

Solución a la práctica 6 con Eviews

El siguiente modelo de regresión relaciona la nota media que obtienen los alumnos en matemáticas (*nota*) en un centro, con el número de profesores disponibles en el centro (*profesores*), el porcentaje de repetidores (*repetidores*) sobre el total de alumnado del centro y el porcentaje de recursos con los que se financia el centro (*recursos*) que proceden de fondos públicos

$$nota_i = \beta_0 + \beta_1 \text{profesores}_i + \beta_2 \text{repetidores}_i + \beta_3 \text{recursos}_i + \varepsilon_i.$$

Usando las bases de datos de la prueba PISA¹ para centros educativos (PISA 2013) se han obtenido datos de una muestra de 225 centros escolares españoles que se encuentran en el fichero *practica6.wf1*. Sabiendo que la puntuación PISA es una puntuación normalizada tal que la media de los países de la OCDE en PISA se establece en 500 y la desviación típica en 100, responda a las siguientes cuestiones:

1) Estime el modelo e interprete los coeficientes

ls *nota c profesores repetidores recursos*

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: NOTA									
Method: Least Squares									
Date: 01/11/16 Time: 22:28									
Sample: 1 225									
Included observations: 225									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
C	511.8863	8.980046	57.00263	0.0000					
PROFESORES	0.402887	0.077305	5.211654	0.0000					
REPETIDORES	-1.770264	0.300842	-5.884366	0.0000					
RECURSOS	-0.322902	0.103628	-3.115961	0.0021					
R-squared	0.318169	Mean dependent var	488.3494						
Adjusted R-squared	0.308914	S.D. dependent var	34.59109						
S.E. of regression	28.75613	Akaike info criterion	9.573197						
Sum squared resid	182748.2	Schwarz criterion	9.633927						
Log likelihood	-1072.985	Hannan-Quinn criter.	9.597708						
F-statistic	34.37582	Durbin-Watson stat	1.488302						
Prob(F-statistic)	0.000000								

La nota media en matemáticas se relaciona positivamente con el número de profesores de los centros y negativamente con el porcentaje de repetidores y con el porcentaje de recursos públicos con los que se financia el centro. En concreto, un colegio con un profesor más que otro colegio con las mismas características obtendría aproximadamente 0.4 puntos más de nota. Además, un punto adicional en el porcentaje de repetidores o en el porcentaje que representan recursos públicos disminuyen la nota media en aproximadamente 1.8 y 0.3 puntos, respectivamente.

2) Realice los diagramas de dispersión entre el cuadrado de los residuos y cada una de las variables explicativas, y entre el cuadrado de los residuos y la variable dependiente estimada por el modelo. Interprete los resultados.

El cuadrado de los residuos constituye una aproximación simple (y la única posible con la información disponible) de la varianza de las perturbaciones. Por tanto, si hay homoscedasticidad, se espera que el comportamiento general del cuadrado de los

¹ Datos disponibles en el Instituto Nacional de Evaluación Educativa

<http://www.mecd.gob.es/inee/Bases-de-datos.html>

residuos sea el mismo a lo largo de todo el recorrido de valores de la variable con la que se compara, sea esta una de las explicativas del modelo o las predicciones de la explicada. Para obtener los residuos, sus cuadrados y la variable dependiente estimada por el modelo se parte de la regresión MCO

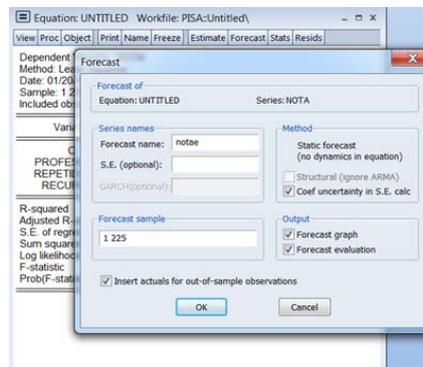
ls nota c profesores repetidores recursos

Los residuos de la regresión se obtienen seleccionando

Procs/make residual series/... se llamarán *e*

Para obtener la variable dependiente estimada por el modelo

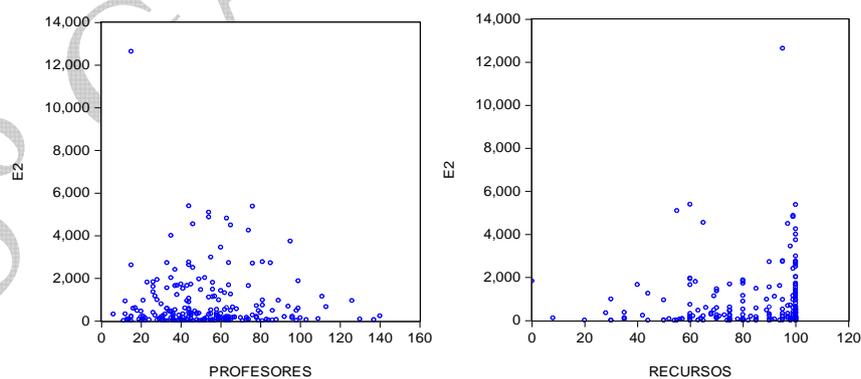
Forecast/... se llamará *notae*

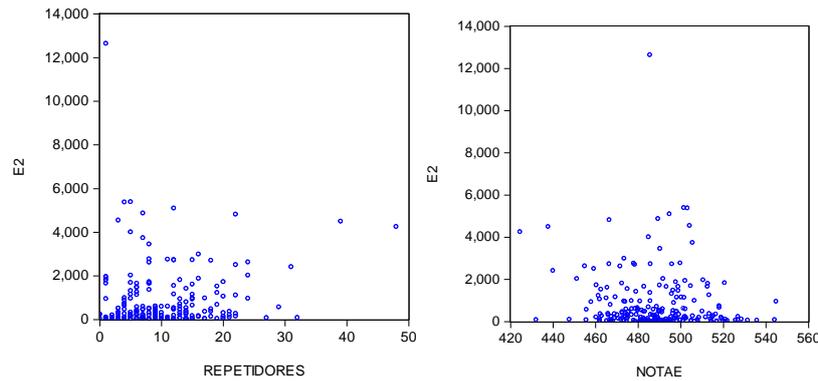


Por último, se obtiene el cuadrado de los residuos anteriores que se llamará *e2*:

Genr e2=e^2

Ahora se obtienen los diagramas de dispersión





En los diagramas de dispersión se intuye que podría existir una cierta relación entre el cuadrado de los residuos, que es una aproximación de la varianza de los errores, y la variable explicativa recursos (en menor medida con los profesores). Sin embargo, no se encuentra esa relación con la estimación de las notas, que recoge el efecto de las tres variables explicativas al mismo tiempo. Por tanto, podría existir un problema de heteroscedasticidad en este modelo. En cualquier caso, la decisión sobre la posible presencia de heteroscedasticidad tiene que apoyarse en el resultado de un contraste.

3) Utilice el contraste de White para detectar la presencia de heteroscedasticidad.

Este contraste no precisa conocer la estructura de la heteroscedasticidad; ni siquiera requiere imponer qué variable(s) explicativa(s) genera(n) los problemas de heteroscedasticidad. Se aplica a través de los siguientes pasos.

a) Se parte de una estimación de la varianza de los residuos al cuadrado (ya los hemos generado en el apartado anterior y los hemos llamado e_2) y se plantea una regresión auxiliar que la explique regresando sobre una constante, las explicativas del modelo inicial, sus cuadrados, y sus productos cruzados, cuidando de que no se repita ninguna explicativa en esta regresión auxiliar. Por tanto, la aproximación que se hace de la varianza es muy flexible por lo que debería ser capaz de capturar la posible presencia de heteroscedasticidad. Es decir, se propone la regresión auxiliar siguiente:

*ls e2 c profesores repetidores recursos profesores^2 repetidores^2 recursos^2 profesores*repetidores profesores*recursos repetidores*recursos*

y se obtiene el R^2 de esta regresión que vale $R_e^2 = 0.0961$.

b) El contraste consiste en

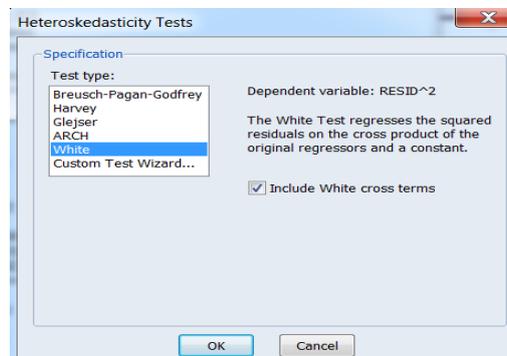
$$\left. \begin{array}{l} H_0 \text{ homoscedasticidad} \\ H_a \text{ heteroscedasticidad} \end{array} \right\} \text{ y se contrasta con el estadístico } NR_e^2 \text{ que bajo la nula sigue una } \chi_q^2, \text{ donde } q \text{ es el número de explicativas que usamos en la regresión auxiliar. El valor crítico con el que hay que comparar el estadístico es el valor de } \chi_{q,\alpha}^2. \text{ Por tanto, se rechaza la nula de homoscedasticidad si } NR_e^2 > \chi_{q,\alpha}^2.$$

En este caso, $NR_e^2 = 225 \cdot 0.0961 = 21.61$ y $q=9$, por lo que se rechaza la nula de homoscedasticidad si $21.61 > \chi_{9,\alpha}^2$. Por ejemplo, si se trabaja con $\alpha = 0.05$ el valor de $\chi_{9,0.05}^2$ es 16.92 -se puede calcular mediante la función de Eviews $@qchisq(0.95,9)$ - y rechazaríamos la nula de homoscedasticidad. Sin embargo, se debe notar que si $\alpha = 0.01$ entonces $\chi_{9,0.01}^2 = 21.65$ -se puede calcular mediante la función de Eviews $@qchisq(0.99,9)$ - y no se rechazaría.

El contraste de White se puede hacer de *forma automática* con Eviews. Para ello en la ventana de estimación se pincha en

View/Residual Diagnosis/Heteroskedasticity Tests...

y en el **Test Type** se elige la opción de **White**.



En nuestro caso, el resultado es:

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Heteroskedasticity Test: White									
F-statistic		2.538862	Prob. F(9,215)		0.0087				
Obs*R-squared		21.61531	Prob. Chi-Square(9)		0.0102				
Scaled explained SS		29.64963	Prob. Chi-Square(9)		0.0005				
Test Equation:									
Dependent Variable: RESID^2									
Method: Least Squares									
Date: 01/11/16 Time: 22:48									
Sample: 1 225									
Included observations: 225									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
C	1283.151	1160.320	1.105859	0.2700					
PROFESORES	-7.278167	21.37135	-0.340557	0.7338					
PROFESORES^2	0.012121	0.101818	0.119045	0.9054					
PROFESORES*REPETIDORES	1.111339	0.664382	1.672741	0.0958					
PROFESORES*RECURSOS	-0.068854	0.219791	-0.313269	0.7544					
REPETIDORES	-83.36071	102.3002	-0.814863	0.4161					
REPETIDORES^2	3.263554	1.080539	3.020302	0.0028					
REPETIDORES*RECURSOS	-0.710359	1.170957	-0.606648	0.5447					

que, como vemos, arroja un resultado de $NR_e^2 = 225 \cdot 0.0961 = 21.61$ como ya se había obtenido en el cálculo manual. Además Eviews presenta el p -valor del contraste, que en este caso es 0.0102, lo cual implica rechazar H_0 y aceptar la existencia de heteroscedasticidad si se trabaja con $\alpha > 0.0102$, como ocurre si $\alpha = 0.05$.

4) Lleve a cabo la estimación del modelo siguiendo el método propuesto por White. ¿Qué ventaja tiene aplicar este método?

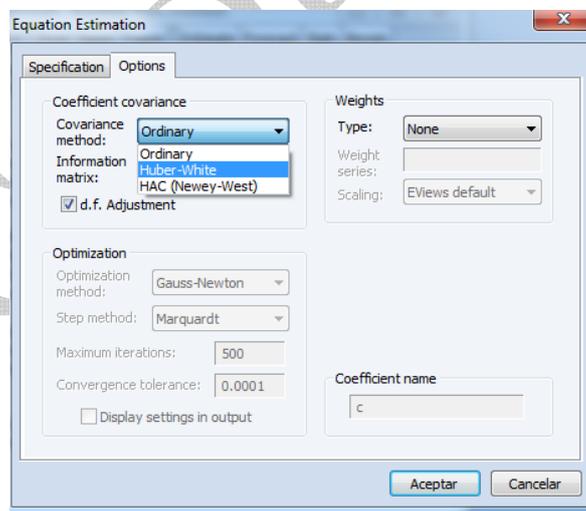
Como se ha detectado la presencia de heteroscedasticidad en los datos, se puede estimar el modelo usando el “método de White”. La ventaja de este método frente a MCO recae en que, aunque usa como estimación de los parámetros la que se obtienen por MCO (que no pierde su propiedad de consistencia ni de insesgadez en presencia de heteroscedasticidad), propone una estimación alternativa de varianzas de los coeficientes que permite hacer inferencia en caso de que exista heteroscedasticidad. En efecto, el estimador MCO habitual de estas varianzas es sesgado e inconsistente en presencia de heteroscedasticidad, mientras que la corrección de White nos proporciona un estimador consistente.

Por tanto la estimación de los coeficientes propuesta se obtiene de la estimación MCO que aparece en el primer apartado. Pero las varianzas se deben obtener mediante otra expresión distinta a la de MCO.

Utilizando Eviews, la estimación correcta de las desviaciones típicas de los parámetros del modelo según el método propuesto por White se obtiene de la siguiente manera. Se parte de la regresión MCO del modelo y entre las **opciones** que propone Eviews en el cuadro de regresión, se elige el método de White

ls nota c profesores repetidores recursos

y se pincha en la pestaña **Options** y luego en **Covariance Method** se elige la opción **Huber-White**.



El cuadro de regresión del método de White que se obtiene es:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	511.8863	8.386070	61.04006	0.0000
PROFESORES	0.402887	0.072959	5.522102	0.0000
REPETIDORES	-1.770264	0.444156	-3.985682	0.0001
RECURSOS	-0.322902	0.105347	-3.065139	0.0024
R-squared	0.318169	Mean dependent var	488.3494	
Adjusted R-squared	0.308914	S.D. dependent var	34.59109	
S.E. of regression	28.75613	Akaike info criterion	9.573197	
Sum squared resid	182748.2	Schwarz criterion	9.633927	
Log likelihood	-1072.985	Hannan-Quinn criter.	9.597708	
F-statistic	34.37582	Durbin-Watson stat	1.488302	
Prob(F-statistic)	0.000000			

Se observa que las estimaciones de los parámetros son las mismas que las que se obtuvieron aplicando el método MCO al modelo original, pero que las estimaciones de las desviaciones típicas de los coeficientes han cambiado. El modelo estimado según el enfoque de White quedaría como

$$\hat{n}otas_i = 511.88 + 0.40 \text{ profesores}_i - 1.77 \text{ repetidores}_i - 0.32 \text{ recursos}_i$$

(8.38)
(0.07)
(0.44)
(0.10)

Usando estos resultados podemos concluir que todas las variables son significativas.