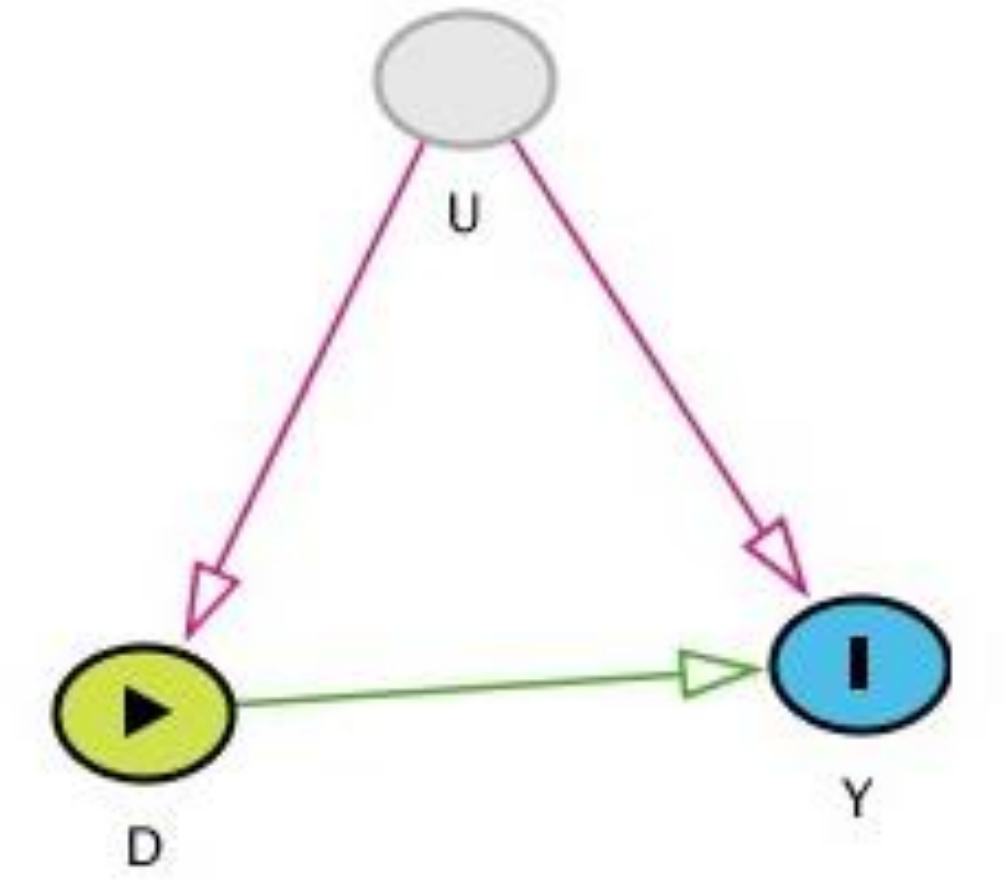
Evaluación de Impacto: IV

Francesco Bogliacino

Tabla de contenido

Selección sobre no observable

- Es el caso donde el tratamiento está atado a no observables
- Para este caso no podemos usar ninguno de los diseños vistos hasta el momento
- Hay un backdoor path D<-U->Y, controlar por D no lo cierra

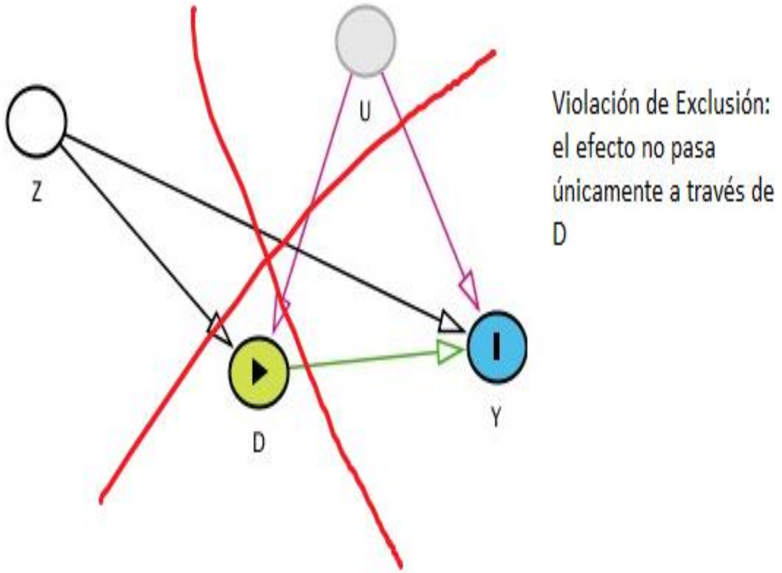


IV

- El diseño de investigación basado en variable instrumental aborda los siguientes problemas:
  - Simultaneidad
  - Causalidad invertida
  - Variable omitida
  - Error de medición
  - RCT con cumplimiento imperfecto

IV

- **El IV es raro**



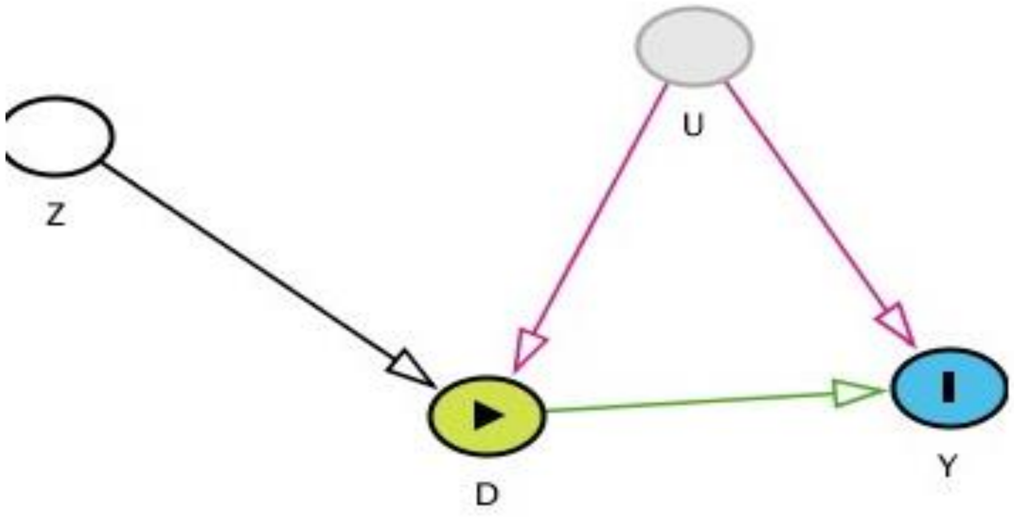
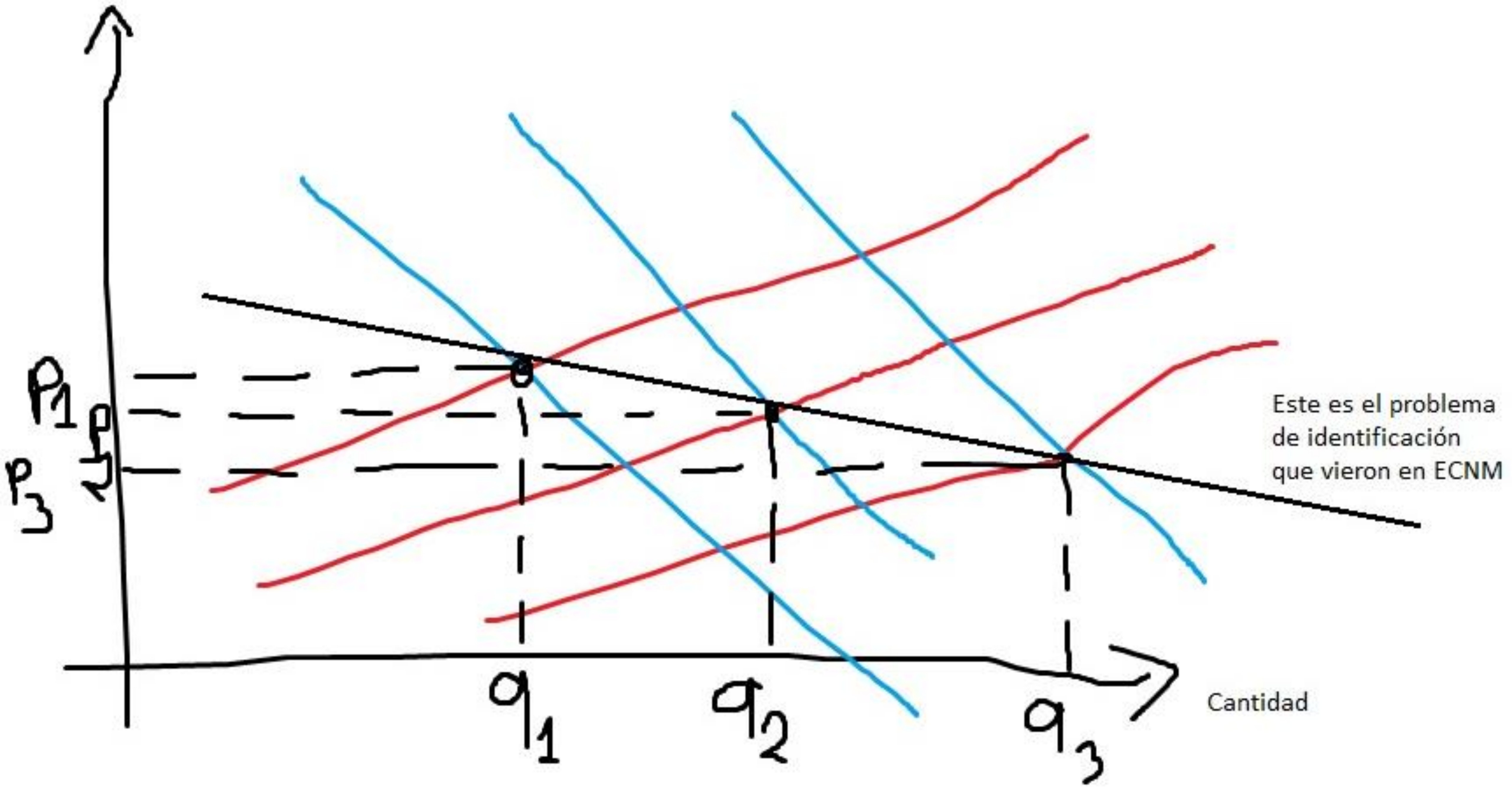


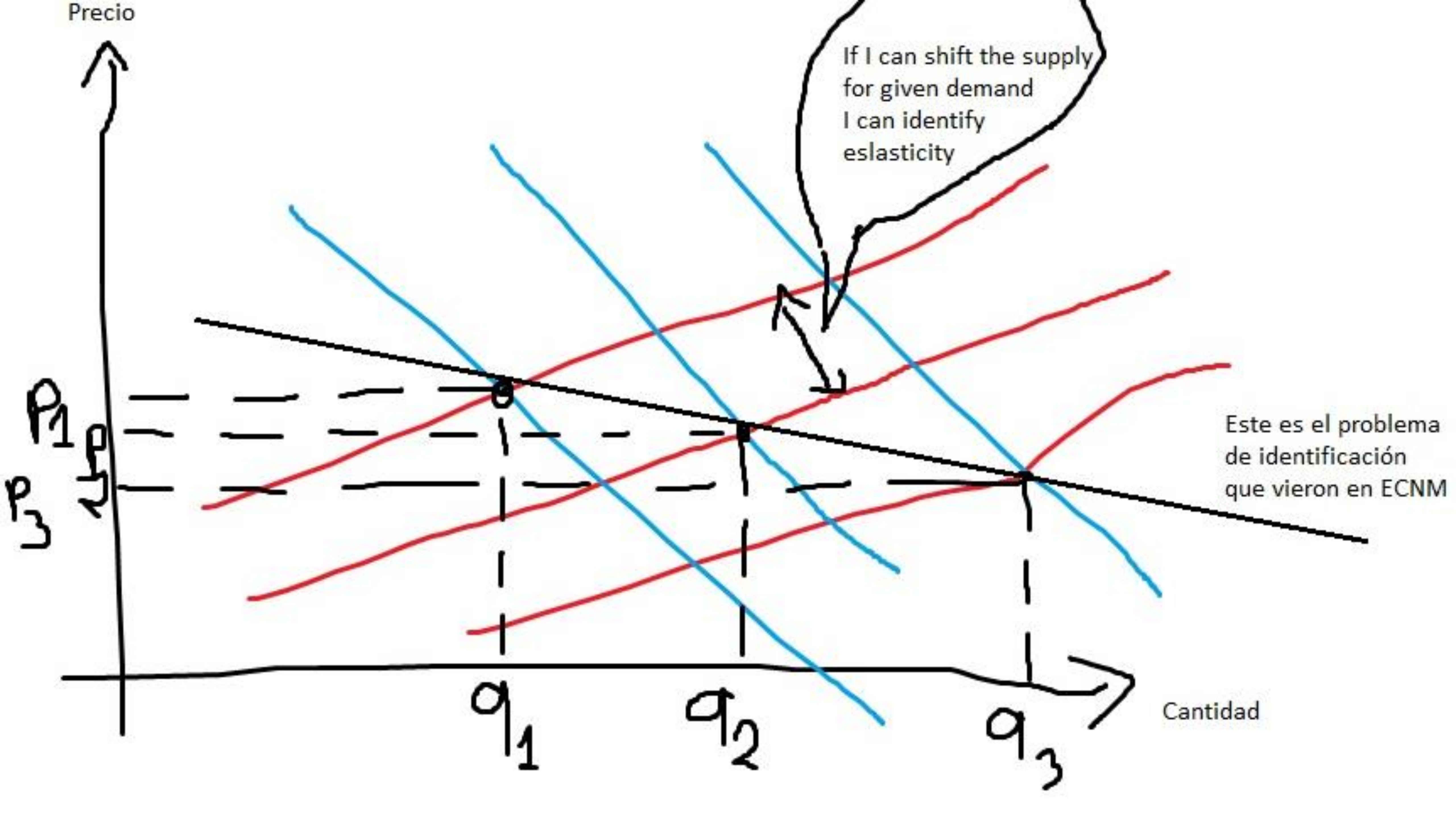
Tabla de contenido

Efecto homogéneo

- Efecto tratamiento constante
- Es equivalente al caso ATE=ATT=ATU
- Veremos el caso de heterogeneidad después [LATE]
- En general hablamos de un efecto β, que sería igual al ATE para todos

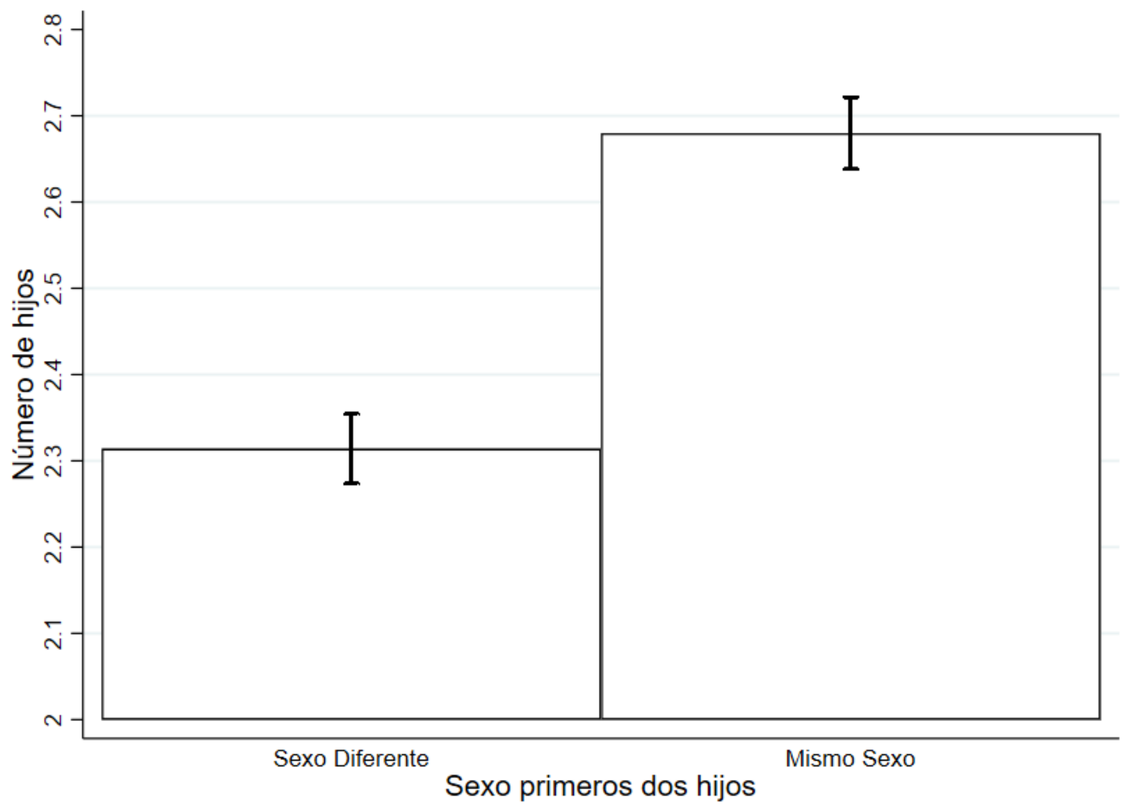


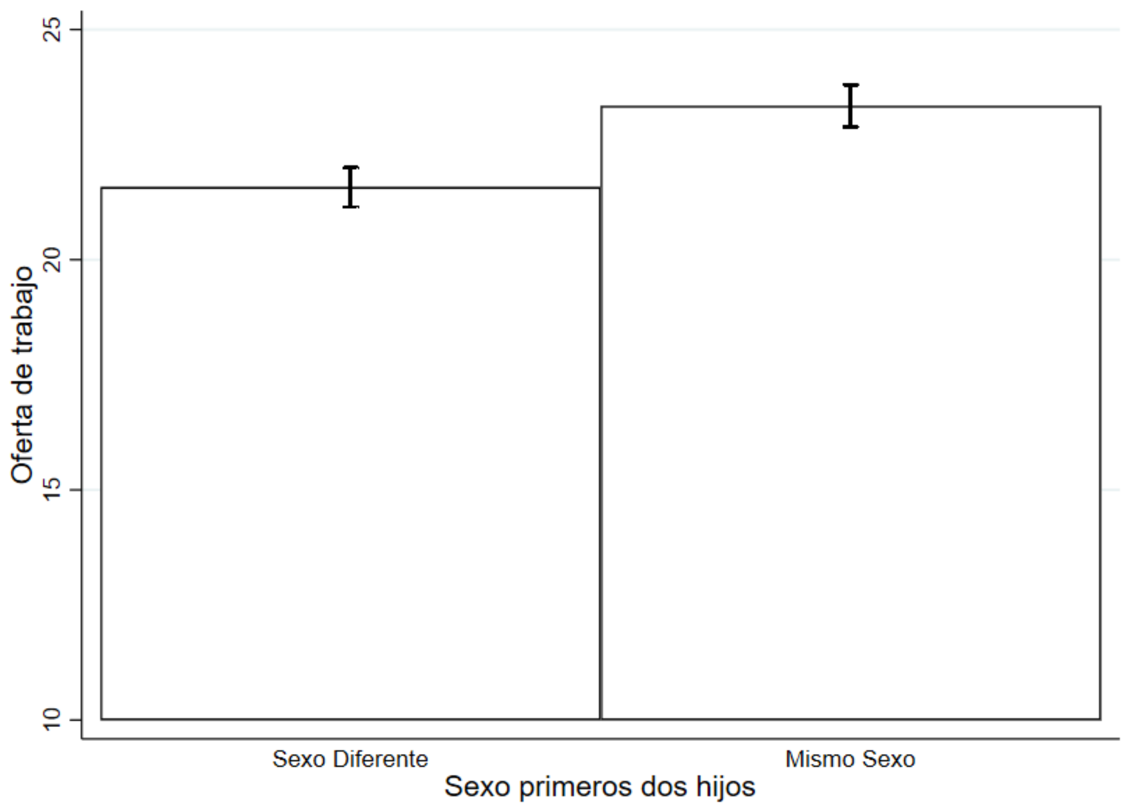
Este es el problema de identificación que vieron en ECNM
Cantidad



Precio
If I can shift the supply
for given demand
I can identify
eslasticity
p_1
p_2
p_3
Este es el problema
de identificación
que vieron en ECNM
q_1
q_2
q_3
Cantidad

Fertilidad y oferta laboral



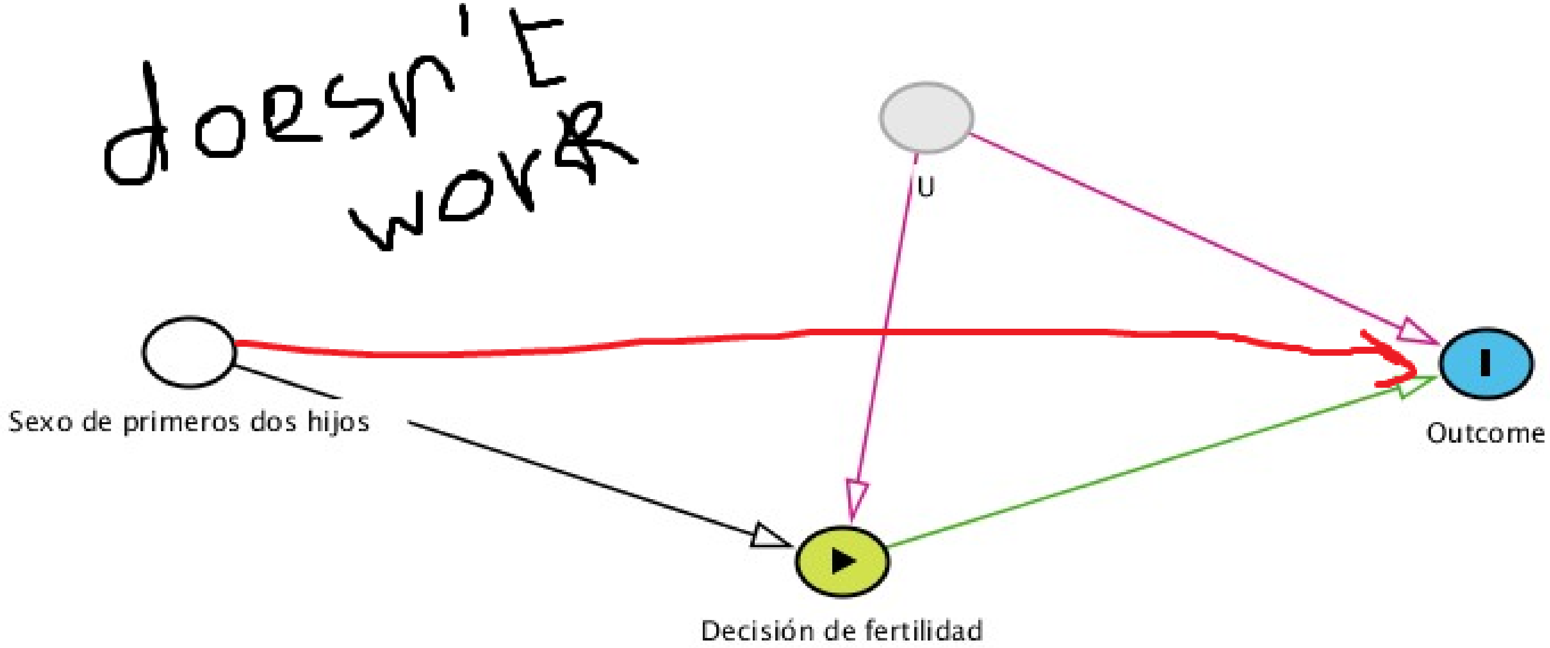


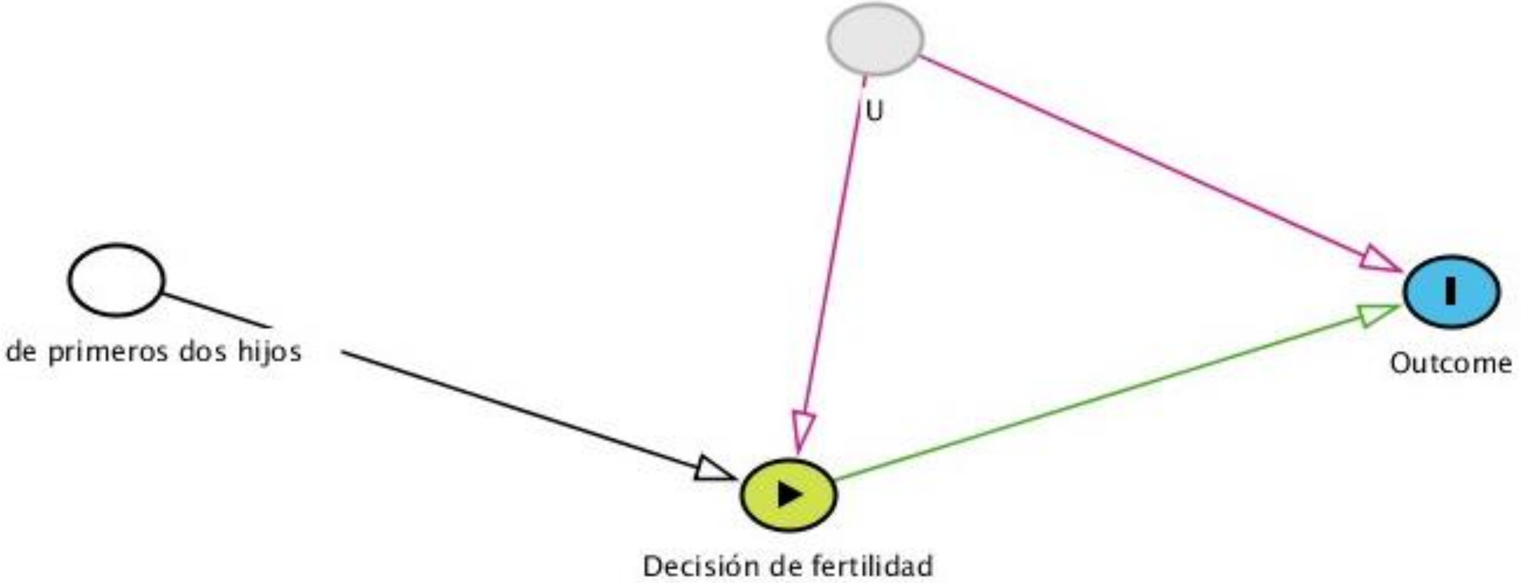
Diseño de investigación por IV

- Tiene que haber un elemento aleatorio
- Tiene que respetar la restricción de exclusión (esa es la parte *rara*)
- Don’t go from the data to the instrument, that’s crap

IV

- Sirve elemento de aleatoriedad
- Hay que interrogarse si la restricción de exclusion es plausible
    - ¿Es posible que Z afecte Y directamente?
    - ¿Es posible que exista otra variable a través de la cual Z afecte Y, que no sea D?





Diseño por IV

$$Y_i = \alpha + \beta S_i + \gamma A_i + \varepsilon_i$$

- A no se observa y es correlacionada con S [clásico caso de variable omitida];
- Existe Z:
  - Z tiene correlación con S
  - Z es “tan buena como aleatoriamente asignada”

Diseño por IV

$$\mathrm{Cov}(Y_i, Z_i) = \mathrm{Cov}(\alpha + \beta S_i + \gamma A_i + \varepsilon_i, Z_i) =$$
$$= \mathrm{Cov}(\alpha, Z_i) + \beta \mathrm{Cov}(S_i, Z_i) + \gamma \mathrm{Cov}(A_i, Z_i) + \mathrm{Cov}(\varepsilon_i, Z_i) =$$
$$= \beta \mathrm{Cov}(S_i, Z_i) + \gamma \mathrm{Cov}(A_i, Z_i) + \mathrm{Cov}(\varepsilon_i, Z_i)$$

- Si $\mathrm{Cov}(S_i, Z_i) \neq 0$
- Si $\mathrm{Cov}(S_i, Z_i) \perp A_i, \varepsilon_i$

Dividiendo por $\mathrm{Cov}(S_i, Z_i)$ obtengo β

Supuestos de identificación

- Tiene que existir el "First Stage"
- Tiene que cumplirse la restricción de exclusión

Si se cumplen [la segunda como siempre es no testeable, tiene que ser plausible], puedo estimar $\mathrm{Cov}(S_i, Z_i)$ y $\mathrm{Cov}(Y_i, Z_i)$ [forma reducida y first stage] y estimar el parámetro de interés

NB TSLS, LIML, Wald, Two Sample IV, son todos estimadores; IV es el diseño de investigación

Supuestos de identificación

$$\text{plim}\widehat{\beta^{IV}} = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(S_i, Z_i)} = \beta + \gamma \frac{\text{Cov}(A_i, Z_i)}{\text{Cov}(S_i, Z_i)}$$

Si $\text{Cov}(A_i, Z_i)$ no es cero, o $\text{Cov}(S_i, Z_i)$ es cero, entonces ese último término no se elimina

Tabla de contenido

TWO SAMPLE IV

Dado que $\widehat{\beta^{IV}} = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(S_i, Z_i)}$, puedo sacar dos muestras para calcularme

- $\text{Cov}(Y_i, Z_i)$
- $\text{Cov}(S_i, Z_i)$

Con heterogeneidad, esto puede ser muy dañino, porque no se que esto estimando

TSLS

- TSLS es casi identificado con IV (pero no se confundan), de cierta manera es importante aprenderse el lenguaje porque quedó grabado en la jerga de IV:
  - Modelo estructural $Y_i = \alpha + \beta S_i + \varepsilon_i$
  - First Stage $S_i = \gamma + \delta Z_i + \mu_i$
  - Reduced Form $Y_i = \theta + \rho Z_i + \omega_i$

- El instrumento se puede ver como una lotería en un RCT sin total cumplimiento. Si es ese el caso, el instrumento es la intención a tratar *(intention to treat)*
  - La Reduced Form $Y_i = \theta + \rho Z_i + \omega_i$ se puede interpretar como el modelo con la intención a tratar
  - No nos obliga a defender la restricción de exclusión
  - Pero es menos limpio desde el punto de vista de la interpretación y cuantificación del efecto

TSLS

$$Y_i = \alpha + \beta S_i + \varepsilon_i$$
$$S_i = \gamma + \delta Z_i + \mu_i$$

Necesitamos:

- $\delta \neq 0$
- $\mathrm{Cov}(Z_i, \varepsilon_i) = 0$

$$\widehat{\beta^{2SLS}} = \frac{\mathrm{Cov}(Z_i, Y_i)}{\mathrm{Cov}(Z_i, S_i)} = \frac{\frac{1}{N}\sum (Z_i - \bar{Z}) Y_i}{\frac{1}{N}\sum (Z_i - \bar{Z}) S_i} =$$

TSLS

$$\widehat{\beta^{2SLS}} = \frac{\frac{1}{N}\sum(Z_i - \bar{Z})(\alpha + \beta S_i + \varepsilon_i)}{\frac{1}{N}\sum(Z_i - \bar{Z})S_i} = $$

$$= \beta + \frac{\frac{1}{N}\sum(Z_i - \bar{Z})\varepsilon_i}{\frac{1}{N}\sum(Z_i - \bar{Z})S_i} = $$

Beta más algo que converge a cero

Cov(Z, S) beta, dividido
por Cov(Z, S) es beta

costante: Cov con Z es cero

$$\widehat{\beta^{2SLS}} = \frac{\frac{1}{N}\sum(Z_i - \bar{Z})(\alpha + \beta S_i + \varepsilon_i)}{\frac{1}{N}\sum(Z_i - \bar{Z})S_i} = $$

TSLS

- $\widehat{\beta^{2SLS}} = \dfrac{\text{Cov}(Z_i,Y_i)}{\text{Cov}(Z_i,S_i)} = \dfrac{\frac{\text{Cov}(Z_i,Y_i)}{\text{Var}(Z_i)}}{\frac{\text{Cov}(Z_i,S_i)}{\text{Var}(Z_i)}} = \dfrac{\hat{\rho}}{\hat{\delta}}$

$$\text{Cov}(z,S) = \hat{\delta}\,\text{Var}(Z)$$

$$\hat{\beta}^{2SLS} = \frac{\text{Cov}(Z_i,Y_i)\hat{\delta}}{\text{Cov}(Z_i,S_i)\hat{\delta}} = \frac{\text{Cov}(Z_i,Y_i)\hat{\delta}}{\hat{\delta}^2\,\text{Var}(z)} = \frac{\text{Cov}(\hat{\delta}Z,Y)}{\text{Var}(\hat{\delta}Z)}$$

TSLS

$$\widehat{S_i} = \hat{\gamma} + \hat{\delta} Z_i$$

Calculamos

$$\text{Cov}\left(\widehat{S_i}, Y_i\right) = \text{Cov}\left[\left(\hat{\gamma} + \hat{\delta} Z_i\right), Y_i\right] = \hat{\delta}\text{Cov}(S_i, Y_i)$$
$$\text{Var}\left(\widehat{S_i}\right) = \text{E}\left(\hat{\gamma} + \hat{\delta} Z_i - \hat{\gamma} - \hat{\delta}\overline{Z_i}\right)^2 = \text{Var}\left(\hat{\delta} S_i\right)$$

Entonces

$$\widehat{\beta^{2SLS}} = \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, S_i)} = \frac{\text{Cov}\left(\widehat{S_i}, Y_i\right)}{\text{Var}\left(\widehat{S_i}\right)}$$

Ojo, en la práctica no puedo “hacer las dos etapas”, porque estaría estimando mal los errores estándar e invalidando la inferencia

Además si tuviéramos “huecos” en las dos muestras haciéndola por etapa estaríamos usando datos diferentes

Intuición

- La intuición de las dos etapas es que la primera me sirve para identificar la variabilidad exógena de S, a través del instrumento Z
- Obviamente la variabilidad de S estaba ahí, el instrumento nos identifica esa parte de la variabilidad en la independiente endógena que es random
- Aquí viene el tradeoff:
  - Me reduce la variabilidad total
  - Pero la que queda es random

Wald

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$$
$$= E[\alpha + \beta D_i + \varepsilon_i|Z_i = 1] - E[\alpha + \beta D_i + \varepsilon_i|Z_i = 0] =$$
$$= E\left[Y_i^1\middle|Z_i = 1\right]P(D_i = 1|Z_i = 1) + E\left[Y_i^0\middle|Z_i = 0\right]P(D_i = 0|Z_i = 1)$$
$$- E\left[Y_i^1\middle|Z_i = 1\right]P(D_i = 1|Z_i = 0) - E\left[Y_i^0\middle|Z_i = 0\right]P(D_i = 0|Z_i = 0) =$$
$$E[Y_i^1]\big(P(D_i = 1|Z_i = 1) - P(D_i = 1|Z_i = 0)\big) -$$
$$E[Y_i^0]\big(P(D_i = 0|Z_i = 1) - P(D_i = 0|Z_i = 0)\big) =$$
$$E[Y_i^1]\big(P(D_i = 1|Z_i = 1) - P(D_i = 1|Z_i = 0)\big) -$$
$$E[Y_i^0]\big(1 - P(D_i = 1|Z_i = 1) - 1 + P(D_i = 1|Z_i = 0)\big) =$$
$$(E\left[Y_i^1\right] - E[Y_i^0])\big(P(D_i = 1|Z_i = 1) - P(D_i = 1|Z_i = 0)\big)$$

Wald

$$\left(P(D_i = 1|Z_i = 1) - P(D_i = 1|Z_i = 0)\right)$$
$$= E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$

$$\beta^{\widehat{Wald}} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

- Nos ayuda en la interpretación
  - La forma reducida nos da un efecto que tenemos que rescalar dependiendo del first stage
  - El first stage nos da el *take up rate*, cuantos tomaron el medicamento si fueron sorteados para eso

LIML

$$\widehat{\beta^{2SLS}} = \left(\widehat{\mathrm{X}}'\widehat{\mathrm{X}}\right)^{-1}\widehat{\mathrm{X}}'\mathrm{Y}$$
$$= ((\mathrm{X}(\mathrm{Z}'\mathrm{Z})^{-1}\mathrm{Z}'\mathrm{X})'(\mathrm{X}(\mathrm{Z}'\mathrm{Z})^{-1}\mathrm{Z}'\mathrm{X}))^{-1}(\mathrm{X}'\mathrm{Z}(\mathrm{Z}'\mathrm{Z})^{-1}\mathrm{X}'\mathrm{Y})$$
$$= (\mathrm{X}'(1-\mathrm{M})\mathrm{X})^{-1}(\mathrm{X}'(1-\mathrm{M})\mathrm{Y})$$

$$\widehat{\beta^{LIML}} = (\mathrm{X}'(1-\lambda\mathrm{M})\mathrm{X})^{-1}(\mathrm{X}'(1-\lambda\mathrm{M})\mathrm{Y})$$

JIVE: Jackknife instrumental variables estimation

- Para la observación i, uso la predicción hecha estimando sobre todos los datos menos i

Tabla de contenido

Software

- En R usar –ivreg()
- En stata ivregress 2sls, pero existen mil actualizaciones

Tabla de contenido

Weak Instruments

- IV es consistente, pero no es no sesgado
- Las propiedades en muestra finita pueden ser muy malas
- Si el first stage es muy débil, en práctica no hay ventaja sobre IV

Weak Instrument

$$Y = X\beta + \varepsilon$$

El first stage es (con un conjunto de instrumentos) $\mathbf{Z}$

$$X = \mathbf{Z}'\pi + \eta$$

La matriz de proyección sobre $\mathbf{Z}$ es:

$$P_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}'$$

Weak Instrument

- Sabemos que para OLS:

$$E\left(\widehat{\beta^{OLS}} - \beta\right) = \frac{Cov(\varepsilon, X)}{Var(X)} = \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2}$$

- Sabemos que 2SLS es

$$\widehat{\beta^{2SLS}} = (X'P_Z X)^{-1} X' P_Z Y = \beta + (X'P_Z X)^{-1} X' P_Z \varepsilon$$

Puedo escribir

$$\widehat{\beta^{2SLS}} - \beta = (X'P_Z X)^{-1} X' P_Z \varepsilon$$

Ahora sabiendo que $X = \mathbf{Z}'\pi + \eta$

$$\widehat{\beta^{2SLS}} - \beta = (X'P_Z X)^{-1} \pi' \mathbf{Z}' \varepsilon + (X'P_Z X)^{-1} \eta' P_Z \varepsilon$$

Weak Instrument

- Primero el E[] no pasa a fácilmente a través de esas matrices y eso hace que esa expresión no sea obvia, ni necesariamente exista
- Segundo, el Z no está correlacionado con el error en el modelo causal, pero el valor esperado de $\eta' P_Z \varepsilon$ no será cero, porque esa el la razón pro la cual fui a buscarme un instrumento
- Angrist and Pischke muestran que aproximadamente, el sesgo es:

$$E\left[\widehat{\beta^{2SLS}} - \beta\right] \approx \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2}\left[\frac{E[(\pi'\mathbf{Z}'\mathbf{Z}\pi)/Q]}{\sigma_\eta^2} + 1\right]^{-1}$$

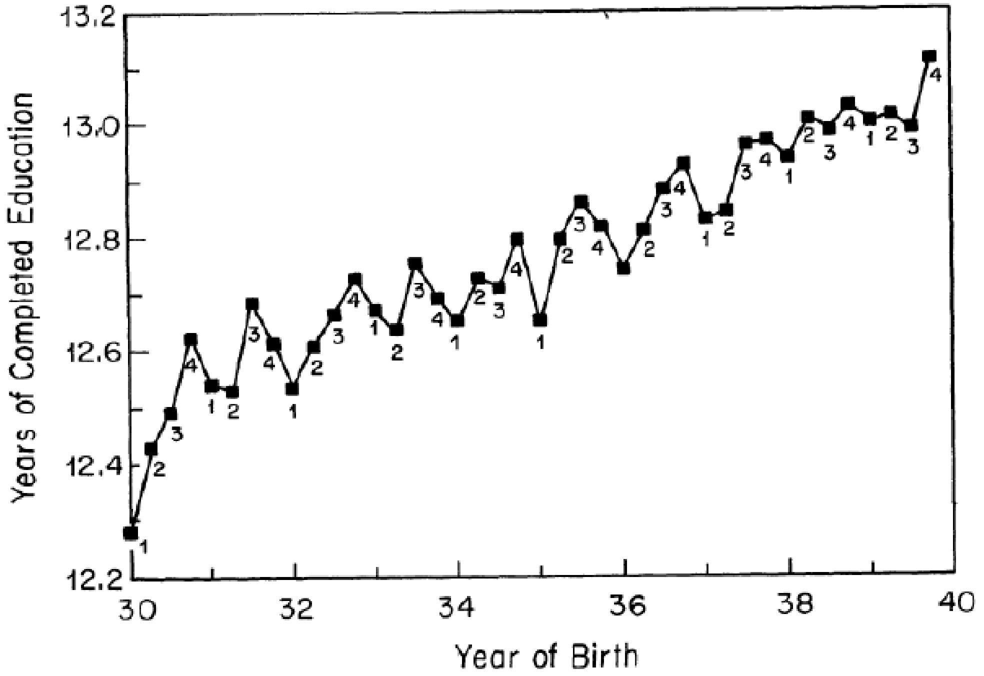
Y eso en parentesis es el F test de la primera etapa

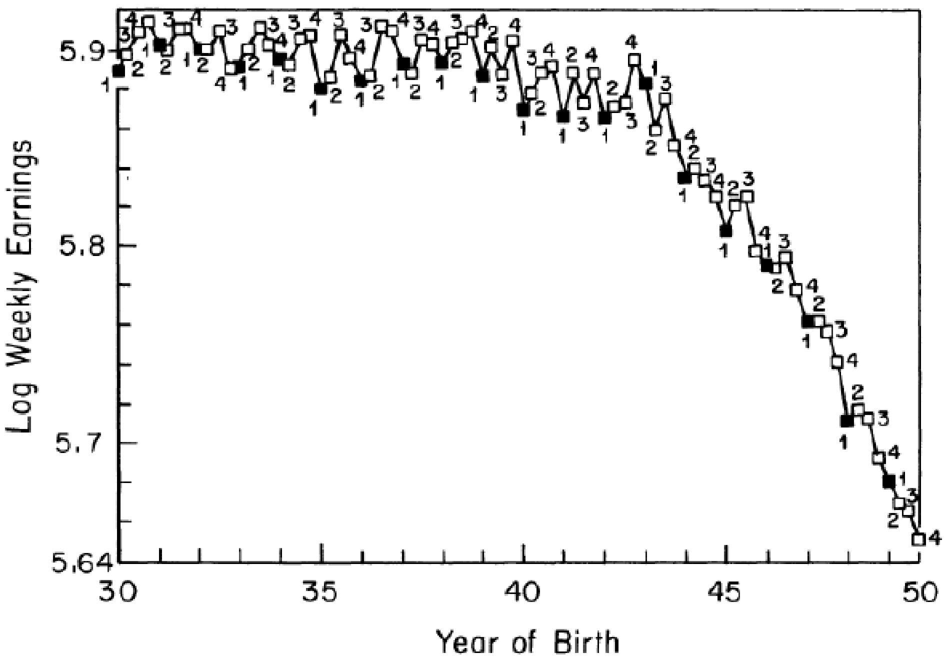
Tabla de contenido

La presentación del IV

- Siempre presentar OLS y IV
- Reportar la primera etapa y el F test
- Mostrar el IV con el instrumento más fuerte que se tiene
- Mostrar la forma reducida
- Usar LIML y JIVE
- Las tablas que incluyen OLS y IV tienes que ser bonitas
- Las gráficas de la forma reducida y de la primera etapa son clave o no van a convencer

Angrist and Krueger





LATE Theorem

- Cuando el efecto es heterogéneo, $Y_i^1 - Y_i^0 = \beta_i$, ¿qué parámetro estamos estimando?
- Volvamos a nuestro mundo de $D, Z \rightarrow \{0, 1\}$

	D=1 is Z=0	**D=0 si Z=0**
D=1 si Z=1	Always takers	Compliers
D=0 si Z=1	Defiers	Never takers

LATE Theorem

- Si vale restricción de exclusión (Y depende de Z via D únicamente), el efecto de hacer el switch del instrumento es cero sobre Always & Never takers
- Quisiéramos quitarnos los Defiers, pero si hay un costo en asumir el tratamiento, tiene sentido que la probabilidad de asumir el D dado Z sea creciente o decreciente para todos *(monotonicidad)*
- Necesitamos SUTVA [los outcome potenciales no dependen de asignaciones a otras unidades]
- First Stage $E[D_i^1 - D_i^0] \neq 0$
- Independencia: $Y_i^1, Y_i^0, D_i^1, D_i^0 \perp Z$

LATE Theorem

$$E[Y_i|Z_i=1]-E[Y_i|Z_i=0]=$$
$$=E[\alpha+\beta_i D_i+\varepsilon_i|Z_i=1]-E[\alpha+\beta_i D_i+\varepsilon_i|Z_i=0]=$$
$$=E\left[\alpha+\beta_i D_i^1\right]-E\left[\alpha+\beta_i D_i^0\right]=$$
$$=E\left[(Y_i^1-Y_i^0)(D_i^1-D_i^0)\right]=$$
$$=E\left[\left(Y_i^1-Y_i^0\right)1|D_i^1-D_i^0=1\right]P\left(D_i^1-D_i^0=1\right)$$
$$+E\left[\left(Y_i^1-Y_i^0\right)0|D_i^1-D_i^0=0\right]P\left(D_i^1-D_i^0=0\right)$$
$$+E\left[\left(Y_i^1-Y_i^0\right)(-1)|D_i^1-D_i^0=-1\right]P\left(D_i^1-D_i^0=-1\right)=$$
$$E\left[\left(Y_i^1-Y_i^0\right)1|D_i^1-D_i^0=1\right]P\left(D_i^1-D_i^0=1\right)$$

A la Wald, divido por $P\left(D_i^1-D_i^0=1\right)=E\left(D_i^1-D_i^0\right)$

LATE Theorem

exclusión

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$$
$$= E[\alpha + \beta_i D_i + \varepsilon_i|Z_i = 1] - E[\alpha + \beta_i D_i + \varepsilon_i|Z_i = 0] =$$
$$= E\left[\alpha + \beta_i D_i^1\right] - E\left[\alpha + \beta_i D_i^0\right] =$$

Independ

$$= E\left[(Y_i^1 - Y_i^0)(D_i^1 - D_i^0)\right] =$$
$$= E\left[\left(Y_i^1 - Y_i^0\right)1|D_i^1 - D_i^0 = 1\right]P\left(D_i^1 - D_i^0 = 1\right)$$

=0

$$+ E\left[\left(Y_i^1 - Y_i^0\right)0|D_i^1 - D_i^0 = 0\right]P\left(D_i^1 - D_i^0 = 0\right)$$
$$+ E\left[\left(Y_i^1 - Y_i^0\right)(-1)|D_i^1 - D_i^0 = -1\right]P\left(D_i^1 - D_i^0 = -1\right) =$$

monot

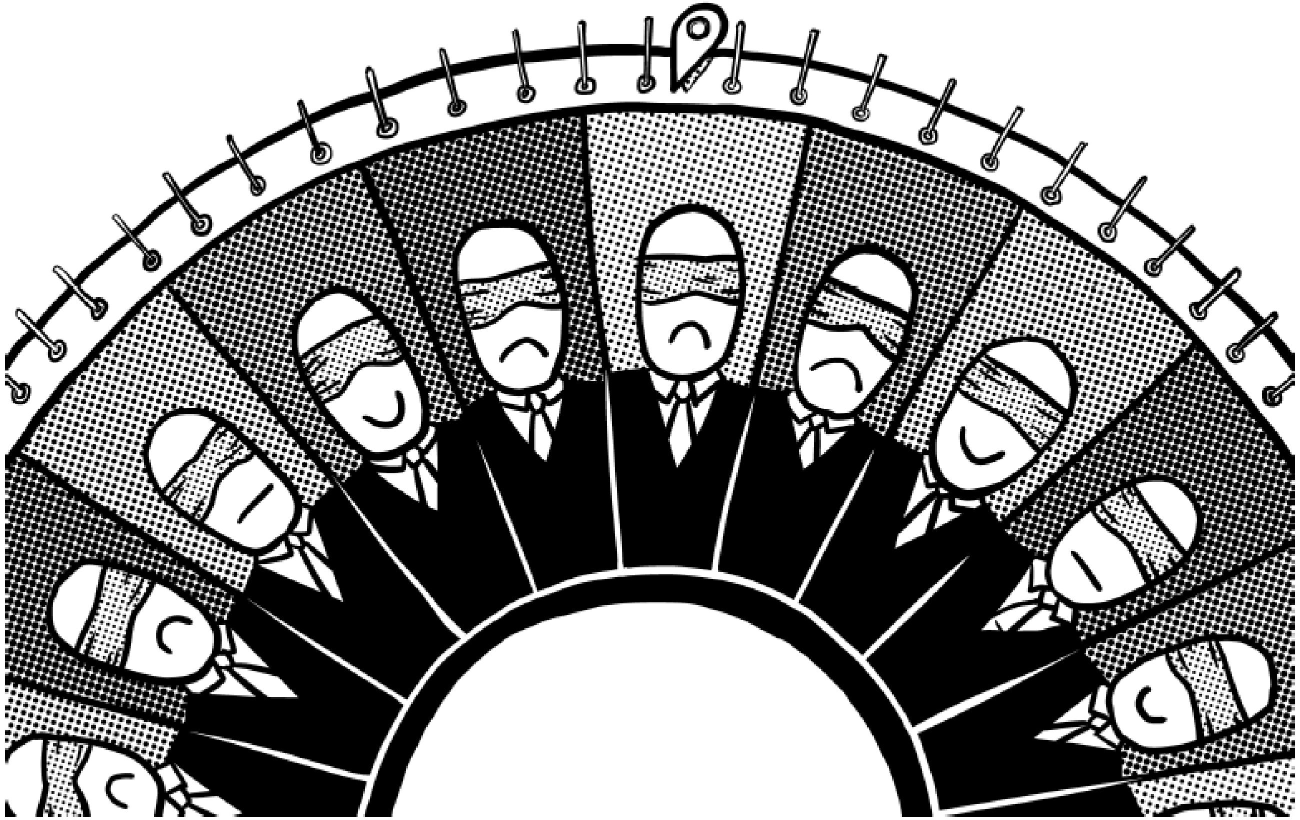
$$E\left[\left(Y_i^1 - Y_i^0\right)1|D_i^1 - D_i^0 = 1\right]P\left(D_i^1 - D_i^0 = 1\right)$$

A la Wald, divido por $P\left(D_i^1 - D_i^0 = 1\right) = \underline{E\left(D_i^1 - D_i^0\right)}$

$\rightarrow$ 1st stage

Tabla de contenido

Judge Fixed Effect



Source:
Cunningham 2018

Judge Fixed Effect

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

$$Z_{j(i)} = \frac{1}{n_{j(i)} - 1} \sum_{k \neq i}^{n_{j(i)}-1} JI_k$$

- Es casi un efecto fijo porque la propensión es difícilmente igual entre dos jueces
- Normalmente se usa JIVE

Bartik Instrument

- Asumamos que el modelo causal es

$$Y_{i,t} = \alpha + \beta X_{i,t} + \varepsilon_{i,t}$$

- X es la inmigración, Y es la demanda de trabajo, i puede ser ciudad o sector industrial
- Obviamente X es endógena
  - Los flujos de inmigración dependen de la probabilidad esperada de conseguir un trabajo que depende de $\varepsilon_{i,t}$

Bartik Instrument

$$X_{i,t} = \gamma + \delta B_{i.t} + \mu_{i,t}$$

Donde $B_{i,t}$ es el instrumento de Bartik, definido como

$$B_{i,t} = \sum_{i}^{K} z_{i,k,0} m_{k,t}$$

Donde k son las categorías (ejemplo los países de origen si son flujos de inmigración), z son las cuotas iniciales *(shares)* y m los cambios en el tiempo específicos de cada país *(shift)*

Predecir los flujos de inmigración usando los flujo por país ponderado con las cuotas iniciales

Bartik Instrument: interpretation

Interpretación 1: *share view*

Si estamos hablando de un choque común (ola de inmigración venezolana) es obvio que la variación te la dan las cuotas iniciales, entonces el supuesto de identificación es que las cuotas sea exógenas

- Esto hay que defenderlo
- Es posible usar los pesos de Rotemberg, que permite usar cada cuota como un instrumento (y puedo concentrarme en los más fuertes)
- Problemas
  - Si son ciudades o sectores, tengo cantidades de instrumentos
  - Restricción de exclusión es difícil y weak instrument es un lío (puedo hacer prueba de sobre-identificación)

Paper clave: Goldsmith-Pinkham, et al. (2020)

Bartik Instrument: interpretation

Interpretación 1: *shift view*

Si tenemos choques temporales, entonces podemos aceptar las cuotas como endógenas y usar las variaciones *(shift)*

Los choques a las variaciones no tienen que ser correlacionadas con los segsos de las cuotas

Paper clave: Borusyak, et al. (2019)