

Clase auxiliar # 9

Endogeneidad y variables instrumentales

¡Recordar!

Causales de endogeneidad: (1) omisión de variables relevantes, (2) existencia de errores de medición no constantes, (3) determinación simultánea de ecuaciones y (4) auto-correlación de los errores (estudiaremos esto cuando llegemos a series de tiempo).

Identificación: se dice que una ecuación está **sub/sobre/exactamente-identificada** cuando la cantidad de variables presentes en el modelo pero ausentes en la ecuación es **menor/mayor/igual** a la cantidad de variables endógenas del modelo menos 1. $k_{\text{ecuación}} = g_{\text{-ecuación}} + e_{\text{-ecuación}} (< / > / =) g - 1$

Mínimos Cuadrados en 2 Etapas (MC2E): supongamos que se quiere estimar una ecuación de la forma $Y = X_1\beta_1 + X_2\beta_2 + U$, donde X_1 representa uno o más regresores endógenos. Además se tiene uno o más instrumentos Z que cumplen los supuestos de relevancia y exclusión para la ecuación que se quiere estimar. Para poder usar este método se requiere el la ecuación esté al menos exactamente identificada.

- **Etapa 0:** darse cuenta del problema de endogeneidad dado por X_1 .
- **Etapa 1:** llamaremos “matriz de instrumentos expandida” a $\tilde{Z} = [X_2', Z']$. Regresionar $X_1 = \tilde{Z}\delta + \varepsilon$ para obtener una proyección \hat{X}_1 de X_1 ortogonal a U .
- **Etapa 2:** reemplazar en la regresión original X_1 por \hat{X}_1 , regresionando $Y = \hat{X}_1\tilde{\beta}_1 + X_2\tilde{\beta}_2 + \tilde{U}$, obteniendo el estimador $\hat{\beta}_{MC2E}$.

Supuestos de variables instrumentales: los instrumentos Z deben cumplir los supuestos de relevancia/fortaleza/no-debilidad ($Cov(Z, X_1) \neq 0$) y exclusión/validez ($Cov(Z, U) = 0$).

Estimador de variables instrumentales: Siguiendo la misma notación que en el punto anterior $\hat{\beta}_{VI} = (\tilde{Z}'X)^{-1}\tilde{Z}'Y$. Este estimador solo se puede obtener cuando la ecuación de interés está exactamente identificada¹ En dicho caso $\hat{\beta}_{VI} = \hat{\beta}_{MC2E}$

1. Endogeneidad

1. Muestre que existe endogeneidad en estas tres situaciones:
 - Cuando hay variables omitidas relevantes.
 - Cuando existen errores de medición no constantes. Por simplicidad puede suponer errores de medición exógenos para esta parte.
 - 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.

¹En el caso sobreidentificado no cuadran las dimensiones de las matrices.

2. Identificación, mínimos cuadrados indirectos y variables instrumentales

Considere el mercado de las papas. La curva de oferta es $Q_t^s = \alpha_0 + \alpha_1 P_t + \alpha_2 L_t + U_t$ y la curva de demanda es $Q_t^d = \beta_0 + \beta_1 P_t + \beta_2 I_t + V_t$, donde L_t es la lluvia caída e I_t el ingreso nominal (ambas son variables exógenas). Asuma que el mercado de las papas se rige por los supuestos de competencia perfecta y que las muestras que se tienen fueron tomadas en equilibrio ($Q_t^d = Q_t^s$).

- Muestre que cantidades y precios son variables endógenas.
- ¿Las ecuaciones están sobre, sub o exactamente identificado?
- Dada su respuesta en la parte (b), ¿es posible estimar los parámetros (α, β) sin usar variables instrumentales?
- Explique cómo construir el estimador $\hat{\beta}_{VI}$ y señale qué supuestos validarían el uso del estimador propuesto.

3. Variables instrumentales: computacional

En el archivo Auxiliar 09 - 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. En una revista de salud, de dudosa procedencia, se estima mediante MCO el siguiente modelo de que busca explicar el gasto médico de los ciudadanos estadounidenses:

$$\ln(\text{gastomed})_i = \beta_0 + \beta_1 \text{segurosalud}_i + \beta_2 \text{enfermedad}_i + \beta_3 \text{edad}_i + \beta_4 \ln(\text{ingreso})_i + u_i$$

- Usted le comenta esto a una amiga, quien le dice “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?
- Realice la estimación de Mínimos Cuadrados en 2 Etapas en *Stata* realizando las regresiones de manera secuencial.
- 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.
- Obtenga el estimador $\hat{\beta}_{VI} = (\bar{Z}^T X)^{-1} \bar{Z}^T Y$ en *Matlab*.