

Taller 3 - Econometría II - Universidad Santo Tomás

Nicolás Ronderos

En todos los casos muestre las salidas de STATA

1. En la base de datos *pubexp* puede encontrar información del gasto publico en educación, PIB en millones de dolares de 1980 y población de 34 países.
 - a. Obtenga estadísticas descriptivas e histogramas de todas las variables y estime la siguiente regresión por MCO $ee_i = \beta_0 + \beta_1 gdp_i + u_i$. Comente sobre sus resultados.
 - b. Calcule los residuales de la anterior regresión y con ellos realice un scatter plot entre los residuales elevados al cuadrado y el PIB. ¿existen indicios de heterocedasticidad? ¿por qué?
 - c. Calcule las pruebas de Breusch-Pagan, White, Harvey y Gleijser. Muestre sus resultados ¿que concluye?
 - d. Estime la anterior regresión usando MCG y errores robustos a la heterocedasticidad. Utilice como ponderaciones la población de los países. Muestre sus resultados y comente.
2. El archivo *stockton96* contiene información de 940 observaciones de ventas de vivienda en Stockton. Búsquelo y descarguelo del moodle.
 - a. Utilice MCO para estimar una regresión del precio contra el tamaño de la vivienda en pies cuadrados y la edad de la vivienda en años. Muestre e interprete sus resultados.
 - b. Suponga que usted es dueño de dos casas. Una tiene 1400 pies cuadrados y la otra tiene 1800. Ambas tienen 20 años de edad. ¿que precio estima que tendrá cada una de las viviendas?
 - c. Calcule las pruebas de Breusch-Pagan, White, Harvey y Gleijser. Muestre sus resultados ¿que concluye?
 - d. Estime $\sigma_i^2 = e^{\alpha_1 + \alpha_2 SQFT + \alpha_3 Age}$ y calcule la varianza predicha de esta regresión. Para ello debe usar los errores elevados al cuadrado y transformarlos de acuerdo a la ecuación. Interprete sus resultados.
 - e. Estime el mismo modelo usando MCG. Interprete sus resultados. Con estos resultados ¿que precio obtendría por sus dos casas?
3. Usando STATA simule el siguiente proceso para $n = 10,000$.

$$x_i \sim nid(0, 1) \quad \sigma_{1i}^2 = \sigma_u^2 \mid x_i \mid$$

$$\sigma_{2i}^2 = \sigma_u^2 x_i^2 \text{ y } \sigma_{3i}^2 = \sigma_u^2 i$$

$$u_1 = e_i \sigma_{1i}$$

$$u_2 = e_i \sigma_{2i}$$

$$u_3 = e_i \sigma_{3i} \text{ donde } e_i \sim nid(0, 1)$$

$$y_{1i} = x_i + u_1$$

$$y_{2i} = x_i + u_2 \text{ y } y_{3i} = x_i + u_3$$

- a. Muestre las estadísticas descriptivas de y_i para $i = 1, 2, 3$. Interprete sus resultados.
- b. Corra las tres regresiones $y_{ji} = \alpha_0 + \alpha_1 x_i + u_j$ para $j = 1, 2, 3$ y calcule los residuales de cada una. Muestre e interprete sus resultados.
- c. Construya un scatter plot de u_1 y u_2 contra x_i . Y de u_3 contra i . Muestre e interprete sus resultados.

- d. Calcule las pruebas de White y Breusch-Pagan sobre los tres modelos del segundo literal. Muestre e interprete sus resultados.

4. En un estudio sobre los determinantes del salario en una ciudad determinada, se estima el siguiente modelo de regresión:

$$\text{Salario} = b_0 + b_1 \text{Educ} + b_2 \text{Exper} + u_i$$

Sin embargo, se sospecha que una variable relevante, la ocupación (tipo de empleo que desempeña la persona), ha sido omitida del modelo. La omisión de esta variable puede generar un sesgo en la estimación de los coeficientes del modelo.

- a. Explica cómo la omisión de la variable "ocupación" puede generar sesgo en los estimadores de β_1 y β_2 . En particular, ¿cómo podría este sesgo depender de las correlaciones entre "ocupación" y las variables incluidas en el modelo (educación y experiencia)?
- b. Bajo el supuesto de que la ocupación está correlacionada positivamente con el nivel educativo y con la experiencia laboral, ¿qué impacto esperas que tenga la omisión de esta variable sobre los coeficientes estimados para β_1 y β_2 ? ¿Serán sobreestimados o subestimados?
- c. Ahora, supón que la variable ocupación no está correlacionada con la educación, pero sí con la experiencia. ¿Cambiaría el signo y magnitud del sesgo en este caso? Justifica tu respuesta.
- d. Diseña una estrategia para abordar el problema de la omisión de variables en este contexto. ¿Cómo podrías modificar el modelo o los datos para mejorar la precisión de las estimaciones y evitar el sesgo generado por la omisión de la variable ocupación?

5. Se desea estudiar la relación entre el precio de un bien y la cantidad demandada del mismo en una ciudad. El investigador propone el siguiente modelo de regresión:

$$\text{CantidadDemandada}_i = a_0 + a_1 \text{Precio}_i + u_i$$

Sin embargo, surge una preocupación: el precio no es exógeno, ya que el precio y la cantidad demandada se determinan simultáneamente en el mercado. Esto podría generar un sesgo de simultaneidad en la estimación de a_1 el efecto del precio sobre la cantidad demandada.

- a. Explica por qué el problema de simultaneidad puede generar un sesgo en la estimación de a_1 . ¿Qué condiciones deben cumplirse para que el precio sea considerado endógeno en este contexto?
- b. Supón que existe una variable no observada, como el ingreso promedio de los consumidores, que afecta tanto el precio como la cantidad demandada. ¿Cómo afectaría esto a la estimación de a_1 ? ¿Esperas que el sesgo sea positivo o negativo? Justifica tu respuesta.
- c. Ahora, supón que el bien estudiado es un bien de lujo. ¿Cómo podría esto cambiar el análisis del sesgo de simultaneidad, considerando la elasticidad-precio de la demanda en este tipo de bienes?