

También analizamos los resultados educativos y encontramos que el grupo tratado tiene un 10 p.p. más de asistencia y terminó el año escolar con un 5 p.p. más de probabilidad de poder leer al nivel adecuado.

Ahora, queremos tomar información de este estudio para aprender sobre la relación general entre la salud y los resultados educativos.

Análisis de mediación

Resultados del programa:

Variable	Efecto
Salud	0.5
Asistencia	0.10
Lectura	0.05

Si asumimos que **el programa afecta la escolarización solo a través de la salud de los niños**, ¿cuál es el efecto de aumentar el índice de salud en 1 sobre la asistencia y la lectura?

Análisis de mediación

Asumiendo que los efectos en la escolarización ocurren solo a través de la salud, podemos concluir que, si el impacto en la salud hubiera sido el doble (si el índice aumentara en 1), entonces el efecto en la escolarización también sería el doble.

Por lo tanto, aumentar la salud en 1 aumenta la asistencia en 20 p.p. y la lectura en 10 p.p.

Mecánicamente, hicimos el mismo tipo de análisis que cuando estudiamos el cumplimiento imperfecto.

Concepto	Cumplimiento Imperfecto	Mediación
Z	Fomento	Programa de dinero
X	Vacunación	Salud
Y	Infección	Educación

Análisis de mediación

Relaciones:

ITT: $E[Y|Z=1]-E[Y|Z=0]$

- El efecto de la carta en las infecciones / El efecto del programa en educacion.

Primera Etapa: $E[X|Z=1]-E[X|Z=0]$

- El efecto de la carta en la vacunación / El efecto del programa en salud.

LATE: $\frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]}$

- El efecto de la vacunación en las infecciones / El efecto de la salud en la educación.

Análisis de mediación

El ITT en este caso es el efecto del programa sobre la educación. La Primera Etapa es el efecto del programa sobre la salud. Y el LATE es el efecto de la salud sobre la educación.

Tenga en cuenta que, en el caso de cumplimiento imperfecto, la participación es una variable ficticia, 0 o 1. Pero la salud es un indicador continuo.

Veamos cómo funcionan los supuestos en este caso:

1. Exogeneidad: Z es aleatorio.

Válido si el experimento fue aleatorio. No hay diferencia.

Veamos cómo funcionan los supuestos en este caso:

2. Exclusión: El experimento solo puede afectar la asistencia a través de la salud.

Esto puede ser problemático.

- Tal vez el dinero en efectivo permite a la familia pagar la escuela, o el transporte para la escuela.
- Tal vez el niño tuvo que trabajar para ayudar a mantener a la familia, pero ya no lo necesita.

No hay forma de probarlo directamente: depende de la teoría causal.

3. Monotonicidad (sin *defiers*)

El supuesto que necesitamos es que no hay ninguna persona que tenga peor salud con el programa de dinero en efectivo que sin él.

Por ejemplo, si encontráramos que el dinero en efectivo permitió a las familias comprar alcohol o drogas y eso tuvo un impacto negativo en la salud del niño, entonces nuestro diseño tendría un problema.

Análisis de mediación

Si confiamos en estos supuestos, el método de IV es muy simple.

Esencialmente, es solo un **reescalado**: tomamos el ITT y lo dividimos por la Primera Etapa para que el "denominador" esté en términos de X en lugar de Z.

- ITT: Pasar de No Tratado (Z=0) a Tratado (Z=1) aumenta la Asistencia en 0.10.
  - Así que 1 unidad en Z -> 0.1 unidades en Y.
- IV: Pasar de Salud 6.5 a Salud 7 aumenta la Asistencia en 0.10.
  - Así que 1 unidad en X -> 0.2 unidades en Y.

Vimos que el LATE debe interpretarse como el efecto **sobre los cumplidores**.

¿Qué significa eso en este caso?

Significa que estamos estimando el efecto sobre las personas cuya salud mejoró debido al programa.

Tal vez algunos niños ya estaban bastante sanos y no mejoraron. El efecto causal para ellos no está incluido en la estimación.

En general, IV estima el efecto **sobre la población sensible al instrumento**.

Es importante señalar que la Primera Etapa está en el denominador, por lo que si el efecto es pequeño, las estimaciones tenderán a ser muy ruidosas.

Si el programa no cambia la salud en absoluto, entonces no podemos usarlo para estudiar el efecto de la salud en la escolarización.

$$IV = \frac{\text{efecto de Z sobre Y}}{\text{efecto de Z sobre X}}$$

Si el efecto de Z sobre X es 0, IV no está definido.

Así que, añadamos otro supuesto:

- La primera etapa es **fuerte**.

Por “fuerte” entendemos que el efecto es estadísticamente significativo y “grande”.

Si la Primera Etapa es débil, nuestras estimaciones serán muy ruidosas, probablemente inútiles.

Ejercicio 2:

Un programa de microcrédito tiene los siguientes efectos:

Aumenta los ahorros familiares en $200 mensuales Aumenta la inversión en educación en $50 mensuales Aumenta la inversión en salud en $30 mensuales

Pregunta: Si asumimos que el programa solo afecta las inversiones a través de los ahorros, ¿cuánto aumentarían las inversiones si los ahorros aumentaran en $100?

Recapitulación

Las Variables Instrumentales pueden aplicarse al análisis de mediación al igual que al cumplimiento imperfecto.

Incluso cuando hay un cumplimiento perfecto con el tratamiento inicial (Z), si ese tratamiento afecta a una variable intermedia (X) que luego afecta a nuestro resultado final (Y), podemos usar el tratamiento inicial como un instrumento para esa variable intermedia.

IV se trata de encontrar una ‘palanca’ válida (Z) que solo afecte nuestro resultado (Y) a través de la variable de interés (X). Los supuestos siguen siendo cruciales, especialmente la restricción de exclusión, que requiere una consideración cuidadosa.

IV Continua

IV Continua

Josh Angrist estaba interesado en cómo el servicio militar afectaba los salarios y el empleo de los hombres en el futuro.

Servir en el ejército generalmente impide la formación de capital humano cuando los jóvenes podrían ir a la universidad o comenzar sus carreras.

Por otro lado, el ejército enseña algunas habilidades potencialmente útiles, y los soldados hacen contactos valiosos.

IV Continua

¿Por qué no simplemente comparar a las personas que sirvieron con las que no lo hicieron?

Muchos posibles factores de confusión: los militares pueden haber tenido peores opciones en el mercado laboral, o pueden ser más adecuados para la vida militar.

Entonces, ¿cómo encontramos una "variación limpia en el servicio militar"?

IV Continua

Angrist utilizó la **lotería del reclutamiento**: durante la guerra de Vietnam, a cada hombre de un determinado grupo de edad se le asignó un número aleatorio único.

En cada comunidad, los hombres con los números más bajos fueron reclutados antes que otros.

El número no predice el servicio **perfectamente**:

- Diferentes áreas reclutaron diferentes números de reclutas.
- Las personas podían ser dadas de baja por varias razones (por ejemplo, médicas, proveedor único).
- Algunas personas simplemente no se alistaron, arriesgándose a la prisión.
- Otros pueden haberse alistado incluso sin el reclutamiento.

Pero tener un número bajo es un fuerte predictor de servir.

IV Continua

Si tuviéramos una lotería que asignara “servicio” y “no servicio” a los jóvenes, esto sería similar a un experimento con cumplimiento imperfecto.

- No todos siguieron la asignación, pero el reclutamiento “fomentó fuertemente la participación”.

Sin embargo, tenemos un número continuo que afecta la probabilidad de servicio.

¿Cómo lidiamos con eso?

IV Continua

Resulta que podemos usar la misma idea de estimar el efecto de la asignación en Y y dividirlo por el efecto en X.

$$IV = \frac{\text{efecto de Z sobre Y}}{\text{efecto de Z sobre X}}$$

La única diferencia es que "el efecto" aquí no es una diferencia de medias, sino el coeficiente de una regresión.

IV Continua

Entonces, podemos estimar:

$$\text{servicio}_i = \alpha_0 + \alpha_1 \text{número\_reclutamiento}_i + u_i$$

y:

$$\text{salarios}_i = \gamma_0 + \gamma_1 \text{número\_reclutamiento}_i + v_i$$

De modo que la estimación IV del efecto del servicio en los salarios es:

$$\hat{\beta} = \frac{\text{efecto de Z sobre Y}}{\text{efecto de Z sobre X}} = \frac{\hat{\gamma}_1}{\hat{\alpha}_1}$$

IV Continua

Digamos que estimamos:

$$E[\text{servicio}_i] = 0.350 - 0.00032\text{número\_reclutamiento}_i$$

Un número 100 más alto significa un 3.2% menos de probabilidad de servir.

$$E[\text{salario}_i] = 16000 + 1.25\text{número\_reclutamiento}_i$$

Un número 100 más alto también significa \$125 más de ingresos en promedio. Entonces,

$$\hat{\beta} = \frac{+1.25}{-0.00032} = -3906$$

Eso significa que servir causa que los salarios sean casi 4 000 dólares más bajos de lo que serían de otra manera.

IV Continua

Lo que hicimos es la aplicación práctica de IV usando variables continuas.

Estimamos dos regresiones simples:

- La Primera Etapa: vincula el instrumento (número de reclutamiento) con la variable endógena (servicio militar).
- "La Forma Reducida": vincula el instrumento con el resultado (salarios).

La razón de estos dos coeficientes nos da nuestra estimación IV, que nos dice el efecto causal del servicio militar en los salarios para aquellos cuyo servicio fue determinado por el reclutamiento.

Esta es precisamente la misma intuición que nuestros cálculos anteriores de LATE, solo que implementada usando regresión lineal.

Pensemos en los supuestos en este caso:

1. Exogeneidad: Z es aleatorio.

Necesitamos que el número de reclutamiento no esté relacionado con las ganancias futuras. En este caso, fue realmente aleatorio, así que está bien.

IV Continua

Pensemos en los supuestos en este caso:

2. Exclusión: el número de reclutamiento puede afectar los salarios *solo* a través del servicio militar.

Si el mismo número se usara para otros fines, podría ser un problema.

Otro problema potencial son las sentencias de prisión para los que evaden el reclutamiento: Alguien que se opone a la guerra podría haber sido arrestado si tenía un número bajo, pero simplemente no se alistó si tenía un número alto.

Así que, este es otro canal potencial. Si esto es relevante, nuestra estimación no sería solo el efecto de servir, sino también de ir potencialmente a prisión.

IV Continua

El tercer supuesto fue: 3. No hay *defiers* (incumplidores inversos).

Pero con un número continuo, es difícil definir un *defier*.

Llamemos a la probabilidad de que alguien sirva dado su número de reclutamiento como P(servicio|número_reclutamiento), o P(X|Z).

Entonces podemos definir el supuesto de monotonicidad:

Para cada individuo i:

$$Z_i \leq Z_i' \leftrightarrow P(X_i = 1|Z_i) \geq P(X_i = 1|Z_i')$$

Lo importante es que tener un número de lotería más alto nunca **disminuye** la probabilidad de servir para nadie.

IV Continua

En general, la Monotonicidad requiere que el instrumento solo aumente o solo disminuya la variable endógena.

En este caso, la monotonicidad parece plausible: un número de lotería más alto nunca debería aumentar la probabilidad de ser reclutado.

IV Continua

¿Qué hay de la interpretación como el efecto sobre los cumplidores?

Esto sigue siendo válido. En este caso, estamos estimando el efecto sobre las personas que sirvieron porque obtuvieron un número bajo, pero no habrían servido con un número bajo.

Para las personas que habrían servido de cualquier manera, o que no habrían servido en cualquier caso, sus efectos de tratamiento no están incluidos.

Pregunta 3

Para cada escenario, identifique qué supuesto de IV se viola:

Escenario A: Usamos la lotería del servicio militar como instrumento para estudiar el efecto del servicio en los ingresos. Sin embargo, los números bajos de lotería también se usaron para determinar elegibilidad para becas universitarias.

Escenario B: Usamos la proximidad a una universidad como instrumento para educación. Pero las personas que viven cerca de universidades tienden a vivir en áreas más prósperas.

Mínimos Cuadrados en Dos Etapas

Mínimos Cuadrados en Dos Etapas

Por ahora, vimos un estimador de IV que consiste en una división. Podemos llamar ese estimador de "Estimador de Wald".

En la práctica, no es siempre mejor la mejor manera de estimar el IV.

- (Veremos por qué más adelante)

Una forma más sencilla, pero equivalente de estimar el coeficiente de IV es la siguiente:

$$\hat{\beta}_{IV} = \frac{\text{Cov}(Y, Z)}{\text{Cov}(X, Z)}$$

Mínimos Cuadrados en Dos Etapas

$$\hat{\beta}_{IV} = \frac{\text{Cov}(Y, Z)}{\text{Cov}(X, Z)}$$

Para ver que es lo mismo, recordemos que:

- El coeficiente de regresión entre Y y Z es $\hat{\gamma}_1 = \frac{\text{Cov}(Y,Z)}{\text{Var}(Z)}$,
- y entre X y Z es $\hat{\alpha}_1 = \frac{\text{Cov}(X,Z)}{\text{Var}(Z)}$

Así, cuando dividimos, el $Var(Z)$ se cancela.

Mínimos Cuadrados en Dos Etapas

La forma más común de estimar IV se llama Mínimos Cuadrados en Dos Etapas (2SLS). Funciona así:

- Usamos el instrumento Z para explicar el tratamiento X.
- Removemos toda variación del tratamiento que *no es* explicada por el instrumento.
- Miramos la relación entre el resultado Y y la parte restante del tratamiento.

O, con más detalles:

- Regredimos X en Z: $X = \alpha_0 + \alpha_1 Z + \varepsilon_x$
- Obtenemos $\hat{X} = \hat{\alpha}_1 Z$.
- Regredimos Y en $\hat{X}$.
- El parametro estimado es el estimador de β: $\hat{\beta}_{iv}$

Mínimos Cuadrados en Dos Etapas

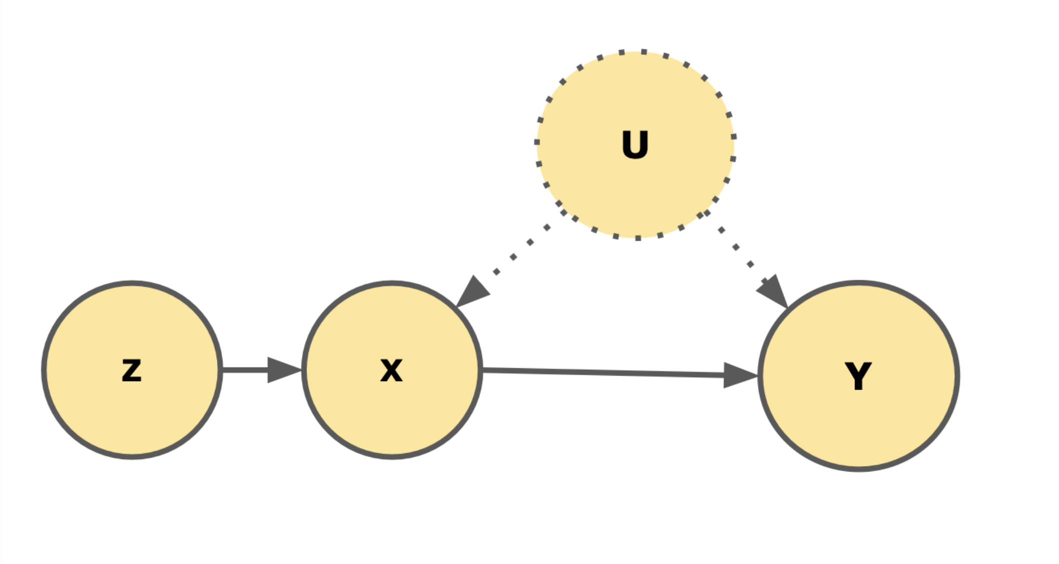
De cierta forma, similar a como funciona una regresión múltipla, pero “en revés.”

- Regresión múltipla: utilizamos los residuoes de la regresión de X en las otras variables, o sea, la variación en X que **no es** explicada por las otras variables.
- Variables instrumentales: utilizamos solamente la variación que **sí es** explicada por los instrumentos.

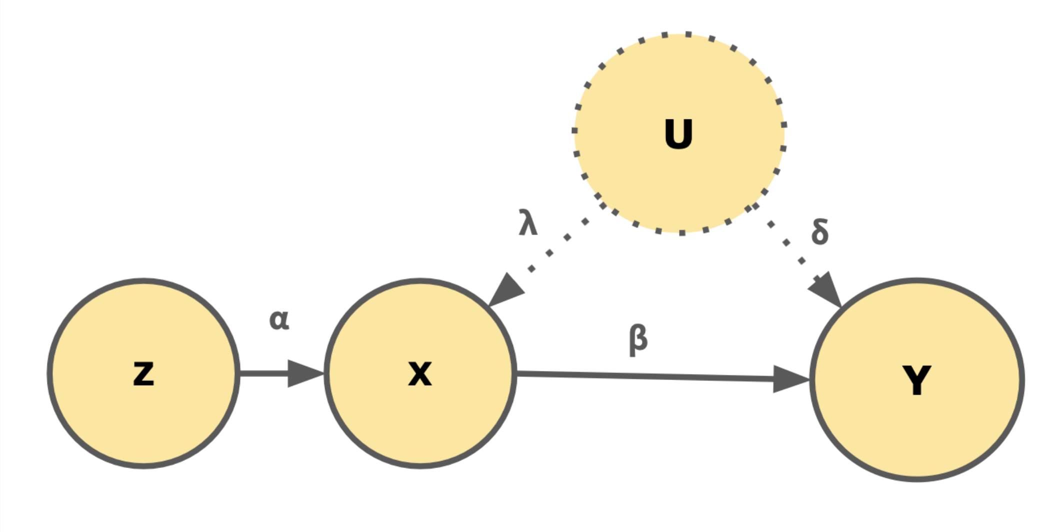
¿Por que funciona?

Vamos mirar las ecuaciones para compreender. Para simplificar, vamos asumir que los efectos causales son **constantes**, o sea, todos tienen el mismo efecto causal.

¿Por que funciona?



¿Por que funciona?



¿Por que funciona?

Primero, vamos a ver por qué el estimador de MCO es sesgado en ese caso.

Se acuerda que el parámetro estimado en una regresión de A en B es (en límite): $\frac{cov(A,B)}{var(B)}$.

¿Por que funciona?

Entonces, si regredimos Y en X, tenemos:

$$
\begin{aligned}
C(Y, X) &= C(\beta X + \delta U + \varepsilon_Y, X) \\
&= \beta V(X) + \delta C(U, X) + C(\varepsilon_Y, X) \\
&= \beta V(X) + \delta C(U, \alpha_0 + \alpha_1 Z + \lambda U + \varepsilon_X) \\
&= \beta V(X) + \delta \lambda V(U) + \delta \alpha_1 C(U, Z) + \delta C(U, \varepsilon_X) + C(\varepsilon_Y, X)
\end{aligned}
$$

Asumiendo $C(U, \varepsilon_X) = 0$ y $C(\varepsilon_Y, X) = 0$:

$$
C(Y, X) = \beta V(X) + \delta \lambda V(U) + \delta \alpha_1 C(U, Z)
$$

¿Por que funciona?

Luego, si $C(U, Z) = 0$, la regresión resulta:

$$\beta_{MCO} \to \frac{\beta V(X) + \delta\lambda V(U)}{V(X)} = \beta + \delta\lambda\frac{V(U)}{V(X)}$$

Está sesgado porque existe un confusor.

¿Por que funciona?

Ahora vamos a ver el estimador de IV.

La primera etapa es la regresión de X en Z.

$$
\begin{aligned}
\alpha_1 \rightarrow & \frac{C(X,Z)}{V(Z)} \\
& = \frac{C(\alpha_0 + \alpha_1 Z + \lambda U + \varepsilon_X, Z)}{V(Z)} \\
& = \alpha_1 + \lambda \frac{C(U,Z)}{V(Z)} + \frac{C(\varepsilon_X, Z)}{V(Z)}
\end{aligned}
$$

Asumiendo $C(U,Z) = 0$ y $C(\varepsilon_X, Z) = 0$:

$$
\alpha_1 \rightarrow \alpha_1
$$

Podemos escribir: $\hat{X} = \hat{\alpha}_0 + \hat{\alpha}_1 Z$

¿Por que funciona?

Ahora, la segunda etapa (tomando los interceptos como constantes):

$$
\begin{aligned}
C(Y,\hat{X}) &= C(\beta_0 + \beta_1 X + \delta U + \varepsilon_Y, \hat{\alpha}_0 + \hat{\alpha}_1 Z) \\
C(Y,\hat{X}) &\approx C(\beta_0 + \beta_1 X + \delta U + \varepsilon_Y, \alpha_0 + \alpha_1 Z) \\
&= \beta_1 C(X, \alpha_1 Z) + \delta C(U, \alpha_1 Z) + C(\varepsilon_Y, \alpha_1 Z) \\
&= \beta_1 \alpha_1 C(X, Z) + \delta \alpha_1 C(U, Z) + \alpha_1 C(\varepsilon_Y, Z)
\end{aligned}
$$

Sustituyendo $X = \alpha_0 + \alpha_1 Z + \lambda U + \varepsilon_X$ y asumiendo $C(U,Z) = 0$, $C(\varepsilon_X, Z) = 0$, $C(\varepsilon_Y, Z) = 0$:

$$
\begin{aligned}
C(Y,\hat{X}) &\approx \beta_1 \alpha_1 (\alpha_1 V(Z) + \lambda C(U,Z) + C(\varepsilon_X, Z)) + \delta \alpha_1 C(U,Z) + \alpha_1 C(\varepsilon_Y, Z) \\
&= \beta_1 \alpha_1 (\alpha_1 V(Z) + \lambda(0) + (0)) + \delta \alpha_1 (0) + \alpha_1 (0) \\
&= \beta_1 \alpha_1^2 V(Z)
\end{aligned}
$$

¿Por que funciona?

Si asumimos que $C(U, Z) = 0$, y nos acordamos que $V(\hat{X}) \approx V(\alpha_1 Z) = \alpha_1^2 V(Z)$:

$$\beta_{IV} \rightarrow \frac{C(Y, \hat{X})}{V(\hat{X})} = \frac{\beta_1 \alpha_1^2 V(Z)}{\alpha_1^2 V(Z)} = \beta_1$$

Así, el estimador de Mínimos Cuadrados en Dos Etapas (2SLS) es consistente para β_1.

Una ventaja de definir el estimador de esa forma es que ahora podemos trabajar con *multiples instrumentos*.

Si tenemos dos o más variables Z que podrían ser instrumentos válidos, podemos utilizar todos a la vez.

Para eso, simplemente usamos todos para prever X en la primera etapa.

Vamos a ver en seguida un ejemplo práctico de cuando se utiliza eso.

Educación y Semestre de Nacimiento

Educación y Semestre de Nacimiento

La relación entre educación y salarios es una cuestión clásica en economía laboral.

Normalmente es difícil de responder porque existen factores de confusión que afectan tanto la elección de la educación como los salarios.

¿Por ejemplo?

Educación y Semestre de Nacimiento

Si queremos utilizar el método de variables instrumentales, necesitamos algo que afecte la educación, pero que no afecte los salarios directamente.

Angrist y Krueger proponen: el trimestre de nacimiento.

¿Por qué?

Educación y Semestre de Nacimiento

La lógica proviene de la combinación de dos políticas:

- Por ley, una persona debe estar en la escuela hasta los 16 años de edad.
- Normalmente un niño empieza la escuela si tiene 6 años de edad el 1 de enero.

Educación y Semestre de Nacimiento

Eso significa que dos personas nacidas en fechas distintas empiezan la escuela a edades ligeramente diferentes y cumplen 16 años en momentos distintos.

A los 16 años pueden elegir dejar la escuela e ir a trabajar.

Educación y Semestre de Nacimiento

Por ejemplo:

- Ana nació el 2 de enero. Como las clases empiezan en septiembre, el año en que empieza la escuela ella tiene 7 años y 9 meses.
- Bernardo nació el 28 de diciembre. Él empezará la escuela cuando tenga 6 años y 9 meses.
- Ana cumple 16 años después de 8 años y 3 meses de educación. Bernardo cumple 16 años con 9 años y 3 meses. Bernardo debe cursar un año adicional antes de poder salir.

Educación y Semestre de Nacimiento

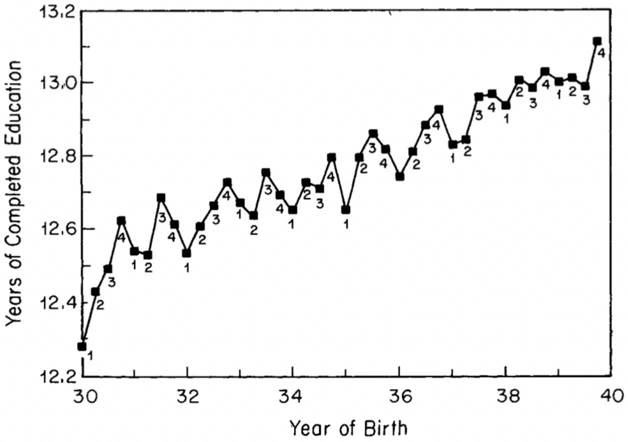


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

Educación y Semestre de Nacimiento

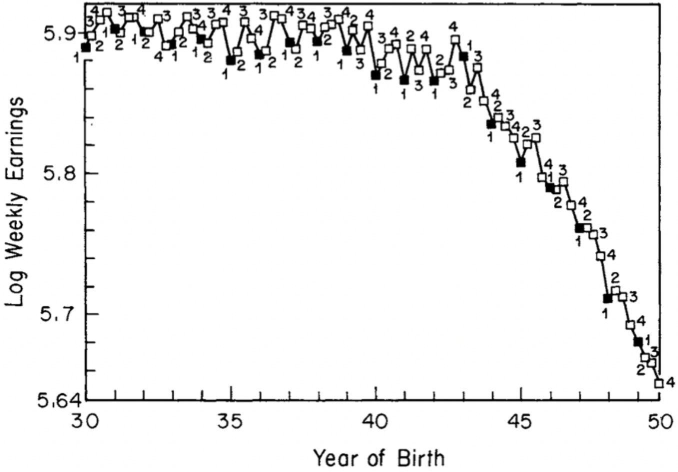


FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

Educación y Semestre de Nacimiento

Entonces, el estimador de Mínimos Cuadrados en Dos Etapas consiste en una primera etapa que estima:

$$E_i = \alpha_0 + \alpha_1 T_{1i} + \alpha_2 T_{2i} + \alpha_3 T_{3i} + \varepsilon_i$$

y una segunda etapa que estima:

$$W_i = \beta_0 + \beta_1 \hat{E}_i + u_i$$

Educación y Semestre de Nacimiento

En la práctica, el artículo estima algo muy parecido:

$$E_i = \pi C_i + \sum_c \delta_c Y_c + \sum_c (\theta_{1c} T_{1i} + \theta_{2c} T_{2i} + \theta_{3c} T_{3i}) + \epsilon_i$$

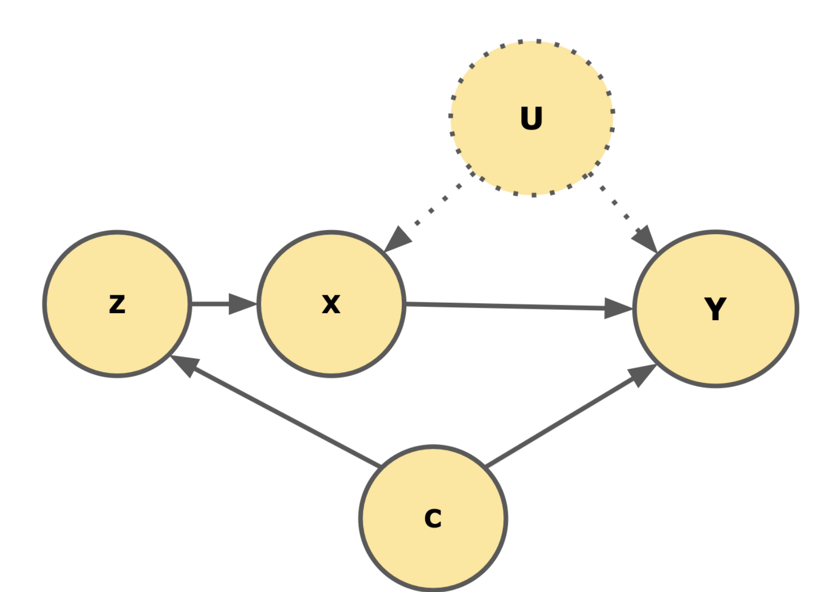
$$\ln W_i = \beta C_i + \sum_c \xi_c Y_c + \sum_c (\theta_{1c} T_{1i} + \theta_{2c} T_{2i} + \theta_{3c} T_{3i}) + \epsilon_i$$

Educación y Semestre de Nacimiento

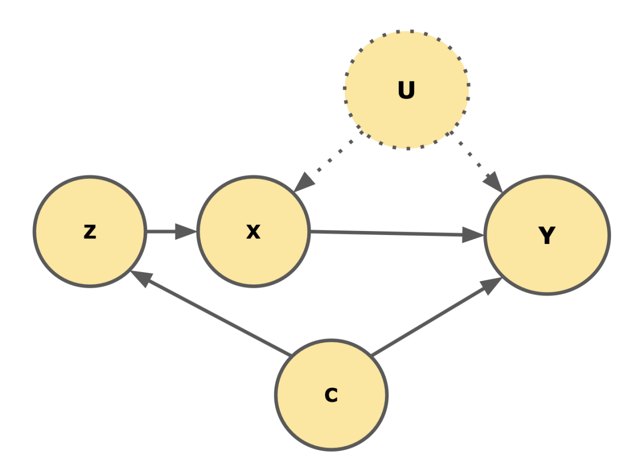
¿Por qué incluyen C_i?

1. Igual que en MCO, incluir covariables importantes para explicar Y ayuda a reducir los errores.
2. En ciertos casos podemos creer que el instrumento es válido *condicional en* C.

Educación y Semestre de Nacimiento



Educación y Semestre de Nacimiento



Z es una variable instrumental válida **si** condicionamos en C.
Es decir, es necesario incluir C como control.

Educación y Semestre de Nacimiento

Volvamos a las hipótesis básicas:

1. **Exogeneidad**

¿Qué significa en este caso?

1. **Exogeneidad**

Significa que el trimestre de nacimiento no puede verse afectado por variables que también afecten educación y salarios.

Por ejemplo, si profesionales con carreras muy demandantes programan su fertilidad para tener hijos durante las vacaciones, habría una asociación espuria.

Educación y Semestre de Nacimiento

2. **Exclusión**

¿Qué significa en este caso?

Educación y Semestre de Nacimiento

2. **Exclusión**

El trimestre de nacimiento puede afectar los salarios solo a través de la educación.

Se violaría si, por ejemplo, los niños nacidos en invierno tuvieran peor salud.

3. **Monotonicidad**

Aquí se vuelve más complejo.

Significa que un nacimiento en el primer trimestre no puede provocar que alguien obtenga *más* educación que si hubiera nacido en otro trimestre.

Educación y Semestre de Nacimiento

¿Quiénes son los cumplidores? ¿Cómo interpretamos el LATE?

Educación y Semestre de Nacimiento

Los cumplidores son las personas que tuvieron más educación porque nacieron más hacia el final del año (o viceversa).

Son quienes considerarían dejar la escuela a los 16 años, pero esperan un poco más para terminar un año adicional.

Es decir, probablemente no son personas que alcanzarían niveles de educación muy altos. Las estimaciones valen para un grupo en riesgo de no terminar la educación secundaria.

Educación y Semestre de Nacimiento

Este es un ejemplo de lo que llamamos "experimentos naturales".

No es una intervención diseñada para estudiar un efecto, sino una variación creada naturalmente por la política educativa.

Pero genera una variación "casi aleatoria" en años de educación, lo que permite estudiar sus efectos.

Educación y Semestre de Nacimiento

Veremos enseguida otro caso de un instrumento que ocurre naturalmente.

Oferta Laboral y el Tamaño Familiar

Oferta Laboral y el Tamaño Familiar

¿Cómo influye el número de hijos en el trabajo de las madres?

Esta es una pregunta importante para entender la oferta laboral femenina. También es clave para explicar el aumento del trabajo femenino conforme disminuyó el tamaño de las familias.

Oferta Laboral y el Tamaño Familiar

Si simplemente comparamos el empleo de mujeres con 1 hijo frente a mujeres con 3 hijos, ¿qué otros factores pueden diferir?

Oferta Laboral y el Tamaño Familiar

Al comparar mujeres con 1 hijo y con 3 hijos podrían diferir:

- Probablemente iniciaron la maternidad más jóvenes $\rightarrow$ menos educación.
- Las mujeres con más hijos quizá tengan menos interés en un empleo formal y se dediquen más a la familia.

Oferta Laboral y el Tamaño Familiar

Para usar una estrategia de variables instrumentales, necesitamos algo “aleatorio” que haga que una familia tenga más hijos. ¿Qué podría ser?

Oferta Laboral y el Tamaño Familiar

Angrist y Evans (1998) sugieren: que los dos primeros hijos de una familia tengan el mismo sexo.

Oferta Laboral y el Tamaño Familiar

Es muy común que los padres prefieran tener al menos un hijo y al menos una hija.

Muchas familias que tienen 2 hijos del mismo sexo eligen tener uno más para intentar obtener ambos sexos.

Como el sexo del segundo hijo es aleatorio, eso genera un "incentivo" para tener más hijos en algunas familias, pero no en otras.

Oferta Laboral y el Tamaño Familiar

Sex of first two children in families with two or more children	1980 PUMS (394,835 observations) Fraction of sample	Fraction that had another child
one boy, one girl	0.494	0.372 (0.001)
two girls	0.242	0.441 (0.002)
two boys	0.264	0.423 (0.002)
(1) one boy, one girl	0.494	0.372 (0.001)
(2) both same sex	0.506	0.432 (0.001)
difference (2) − (1)	—	0.060 (0.002)

Oferta Laboral y el Tamaño Familiar

Las familias que tuvieron un hijo y una hija tienen 37.2 % de probabilidad de tener otro hijo.

Las familias con hijos del mismo sexo tienen 43.2 % de probabilidad de tener otro hijo.

Entonces, este “instrumento” aumenta la fertilidad en 6 puntos porcentuales (o más).

Oferta Laboral y el Tamaño Familiar

Por lo tanto, para estimar el efecto del tamaño familiar sobre la oferta laboral, estimamos:

Primera etapa, donde x_i es “tener 3 hijos o más”:

$$x_i = \pi_0 w_i + \pi_1 s_{1i} + \pi_2 s_{2i} + \gamma \text{MismoSexo}_i + \eta_i$$

Segunda etapa:

$$y_i = \alpha_0 w_i + \alpha_1 s_{1i} + \alpha_2 s_{2i} + \beta \hat{x}_i + \varepsilon_i$$

TABLE 7—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1980 CENSUS DATA

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*
Dependent variable:									
Worked for pay	−0.176	−0.120	−0.113	−0.167	−0.120	−0.113	−0.008	0.004	0.001
	(0.002)	(0.025)	(0.025)	(0.002)	(0.028)	(0.028)	(0.001)	(0.009)	(0.008)
			[0.013]			[0.013]			[0.013]
Weeks worked	−8.97	−5.66	−5.37	−8.05	−5.40	−5.16	−0.82	0.59	0.45
	(0.07)	(1.11)	(1.10)	(0.09)	(1.20)	(1.20)	(0.04)	(0.60)	(0.59)
			[0.017]			[0.071]			[0.030]
Hours/week	−6.66	−4.59	−4.37	−6.02	−4.83	−4.61	0.25	0.56	0.50
	(0.06)	(0.95)	(0.94)	(0.08)	(1.02)	(1.01)	(0.05)	(0.70)	(0.69)
			[0.030]			[0.049]			[0.71]
Labor income	−3768.2	−1960.5	−1870.4	−3165.7	−1344.8	−1321.2	−1505.5	−1248.1	−1382.3
	(35.4)	(541.5)	(538.5)	(42.0)	(569.2)	(565.9)	(103.5)	(1397.8)	(1388.9)
			[0.126]			[0.703]			(0.549)

Otra vez, repasemos las hipótesis básicas:

1. **Exogeneidad**

1. **Exogeneidad**

Se basa en el hecho biológico de que, en promedio, el 50 % de los nacimientos son de cada sexo.

¿Qué la podría violar? Por ejemplo, si existe una alta tasa de abortos selectivos (o infanticidio) según el sexo, las familias podrían sesgar el resultado, asociando la capacidad de conocer el sexo del bebé y el instrumento.

2. **Exclusión**

¿Qué sería?

Oferta Laboral y el Tamaño Familiar

2. **Exclusión**

Tener dos hijos del mismo sexo puede afectar el empleo **solo** a través de la decisión de tener más hijos.

¿Cómo se podría violar? Por ejemplo, si el hecho de tener dos hijos varones cambiara las expectativas de ingreso futuro de la madre independientemente del número total de hijos.

3. **Monotonicidad**

¿Qué sería?

3. **Monotonicidad**

Nadie puede decidir tener **otro** hijo **porque** ya tiene uno de cada sexo. Nadie puede decidir **no tener más** hijos **porque** tiene dos hijos del mismo sexo.

Oferta Laboral y el Tamaño Familiar

Interpretación del LATE:

1. Supone que la familia tiene cierto control sobre la fertilidad; una familia en un contexto sin planificación familiar no está incluida en la población de interés.
2. Más importante: solo vale para familias con 2 hijos. Tiene gran valor histórico, pero quizá sea menos aplicable hoy.

Oferta Laboral y el Tamaño Familiar

Aquí vemos otro ejemplo de "experimento natural" que utiliza la variación que ocurre naturalmente para aislar un efecto causal.

Como en toda aplicación de IV, el conocimiento del contexto es crucial para evaluar si las hipótesis son válidas.

Diseño del juez

Una cuestión central en la economía del crimen es el efecto causal del encarcelamiento sobre los resultados de las personas.

- ¿Reduce la reincidencia un encarcelamiento (más largo)?
- ¿Cómo afecta un antecedente penal a las oportunidades laborales?
- ¿Cuáles son los efectos para una familia cuando el padre es encarcelado?

Diseño del juez

Claramente, no es adecuado comparar los resultados de la población encarcelada con los de los no criminales.

Supongamos que regresamos la probabilidad de empleo sobre tener antecedentes penales. ¿Qué sesgo obtendríamos?

Diseño del juez

El enfoque de variables instrumentales para este problema consiste en encontrar una variable que afecte el encarcelamiento "exógenamente". Es decir, que no esté directamente relacionada con el comportamiento de la persona.

El instrumento propuesto es una medida de cuán estricto o indulgente es el juez asignado a cada caso.

- Las personas asignadas a jueces más duros tendrán más probabilidad de ir a prisión o de recibir sentencias más severas.
- Quienes sean asignados a un juez más indulgente recibirán sentencias más leves.

Diseño del juez

Una forma de implementar esta estrategia es la siguiente:

- La primera etapa mide cómo cambia la probabilidad de condena según el juez:

$$prison_i = \alpha_0 + \alpha_1 Judge\_1_i + \alpha_2 Judge\_2_i + ... + \alpha_n Judge\_n_i + \varepsilon_i$$

(se omite un juez como referencia)

- En la segunda etapa estimamos la ecuación de interés:

$$Employment_i = \beta_0 + \beta_1 prison_i + u_i$$

Diseño del juez

De manera alternativa, podemos estimar directamente qué tan "duro" es cada juez, en función de la frecuencia con que condena.

- La primera etapa es:

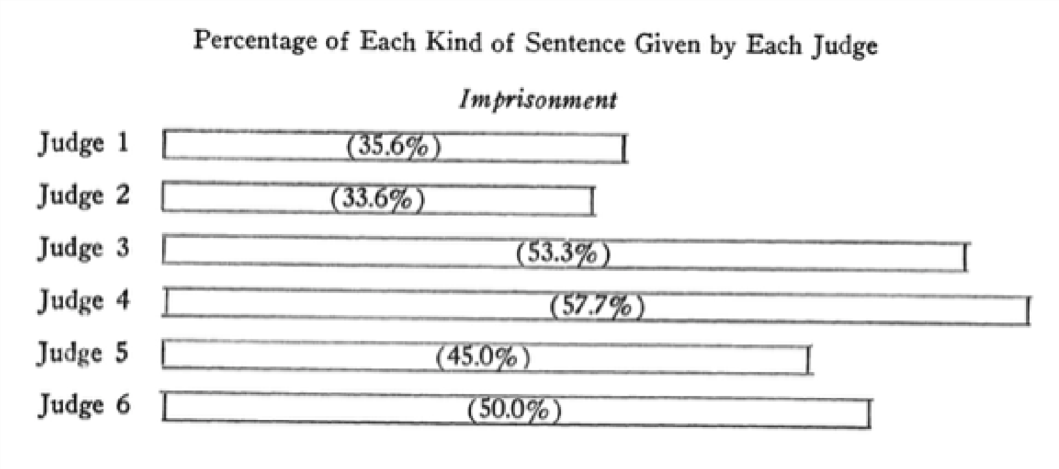
$$prison_i = \alpha_0 + \alpha_1 P(conviction|Judge)_i + \varepsilon_i$$

donde $P(conviction|Judge)_i$ es simplemente la tasa media de condena del juez asignado al individuo i, estimada en *otros* casos.

- La segunda etapa es la misma:

$$Employment_i = \beta_0 + \beta_1 prison_i + u_i$$

Diseño del juez



Diseño del juez

Es importante que (al menos en algunos contextos) los casos se asignen a los jueces aleatoriamente o por simple orden de llegada.

Si el sistema judicial pudiera, por ejemplo, asignar a los peores infractores a los jueces más duros, ¿qué supuesto se violaría?

Diseño del juez

¿Y qué hay del supuesto de exclusión? ¿Podría violarse en este caso?

Diseño del juez

En este diseño, la hipótesis de monotonicidad puede ser problemática, porque necesitamos que cada juez sea más duro o más blando para *todos* los casos.

Por ejemplo, supongamos que hay un juez muy indulgente, salvo en casos de delitos violentos. O en casos que involucren drogas. O cuando el acusado es afroamericano.

Cada uno de estos ejemplos es una violación, pues *algunas* personas tendrían menor probabilidad de encarcelamiento mientras que *otras* tendrían mayor probabilidad.

Diseño del juez

¿Qué se está estimando?

Se acuerden de la interpretación del LATE: el efecto para los que son afectados por el instrumento.

En ese caso, los casos más graves puden resultar en condenación con cualquier juez, y otros van a ser liberados en todo caso. Eses no hacen parte de los “compliers”.

El efecto estimado es sobre los casos que podrían ser condenados por ciertos jueces y liberados por otros.

Diseño del juez

Este diseño puede adaptarse siempre que haya distintas personas que deciden resultados para una población determinada y la asignación sea aleatoria.

Por ejemplo: una plataforma de comercio electrónico quiere aprender sobre su programa de reembolsos. Cuando un cliente tiene un problema con una entrega, puede recibir reembolsos más o menos generosos. ¿Cómo afecta eso su probabilidad de volver a comprar en la misma tienda?

Se puede usar la misma idea, basándose en los profesionales del call-center que atienden el proceso.

Diseño del juez

Por último, analicemos una variable que a menudo se usa como instrumento, pero que hoy no es muy popular.

Lluvia

Lluvia

La lluvia es muy importante para la agricultura, que es una actividad económica fundamental.

Además, está determinada mayormente por factores físicos con poca conexión con la actividad económica, de modo que podría ser un candidato interesante para instrumento.

¿Pero realmente lo es?

Lluvia

Revisemos dos aplicaciones de la lluvia como instrumento y pensemos si los supuestos tienen sentido.

- Ingresos y trabajo infantil
- Participación electoral

Lluvia, ingresos y trabajo infantil

Una investigadora quiere estudiar cómo varía la prevalencia de trabajo infantil con el ingreso del hogar.

Para identificar el efecto causal, necesita un instrumento para el ingreso del hogar.

Propone usar la lluvia: en un contexto agrícola y pobre, una mejor temporada de lluvias tiene un fuerte efecto sobre la cosecha y, por ende, sobre el ingreso del hogar. La lluvia varía aleatoriamente de un año a otro, por lo que la exogeneidad parece válida.

Lluvia, ingresos y trabajo infantil

La investigadora propone estimar:

$$income_{it} = \alpha_0 + \alpha_1 rainfall_{it} + \varepsilon_{it}$$

$$child\_labor_{it} = \beta_0 + \beta_1 income_{it} + u_{it}$$

¿Qué opinas?

Lluvia, ingresos y trabajo infantil

En este caso, la exogeneidad quizá se cumpla, pero la **exclusión** probablemente no. Precisamente porque la lluvia afecta muchas cosas:

- Más lluvia puede aumentar la *demanda laboral*, incrementando directamente el trabajo infantil,
- Puede afectar la *prevalencia de enfermedades*,
- Las lluvias fuertes pueden destruir infraestructura, incluidas escuelas, reduciendo las opciones no laborales.

Lluvia, ingresos y trabajo infantil

La monotonicidad también puede ser problemática, porque en algunos contextos la lluvia **excesiva** es perjudicial para los cultivos. También puede causar inundaciones y destrucción de infraestructura.

Así, para ciertos niveles la lluvia mejora el ingreso, pero para otros lo perjudica.

Para entenderlo, habría que conocer las características agronómicas del cultivo y del suelo.

Lluvia, ingresos y trabajo infantil

En general, para preguntas relacionadas con capital humano o ingresos, la lluvia puede tener demasiadas vías causales para ser un instrumento limpio.

En algunos casos, podemos intentar cerrar algunos de esos canales, ¡pero es difícil!

Lluvia, participación electoral y resultados

Una variable importante en política electoral es la participación de los votantes.

Las políticas que facilitan el voto pueden beneficiar al partido más asociado con las clases bajas (usualmente la izquierda).

Pero en algunos casos, los votantes menos comprometidos pueden tener opiniones más de derecha sobre inmigración, etc.

Lluvia, participación electoral y resultados

Para estudiar esta cuestión, quisiéramos aumentar o disminuir la participación electoral de forma aleatoria.

Una opción es usar la lluvia *el día de la elección*. La idea es que la lluvia puede interrumpir el transporte o simplemente aumentar el costo de ir a votar (¡es desagradable hacer fila bajo la lluvia!).

Si tenemos datos de participación para lugares con y sin lluvia el mismo día, podemos compararlos.

Lluvia, participación electoral y resultados

Debemos tener cuidado de controlar el nivel *baseline* de lluvia.

En cualquier elección, es mucho menos probable que haya llovido en Arica que en Puerto Aysén.

Pero quizá Arica y Puerto Aysén también tengan preferencias políticas promedio distintas.

Una mejor comparación serían dos comunas cercanas a Concepción, donde una tuvo lluvia y la otra no, pero ambas tienen la misma lluvia promedio anual.

Lluvia, participación electoral y resultados

Un posible diseño sería:

$$turn\_out_{ct} = \alpha_0 + \alpha_1 rainfall\_election\_day_{ct} + \alpha_2 avg\_rainfall_{ct} + \alpha_3 region_c + \varepsilon_{ct}$$

$$share\_left_{ct} = \beta_0 + \beta_1 turnout_{ct} + \beta_2 avg\_rainfall_{ct} + \beta_3 region_c + u_{ct}$$

Lluvia, participación electoral y resultados

En este diseño, la restricción de exclusión es mucho más creíble.

Como la variación proviene de la lluvia en un único día, y no a lo largo de toda la temporada, hay menos margen para que la lluvia afecte otras cosas.

Puesto que podemos usar simplemente una variable binaria de si llovió o no, la monotonicidad es más fácil de cumplir.

Repaso de IV

El método de variables instrumentales se basea en encontrar una variable (Z) que afecta la variable endogena (X), pero es exógena.

Estimación:

- Primera etapa: regresion de X en Z
- Segunda etapa: regresion de Y en $\hat{X}$

Hipótesis:

- Exogeneidad
- Exclusión
- Monotonicidad

Interpretación de LATE: Efecto promedio para los que son afectados por la primera etapa.