

Econometría II, Universidad de los Andes

Taller 1

Danilo Aristizabal y Jorge Montero

Fecha de entrega: 14 de Junio 2024

Reglas de juego:

- Las pautas generales de entrega de las tareas se encuentran en Bloque Neón.
- Esta actividad se debe realizar en grupos de dos (2) personas. Las respuestas las debe subir una persona del grupo, indicando el nombre, código y aclarar si se está en el pensum de 3 o 4 créditos para cada una de las personas que hacen parte del grupo. No mencionar esta información o entregar información incompleta se penalizará con un punto (1.0) sobre la nota final del taller para todos los miembros del grupo.
- Para la parte 01 de esta actividad, adjunte uno o varios archivos con su explicación ESCRITA A MANO. Para la parte 02, incluya archivos .do (Stata) con los que obtuvieron sus resultados – debe ser posible replicar sus resultados – y un archivo de texto (.doc, .docx, PDF) con sus respuestas a las preguntas. Respuestas que no sigan esas pautas o respuestas ilegibles no serán tenidas en cuenta.
- La fecha límite para entregar el trabajo es el viernes 14 de junio al inicio de la complementaria 7:00:00 am. Entregas posteriores de esa hora serán penalizadas con una nota máxima de 3 hasta las 10:00:00 am del viernes 14 de junio. Después de esta hora, la actividad no estará disponible y la nota de la actividad será 0.0.
- Muestre su trabajo (respuestas sin procedimiento/explicación no son válidas)

Parte teórica (entregar respuestas escritas a mano)

1. Variables instrumentales

Existen varias formas para solucionar el problema de endogeneidad. Una opción es el uso de **variables instrumentales**. Este tipo de aproximación identifica β a partir de variables que afectan el comportamiento de las variables endógenas y no están relacionadas con el error.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i = X_i' \beta + u_i$$

Suponga que en el modelo de regresión anterior, $E(u_i|X_{1i}) = E(u_i|X_{2i}) = E(u_i) = 0$ y $E(u_i|X_{3i}) = a \neq 0$, usted tiene un instrumento Z_i tal que $Cov(Z_i, X_{3i}) \neq 0$, $Cov(Z_i, X_{1i}) < \infty$, y $E(u_i|Z_i) = 0$

- (a) Sea $Z_i = (1, X_{1i}, X_{2i}, Z_i)'$. Muestre que, bajo los supuestos anteriores, $E(Z_i u_i) = 0$
- (b) Usando la condición $E(Z_i u_i) = 0$ y los supuestos adicionales sobre Z_i , muestre que β se encuentra identificado como

$$\beta = [E(Z_i X_i')]^{-1} E(Z_i Y_i) \quad (1)$$

- (c) Una forma de solucionar el problema de endogeneidad es a través del estimador de variables instrumentales (IV), que es el análogo muestral de β derivado de la ecuación (1),

$$\hat{\beta}_{IV} = \left(\frac{1}{n} \sum_{i=1}^n Z_i X_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i Y_i$$

Muestre que, bajo los supuestos mencionados, $\hat{\beta}_{IV}$ es un estimador insesgado de β . Si requiere hacer supuestos adicionales mencínelos explícitamente.

- (d) Suponga que $u_i | Z_i \sim N(0, \sigma^2)$. Encuentre la distribución muestral de $\hat{\beta}_{IV}$.

2. Métodos de estimación

Suponga que tiene una muestra aleatoria i.i.d X_1, X_2, \dots, X_n tal que $X_i \sim \text{Poisson}(\lambda)$. Usando esta información, responda las siguientes preguntas.

- (a) Proponga un estimador de λ por el método de momentos.
- (b) Plantee la función de verosimilitud y la función de log verosimilitud del parámetro λ dada la muestra X_1, \dots, X_n
- (c) Muestre que el estimador de máxima verosimilitud del parámetro λ es $\hat{\lambda} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$
- (d) ¿Es el estimador de λ , $\hat{\lambda}$, insesgado?
- (e) Muestre que $\text{Var}(\hat{\lambda}) = \frac{\lambda}{n}$

3. Endogeneidad [PREGUNTA PARA PERSONAS QUE SE ENCUENTRAN EN EL PENSUM DE 4 CREDITOS]

- (a) Usted está interesado en el efecto del tamaño del aula (número de estudiantes) sobre el desempeño promedio de los estudiantes. Considere el siguiente modelo

$$\text{Desempeño}_i = \beta_0 + \beta_1 \text{Tamaño}_i + \beta_2 X_i + u_i \quad (2)$$

Donde X_i son controles sociodemográficos del estudiante como su edad, grado, sexo, y educación de los padres.

¿Por qué razón podría ser Tamaño_i endógena en este modelo?

- (b) Usted está interesado en ver el efecto de la migración sobre el bienestar del individuo. Para esto estima el siguiente modelo:

$$\text{Bienestar}_i = \beta_0 + \beta_1 \text{migración}_i + \beta_2 X_i + u_i \quad (3)$$

Donde X_i es un vector de controles sociodemográficos. ¿Es esta estrategia apropiada para ver el efecto de migración sobre bienestar? ¿Por qué?

Parte práctica

1. Ecuación de Mincer (Corte transversal repetido)

En este ejercicio vamos a estudiar el sesgo de especificación, variables proxy y valores atípicos desde un punto de vista práctico. Para esto vamos a usar la información de la Gran Encuesta Integrada de Hogares (GEIH) para los meses de enero a diciembre del año 2017 de las personas ocupadas. Usaremos la información de salarios por hora, años de educación y edad de las 23 ciudades principales del país.

- (a) Cargue la base de datos 20240607_Tarea.01.dta. Genere la variable *lsalarios* definida como el logaritmo natural de la variable *salarios* por hora. Muestre un histograma de *lsalarios*. En el histograma, utilice la opción *xline* para agregar una línea vertical en el logaritmo del salario mínimo legal por hora (el salario mínimo mensual en Colombia en 2017 era 737,717). Basado en la evidencia gráfica ¿En qué rango de valores se concentra la distribución de salarios de los ocupados en las 23 ciudades principales?
- (b) Estime el modelo de regresión lineal simple, $Y = \beta_0 + \beta_1 X + u$: Uno en el que la variable dependiente es *salarios* por hora y la variable independiente es *esc* y otro en el que la variable dependiente es *lsalarios* y la variable independiente es *esc*. Interprete los coeficientes estimados asociados a años de escolaridad.
- (c) Presente las estadísticas descriptivas de salarios por hora. ¿Entre qué rango de valores se encuentran los salarios por hora? Elimine los datos atípicos de salario por hora, para esto elimine el percentil 1 y 99 de la variable *salario* por hora (tenga en cuenta los factores de expansión: `[fex_c.2011]`). Ahora ¿entre qué rango se encuentran los salarios por hora?
- (d) La literatura económica sugiere que el modelo correcto que se debería estimar para ver la relación de salarios y escolaridad es:

$$\text{salarios}_x \text{hora}_i = \gamma_0 + \gamma_1 \text{esc}_i + \gamma_2 \text{exper}_i + \gamma_3 \text{exper}_i^2 + u \quad (4)$$

¿Qué implicaciones tendría omitir la variable experiencia sobre $\hat{\gamma}_1$? Justifique.

- (e) Sin embargo, experiencia es una variable que no tenemos en nuestra base de datos. Una proxy de experiencia es la edad del individuo menos los años de escolaridad menos 6 (edad a la que un niño/a empieza la educación primaria). Construya la variable experiencia y estime y presente sus resultados de la siguiente regresión

$$\text{salarios}_x \text{hora}_i = \gamma_0 + \gamma_1 \text{esc}_i + \gamma_2 \text{exper}_i + \gamma_3 \text{exper}_i^2 + u \quad (5)$$

¿Usando $\hat{\gamma}_2$ y $\hat{\gamma}_3$, y el valor promedio de experiencia, ¿cuál es el rendimiento de un año más de experiencia para un individuo promedio de su muestra?