ECONOMETRÍA II
AA. 2022-2023
EJERCICIOS UNIDAD 2.1 – HETEROSCEDASTICIDAD

1.- Una empresa de autobuses desea estimar la demanda de billetes (Y_t) en función de una constante, del precio de los billetes (X_{2t}) y de la calidad del servicio, evaluada a través de los gastos que la empresa realiza para la mejora del mismo (X_{3t}). Se dispone de 50 observaciones, ordenadas en orden creciente según la variable X_{2t} y de la estimación MCO de las siguientes ecuaciones:

$$(1): \quad Y_t = 0.68 - 0.30X_{2t} + 0.33X_{3t} + e_t; \qquad \widehat{\sigma} = 184;$$

$$(2): \quad e_t^2 = \widehat{\beta}_1 + \widehat{\beta}_2 X_{2t} + \widehat{\beta}_3 X_{3t} + \widehat{\beta}_4 X_{2t}^2 + \widehat{\beta}_5 X_{3t}^2 + \widehat{\beta}_6 X_{2t}X_{3t} + \widehat{v}_t; \quad R^2 = 0.053;$$

$$(3): \quad e_t^2 = \widehat{\gamma}_1 + \widehat{\gamma}_2 e_{t-1}^2 + \widehat{w}_t; \qquad R^2 = 0.025;$$

$$(4): \quad \widehat{Y}_t = -0.10 - 0.93X_{2t} + 1.48X_{3t}; \qquad \widehat{\sigma} = 204;$$

$$(5): \quad \widehat{Y}_t = 0.42 - 1.85X_{2t} + 1.66X_{3t}; \qquad \widehat{\sigma} = 150.$$

Las tres primeras ecuaciones se estimaron con las 50 observaciones; la cuarta se estimó con las 20 primeras observaciones y la quinta con las 20 últimas (odenadas según la variable X_{2t}). Supondremos que los errores del modelo son normales y que las aproximaciones asintóticas son válidas.

a) Utilice el contraste de White para contrastar el supuesto de homocedasticidad de los errores en la ecuación (1).

b) Contraste el supuesto de homocedasticidad en el modelo estimado en la ecuación (1) con el contraste de Goldfeld-Quandt, eliminando las 10 observaciones centrales.

2. Se quiere estudiar si los beneficios de las empresas dependen del gasto en I+D que realizan. Para estimar la ecuación

$$beneficios_t = \beta_1 + \beta_2 ventas_t + \beta_3 gID_t + u_t \qquad (1)$$

se dispone de una muestra de 70 empresas, de las que se conocen sus beneficios, ventas y gastos en I+D (gID) en un determinado periodo. La mayor parte de estas 70 empresas fabrica un solo producto y se compone, por tanto, de una única unidad de producción. Sin embargo, algunas empresas fabrican dos productos y poseen dos unidades de producción. Cuando esto sucede, los beneficios de la empresa, las ventas y los gastos I+D se obtienen sumando los beneficios, las ventas y los gastos en I+D respectivamente de las dos unidades de producción que la componen. Los beneficios de la unidad de producción p en la empresa t están determinados por el modelo

$$beneficios_{t_p} = \alpha_1 + \beta_2 ventas_{t_p} + \beta_3 gID_{t_p} + u_{t_p}, \qquad (2)$$

que satisface todas las hipótesis del MLG, en concreto: $E\left[u_{t_p}\right] = 0$, $var\left(u_{t_p}\right) = \sigma^2$, $cov\left(u_{t_p}, u_{s_q}\right) = 0 \ \forall t \neq s$ y cualquier par (p, q), y $cov\left(u_{t_1}, u_{t_2}\right) = 0$.

a) ¿Satisfacen los errores de la ecuación (1) las hipótesis del MLG? Justifique su respuesta.

b) El modelo de la ecuación (1) se estima por MCO obteniéndose

$$beneficios_t = -1,253 + 0,655 ventas_t + 0,120 gID_t + e_t.$$

Se estima además por MCO la regresión auxiliar

$$e_t^2 = \hat{\gamma}_1 + \hat{\gamma}_2 ventas_t + \hat{\gamma}_3 gID_t + \hat{\gamma}_4 ventas_t^2 + \hat{\gamma}_5 gID_t^2 + \hat{\gamma}_6 ventas_t * gID_t + \hat{v}_t, \qquad R^2 = 0,23.$$

¿Confirman los datos la conclusión que obtuvo en el apartado a)? Justifique su respuesta.

c) Sabiendo que el estimador de White de la varianza de $\hat{\beta}_3$ es 0.004624, contrate la hipótesis nula de que el gasto en I+D no afecta a los beneficios de las empresas frente a la alternativa de que los beneficios empresariales dependen positivamente del gasto en I+D.

3.- Considere el modelo de regresión simple sin constante:

$$y_t = \beta x_t + u_t, \quad t = 1, ..., T,$$

donde x_t sólo toma valores positivos y $var(u_t) = \sigma^2 x_t$. Se dispone de la siguiente información:

$$\begin{array}{lll}
\sum_{t=1}^{T} x_t = 1977 & \sum_{t=1}^{T} y_t = 1578 & \sum_{t=1}^{T} x_t^4 = 41630 \\
\sum_{t=1}^{T} x_t^2 = 4897 & \sum_{t=1}^{T} x_t^2 e_t^2 = 6121341 & \sum_{t=1}^{T} \widehat{u}_t^2 = 260125 \\
\sum_{t=1}^{T} x_t y_t = 3851 & \sum_{t=1}^{T} e_t^2 = 380560 & T = 1000
\end{array}$$

donde e_t son los residuos basados en el estimador MCO y $\widehat{u}_t$ son los residuos basados en el estimador MCG

a) Contraste, utilizando el estimador MCO, $H_0 : \beta = 0$ frente a $H_1 : \beta \neq 0$

b) Obtenga el estimador de mínimos cuadrados generalizados de β y calcule su varianza estimada

4.- Considere el modelo lineal con una sola variable explicativa

$$Y_t = \beta X_t + u_t$$

donde la variable X_t siempre es positiva. Se consideran los siguientes estimadores alternativos del parámetro unidimensional β :

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^T Y_t}{\sum_{t=1}^T X_t}; \quad \widehat{\beta}_2 = \frac{\sum_{t=1}^T X_t Y_t}{\sum_{t=1}^T X_t}; \quad \widehat{\beta}_3 = \frac{\sum_{t=1}^T X_t Y_t}{\sum_{t=1}^T X_t^2}; \quad \widehat{\beta}_4 = \frac{1}{T}\sum_{t=1}^{T}\frac{Y_t}{X_t}$$

a) Si se verifican todos los supuestos del MLG, determine la esperanza y la varianza de cada uno de estos estimadores y seleccione, entre los que sean insesgados, aquel que sea óptimo.

b) Supongamos que en realidad no se cumplen las hipótesis del MLG porque falla el supuesto de homocedasticidad, ya que las perturbaciones son independientes con media 0 y $Var(u_t) = \sigma^2 \lambda_t$.

b1) Determine cuál es ahora el estimador lineal insesgado óptimo de β y calcule su esperanza y su varianza.

5.- Considere el modelo de regresión simple sin constante:

$$y_t = \beta x_t + u_t, \quad t = 1, ..., T,$$

que satisface todos los supuestos básicos salvo el de homocedasticidad. Se dispone de la siguiente información:

$$\begin{array}{ccc} \sum_t x_t = 1977 & \sum_t y_t = 1842 & \sum_t x_t^4 = 41629 \\ \sum_t x_t^2 = 4897 & \sum_t x_t^2 e_t^2 = 221193 & T = 1000 \\ \sum_t x_t y_t = 4445 & \sum_t e_t^2 = 26295 & \end{array}$$

$e_t = y_t - \widehat{\beta} x_t$, donde $\widehat{\beta}$ es el estimador MCO de β.

a) Contraste $H_0 : \beta = 0$ frente a $H_1 : \beta \neq 0$.

b) Sabiendo ahora que

$$\sum_t \frac{x_t y_t}{\sigma_t^2} = 165 \text{ y } \sum_t \frac{x_t^2}{\sigma_t^2} = 176,$$

obtenga el estimador de mínimos cuadrados generalizados y proponga un estimador de su varianza.

c) Suponiendo que $\sigma_t^2 = \gamma_1 + \gamma_2 x_t^2$, donde γ_1 y γ_2 son parámetros desconocidos, obtenga estimadores consistentes de los parámetros γ_1 y γ_2. Explique cómo obtendría un estimador del parámetro β asintóticamente eficiente a partir de estas estimaciones.

6. El modelo con una sola variable explicativa:

$$Y_t = \beta X_t + u_t \qquad t = 1, 2, ..., 60 \quad (1)$$

cumple todas las hipótesis del MLG excepto la hipótesis de homocedasticidad. Se sabe que las observaciones están ordenadas según el valor de X y que para las 20 primeras observaciones (grupo 1) $Var(u_t) = 2$, mientras que para las 40 observaciones restantes (grupo 2) $Var(u_t) = \sigma^2$, donde σ^2 es desconocido.

a) Explique cómo podría obtener un estimador asintóticamente eficiente de β y proporcione una expresión para este estimador.

b) Indique cómo contrastar la hipótesis de que X es relevante para explicar Y en el modelo de la ecuación (1). (Indique el estadístico que debería utilizar y su distribución).

c) Suponga que desea contrastar la hipótesis de homocedasticidad en el modelo original utilizando el contraste de Goldfeld-Quandt. Desafortunadamente, un investigador un poco patoso ha reorganizado las observaciones de manera aleatoria. Suponiendo que las perturbaciones son heterocedásticas, ¿podría este contraste detectar la presencia de heterocedasticidad en los datos después de la intervención del investigador patoso?. Justifique la respuesta.

7.- El número de desempleados de un país en un periodo de tiempo (D_t) depende de la población activa que exista en dicho periodo (A_t) y del nivel de producción (P_t), según el modelo

$$D_t = \beta_1 + \beta_2 A_t + \beta_3 P_t + u_t$$

Se dispone de 78 observaciones de estas variables para la economía española. En base a esa muestra se ha estimado por MCO el modelo obteniendose los siguientes resultados:

$$(\text{Ec. } 1): \quad D_t = -34241 + \underset{\substack{(0.558) \\ [0.402]}}{3.75}\, A_t - \underset{\substack{(0.377) \\ [0.281]}}{1.98}\, P_t + e_t.$$

Entre paréntesis figuran los errores estándar de los estimadores MCO válidos cuando se verifican todas las hipótesis del MLG y entre corchetes los errores estándar robustos a heterocedasticidad (basados en la matriz de varianzas de White). Supondremos que este modelo satisface todas la hipótesis del MLG con normalidad salvo, quizá, la de homocedasticidad. Se dispone de los sigientes resultados basados en estimaciones MCO auxiliares (entre paréntesis se dan los errores estándar MCO):

- (Ec. 2) $e_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 A_t + \hat{\alpha}_2 P_t + \hat{\alpha}_3 A_t^2 + \hat{\alpha}_4 P_t^2 + \hat{\alpha}_5 A_t P_t +$residuo; $R^2 = 0.45$;
- (Ec. 3) $e_t = \hat{\delta}_0 + \hat{\delta}_1 A_t + \hat{\delta}_2 P_t + \hat{\delta}_3 A_t^2 + \hat{\delta}_4 P_t^2 + \hat{\delta}_5 A_t P_t +$residuo; $R^2 = 0.05$;
- (Ec. 4) $e_t^2/\hat{\sigma}^2_{\mathrm{MV}} = \hat{\gamma}_1 + \hat{\gamma}_2 A_t + \hat{\gamma}_3 P_t + \hat{v}_t$; $R^2 = 0.24$, $\hat{v}'\hat{v} = 135.4$;
- (Ec. 5) $e_t^2/\hat{\sigma}^2_{\mathrm{MV}} = \hat{\lambda}_1 + \hat{\lambda}_2 A_t + \hat{\lambda}_3 P_t + \hat{\lambda}_4 A_t P_t + \hat{r}_t$; $R^2 = 0.29$, $\hat{r}'\hat{r} = 126.9$;
- (Ec. 6) $e_t/\hat{\sigma}^2_{\mathrm{MV}} = \hat{\theta}_1 + \hat{\theta}_2 A_t + \hat{\theta}_3 P_t + \widehat{\omega}_t$; $R^2 = 0.01$, $\widehat{\omega}'\widehat{\omega} = 34587$;
- (Ec. 7) $D_t^* = D_t/\sqrt{A_t}$; $A_t^* = A_t/\sqrt{A_t}$; $P_t^* = P_t/\sqrt{A_t}$;

$$\begin{array}{ccccccccccc} D_t^* & = & 562 & + & 6.02 & A_t^* & - & 1.92 & P_t^* & + & e_t^*; \\ & & (84.8) & & (0.94) & & & (0.39) & & & \end{array}$$

- (Ec. 8) $D_t^* = D_t/\sqrt{A_t}$; $C_t^* = 1/\sqrt{A_t}$; $A_t^* = A_t/\sqrt{A_t}$; $P_t^* = P_t/\sqrt{A_t}$;

$$\begin{array}{ccccccccccc} D_t^* & = & -34449 & C_t^* & + & 3.77 & A_t^* & - & 1.99 & P_t^* & + & e_t^*; \\ & & (4965) & & & (0.059) & & & (0.39) & & & \end{array}$$

- (Ec. 9) $D_t^* = D_t/A_t$; $C_t^* = 1/A_t$; $P_t^* = P_t/A_t$;

$$\begin{array}{ccccccccccc} D_t^* & = & 3.78 & - & 34634 & C_t^* & - & 1.99 & P_t^* & + & e_t^*. \\ & & (0.60) & & (5133) & & & (0.40) & & & \end{array}$$

Con esta información:

a) Contraste, de todas las formas que pueda, que la varianza de las perturbaciones u_t depende de las variables A_t y P_t.

b) Utilizando el estimador MCO, contraste la significatividad individual de la variable A_t.

c) Un analista llega a la conclusión de que $Var(u_t) = \sigma^2 A_t$ y realiza una estimación eficiente del modelo basada en esa información. Indique cómo debería contrastar este analista la significatividad individual de la variable A_t y qué resultado obtendría.