

Taller #5
Respuestas Sugeridas
Econometría 06169
Grupo 1

Profesor: Julio César Alonso

Nota: Este taller debe ser entregado en papel y escrito en computador. No se revisarán trabajos escritos a mano.

1. Un investigador desea investigar la relación entre el ingreso per cápita (I) y el gasto per cápita en educación (G) en diferentes estados de Estados Unidos para el año 1979. Los datos en dólares per cápita se encuentran en la siguiente tabla.

Ingresos y Gastos en Educación (Dólares per cápita) para los Estados de Estados Unidos (1979)

Estado	Gasto	Ingresos	Estado	Gasto	Ingresos	Estado	Gasto	Ingresos
Ala.	275	6247	Alaska	821	10851	Ariz.	339	7374
Ark.	275	6183	Cal.	387	8850	Colo.	452	8001
Ct.	531	8914	Del.	424	8604	D.C.	428	10022
Fla.	316	7505	Ga.	265	6700	Hawaii	403	8380
Idaho	304	6813	Ill	437	8745	Ind.	345	7696
Iowa	431	7873	Kans.	355	8001	Ky.	260	6615
La.	316	6640	Maine	327	6333	Md.	427	8306
Mass.	427	8063	Mich.	466	8442	Minn.	477	7847
Miss.	259	5736	Mo.	274	7342	Mont.	433	7051
Nebr.	294	7391	Nev.	359	9032	N.H.	279	7277
N.J.	423	8818	N.Mex.	388	6505	N.Y.	447	8267
N.C.	335	6607	N.Dak.	311	7478	Ohio	322	7812
Okla.	320	6951	Oreg.	397	7839	Pa.	412	7733
R.I.	342	7526	S.C.	315	6242	S.Dak.	321	6841
Tenn.	268	6489	Texas	315	7697	Utah	417	6622
Vt.	353	6541	Va.	356	7624	Wash.	415	8450
W.Va.	320	6456				Wyo.	500	9096

Fuente: U.S. Department of Commerce

- a) Estime el siguiente modelo por el método de MCO: $G_i = b_1 + b_2 I_i + b_3 I_i^2 + e_i$.
 Presente sus resultados en una tabla y discuta brevemente el significado de cada uno de los coeficientes estimados y otros resultados que crea pertinente.

El modelo

$$G_i = b_1 + b_2 I_i + b_3 I_i^2 + e_i \quad (1)$$

fue estimado con EasyReg obteniendo los resultados reportados en la Tabla 1. Noten que $\hat{b}_1 = 832.91$ implica que de no existir ningún ingreso per cápita; el gasto per cápita promedio en educación en los Estados de E.U. era de \$832.91 en 1979. La interpretación de \hat{b}_2 y \hat{b}_3 no es sencilla, pues $\frac{\partial G_i}{\partial I_i} = \hat{b}_2 + 2\hat{b}_3 I_i$. Así, $\hat{b}_2 = -0.18342$ es el efecto marginal en el gasto del primer dólar per cápita de ingreso, i.e., $\left. \frac{\partial G_i}{\partial I_i} \right|_{I_i=0} = \hat{b}_2 + 2\hat{b}_3(0) = \hat{b}_2$. Similarmente, $\hat{b}_3 = 0.00002$ es la mitad del aumento en el efecto marginal, sobre el gasto en educación per cápita, de un aumento en un dólar en el ingreso per cápita.

Nota: Esta interpretación o era fácil. Mi intención era que vieran cuidadosamente lo que significa cada uno de los betas estimados y no hicieran una interpretación mecánica. Recuerden que la interpretación es mucho más importante que el propio cálculo.

Otros resultados pertinentes, serían el R^2 relativamente alto, pues tenemos que la regresión explica un 65.3% de la variabilidad de la variable dependiente. Noten que el F global no tiene sentido si estamos en presencia de Heteroscedasticidad. En este caso tenemos altas sospechas que exista Heteroscedasticidad, pues estamos estudiando datos de corte transversal, así analizar el F global no tiene sentido.

Tabla 1. Ecuaciones Estimadas.

	VARIABLE DEPENDIENTE: G_i	
	Estadísticos t entre paréntesis	
	Ecuación 1	Ecuación 2
	MCO	MCO
constante	832,9144 (2,55)	832,9144 (1,81) *
I_i	-0,1834 (-2,21) **	-0,1834 (-1,48)
I_i^2	0,00002 (3,06) ***	0,00002 (1,91) *
Goldfeld y Quandt	1,61700	NA
Breusch-Pagan	10,45800 ***	NA
White	21,15900 ***	NA
R^2	0,65534	0,65534
# de Obs.	50	50

(*) nivel de significancia: 10%

(**) nivel de significancia: 5%

(***) nivel de significancia: 1%

MCO: Mínimos Cuadrados Ordinarios

NA: No Aplica

- b) Discuta si existe algún tipo de problema de multicolinealidad en el modelo estimado en la parte a) (Si encuentra algún tipo de problema, no lo solucione),

A primera vista, uno esperaría que no existiera ningún problema de multicolinealidad, pues aunque existe una relación clara entre las variables independientes I_i e I_i^2 , esta relación no es lineal; es cuadrática. Pero de todas maneras examines si existe o no el problema. Primero noten que los síntomas no están claramente presentes, el R^2 es relativamente alto, pero no tanto como para despertar sospechas. Noten que ni el F-global ni los t-calculados son buenos indicadores de la presencia de multicolinealidad, pues como estamos sospechando que exista heteroscedasticidad entonces la matriz de varianzas y covarianzas de los estimadores MCO estará sesgada y por tanto el F-global y los t-calculados no serán los correctos.

Exploremos ahora la matriz de correlaciones entre las variables independientes. En este caso tenemos $R = \begin{bmatrix} 1 & 0.9956694 \\ 0.9956694 & 1 \end{bmatrix}$, con un determinante de 0.0086. Este determinante es relativamente bajo, evidenciando un problema serio de multicolinealidad.

Finalmente, noten que no hay necesidad de estudiar la matriz de varianzas y covarianzas de los coeficientes estimados, pues como esperamos que exista un problema de Heteroscedasticidad, esta matriz será sesgada y por tanto la información que derivemos de ella no es confiable.

- c) Grafique los residuos estimados versus el ingreso per cápita y el ingreso per cápita al cuadrado (haga dos gráficos por separado). Reporte su gráfico y discuta lo que observa. Recuerde que cada grafico necesita tener su respectivo título.

A continuación se reportar los gráficos de los residuos estimados versus el ingreso per cápita y el ingreso per cápita al cuadrado

Gráfico 1 Residuos Estimados versus el Ingreso per cápita

$$Y = \text{OLS Residual of Gasto } (t=1 \rightarrow 50)$$

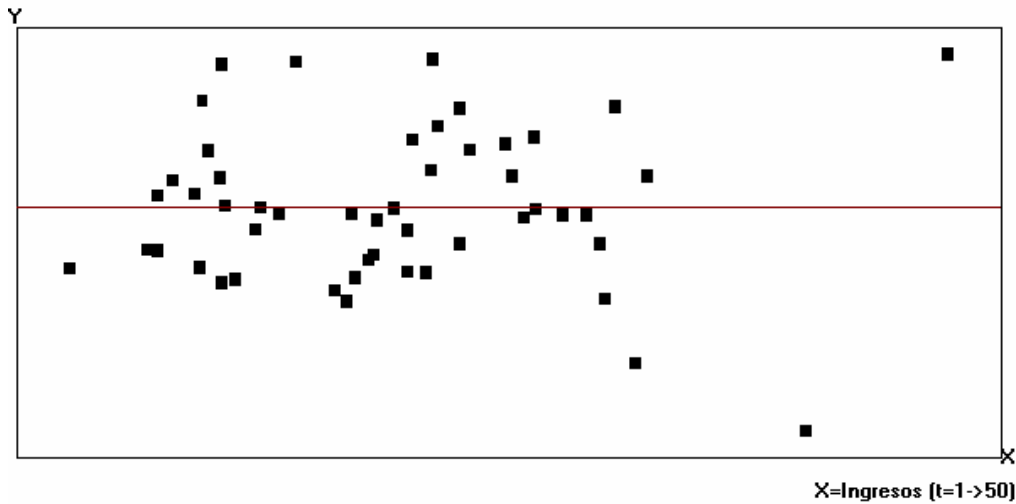
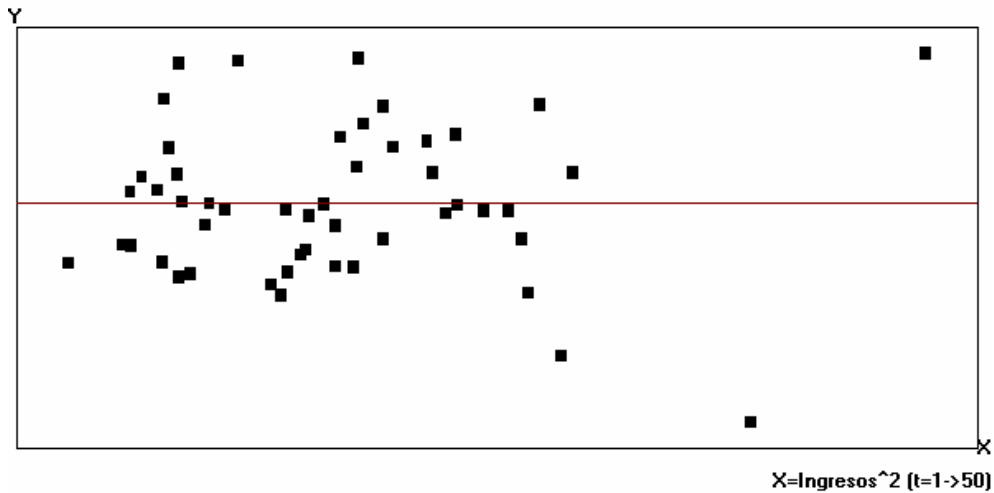


Gráfico 2 Residuos Estimados versus el Ingreso per cápita al cuadrado

$$Y = \text{OLS Residual of Gasto } (t=1 \rightarrow 50)$$



Noten que ambos gráficos muestran que los errores tienden a tener mas variabilidad cuando el ingreso per cápita aumenta. En todo caso es claro que la variabilidad del error no es constante.

2. Empleando la misma información del punto anterior, responda las siguientes preguntas:

- a) Efectúe el test de Golfeldt y Quandt. Escriba claramente los pasos seguidos y cual es su decisión.

Recuerde que este test involucra, en este caso, los siguientes pasos:

1. Ordenar los datos de menor a mayor de acuerdo a la variable independiente I_i (pues de acuerdo a los gráficos se espera que esta variable afecte a la varianza del error)
2. Omite los $d=8$ datos de la mitad ($d < \frac{1}{5}50 = 10$)
3. Corra dos regresiones, una con los datos asociados a los valores bajos de I_i (esta será la regresión 1 que se correrá con los datos $i = 1, 2, \dots, 21$) y otra regresión con los datos asociados con los valores altos de I_i (esta será la regresión 2 que se correrá con los datos $i = 30, 31, \dots, 50$)
4. Encuentre $SSE_1=42755.37$ y $SSE_2.= 69138.9$
5. Calcule $F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{69138.960641}{42755.370120} = 1.617$
6. Compare F_{GQ} con $F_{(n-d-2k, n-d-2k)} = F_{(36,36),0.01} = 2.205$.

Claramente, no podemos rechazar la hipótesis nula de no Heteroscedasticidad.

- b) Efectúe el test de Breusch-Pagan. Escriba claramente los pasos seguidos y cual es su decisión.

Los pasos de este test para comprobar la $H_A : \mathbf{s}_i^2 = f(\mathbf{g} + \mathbf{d}I_i)$ son los siguientes:

1. Corra el modelo original $G_i = \mathbf{b}_1 + \mathbf{b}_2 I_i + \mathbf{b}_3 I_i^2 + \mathbf{e}_i$ y encuentre la serie de los residuos.
2. Calcule $\hat{\mathbf{s}}^2 = \frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{n} = \frac{150985.503368}{50} = 3019,71006736$.
3. Estime la siguiente regresión auxiliar: $\frac{\hat{\mathbf{e}}_i^2}{\hat{\mathbf{s}}^2} = \mathbf{g} + \mathbf{d}I_i + \mathbf{m}_i$.
4. Construya los $SSR = 119.387059 - 98.471986 = 20.915073$ de la última regresión.
5. Calcule $BP = \frac{SSR}{2} = \frac{20.915073}{2} = 10.458$.
6. Compare BP con $\mathbf{c}_{1,01}^2 = 6.63$

Claramente, EL estadístico de Breusch-Pagan es mayor que el valor crítico. Por tanto existe suficiente evidencia para rechazar la hipótesis nula de no heteroscedasticidad.

- c) Efectúe el test de White. Escriba claramente los pasos seguidos y cual es su decisión.

En este caso el test de White implica los siguientes pasos:

1. Corra el modelo original y encuentre la serie de los residuos.
2. Ahora corra la siguiente regresión auxiliar:

$$\hat{e}_i^2 = \mathbf{g} + \sum_{m=1}^k \sum_{j=1}^k \mathbf{d}_s X_{mi} X_{ji} + \mathbf{m}_i = \mathbf{g} + \mathbf{d}_1 I_i + \mathbf{d}_2 I_i^2 + \mathbf{d}_3 I_i^3 + \mathbf{d}_4 I_i^4 + \mathbf{m}_i$$

3. Calcule el R^2 de la última regresión, en este caso $R^2 = 0.423188$
4. Calcule $W_a = nR^2 = 21.159$
5. Compare W_a con $c_{4,0.01}^2 = 13.28$

Claramente podemos rechazar la hipótesis nula de la no presencia de Heteroscedasticidad. Así podemos concluir que existe Heteroscedasticidad.

3. Empleando la misma información del punto anterior, responda las siguientes preguntas:

- a) Calcule la matriz de varianzas y Covarianzas de los estimadores MCO por el método sugerido por White (1980). Y discuta la significancia de los coeficientes.

Como lo habíamos discutido, esta matriz es calculada automáticamente por el EasyReg. En este caso tenemos que:

$$EstVar[\hat{\mathbf{b}}] = EstVar \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \hat{\mathbf{b}}_3 \end{bmatrix} = \begin{bmatrix} 10621056.2656579 & -2858.4959261 & 0.1897037 \\ & 0.7725779 & -0.0000515 \\ & & 0.0000000 \end{bmatrix}$$

En la Tabla 1, bajo la denominación de Ecuación 2 se muestran los t-calculados teniendo en cuenta la corrección de White por Heteroscedasticidad.

Como se ve en la Tabla 1, el coeficiente asociado al ingreso per cápita no es significativo, mientras que los otros dos coeficientes si son significativos, con un nivel de confianza del 10%.

- b) Pruebe la Hipótesis nula $\mathbf{b}_2 = \mathbf{b}_3$. Discuta su decisión.

El estadístico de Wald, teniendo en cuenta la corrección por Heteroscedasticidad, para probar la hipótesis nula $\mathbf{b}_2 = \mathbf{b}_3$ ($\mathbf{b}_2 - \mathbf{b}_3 = 0$) es 2.18, que es menor que el estadístico Chi-cuadrado con un grado de libertad a los niveles de significancia convencionales. Así, podemos aceptar la hipótesis nula.

4. Suponga que el investigador sabe que el término de error del modelo $G_i = \mathbf{b}_1 + \mathbf{b}_2 I_i + \mathbf{b}_3 I_i^2 + \mathbf{e}_i$ tiene $Var(\mathbf{e}_i) = \mathbf{s}^2 I_i^4$. ¿Cómo obtendría usted estimadores MELI para este modelo? Demuestre que el procedimiento sugerido si produce estimadores MELI. (Usted no necesita estimar este modelo).

En clase vimos que una opción para solucionar el problema de heteroscedasticidad es emplear los mínimos cuadrados ponderados. En este caso podemos transformar el modelo original en uno homoscedástico dividiendo a ambos lados del modelo por I_i^2 . Es decir:

$$G_i = \mathbf{b}_1 + \mathbf{b}_2 I_i + \mathbf{b}_3 I_i^2 + \mathbf{e}_i \quad (2)$$

$$\frac{G_i}{I_i^2} = \mathbf{b}_1 \frac{1}{I_i^2} + \mathbf{b}_2 \frac{I_i}{I_i^2} + \mathbf{b}_3 \frac{I_i^2}{I_i^2} + \frac{\mathbf{e}_i}{I_i^2} \quad (3)$$

$$\frac{G_i}{I_i^2} = \mathbf{b}_1 \frac{1}{I_i^2} + \mathbf{b}_2 \frac{1}{I_i} + \mathbf{b}_3 + \frac{\mathbf{e}_i}{I_i^2} \quad (4)$$

$$G_i^* = \mathbf{b}_1 z_{1i} + \mathbf{b}_2 z_{2i} + \mathbf{b}_3 z_{3i} + \mathbf{e}_i^* \quad (5)$$

donde $G_i^* = \frac{G_i}{I_i^2}$, $z_{1i} = \frac{1}{I_i^2}$, $z_{2i} = \frac{1}{I_i}$, $z_{3i} = 1$, y $\mathbf{e}_i^* = \frac{\mathbf{e}_i}{I_i^2}$. Noten que el nuevo modelo ya es

homoscedástico pues $Var[\mathbf{e}_i^*] = Var\left[\frac{\mathbf{e}_i}{I_i^2}\right] = \frac{1}{I_i^4} Var[\mathbf{e}_i] = \frac{1}{I_i^4} \mathbf{s}^2 I_i^4 = \mathbf{s}^2$. Entonces el modelo (5) es homoscedástico y por tanto los estimadores MCO aplicados al modelo (5) serán MELI.