IN4402 - Aplicaciones de Probabilidades y Estadística en Gestión
Profesor: Benjamín Villena Roldán
Profesor auxiliar: Ronald Leblebici Garo

INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Clase auxiliar # 7

Regresión discontinua difusa, endogeneidad y variables instrumentales

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

1. Endogeneidad

1. Muestre que existe endogeneidad en estas tres situaciones:

   - Cuando hay variables omitidas relevantes.

     - **Modelo verdadero:** $Y_i = \alpha + \beta x_i + \gamma z_i + u_i$.
     - **Modelo estimado:** $Y_i = \alpha + \beta x_i + v_i$.

     Notemos que existe endogeneidad al omitir variables relevantes:

     $$Cov(x_i, v_i) = Cov(x_i, \gamma z_i + u_i) = \gamma \underbrace{Cov(x_i, z_i)}_{\neq 0} + \underbrace{Cov(x_i, u_i)}_{=0} \neq 0$$

     Es así como la omisión de variables omitidas es un caso particular de endogeneidad.

   - Cuando existen errores de medición no constantes.

     - **Modelo verdadero:** $Y_i = \alpha + \beta x_i + \epsilon_i$.
     - **Modelo estimado:** $Y_i = \alpha + \beta x_i^* + u_i$, donde $x_i^* = x_i + \delta_i$.

     $$Y_i = \alpha + \beta x_i^* + u_i = \alpha + \beta(x_i + \delta_i) + u_i = \alpha + \beta x_i + \underbrace{(\beta\delta_i + u_i)}_{\epsilon_i}$$

     $$\begin{aligned} Cov(x_i^*, u_i) &= Cov(x_i + \delta_i, \epsilon_i - \beta\delta_i) \\ &= Cov(x_i + \delta_i, -\beta\delta_i) \\ &= -\beta Cov(x_i + \delta_i, \delta_i) \\ &= -\beta[Cov(x_i, \delta_i) + \mathbb{V}(\delta_i)] \neq 0 \end{aligned}$$

   - Cuando dos variables se determinan simultáneamente.

     - **Modelo 1:** $y_i = \alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i$.

- **Modelo 2:** $z_i = \beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i$.

$$\begin{aligned}
Cov(z_i, u_i) &= Cov(\beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i, u_i) \\
&= \beta_1 Cov(y_i, u_i) \\
&= \beta_1 Cov(\alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i, u_i) \\
&= \beta_1 [\alpha_1 Cov(z_i, u_i) + \mathbb{V}(u_i)]
\end{aligned}$$

$$\Rightarrow Cov(z_i, u_i) = \frac{\mathbb{V}(u_i)}{1 - \alpha_1 \beta_1}$$

2. Muestre que en presencia de endogeneidad $\hat{\beta}$ es sesgado.

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X\beta + U) = \beta + (X^T X)^{-1} X^T U$$
$$\Rightarrow \mathbb{E}(\hat{\beta}|X) = \beta + \underbrace{(X^T X)^{-1} X^T \mathbb{E}(U|X)}_{\text{sesgo}}$$

Ahora debemos demostrar que el sesgo calculado es distinto de cero.

Demostremos que $\mathbb{E}(X^T U) \neq 0 \Rightarrow \mathbb{E}(U|X) \neq 0$.

Por contrarecíproca demostraremos que $\mathbb{E}(U|X) = 0 \Rightarrow \mathbb{E}(X^T U) = 0$.

Utilizando la ley de las esperanzas iteradas, se obtiene esta propiedad.

$$\mathbb{E}(U|X) = 0 \Rightarrow \mathbb{E}(X^T U) = \mathbb{E}[\mathbb{E}(X^T U|X)] = \mathbb{E}[X^T \underbrace{\mathbb{E}(U|X)}_{=0}] = 0$$

$$\mathbb{E}(U|X) \neq 0 \Rightarrow \underbrace{(X^T X)^{-1} X^T \mathbb{E}(U|X)}_{\text{sesgo}} \neq 0$$

3. Mostrar que en presencia de endogeneidad $\hat{\beta}$ es inconsistente.

$$\hat{\beta} = \beta + \left(\frac{1}{N} X^T X\right)^{-1} \left(\frac{1}{N} X^T U\right)$$

Utilizando los teoremas de Slutzky y de mapeo continuo...

$$\hat{\beta} \xrightarrow{p} \beta + \mathbb{E}(X^T X)^{-1} \underbrace{\mathbb{E}(X^T U)}_{\neq 0} \neq \beta$$

$$\hat{\beta} =$$

2. 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 cuociente 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:

   $$\texttt{log\_gasto\_med}_i = \beta_0 + \beta_1 \texttt{seguro\_salud}_i + \beta_2 \texttt{enfermedad}_i + \beta_3 \texttt{edad}_i + \beta_4 \texttt{log\_ingreso}_i + u_i$$

   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?

   La amiga en cuestión se está refiriendo a un problema de simultaneidad, que –como vimos en la pregunta anterior– genera endogeneidad en el modelo

2. Realice la estimación de Mínimos Cuadrados en 2 Etapas en *Stata* realizando las regresiones de manera secuencial.

   Ver archivo `Auxiliar 07 - Código.do`.

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.

   Ver archivo `Auxiliar 07 - Código.do`.

4. Obtenga el estimador $\hat{\beta}_{VI} = (\bar{Z}^T X)^{-1} \bar{Z}^T Y$ en *Matlab*.

   Ver archivo `Auxiliar 07 - Matlab.rar`.