

Clase auxiliar # 7

Endogeneidad y variables instrumentales

Canción de hoy: Dark necessities - Red Hot Chili Peppers.

1. ¡Recordar!

El método de mínimos cuadrados en 2 etapas (MC2E) permite una estimación consistente en presencia de endogeneidad.

- **Etapla 0:** darse cuenta de que en un modelo de la siguiente forma hay una variable endógena x_i .

$$y_i = \alpha_0 + \alpha_1 w_{1i} + \dots + \alpha_p w_{pi} + \beta x_i + u_i$$

- **Etapla 1:** obtener una predicción $\hat{x}$ ortogonal a u mediante la siguiente regresión

$$x_i = \gamma_0 + \gamma_1 w_{1i} + \dots + \gamma_p w_{pi} + \varepsilon_i$$

- **Etapla 2:** reemplazar x por $\hat{x}$ en el modelo original (etapa 0), estimando β'

$$y_i = \alpha_0 + \alpha_1 w_{1i} + \dots + \alpha_p w_{pi} + \beta' \hat{x}_i + u_i$$

2. Endogeneidad

1. Muestre que existe endogeneidad en estas tres situaciones:
 - Cuando hay variables omitidas relevantes.
 - Cuando existen errores de medición no constantes.
 - Cuando dos variables se determinan simultáneamente.
2. Muestre que en presencia de endogeneidad $\hat{\beta}$ es sesgado.
3. Mostrar que en presencia de endogeneidad $\hat{\beta}$ es inconsistente.

3. Variables instrumentales

En el archivo Auxiliar 07 - BD Variables Instrumentales.dta contenido en Material Docente se encuentra una base de datos que contiene información sobre aspectos económicos y de salud de distintos ciudadanos estadounidenses. Las variables contenidas se describen a continuación:

- `log_gasto_med`: logaritmo del gasto médico de la persona.
- `seguro_salud`: variable dummy que señala si la persona tiene o no seguro de salud.
- `enfermedad`: cantidad de enfermedades que tiene la persona.
- `edad`: edad de la persona.
- `log_ingreso`: logaritmo del ingreso de la persona.
- `ssi`: el Seguro Suplementario al Ingreso (SSI) es un programa del gobierno de EEUU que proporcion pagos a personas de bajos ingresos que tienen 65 años o más, que sean ciegos o discapacitados. Esta variable contiene el cociente entre dicho aporte y el ingreso de la persona.

1. En la misma revista de salud, de dudosa procedencia, de la cual hablamos en la clase auxiliar 5, se propone el siguiente modelo de regresión que busca explicar el gasto médico de los ciudadanos estadounidenses mediante el siguiente modelo:

$$\log_gasto_med_i = \beta_0 + \beta_1 \text{seguro_salud}_i + \beta_2 \text{enfermedad}_i + \beta_3 \text{edad}_i + \beta_4 \log_ingreso_i + u_i$$

Al ver esto, usted recurre extasiado a una amiga suya y le muestra el modelo, a lo que ella responde "hay algo que me hace ruido: una persona puede contratar un seguro de salud porque sus gastos médicos son altos, así como puede tener altos gastos médicos gracias a que contrata un seguro de salud". ¿A qué posible fuente de endogeneidad se puede estar refiriendo su amiga?

2. Realice la estimación de Mínimos Cuadrados en 2 Etapas en *Stata* realizando las regresiones de manera secuencial.
3. Realice la estimación de Mínimos Cuadrados en 2 Etapas en *Stata* mediante el comando `ivregress` y compare sus resultados con los de la parte anterior.
4. Obtenga el estimador $\hat{\beta}_{VI} = (\bar{Z}^T X)^{-1} \bar{Z}^T Y$ en *Matlab*.