

Clase auxiliar 7

Aplicaciones de probabilidades y estadística en gestión
Departamento de Ingeniería Civil Industrial
Universidad de Chile

Ronald Leblebici Garo

31 de octubre de 2017

Antes de imprimir esta presentación, piense si es realmente necesario.

¿Qué vimos la clase pasada?

- Introducción a datos de corte transversal, series de tiempo y datos de panel
- Dif-in-Dif
- Test Reset de Ramsey (selección de modelos)

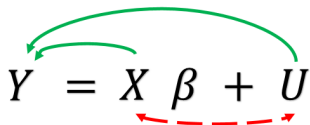
¿Qué vimos la clase pasada

- Existen distintos tipos de bases de datos: de **corte transversal** (muchos individuos en un periodo), **series de tiempo** (un individuo en muchos periodos) y **datos de panel** (muchos individuos en muchos periodos).
- Los **datos de panel** pueden ser útiles para aplicar el método **Dif-in-Dif**, el cual permite estimar el efecto de un tratamiento que se aplica para una parte de la muestra en un intervalo de tiempo.
- **Dif-in-dif en múltiples periodos** nos permite controlar heterogeneidades entre individuos que no varían en el tiempo.
- El **test Reset de Ramsey** permite evaluar si es “correcta” la forma funcional de nuestro modelo. Nos da una señal de si es que son significativas –en su conjunto– algunas combinaciones no lineales de los regresores ya considerados.

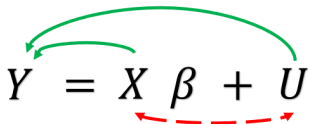
¿Qué veremos hoy?

- Regresión discontinua difusa.
- Endogeneidad.
- Variables instrumentales y mínimos cuadrados en 2 etapas.

¿Qué es la endogeneidad?

$$Y = X\beta + U$$


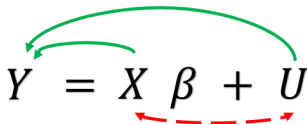
¿Qué es la endogeneidad?



The diagram shows the regression equation $Y = X\beta + U$. A solid green curved arrow points from the error term U to the regressor X , indicating that the error term influences the regressor. A dashed red curved arrow points from the regressor X to the error term U , indicating that the regressor influences the error term. These two arrows together represent the endogeneity of the regressor X .

- Es cuando uno o más regresores (o variables observadas) están correlacionados con el error (o variables no observadas). En otras palabras, es un problema de colinealidad entre lo observable y lo no observable.

¿Qué es la endogeneidad?



The diagram shows the regression equation $Y = X\beta + U$. A solid green curved arrow points from U to X , and a shorter solid green curved arrow points from X to Y . A dashed red curved arrow points from U to U , indicating a correlation between the regressor X and the error term U .

- Es cuando uno o más regresores (o variables observadas) están correlacionados con el error (o variables no observadas). En otras palabras, es un problema de colinealidad entre lo observable y lo no observable.
- **Intuitivamente:** si lo observable y lo no observable se mueven juntos y ambos afectan Y , ¿cómo sé a quién echarle la culpa –y en qué magnitud– de los cambios en Y ? Esto se llama **problema de identificación**.

¿Qué es la endogeneidad?

The diagram shows the regression equation $Y = X\beta + U$. A solid green curved arrow points from the error term U to the regressor X , indicating that U is correlated with X . A dashed red curved arrow points from the regressor X back to the error term U , indicating that X is correlated with U . These two arrows together represent the endogeneity of X .

- Es cuando uno o más regresores (o variables observadas) están correlacionados con el error (o variables no observadas). En otras palabras, es un problema de colinealidad entre lo observable y lo no observable.
- **Intuitivamente:** si lo observable y lo no observable se mueven juntos y ambos afectan Y , ¿cómo sé a quién echarle la culpa –y en qué magnitud– de los cambios en Y ? Esto se llama **problema de identificación**.
- **Matemáticamente:** $\hat{\beta}$ se vuelve sesgado e inconsistente. Estaremos estimando mal β y lo seguiremos estimando mal a pesar de tener una muestra grande.

¿Qué es la endogeneidad?

Diremos que X es una variable endógena si $\text{Cov}(X, U) \neq 0$.

Notemos que $\text{Cov}(X, U) = \mathbb{E}(X^T U)$. En efecto:

$$\text{Cov}(X, U) = \mathbb{E}[(X - \mathbb{E}(X))^T (U - \underbrace{\mathbb{E}(U)}_{=0})]$$

$$= \mathbb{E}[X^T U - \mathbb{E}(X) U]$$

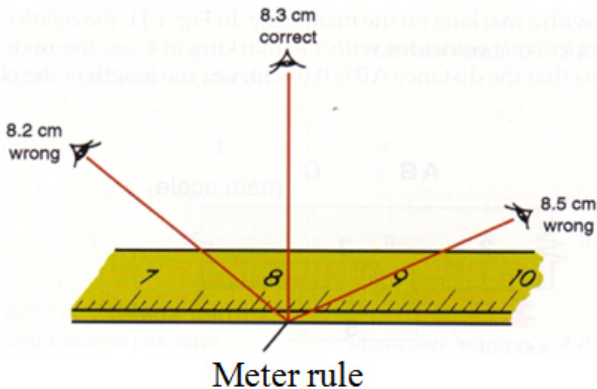
$$= \mathbb{E}(X^T U) - \mathbb{E}(X) \underbrace{\mathbb{E}(U)}_{=0}$$

$$= \mathbb{E}(X^T U)$$

Fuentes de endogeneidad

- Errores de medición
- Variables omitidas
- Ecuaciones simultáneas
- Rezagos y correlación serial

Errores de medición



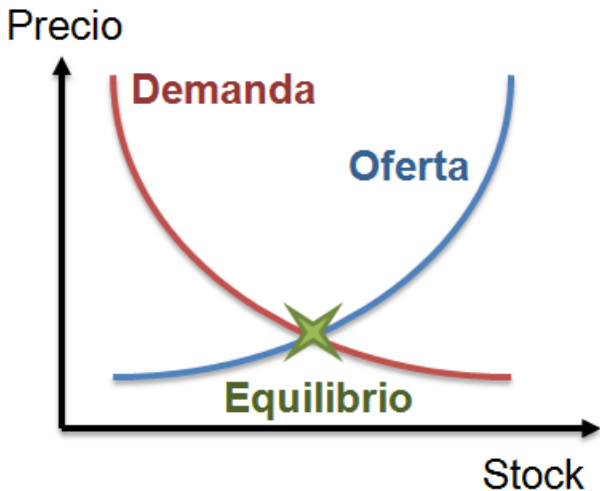
Variables omitidas



Si observo la cantidad de helados vendidos en función del clima puede ser que haya endogeneidad.

El señor de los helados puede fijar sus precios dependiendo del clima, ¿no?

Simultaneidad



Clásico ejemplo de variables que se determinan mutuamente.

P1: Endogeneidad al omitir variables

- **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:

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

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

P1. Endogeneidad al error en la medición

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

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

P1. Endogeneidad al error en la medición

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

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

$$\text{Cov}(x_i^*, u_i) = \text{Cov}(x_i + \delta_i, \epsilon_i - \beta \delta_i)$$

P1. Endogeneidad al error en la medición

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

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

$$\begin{aligned} \text{Cov}(x_i^*, u_i) &= \text{Cov}(x_i + \delta_i, \epsilon_i - \beta \delta_i) \\ &= \text{Cov}(x_i + \delta_i, -\beta \delta_i) \end{aligned}$$

P1. Endogeneidad al error en la medición

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

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

$$\begin{aligned} \text{Cov}(x_i^*, u_i) &= \text{Cov}(x_i + \delta_i, \epsilon_i - \beta \delta_i) \\ &= \text{Cov}(x_i + \delta_i, -\beta \delta_i) \\ &= -\beta \text{Cov}(x_i + \delta_i, \delta_i) \end{aligned}$$

P1. Endogeneidad al error en la medición

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

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

$$\text{Cov}(x_i^*, u_i) = \text{Cov}(x_i + \delta_i, \epsilon_i - \beta \delta_i)$$

$$= \text{Cov}(x_i + \delta_i, -\beta \delta_i)$$

$$= -\beta \text{Cov}(x_i + \delta_i, \delta_i)$$

$$= -\beta [\text{Cov}(x_i, \delta_i) + \mathbb{V}(\delta_i)] \neq 0$$

P1. Endogeneidad en simultaneidad

- **Modelo 1:** $y_i = \alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i$.
- **Modelo 2:** $z_i = \beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i$.

$$\text{Cov}(z_i, u_i) = \text{Cov}(\beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i, u_i)$$

P1. Endogeneidad en simultaneidad

- **Modelo 1:** $y_i = \alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i$.
- **Modelo 2:** $z_i = \beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i$.

$$\begin{aligned} \text{Cov}(z_i, u_i) &= \text{Cov}(\beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i, u_i) \\ &= \beta_1 \text{Cov}(y_i, u_i) \end{aligned}$$

P1. Endogeneidad en simultaneidad

- **Modelo 1:** $y_i = \alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i$.
- **Modelo 2:** $z_i = \beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i$.

$$\text{Cov}(z_i, u_i) = \text{Cov}(\beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i, u_i)$$

$$= \beta_1 \text{Cov}(y_i, u_i)$$

$$= \beta_1 \text{Cov}(\alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i, u_i)$$

P1. Endogeneidad en simultaneidad

- **Modelo 1:** $y_i = \alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i$.
- **Modelo 2:** $z_i = \beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i$.

$$\text{Cov}(z_i, u_i) = \text{Cov}(\beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i, u_i)$$

$$= \beta_1 \text{Cov}(y_i, u_i)$$

$$= \beta_1 \text{Cov}(\alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i, u_i)$$

$$= \beta_1 [\alpha_1 \text{Cov}(z_i, u_i) + \mathbb{V}(u_i)]$$

P1. Endogeneidad en simultaneidad

- **Modelo 1:** $y_i = \alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i$.
- **Modelo 2:** $z_i = \beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i$.

$$\text{Cov}(z_i, u_i) = \text{Cov}(\beta_0 + \beta_1 y_i + \beta_2 x_{2i} + v_i, u_i)$$

$$= \beta_1 \text{Cov}(y_i, u_i)$$

$$= \beta_1 \text{Cov}(\alpha_0 + \alpha_1 z_i + \alpha_2 x_{1i} + u_i, u_i)$$

$$= \beta_1 [\alpha_1 \text{Cov}(z_i, u_i) + \mathbb{V}(u_i)]$$

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

P1. Endogeneidad y sesgo

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

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X\beta + U) = \beta + (X^T X)^{-1} X^T U$$

P1. Endogeneidad y sesgo

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

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X\beta + U) = \beta + (X^T X)^{-1} X^T U$$

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

Debemos demostrar que el sesgo calculado es distinto de cero.

P1. Endogeneidad y sesgo

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

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X\beta + U) = \beta + (X^T X)^{-1} X^T U$$

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

Debemos demostrar que el sesgo calculado es distinto de cero.

Demostremos que

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

P1. Endogeneidad y sesgo

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

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X\beta + U) = \beta + (X^T X)^{-1} X^T U$$

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

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 contrarecípoca demostraremos que

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

P1. Endogeneidad y sesgo

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

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

P1. Endogeneidad y sesgo

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

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

Esto demuestra que

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

P1. Endogeneidad y sesgo

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

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

Esto demuestra que

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

Finalmente

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

P1. Endogeneidad e inconsistencia

Para que MCO sea consistente, $\hat{\beta}$ debe converger en probabilidad a β .

P1. Endogeneidad e inconsistencia

Para que MCO sea consistente, $\hat{\beta}$ debe converger en probabilidad a β .

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

P1. Endogeneidad e inconsistencia

Para que MCO sea consistente, $\hat{\beta}$ debe converger en probabilidad a β .

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

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

P1. Endogeneidad e inconsistencia

Para que MCO sea consistente, $\hat{\beta}$ debe converger en probabilidad a β .

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

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

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

P1. Endogeneidad e inconsistencia

Para que MCO sea consistente, $\hat{\beta}$ debe converger en probabilidad a β .

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

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

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

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

P1. Test de endogeneidad

- No se puede testear formalmente la endogeneidad de alguna de las variables incluidas en el modelo.

P1. Test de endogeneidad

- No se puede testear formalmente la endogeneidad de alguna de las variables incluidas en el modelo.
- En un eventual test, la hipótesis nula sería $\mathbb{E}(X^T U) = 0$. Sin embargo, U no es observable (no puede estar en nuestra base de datos).

P1. Test de endogeneidad

- No se puede testear formalmente la endogeneidad de alguna de las variables incluidas en el modelo.
- En un eventual test, la hipótesis nula sería $\mathbb{E}(X^T U) = 0$. Sin embargo, U no es observable (no puede estar en nuestra base de datos).
- Se podría pensar que se puede utilizar en su lugar el estimador $\hat{U}$, sin embargo esto no tendría sentido pues por construcción $X^T U = 0$ (condición de primer orden de MCO).

P1. Test de endogeneidad

Recordemos la condición de primer orden de la estimación MCO.

$$\begin{aligned}\sum_{i=1}^N \hat{U}_i^2(\hat{\beta}) &= \hat{U}^T \hat{U} \\ &= Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta} X^T X \hat{\beta}\end{aligned}$$

P1. Test de endogeneidad

Recordemos la condición de primer orden de la estimación MCO.

$$\begin{aligned}\sum_{i=1}^N \hat{U}_i^2(\hat{\beta}) &= \hat{U}^T \hat{U} \\ &= Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta} X^T X \hat{\beta}\end{aligned}$$

$$\nabla_{\hat{\beta}} \left(\sum_{i=1}^N \hat{U}_i^2(\hat{\beta}_{MCO}) \right) = -2X^T Y + 2X^T X \hat{\beta} = \vec{0}_{K \times 1}$$

P1. Test de endogeneidad

Recordemos la condición de primer orden de la estimación MCO.

$$\begin{aligned}\sum_{i=1}^N \hat{U}_i^2(\hat{\beta}) &= \hat{U}^T \hat{U} \\ &= Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta} X^T X \hat{\beta}\end{aligned}$$

$$\begin{aligned}\nabla_{\hat{\beta}} \left(\sum_{i=1}^N \hat{U}_i^2(\hat{\beta}_{MCO}) \right) &= -2X^T Y + 2X^T X \hat{\beta} = \vec{0}_{K \times 1} \\ \Rightarrow -X^T Y + X^T X \hat{\beta} &= \vec{0}_{K \times 1}\end{aligned}$$

P1. Test de endogeneidad

Recordemos la condición de primer orden de la estimación MCO.

$$\begin{aligned}\sum_{i=1}^N \hat{U}_i^2(\hat{\beta}) &= \hat{U}^T \hat{U} \\ &= Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta} X^T X \hat{\beta}\end{aligned}$$

$$\nabla_{\hat{\beta}} \left(\sum_{i=1}^N \hat{U}_i^2(\hat{\beta}_{MCO}) \right) = -2X^T Y + 2X^T X \hat{\beta} = \vec{0}_{K \times 1}$$

$$\Rightarrow -X^T Y + X^T X \hat{\beta} = \vec{0}_{K \times 1}$$

$$\Rightarrow X^T \underbrace{(Y - X\hat{\beta})}_{\hat{\varepsilon}} = 0$$

Variables instrumentales

Variable Z que no pertenece a la ecuación explicativa de una variable dependiente Y y se correlaciona con al menos una variable endógena X . Deben cumplir dos supuestos: relevancia y exclusión.

$$Y = X\beta + U$$


Relevancia / No debilidad
 $Cov(Z, U) \neq 0$

Validez / Exclusión
 $Cov(Z, U) \neq 0$

Variables instrumentales

Encontrar una variable instrumental no suele ser tan directo. Se requiere imaginación.



Mínimos cuadrados en 2 etapas (MC2E)

Permite una estimación consistente en presencia de endogeneidad.

Mínimos cuadrados en 2 etapas (MC2E)

Permite una estimación consistente en presencia de endogeneidad.

- **Etapas 0:** darse cuenta de que en un modelo de la siguiente forma hay x_i es endógena

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

Mínimos cuadrados en 2 etapas (MC2E)

Permite una estimación consistente en presencia de endogeneidad.

- **Etapa 0:** darse cuenta de que en un modelo de la siguiente forma hay x_i es endógena

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

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

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

Mínimos cuadrados en 2 etapas (MC2E)

Permite una estimación consistente en presencia de endogeneidad.

- **Etapa 0:** darse cuenta de que en un modelo de la siguiente forma hay x_i es endógena

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

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

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

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

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

P2: Mínimos cuadrados en 2 etapas

$$\log_gasto_med_i = \beta_0 + \beta_1 seguro_salud_i + \beta_2 enfermedad_i + \beta_3 edad_i \\ + \beta_4 \log_ingreso_i + u_i$$

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?

P3: Mínimos cuadrados en 2 etapas

- 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.

Estimador $\hat{\beta}_{VI}$

Consideremos el siguiente modelo:

$$Y = X_1\beta_1 + X_2\beta_2 + U$$

- $X_1 \in \mathbb{R}^{N \times K_1}$: matriz de regresores endógenos.
- $X_2 \in \mathbb{R}^{N \times K_2}$: matriz de regresores exógenos.
- $Z \in \mathbb{R}^{N \times J}$: matriz de instrumentos.

Estimador $\hat{\beta}_{VI}$

Consideremos el siguiente modelo:

$$Y = X_1\beta_1 + X_2\beta_2 + U$$

- $X_1 \in \mathbb{R}^{N \times K_1}$: matriz de regresores endógenos.
- $X_2 \in \mathbb{R}^{N \times K_2}$: matriz de regresores exógenos.
- $Z \in \mathbb{R}^{N \times J}$: matriz de instrumentos.
- Diremos que el modelo está sobre / exactamente / sub - identificado, dependiendo si $J > K_1$, $J = K_1$ o $J < K_1$. Debemos procurar que el modelo **no** esté sub-identificado.

Estimador $\hat{\beta}_{VI}$

Consideremos el siguiente modelo:

$$Y = X_1\beta_1 + X_2\beta_2 + U$$

- $X_1 \in \mathbb{R}^{N \times K_1}$: matriz de regresores endógenos.
- $X_2 \in \mathbb{R}^{N \times K_2}$: matriz de regresores exógenos.
- $Z \in \mathbb{R}^{N \times J}$: matriz de instrumentos.
- Diremos que el modelo está sobre / exactamente / sub - identificado, dependiendo si $J > K_1$, $J = K_1$ o $J < K_1$. Debemos procurar que el modelo **no** esté sub-identificado.

Nos interesa estimar β_1 . Para hacer esto, partiremos definiendo $\bar{Z} = (X_2, Z)$.

Estimador $\hat{\beta}_{VI}$

Consideremos el siguiente modelo:

$$Y = X_1\beta_1 + X_2\beta_2 + U$$

- $X_1 \in \mathbb{R}^{N \times K_1}$: matriz de regresores endógenos.
- $X_2 \in \mathbb{R}^{N \times K_2}$: matriz de regresores exógenos.
- $Z \in \mathbb{R}^{N \times J}$: matriz de instrumentos.
- Diremos que el modelo está sobre / exactamente / sub-identificado, dependiendo si $J > K_1$, $J = K_1$ o $J < K_1$. Debemos procurar que el modelo **no** esté sub-identificado.

Nos interesa estimar β_1 . Para hacer esto, partiremos definiendo $\bar{Z} = (X_2, Z)$.

Luego pre-multiplicamos el modelo por $\bar{Z}^T$

Estimador $\hat{\beta}_{VI}$

Consideremos el siguiente modelo:

$$Y = X_1\beta_1 + X_2\beta_2 + U$$

- $X_1 \in \mathbb{R}^{N \times K_1}$: matriz de regresores endógenos.
- $X_2 \in \mathbb{R}^{N \times K_2}$: matriz de regresores exógenos.
- $Z \in \mathbb{R}^{N \times J}$: matriz de instrumentos.
- Diremos que el modelo está sobre / exactamente / sub-identificado, dependiendo si $J > K_1$, $J = K_1$ o $J < K_1$. Debemos procurar que el modelo **no** esté sub-identificado.

Nos interesa estimar β_1 . Para hacer esto, partiremos definiendo $\bar{Z} = (X_2, Z)$.

Luego pre-multiplicamos el modelo por $\bar{Z}^T$

Notar que $\bar{Z}$ son los regresores de la primera etapa, en la que obtenemos $\hat{X}$. ¿Qué significa esto?

Estimador $\hat{\beta}_{VI}$

$$\Rightarrow \bar{Z}^T Y = \bar{Z}^T X_1 \beta_1 + \bar{Z}^T X_2 \beta_2 + \bar{Z}^T U$$

Estimador $\hat{\beta}_{VI}$

$$\Rightarrow \bar{Z}^T Y = \bar{Z}^T X_1 \beta_1 + \bar{Z}^T X_2 \beta_2 + \bar{Z}^T U$$

$$\Rightarrow \bar{Z}^T Y = \bar{Z}^T \underbrace{(X_1 \ X_2)}_X \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \bar{Z}^T U$$

Estimador $\hat{\beta}_{VI}$

$$\Rightarrow \bar{Z}^T Y = \bar{Z}^T X_1 \beta_1 + \bar{Z}^T X_2 \beta_2 + \bar{Z}^T U$$

$$\Rightarrow \bar{Z}^T Y = \bar{Z}^T \underbrace{(X_1 \ X_2)}_X \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \bar{Z}^T U$$

$$\hat{\beta}_{VI} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (\bar{Z}^T X)^{-1} \bar{Z}^T Y$$

P3: Mínimos cuadrados en 2 etapas

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

Resumen de la clase

- La endogeneidad es cuando hay correlación entre variables observables y no observables de un modelo.

Resumen de la clase

- La endogeneidad es cuando hay correlación entre variables observables y no observables de un modelo.
- Es un problema común en economía y ciencias sociales, que se puede deber a omisión de variables, simultaneidad (ej: oferta y demanda), errores de medición y algunas situaciones que estudiaremos cuando veamos series de tiempo.

Resumen de la clase

- La endogeneidad es cuando hay correlación entre variables observables y no observables de un modelo.
- Es un problema común en economía y ciencias sociales, que se puede deber a omisión de variables, simultaneidad (ej: oferta y demanda), errores de medición y algunas situaciones que estudiaremos cuando veamos series de tiempo.
- En presencia de endogeneidad $\hat{\beta}_{MCO}$ se vuelve sesgado e inconsistente.

Resumen de la clase

- La endogeneidad es cuando hay correlación entre variables observables y no observables de un modelo.
- Es un problema común en economía y ciencias sociales, que se puede deber a omisión de variables, simultaneidad (ej: oferta y demanda), errores de medición y algunas situaciones que estudiaremos cuando veamos series de tiempo.
- En presencia de endogeneidad $\hat{\beta}_{MCO}$ se vuelve sesgado e inconsistente.
- La estimación de mínimos cuadrados en 2 etapas (M2CE) sirve para solucionar estos problemas.

Resumen de la clase

- La endogeneidad es cuando hay correlación entre variables observables y no observables de un modelo.
- Es un problema común en economía y ciencias sociales, que se puede deber a omisión de variables, simultaneidad (ej: oferta y demanda), errores de medición y algunas situaciones que estudiaremos cuando veamos series de tiempo.
- En presencia de endogeneidad $\hat{\beta}_{MCO}$ se vuelve sesgado e inconsistente.
- La estimación de mínimos cuadrados en 2 etapas (M2CE) sirve para solucionar estos problemas.
- Para implementar M2CE, requerimos de al menos tantos instrumentos como variables endógenas tengamos.

Resumen de la clase

- La endogeneidad es cuando hay correlación entre variables observables y no observables de un modelo.
- Es un problema común en economía y ciencias sociales, que se puede deber a omisión de variables, simultaneidad (ej: oferta y demanda), errores de medición y algunas situaciones que estudiaremos cuando veamos series de tiempo.
- En presencia de endogeneidad $\hat{\beta}_{MCO}$ se vuelve sesgado e inconsistente.
- La estimación de mínimos cuadrados en 2 etapas (M2CE) sirve para solucionar estos problemas.
- Para implementar M2CE, requerimos de al menos tantos instrumentos como variables endógenas tengamos.
- Los instrumentos deben cumplir 2 supuestos: relevancia y exclusión.