Semana 6. Variables Instrumentales Clásicas (IV)

Equipo Econometría Avanzada

Universidad de los Andes

16 de septiembre de 2022

Contenido

Contexto

Angrist y Krueger (1991) utilizan datos de niños y niñas de Estados Unidos para estimar los retornos a la educación.

- Los niños debían tener 6 años al 1ro de Enero del año escolar (que inicia en Agosto) al que ingresaban.
- Las leyes educativas requerían que los jóvenes continuaran con la escuela al menos hasta sus 16-17 años.

Pregunta de investigación

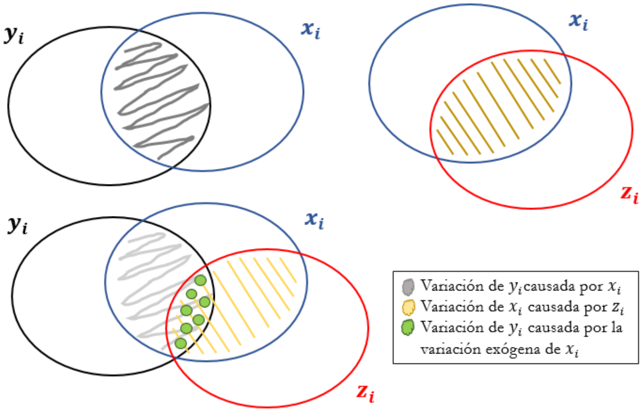
¿Cuál es el efecto de la educación $(educ_i)$ sobre el salario (y_i)?

$$ln(y_i) = \beta_0 + \beta_1 educ_i + u_i$$

- ¿Sería $\widehat{\beta}_1^{OLS}$ un estimador consistente del efecto de interés?

Variables instrumentales

Lidia con cierto tipo de endogeneidad ($E(u|x) \neq 0$) empleando la variación exógena de x causada por Z para identificar un efecto causal de x sobre y.



Instrumento para los años de educación

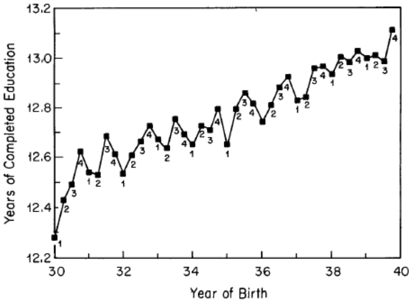


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

- Si se tiene un instrumento Z continuo para *educ* y se quiere incluir $educ^2$, ¿hay que buscar otro instrumento Z' para $educ^2$?
- Tarea: En aeioTU, ¿con qué podríamos instrumentar la asistencia efectiva?

Instrumento para los años de educación

La metodología de estimación es **Mínimos Cuadrados en Dos Etapas (MC2E)**

Suponga que el proceso generador de los datos viene dado por:

$$educ_i = \alpha_0 + \alpha_1 Trim1_i + x_i\alpha_2 + e_i \tag{1}$$

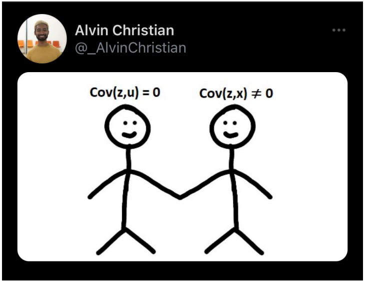
$$ln(y_i) = \beta_0 + \beta_1 educ_i + x_i\beta_2 + u_i \tag{2}$$

1. Estime por MCO la ecuación (1) y recupere $\widehat{educ}_i$.
2. Estime por MCO

$$ln(y_i) = \beta_0 + \beta_1 \widehat{educ}_i + x_i\beta_2 + \tilde{u}_i \tag{3}$$

Supuestos de identificación

Bajo el supuesto de efectos de tratamiento homogeneos los supuestos de identificación son:



(1) Exogeneidad de z ($Cov(Trim1_i, u_i) = 0$) y (2) relevancia de z ($\alpha_1 \neq 0$).

- Bajo (1) y (2), ¿$\widehat{\beta}_1$ es un estimador insesgado y consistente de β_1?

¿Nos basta la relevancia?

En el caso univariado tenemos

$$x_i = \gamma_0 + \gamma_1 z_i + \eta_i$$
$$y_i = \delta_0 + \delta_1 x_i + \varepsilon_i$$

Luego,

$$cov(y_i, z_i) = cov(\delta_0 + \delta_1 x_i + \varepsilon_i, z_i) = \delta cov(x_i, z_i) + cov(\varepsilon_i, z_i)$$

Usando que $\delta_{IV} = \frac{cov(y_i, z_i)}{cov(x_i, z_i)}$,

$$\delta_{IV} = \delta + \frac{cov(\varepsilon_i, z_i)}{cov(x_i, z_i)}$$

Luego, si el instrumento es débil ($cov(x_i, z_i) \to 0$), el sesgo se hace cada vez mayor. De hecho, en el caso general,

$$\mathbb{E}[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2} \frac{1}{1 + F}$$

Donde F es el estadístico F que resulta de la prueba de significancia conjunta de los instrumentos en la primera etapa.

Problemas de tener instrumentos débiles

Deseamos entonces que $F \to \infty$... Pero: ¿en la práctica qué se considera un F grande?

- Para poder hacer inferencia estadísticamente válida mediante una distribución normal (la manera usual) se necesita un F de al menos 104.7 (Ver link).
- Realizar inferencia cuando la F está entre 16.4 y 104.7 puede resultar en hasta un 10 % más de probabilidad de cometer error tipo I relativo al fijado por el nivel de significancia.
- Para F más pequeños no se puede predecir el comportamiento del estimador (se sugiere evitar usar MC2E).

Pero además, en presencia de instrumentos débiles:

- Las pruebas de hipótesis asociadas a MC2E tienen muy poco poder.
- Cuando el sesgo por MCO es positivo, los errores estándar de MC2E son mucho más pequeños que los verdaderos $\to$ En contexto de un tratamiento, es muy difícil detectar efectos negativos si la selección fue positiva (Ej: El efecto de la eduación en el salario).

¿Qué hacer si mi instrumento es (moderadamente) débil?

Si el instrumento es relativamente débil $F \in [10, 104{,}7]$, todavía podemos hacer inferencia exacta de nuestras pruebas de hipótesis. **En estos casos, se recomienda utilizar estadísticos robustos a instrumentos débiles además de MC2E** (Ver link).

Para el caso de un solo instrumento:

- Estadístico de Anderson-Rubin.

Para el caso de múltiples instrumentos:

- El k-estadístico de Fuller/ El estadístico de máxima verosimilitud limitada (LIML).
- Variables instrumentales tipo jacknife (JIVE).
- El estimador de Andrews & Armstrong (2017).

Ojo: Si $F < 10$, los estadísticos robustos a inferencia débil pueden no ser confiables.

¿Cómo efectuar inferencia estadística?

Suponga que estimamos (3) por MCO. ¿Podemos realizar inferencia estadística sobre β_1 a partir de los errores que obtengamos?

- ¿Qué fuente de incertidumbre no estamos considerando?

¿Qué soluciones hay?

Bootstrap

Bootstrap o *Bootstrapping* es una metodología **no paramétrica** que permite, entre otras cosas, calcular los errores estándar asociados a un estimador.

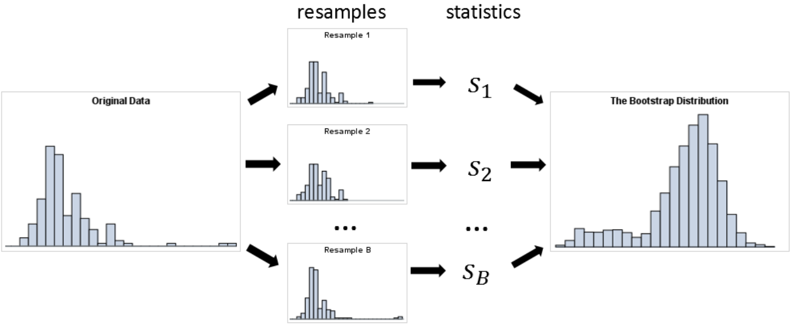


Imagen tomada de SAS (2021).

Al ser una metodología **no paramétrica**, no depende de la distribución de los errores. ¿Cuál es la diferencia entre paramétrico y no paramétrico?

Bootstrap

La metodología consiste en usar remuestreos aleatorios usando *la distribución observada* de las variables, para reestimar múltiples veces el parámetro de interés y así inferir su distribución aproximada. Para implementar un esquema de Bootstrap clásico, se deben seguir los siguientes pasos:

1. Construya una muestra aleatoria de tamaño $N^{Bootstrap}$ de la base de datos original (de tamaño N). La muestra debe ser construida **con remplazo.**
   - Generalmente $N^{Bootstrap} = N$
2. Estime el parámetro de interés en la base remuestreada.
3. Repita los pasos 1 y 2 K veces.
4. A partir de la distribución de los parámetros estimados en las K iteraciones, calcule el estadístico de interés.
   - El error estándar de un parámetro estimado corresponde a la desviación estándar de la distribución.

Esta metodología puede ser adaptada a situaciones en las que hay presencia de correlación a nivel de clúster.

¡Gracias!