Universidad Torcuato Di Tella
Maestrías en Economía y Econometría
2022

Microeconometría II
Práctica 5
Variables instrumentales

---

1. Estimador de Wald

Suponga un modelo de regresión simple

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

donde x_i es potencialmente endógena. Además, suponga que que el instrumento, z_i es una variables binaria. Muestre que el estimador IV en este caso es

$$\beta_1^{IV} = \frac{\overline{y}_1 - \overline{y}_0}{\overline{x}_1 - \overline{x}_0}$$

donde $\overline{y}_1, \overline{x}_1\left(\overline{y}_0, \overline{x}_0\right)$ representan las medias cuando $z = 1\left(z = 0\right).$

Solution:
Como vimos en el ejercicio anterior, el estimador IV es

$$\beta_1^{IV} = \left(Z'X\right)^{-1}\left(Z'y\right)$$

En este caso tenemos (recuerden que el vector de instrumentos también debe tener a las variables exógenas del modelo)

$$X = \begin{bmatrix}\mathbf{1}_n & \mathbf{x}\end{bmatrix} \qquad Z = \begin{bmatrix}\mathbf{1}_n & \mathbf{z}\end{bmatrix}$$

Entonces

$$\beta^{IV} = \left(\begin{bmatrix}\mathbf{1}_n \\ \mathbf{z}\end{bmatrix}\begin{bmatrix}\mathbf{1}_n & \mathbf{x}\end{bmatrix}\right)^{-1}\left(\begin{bmatrix}\mathbf{1}_n \\ \mathbf{z}\end{bmatrix} y\right) = \begin{bmatrix}\mathbf{1}_n'\mathbf{1}_n & \mathbf{1}_n'\mathbf{x} \\ \mathbf{z}'\mathbf{1}_n & \mathbf{z}'\mathbf{x}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{1}_n y' \\ \mathbf{z}'y\end{bmatrix} = \begin{bmatrix}n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n z_i & \sum_{i=1}^n z_i x_i\end{bmatrix}^{-1}\begin{bmatrix}\sum_{i=1}^n y_i \\ \sum_{i=1}^n z_i y_i\end{bmatrix}$$

Llamemos n_1 a la cantidad de observaciones donde $z = 1$ y n_0 a la cantidad de observaciones donde $z = 0$. Como z es binaria tenemos

$$\sum_{i=1}^n z_i = n_1 \qquad \sum_{i=1}^n z_i x_i = \sum_{z_i=1} x_i \qquad \sum_{i=1}^n z_i y_i = \sum_{z_i=1} y_i$$

Reemplazando en β^{IV}

$$\beta^{IV} = \begin{bmatrix}n & \sum_{i=1}^n x_i \\ n_1 & \sum_{z_i=1} x_i\end{bmatrix}^{-1}\begin{bmatrix}\sum_{i=1}^n y_i \\ \sum_{z_i=1} y_i\end{bmatrix}$$

Primero invertimos $Z'X$. Calculamos el determinante

$$det\left(Z'X\right) = n\sum_{z_i=1} x_i - n_1\sum_{i=1}^n x_i$$

Y como la matriz es de 2×2 calculamos su inversa

$$\left(Z'X\right)^{-1} = \frac{1}{n\sum_{z_i=1} x_i - n_1\sum_{i=1}^n x_i}\begin{bmatrix}\sum_{z_i=1} x_i & -\sum_{i=1}^n x_i \\ -n_1 & n\end{bmatrix}$$

Nos interesa nada más calcular el segundo elemento de β^{IV}. El mismo es, entonces (hay que hacer el producto de la segunda fila de la primer matriz con la columna de la segunda matriz)

$$\begin{aligned}
\beta_2^{IV} &= \frac{n\sum_{z_i=1} y_i - n_1 \sum_{i=1}^{n} y_i}{n\sum_{z_i=1} x_i - n_1 \sum_{i=1}^{n} x_i} \\
&= \frac{n\sum_{z_i=1} y_i - n_1 \left(\sum_{z_i=1} y_i + \sum_{z_i=0} y_i\right)}{n\sum_{z_i=1} x_i - n_1 \left(\sum_{z_i=1} x_i + \sum_{z_i=0} x_i\right)} \\
&= \frac{(n-n_1)\sum_{z_i=1} y_i - n_1 \sum_{z_i=0} y_i}{(n-n_1)\sum_{z_i=1} x_i - n_1 \sum_{z_i=0} x_i} \\
&= \frac{n_0\sum_{z_i=1} y_i - n_1 \sum_{z_i=0} y_i}{n_0\sum_{z_i=1} x_i - n_1 \sum_{z_i=0} x_i} \\
&= \frac{n_0 n_1 \bar{y}_1 - n_0 n_1 \bar{y}_0}{n_0 n_1 \bar{x}_1 - n_0 n_1 \bar{x}_0} \\
&= \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}
\end{aligned}$$

2. Estimador de Wald con datos simulados

En este ejercicio se propone extender la simulación del Problem Set 1 a un marco en el que la asignación del tratamiento y quienes resultan tratados no son iguales.

1. Inicialice una muestra con 100 observaciones. Genere resultados potenciales de no recibir el tratamiento como

$$Y_0 \sim \mathcal{N}(100, 30)$$

.

2. Genere ahora un efecto de tratamiento constante e igual a 20 para todos, es decir, $TE_i = 20$ para todo $i = 1, ..., n$. Genere una variable aleatoria normal estándar. Genere una variable de tratamiento D_i igual a 1 para aquellas observaciones que poseen un valor positivo en la variable aleatoria normal.

3. Genere una variable aleatoria uniforme en el intervalo $[0, 1]$. Con ella, genere variables que indiquen el tipo de individuo. Utilice: *always taker* si la variable es menor a 0.25, *never taker* si la variable está entre 0.25 y 0.5, *defier* si la variable está entre 0.5 y 0.75 y *complier* si la variable es mayor a 0.75. *G*enere la variable de si los individuos toman el tratamiento o no dependiendo del grupo en el que están.

4. Genere la variable Y observada como $Y = DY_1 + (1 - D)Y_0$.

5. Estime el LATE y compare con el ATE.

3. Galiani & Schargrodsky (2010)

Lea el artículo *"Property rights for the poor: Effects of land titling"* de Galiani & Schargrodsky.

1. ¿Qué efectos intentan estimar en el paper?

2. ¿Cuál es la estrategia de identificación? ¿Por qué no funciona la diferencia de medias simple?

3. Replique las resultados del paper.