# Trabajo Práctico de Econometría

Martín González-Rozada
(Fecha de Entrega: lunes 18 de marzo de 2024).

1 de diciembre de 2023

## 1. Inferencia en el modelo de regresión lineal

El objetivo de este ejercicio es estudiar la inferencia asintótica en la práctica cuando no se asume normalidad de los errores.

Considere el siguiente modelo:

$$y = \alpha + \beta x + u \tag{1}$$

donde $x$ es un escalar, $\alpha = 3$, $\beta = 1$ y $x \sim N(2, 1)$. Analice cómo cambia la inferencia estadística si no se asume normalidad en el término de error. El *benchmark* de comparación será $u \sim N(0, 12)$, es decir, la varianza de los errores es igual 12.

Para hacer el ejercicio vamos a analizar los siguientes casos:

i La distribución de los errores es uniforme.

ii La distribución de los errores es asimétrica, como una transformación lineal de la distribución Beta o con una transformación de una distribución Bernoulli.

iii La distribución de los errores tiene "colas gruesas" como la distribución t-Student.

iv La distribución *benchmark*.

v La distribución *benchmark* con una observación que es un *outlier*.

Todas las simulaciones se basan en 2000 replicaciones.

1. Si el modelo verdadero cumple que $u \sim_{i.i.d.} U[-6, 6]$, se le pide:

   (a) Genere 2000 muestras de $n = 10$ observaciones. Para cada muestra estime por MCC los parámetros del modelo y contraste la hipótesis nula $H_0 : \beta = 1$ versus una hipótesis alternativa $H_1 : \beta \neq 1$ con un test $t$ y un nivel de significación del 5 %. Luego, reporte el tamaño del test (la proporción de las 2000 muestras en las que se rechaza la hipótesis nula.)

   (b) Repita el inciso anterior con $n = 20$

   (c) Repita el inciso anterior con $n = 100$

(d) Repita el inciso anterior con $n = 200$

(e) Repita el inciso anterior con $n = 500$

(f) Repita el inciso anterior con $n = 1000$

(g) Repita el inciso anterior con $n = 5000$

(h) Repita el inciso anterior con $n = 10000$

(i) Resuma los pasos anteriores en una tabla para mostrar el tamaño del test en cada caso y reportar qué ocurre a medida que el tamaño de muestra va aumentando.

(j) A partir de los resultados obtenidos, diga a qué conclusiones puede arribar respecto a la inferencia en el modelo en la práctica.

2. Repita los incisos anteriores para $u \sim_{i.i.d.} \sqrt{\frac{2352}{5}} \cdot \left(Beta(2,5) - \frac{2}{7}\right)$.

3. Repita los incisos anteriores para $u \sim_{i.i.d.} t_{2,1818}$.

4. Repita los incisos anteriores para $u \sim_{i.i.d.} \sqrt{75} \cdot (Be(0{,}8) - 0{,}8)$.

5. Repita los incisos anteriores para $u \sim_{i.i.d.} N(0, 12)$.

6. Repita los incisos anteriores para $u \sim_{i.i.d.} N(0, 12)$ cuando la primera observación del error toma el valor $u_1 = 500$.

7. Comparar cómo difiere la velocidad de convergencia en cada caso y con respecto a tener errores normales.

Notar que en todos los modelos los errores tienen la misma varianza.

## 2. Propiedades de muestra finita de FGLS

Considere el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, .., 5N; \tag{2}$$

con $\beta_0 = -3$, $\beta_1 = 0{,}8$, $u_j \sim N(0, \Omega \otimes I_{N\times N})$, $\Omega = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 \\ 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 0 & 36 \end{bmatrix}$ y $x_j \sim U[1\ \ 50]$

1. Genere 5000 muestras de $5N = 5$ observaciones de corte transversal a partir del modelo (2).

   a) Para cada muestra estime por FGLS los parámetros del modelo y realice un test de hipótesis para contrastar que $H_0 : \beta_1 = 0{,}8$. Reporte **tamaño del test** al 1 % y 5 % y el **poder del test** cuando $\beta_1 = 0$ y $\beta_1 = 0{,}4$. Adicionalmente, reporte la media, mediana y desvío estándar de las estimaciones de $\beta_0$ y $\beta_1$.

   b) **BONUS:**[1] Utilice la descomposición de Cholesky para encontrar una matriz $P$ que cumpla que se puede escribir a $\Omega = P \cdot P^{\perp}$, aplique la transformación correspondiente a MCG al modelo (2) y estime el modelo por MCC, comente si cambia algún resultado.

2. Repita el punto anterior (sólo el inciso a)) con $5N = 10$.

3. Repita el punto anterior (sólo el inciso a))con $5N = 30$.

4. Repita el punto anterior (sólo el inciso a))con $5N = 100$.

5. Repita el punto anterior (sólo el inciso a)) con $5N = 200$.

6. Repita el punto anterior (sólo el inciso a)) con $5N = 500$.

7. Describa detalladamente las propiedades de muestra finita de FGLS de acuerdo a lo que observó de los cuatro puntos anteriores. En especial, explique cómo cambia el tamaño y la potencia de los tests a medida que aumenta el tamaño de muestra.

[1] No es obligatorio que haga este inciso.

# 3. Corrección de la matriz de varianzas y covarianzas en presencia de heterocedasticidad, White (1980)

En este ejercicio se le pide evaluar el funcionamiento de la corrección sugerida por White para estimar la matriz de varianzas y de covarianzas de los coeficientes estimados por mímos cuadrados clásicos en muestras finitas. El objetivo final del ejercicio será aprender el cálculo correcto de la matriz de varianzas y covarianzas de los coeficientes estimados por mínimos cuadrados clásicos en presencia de heterocedasticidad. Para esto considere el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sqrt{\nu_i} u_i \quad i = 1, 2, \dots, n \tag{3}$$

Para $n = 20$, $x_1$ se determina como una secuención de 18 puntos espaciados uniformemente entre $-1$ y 1 con puntos extremos dados por $-1,1$ y $1,1$ mientras que $x_2$ son cuantiles de una distribución normal estándar elegidos aleatoriamente. Las observaciones se repiten tres veces para obtener una muestra de $n = 60$, cinco veces para obtener una muestra de $n = 100$, diez veces para obtener una muestra de $n = 200$, veinte veces para obtener una muestra de $n = 400$ y treinta veces para obtener una muestra de $n = 600$. La generación de los datos de la variable dependiente se hace con $\beta_0 = \beta_1 = \beta_2 = 1$. Hay dos diseños:

Diseño 1 $u_i \sim N(0,1)$ y $\nu_i = \exp^{0,25x_{1i}+0,25x_{2i}}$ (normalidad y heterocedasticidad)
Diseño 2 $u_i \sim t_5$ y $\nu_i = \exp^{0,25x_{1i}+0,25x_{2i}}$ (no normalidad y heterocedasticidad)

Todas las simulaciones se basan en $5,000$ replicaciones. En el modelo de regresión lineal (3) la estimación de MCC es,

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \left(X'X\right)^{-1} X'y$$

Donde $X$ es la matriz de variables explicativas de dimensión $n \times 3$ e $y$ es el vector de observaciones de la variable independiente de dimensión $n \times 1$. Bajo heterocedasticidad, la matriz de varianzas y covarianzas de los estimadores de MCC es:
$\text{Var}(\hat{\beta}) = (X'X)^{-1} X'\Omega X \,(X'X)^{-1}$ con $\Omega = \text{diag}\left(\sigma_1^2, \dots, \sigma_n^2\right)$.

Una estimación consistente de esta matriz está dada por la denominada matriz de White (White, 1980)

$$\widehat{\text{Var}(\hat{\beta})} = \left(X'X\right)^{-1} X'\hat{\Omega}X \left(X'X\right)^{-1}$$

con $\hat{\Omega} = \text{diag}\left(\hat{u}_1^2, \dots, \hat{u}_n^2\right)$ y $\hat{u}$ es el vector de dimensión $n \times 1$ de residuos de la estimación por MCC. Se le pide:

(a) Para cada uno de los diseños 1 y 2 genere $5,000$ muestras de $n = 20$ observaciones a partir del modelo (3). Para cada muestra estime por MCC los parámetros del modelo y reporte el sesgo relativo de la estimación de las varianzas de $\tilde{\beta}_0$, $\tilde{\beta}_1$ y $\tilde{\beta}_2$ y los sesgos relativos totales. El sesgo relativo $(b_0, b_1, b_2)$ se define como el promedio, a lo largo de las 5000 simulaciones, de la varianza estimada menos la varianza verdadera divididas por la varianza verdadera.El sesgo relativo total es la suma del valor absoluto de $(b_0, b_1, b_2)$. El sesgo relativo total es una medida del sesgo agregdo de las tres varianzas.

(b) Repita el punto anterior con $n = 60$.

(c) Repita el punto anterior con $n = 100$.

(d) Repita el punto anterior con $n = 200$.

(e) Repita el punto anterior con $n = 400$.

(f) Repita el punto anterior con $n = 600$.

(g) Repita los puntos (a) a (f) pero ahora construya la estimación de la matriz de White usando una matriz diagonal con los errores verdaderos elevados al cuadrado en lugar de utilizar la matriz diagonal con los residuos elevados al cuadrado.

(h) Describa detalladamente las propiedades de muestra finita de la estimación de las varianzas de los coeficientes de MCC robustas ante la presencia de heterocedasticidad (con el procedimiento de White) de acuerdo a lo que observado en los puntos anteriores.

## Instrucciones para la entrega:

- El trabajo es **individual**.
- Usar como **seed en el archivo .do** los últimos 4 números de su documento para que los resultados sean replicables.
- El trabajo se puede hacer en Stata, R, Python o Matlab.
- Hay que entregar un archivo .pdf con el reporte y adjuntar el código (**.do**, **.r**, o **.m**) o entregar el **.log** *file* aparte.
- Los archivos deberán tener el apellido de quien entregue y la seed utilizada, es decir, si Caravello usó el seed 1668 los archivos se llamarán **Caravello1668.pdf, Caravello1668.do, Caravello1668.log, Caravello1668.m o Caravello1668.r**
- En el código deberán hacer **comentarios breves** para que se entienda el **procedimiento**.
- La entrega se realizará **vía Campus Virtual hasta el día lunes 18 de marzo de 2024 a las 11.59pm**.