Relacion 3
Heteroscedasticidad

Ejercicio 0.1. Considera el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

donde $i = 1, ..., N$.

1. Escribe el modelo en álgebra matricial, para todas las observaciones, denominando ε al vector de perturbaciones del modelo.
2. Escribe la matriz de covarianzas de ε. ¿Qué elemento es el $(2, 2)$ de esta matriz? ¿Cuál es el elemento $(1, N)$ de esta matriz?
3. ¿Cómo será esta matriz si suponemos que $var(\varepsilon_i) = 2x_i$?

Ejercicio 0.2. La estimación MCO del modelo de regresión

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

ha sido $\widehat{\beta}_0 = 1$ y $\widehat{\beta}_1 = 0.5$. Suponiendo que tanto las observaciones de x y de y se multiplican por 10, obtener la nueva estimación. ¿Supone esto que el modelo es más heterocedástico?

Ejercicio 0.3. Considera el modelo

$$C_i = \beta_0 + \beta_1 PNB_i + \beta_2 D_i + \varepsilon_i,$$

donde C_i es el consumo del país i, PNB_i es el Producto nacional Bruto del país i, y D_i son los gastos en defensa del país i. Para estimar los parámetros β_0, β_1, y β_2, se estiman los siguientes modelos:

$$\begin{aligned}
\widehat{C}_i &= \underset{(2.7)}{26.19} + \underset{(0.006)}{0.62}\, PNB_i - \underset{(0.07)}{0.44} D_i,\ R^2 = 0.9 \\
\frac{\widehat{C}_i}{PNB_i} &= \underset{(2.2)}{25.90} \frac{1}{PNB_i} + \underset{(0.006)}{0.62} - \underset{(0.06)}{0.43} \frac{D_i}{PNB_i},\ R^2 = 0.8
\end{aligned}$$

¿Qué supuesto sobre los errores habrán hecho los autores de las estimaciones anteriores?

Ejercicio 0.4. Considera el modelo

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

donde los errores están independientemente distribuidos, con media 0 y varianza σ_i^2.

1. Explica cómo estimar los parámetros del modelo, suponiendo que $\sigma_i^2 = \lambda x_i$, siendo λ una constante positiva de valor conocido.

2. Contesta a la pregunta anterior suponiendo ahora que $\sigma_i^2 = a + bx_i$, siendo a y b constantes positivas de valor conocido.

3. Contesta a las dos preguntas anteriores, suponiendo que λ, a y b son ahora constantes positivas desconocidas.

Ejercicio 0.5. Suponiendo el modelo

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 w_i + \varepsilon_i,$$

donde $\varepsilon_i \sim N(0, \sigma^2)$. Un investigador cree erróneamente que $var(\varepsilon_i) = 2x_i^2$, por lo que transforma el modelo y aplica MCO al modelo transformado. ¿Qué propiedades tendrá este estimador?

Ejercicio 0.6. Un económetra trata de estimar el consumo regional en función de la renta. Para ello toma datos de 10 regiones y propone el siguiente modelo

$$C_i = \beta_0 + \beta_1 R_i + \epsilon_i,$$

donde C_i y R_i son el consumo y la renta medios de cada región. El investigador supone que el consumo por individuo tiene una varianza constante σ^2 y aplica MCO al modelo propuesto. ¿Ha realizado bien la estimación? En caso de que tu respuesta sea negativa, propón una alternativa.

Ejercicio 0.7. Considere el modelo $y_i = \beta x_i + \varepsilon_i$, con $var(\varepsilon_i) = (kx_i)^2$. Prueba que el estimador MCG de β es igual al promedio muestral del cociente $\frac{y_i}{x_i}$. Halle su varianza.

Ejercicio 0.8. Considere el siguiente modelo de regresión simple:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2) \\ i &= 1, ..., n \end{aligned}$$

Utilizando una muestra, (y_i, x_i), de datos agregados,

$$\begin{array}{ll} y_1 = Y_1 & x_1 = X_1 \\ y_2 = Y_1 + Y_2 & x_2 = X_1 + X_2 \\ y_3 = Y_1 + Y_2 + Y_3 & x_3 = X_1 + X_2 + X_3 \\ \cdots\cdots\cdots & \cdots\cdots\cdots \\ y_n = Y_1 + ... + Y_n & x_n = X_1 + ... + X_n \end{array}$$

halle el estimador ELIO de β_1.

Ejercicio 0.9. Sea el siguiente modelo lineal sin término constante y un solo regresor:

$$\begin{aligned} y_i &= \beta x_i + \varepsilon_i \\ E(\varepsilon_i) &= 0, V(\varepsilon_i) = \sigma^2 x_i \end{aligned}$$

1. Obtener la expresión analítica del estimador MCG.
2. ¿Qué ocurriría si se estimase el modelo por MCO y se utilizase $s^2(X'X)^{-1}$ como matriz de varianzas-covarianzas estimada del estimador MCO? (s^2 es el estimador MCO de la varianza de las perturbaciones, $s^2 = \frac{e'e}{N-(k+1)}$).

Ejercicio 0.10. Para estimar la relación entre las ventas (variable y_i) y los gastos en publicidad (variable x_i) de una cadena de tiendas se propone el siguiente modelo lineal con constante para N observaciones, cuya expresión matricial es:

$$Y = X\beta + \varepsilon, \qquad \text{donde } \varepsilon \sim N(0, V)$$

Conteste a las siguientes preguntas, justificando detalladamente sus respuestas:

1. Si suponemos $V = 9I$, siendo I la matriz identidad $(N \times N)$, calcule la esperanza y la varianza de la estimación MCO de β. ¿Cree que hay homocedasticidad?
2. Si suponemos ahora:

$$V = \begin{pmatrix} 9x_1^2 & & 0 \\ & \ddots & \\ 0 & & 9x_n^2 \end{pmatrix},$$

   calcule la esperanza y la varianza de la estimación MCO de β. ¿Habrá heterocedasticidad? Si es así, halle el estimador MCG de β y calcule su esperanza y su varianza.

Ejercicio 0.11. Dado el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$$

donde $E(\varepsilon_i) = 0$ y $Var(\varepsilon_i) = \sigma^2 X_i^2$

1. Halle los estimadores MCO y MCG de β_1 y β_2, ¿son insesgados? Demuéstrelo.

1 Algunas cuestiones de Verdadero o Falso

Cuestión 1.1. En el modelo lineal clásico, el supuesto de normalidad no es necesario si el objetivo es meramente la estimación.

Cuestión 1.2. En un modelo de regresión lineal clásico la suma de residuos mínimo cuadráticos siempre es cero.

Cuestión 1.3. Bajo los supuestos clásicos, los estimadores MCO son ELIO, independientemente de que las perturbaciones del modelo posean una distribución normal o no.

Cuestión 1.4. En el modelo lineal clásico con constante el coeficiente de determinación corregido es siempre mayor que cero.

Cuestión 1.5. En el modelo de regresión lineal clásico:

$$\begin{aligned} Y &= X\beta + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I) \end{aligned}$$

las distribuciones de los residuos MCO y de las perturbaciones coinciden.

Cuestión 1.6. La estimación MCO de β_1 en dos modelos distintos de la forma:

$$\begin{aligned} Y_i &= \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \\ Y_i &= \beta_1 X_{1i} + \eta_i \end{aligned}$$

es igual siempre y cuando X_1 y X_2 sean ortogonales, es decir, cuando $\sum_{i=1}^{N} X_{1i} X_{2i} = 0$.

Cuestión 1.7. Considere dos modelos del tipo $[M1]$ $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$, y $[M2]$ $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$, entonces si $[M1]$ es un modelo correctamente especificado, el estimador MCO de β_1 en el modelo $[M2]$ es siempre sesgado. Sin embargo, si $[M2]$ es un modelo correctamente especificado, el estimador MCO de β_1 en el modelo $[M1]$ es insesgado.

Cuestión 1.8. Si en un modelo del tipo $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$ se cumplen todos los supuestos del modelo lineal clásico, la varianza estimada del estimador MCO de β_1 será tanto menor cuanto mayor sea la varianza muestral de $X_{i.}$

Cuestión 1.9. Suponga que se desea desarrollar un modelo que explique la conducta del ahorro agregado como una función de los tipos de interés. En ese caso, será mejor obtener una muestra correspondiente a un período de tipos de interés fluctuantes que otra correspondiente a un período de tipos de interés estables.

Cuestión 1.10. Dado el modelo lineal $Y = X\beta + \varepsilon$ donde X es una matriz de dimensión $N \times K+1$, el hecho de que rango$(X) < K+1$ implica multicolinealidad fuerte, lo que afecta a la precisión de los estimadores.

Cuestión 1.11. La multicolinealidad fuerte (no exacta) se debe a una mala especificación del modelo.

Cuestión 1.12. La existencia de multicolinealidad aumenta el riesgo de no rechazar hipótesis falsas en los contrastes.

Cuestión 1.13. El factor común a la mayoría de las soluciones a la multicolinealidad es tratar de encontrar un estimador de los parámetros del modelo con menor varianza que el de MCO, posiblemente en el grupo de estimadores sesgados.

Cuestión 1.14. En un modelo de regresión lineal general, el coeficiente de determinación, R^2, nunca puede ser alto si todos los parámetros son individualmente no significativos porque, en este caso, un gran porcentaje de la variación de la variable endógena queda sin explicar y el R^2 será pequeño.

Cuestión 1.15. La llamada *trampa de las ficticias* afecta a las ficticias aditivas, pero nunca a las ficticias multiplicativas.

Cuestión 1.16. Omitir variables relevantes tiene efectos más perjudiciales sobre la estimación que incluir variables irrelevantes.

Cuestión 1.17. En presencia de heterocedasticidad, la estimación MCO de los parámetros del modelo y de su varianza son insesgadas. No obstante, este estimador no es eficiente.

Cuestión 1.18. Si utilizamos datos de corte transversal para estimar un modelo que explique el comportamiento del consumo en función de la renta de los individuos, probablemente los errores serán heterocedásticos, y por tanto los estimadores MCO serán sesgados.

Cuestión 1.19. Cuando no se conoce nada sobre la heterocedasticidad, en base a los resultados de White, está justificado utilizar MCO para estimar los β del modelo, y como estimador consistente de la matriz de varianzas-covarianzas de $\widehat{\beta}$:

$$\widehat{V}(\widehat{\beta}) = (X'X)^{-1}X'diag(e_i^2)X(X'X)^{-1}$$

Cuestión 1.20. Las soluciones a la heterocedasticidad pasan por construir un nuevo estimador, MCG, que mejore la eficiencia del estimador MCO, ya que, en presencia de este problema el estimador MCO sigue siendo insesgado, pero ya no es el de mínima varianza. Este procedimiento requiere, como mínimo, conocer cuál o cuáles son las causas del problema.

Cuestión 1.21. En un modelo donde detectamos heterocedasticidad, el estimador MCGF es siempre preferible al estimador propuesto por White.

Cuestión 1.22. En modelos heteroscedásticos los estadísticos "t" y "F" habituales ya no siguen una distribución t y F, respectivamente.