Econometría I

4. Inferencia en el MRLS
Distribución de Muestreo de los Estimadores de MCO e Intervalos de Confianza

José Carlos Tello

Universidad Americana

Clase 5 — 25 de agosto
Segundo Semestre 2023

Supuestos Necesarios para Insesgadez de Estimadores de MCO

- Dado que sabemos como "estimar" los parámetros del modelo lineal simple a partir de una muestra, ahora retomamos el modelo poblacional para "decir algo más" de los estimadores de MCO.
- Vamos a enumerar 4 supuestos necesarios para encontrar la media poblacional de los estimadores de MCO.

  S1. Linealidad en los parámetros (asumido en definición FRP).
  S2. Se cuenta con una muestra aleatoria.
  S3. Variable explicativa no constante (¿obvio?).
  S4. Media condicional de los errores u dado X es igual a cero (asumido en definición FRP).

Supuestos Necesarios para Insesgadez de Estimadores de MCO

S1. Linealidad en los parámetros.

- El modelo debe ser lineal en los parámetros, es decir, puede escribirse de la siguiente manera: $y = \beta_0 + \beta_1 x + u$.
- Tanto "y" como "x" pueden ser transformaciones lineales o no lineales de las variables originales (como $\ln x$) pero los coeficientes a estimarse deben ser únicos.

S2. Muestreo aleatorio

- La mayoría de muestras de corte transversal pueden cumplir sin contratiempo este supuesto. No sucede lo mismo con las series de tiempo.
- Para las distintas observaciones, (x_i, y_i) son independientes e idénticamente distribuidas.

Supuestos de MCO

S3. $\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$.

- En caso que falle no sería posible calcular la pendiente del modelo lineal simple (la regresión sería una recta vertical).
- Necesitamos que la variable explicativa varíe para que nuestro análisis sea de interés.

S4. E[$u|x$]=0.

- Dado el S2, el S4 implica $E[u_i|x_i] = 0$ para todo $i = 1, ..., n$.
- También sabemos que este supuesto implica que $E[u] = 0$ y que la covarianza entre x y u es nula.
- Además, este supuesto implica que el componente sistémico y el no sistémico no tengan ninguna relación lineal entre ellos.

Esperanza del Estimador $\hat{\beta}_1$

- Dado el **S1** podemos reexpresar la fórmula del estimador de β_1

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\hat{\beta}_1(x_i - \bar{x}) + u_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Ordenando los términos se obtiene

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Por el supuesto **S2**, **S3** y **S4** se deduce que:

$$E[\hat{\beta}_1] = \beta_1 \tag{1}$$

Esperanza del Estimador $\hat{\beta}_1$ y $\hat{\beta}_0$

- Podemos afirmar que el estimador $\hat{\beta}_1$ es **insesgado**.
- ¿Qué quiere decir? si tuvieramos muchas muestras, del tipo $\{(x_i, y_i) : i = 1, ..., n\}$, los valores calculados de $\hat{\beta}_1$ **en promedio coincidirá** con β_1.
- Dado el estimador $\hat{\beta}_0$, y si se cumple los 4 supuestos se puede demostrar:

$$E[\hat{\beta}_0] = \beta_0 \tag{2}$$

- El estimador del intercepto también es **insesgado**.

Distribución de Muestreo de los Estimadores de MCO

Simulación de Montecarlo

- También conocido como método de Montecarlo, consiste en aproximar el comportamiento de un estadístico o estimador o algún problema matemático al permitir la realización de experimentos con muestreos de números pseudo-aleatorios en una computadora.
- Se trata de un ejercicio teórico pues la aproximación del comportamiento de un estimador consiste en simular la extracción de varias muestras aleatorias de la población.
- El nombre del método hace referencia al Casino de Montecarlo por ser la capital del juego de azar.

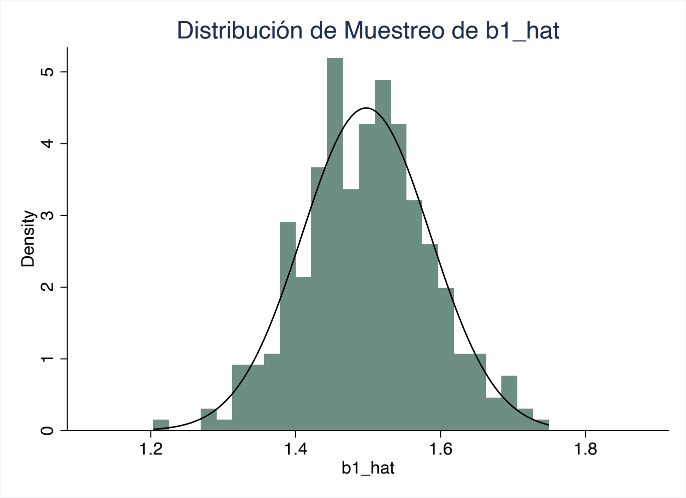
Distribución de Muestreo de los Estimadores de MCO

Simulación de Monte Carlo

- Asumamos que el modelo poblacional es $y = 2 + 1{,}5x + u$,
  donde $u \sim U(-2, 2)$ y $x \sim Gamma(2, 1)$.
- Con esta información podemos extraer (generar) 300 muestras aleatorias, cada una con 100 observaciones.
- Si estimamos por MCO la recta de regresión para cada muestra aleatoria, ¿cómo se comportará esta recta de regresión respecto a la función de regresión poblacional?

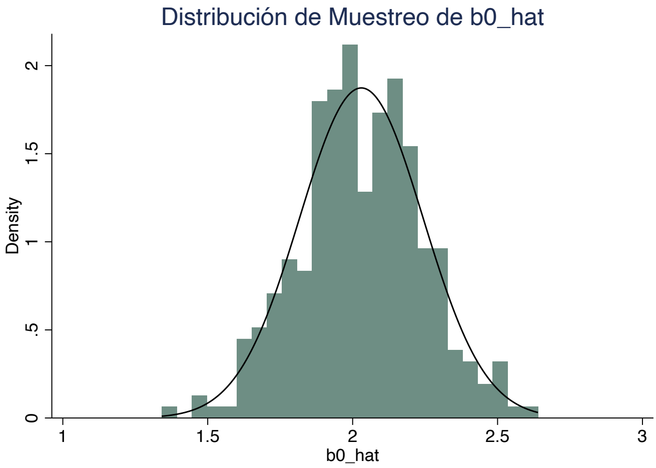
Distribución de Muestreo de los Estimadores de MCO

Simulación de Monte Carlo



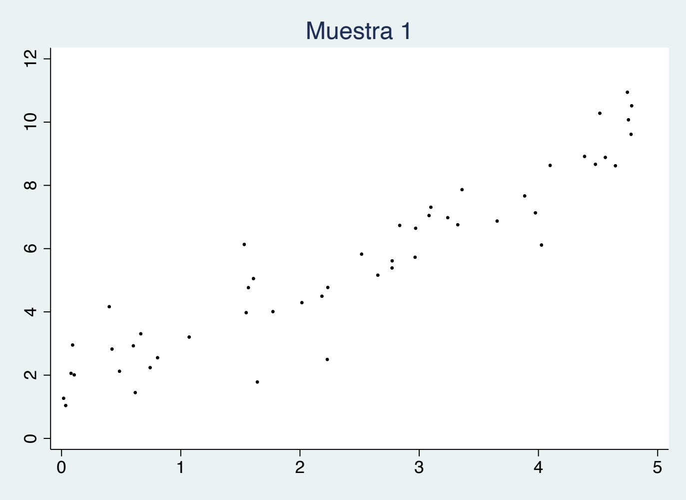
Distribución de Muestreo de los Estimadores de MCO

Simulación de Monte Carlo



Distribución de Muestreo de los Estimadores de MCO

Simulación de Monte Carlo



Distribución Muestral de los Estimadores de MCO

- Cabe destacar que no es necesario hacer supuestos sobre la varianza de los errores ni su distribución para mostrar la distribución muestral de los estimadores.
- En muestra grande (por ejemplo, muestras con más de 100 observaciones) la distribución de los estimadores de MCO se aproximan a una distribución normal.
- Este resultado importante tiene como soporte el Teorema Central del Límite (TCL)

TCL. $y_1, y_2, ..., y_n$ son i.i.d. con $E[y_i] = \mu_y$ y $V[y_i] = \sigma_y^2$ (donde $0 < \sigma_y^2 < \infty$). Si n es muy grande entonces $\frac{(\bar{y}-\mu_y)}{\sqrt{\sigma_{\bar{y}}^2}}$, donde $\sigma_{\bar{y}}^2 = \frac{\sigma_y^2}{n}$ se aproxima a la distribución normal estándar (o sea una normal con media igual a cero y varianza igual a 1).

Distribución Muestral de los Estimadores de MCO

- Volvamos a la ecuación donde se muestra la diferencia entre el estimador $\hat{\beta}_1$ y el parámetro poblacional β_1.

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Multiplicar el numerador y denominador del segundo término de la derecha por $\frac{1}{n}$ no altera la igualdad y se obtiene

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Luego,

$$\hat{\beta}_1 - \beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Distribución Muestral de los Estimadores de MCO

- Se denomina $v_i = (x_i - \bar{x})u_i$.
- Luego se puede deducir que la media muestral $\bar{v} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i$ sería una buena aproximación de $E[v_i]$.
- Además, dado el $S4$ de MCO sabemos que $E[v_i] = 0$.
- Por el $S2$, v_i es idéntica e independientemente distribuida.
- Por tanto, la varianza poblacional de v_i es $\sigma_v^2 \equiv E[(x_i - E[x])u_i)^2]$.
- Ahora se puede ver $\bar{v}$ satisface **Teorema Central del Límite**.

Distribución Muestral de los Estimadores de MCO

- Por lo tanto, $\frac{(\bar{v}-0)}{\sqrt{\sigma_{\bar{v}}^2}}$, en muestras grandes, está distribuida como $N(0,1)$ donde $\sigma_{\bar{v}}^2 = \frac{\sigma_v^2}{n}$.
- También podemos afirmar que la distribución de $\bar{v}$ está bien aproximada por $N(0, \sigma_v^2/n)$
- Dado que el denominador $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ es una buena aproximación de $V[x_i]$ (pues n o $n-1$ es irrelevante en muestras grandes).
- Combinando los resultados anteriores podemos afirmar lo siguiente

$$\hat{\beta}_1 - \beta_1 \cong \frac{\bar{v}}{V[x_i]}$$

- Finalmente, la distribución en muestras grandes de $\hat{\beta}_1 - \beta_1$ es $N\left(0, \frac{\sigma_v^2}{n(V[x_i])^2}\right)$

Distribución Muestral de los Estimadores de MCO

- La aproximación normal del estimador $\hat{\beta}_1$ es una herramienta poderosa.
- Con este resultado se puede realizar inferencias sobre el valor poblacional de β_1 sin ser necesario supuestos sobre distribución de los errores.
- Recordemos que $\hat{\beta}_1$ difiere de una muestra a otra.
- Acabamos de ver que $\hat{\beta}_1 - \beta_1$ presenta una distribución normal, en muestras grandes, con media igual a cero y varianza igual a $\frac{\sigma_v^2}{n(V[x_i])^2}$.
- Sin embargo, la varianza en muestras grandes no puede ser utilizada en el estudio empírico, para esto es necesario contar con un estimador de $\sigma_v^2 \equiv E[((x_i - E[x])u_i)^2]$ pues no se tienen los valores u_i sino de $\hat{u}_i$.

Inferencia en el MRLS

Estimador de la varianza de β_1 y $\hat{\beta}_0$

- Un estimador insesgado y consistente de la varianza de $\hat{\beta}_1$ es

$$\widehat{V}[\hat{\beta}_1] = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

- El estimador insesgado y consistente de la varianza de $\hat{\beta}_0$ es

$$\widehat{V}[\hat{\beta}_0] = \frac{1}{n} \times \left(\frac{\frac{1}{n-2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} \bar{x}^2 + \frac{\frac{1}{n-2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)$$

Inferencia en el MRLS

Estimador del error estándar de β_1 y $\hat{\beta}_0$

- El error estándar (ee) de un estimador es la raíz cuadrada de su varianza estimada.
- Por lo tanto,

$$ee(\hat{\beta}_1) = \sqrt{\widehat{V}[\hat{\beta}_1]}$$

$$ee(\hat{\beta}_0) = \sqrt{\widehat{V}[\hat{\beta}_0]}$$

- Afortunadamente, los programas estadísticos calculan estas fórmulas.

Inferencia en el MRLS

Intervalos de Confianza en muestra grande

- Es un intervalo que presenta una probabilidad del 95 % de contener el verdadero valor de β_1, en otras palabras, en el 95 % **de las posibles muestras que podrían ser seleccionadas**, el intervalo de confianza contendrá el verdadero valor de β_1.
- Debido a que este intervalo contiene el valor real en el 95 % de todas las potenciales muestras, se dice que tiene un nivel de confianza de 95 %.

$$\left[\hat{\beta}_1 - 1{,}96 \times ee(\hat{\beta}_1), \quad \hat{\beta}_1 + 1{,}96 \times ee(\hat{\beta}_1)\right]$$

- Los límites del intervalo no son constantes, los valores dependen de la muestra disponible.
- De la misma manera se construye y analiza el intervalo de confianza para β_0.

$$\left[\hat{\beta}_0 - 1{,}96 \times ee(\hat{\beta}_0), \quad \hat{\beta}_0 + 1{,}96 \times ee(\hat{\beta}_0)\right]$$