

Taller 6: Heteroscedasticidad
Econometría 06216
21-02-2011

Profesor: Carlos Giovanni González Espítia

Monitores: Adriana Caicedo – Jessica Echeverry – Samir Aristizábal

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 28 de febrero de 2011. **Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta** (no se recibirán talleres después de esa hora y fecha límite).

Instrucciones:

- Este taller debe ser escrito en computador. Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo en pareja. Por tanto el taller debe reflejar **únicamente** el trabajo de la pareja.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos **con todos ellos**.

El Dr. Barriga, jefe de la división de planeación del departamento de mercadeo de la cadena de almacenes de ropa Tandamangapio, desea estimar el comportamiento de las ventas en cada una de sus sucursales. Para ello, ha contratado un equipo de reconocidos econométricos encabezados por usted para que realice dicho estudio. Uno de los líderes de la investigación, el Dr. Jirafalles sugiere que la expresión que mejor explica el comportamiento de las ventas es la siguiente:

$$V = \beta_1 + \beta_2 \ln R + \beta_2 \frac{1}{S} + \beta_3 D \quad (1)$$

Donde V representa las ventas diarias en millones de pesos; R refleja el número de reclamos en el día i ; S corresponde al número de ofertas diarias y D representa el porcentaje de descuento ofrecido en el día i . Usted cuenta con la información contenida en el archivo adjunto T6-11-01.xls

Pregunta 1

- a. Escriba correctamente el modelo a estimar.
- b. Interprete los coeficientes a priori y justifique el signo esperado.

Pregunta 2

- a. Estime el modelo (1) y reporte sus resultados en una tabla.
- b. La Profesora Clotilde, quien ha enfatizado sus investigaciones en problemas econométricos sospecha de la existencia de un problema de heteroscedasticidad en la muestra. ¿Cuáles son los síntomas de la presencia de heteroscedasticidad? ¿Existen síntomas de este problema econométrico en este caso?

Pregunta 3

- a. El Gerente de la compañía, el Dr. Chapatí, preocupado por la posibilidad de que el modelo no sea útil en el análisis del comportamiento de las ventas, le pide que compruebe si en realidad existe un problema con el modelo; más concretamente su inquietud es: ¿Existe un problema de Heteroscedasticidad? (Realice las pruebas que considera necesarias). Sea claro en las hipótesis y reglas de decisión de cada una de las pruebas realizadas.
- b. Si encontró un problema de heteroscedasticidad, ¿cómo solucionaría el problema? Explique claramente el por qué de su respuesta. Sea lo más claro y preciso.

Pregunta 4

- a. Uno de sus compañeros de investigación, el Dr. Lucas Tañeda, sugiere que el problema econométrico puede corregirse mediante una serie de transformaciones del modelo. ¿Tiene razón? Corrija (de ser posible) el problema de la forma como usted lo considere correcto y reporte sus resultados.
- b. Si la corrección del problema fue posible, muestre que ahora el modelo no tiene problemas de heteroscedasticidad.

Pregunta 5

Interprete los coeficientes estimados teniendo en cuenta su significancia y discuta si los signos fueron los esperados.

Pregunta 6

- a. El Dr. Fredonio Gordorritúa, asesor principal del líder de la investigación, le solicita que concluya si los regresores empleados explican o no el comportamiento de las ventas de la empresa. Muestre los resultados con los que toma su decisión.
- b. Ya estando usted satisfecho con el buen trabajo que había realizado, llega María Julieta de las Nieves, una reconocida economista que le colaboró en la investigación, y le informa que la naturaleza del problema de heteroscedasticidad si podía conocerse y era: $var(\varepsilon_i) = \sigma^2 D_i$, donde D_i es una variable dummy que toma el valor de uno si la sucursal

quede en el norte del país y cero en caso contrario. Ella afirma que el modelo se puede transformar dividiendo por $\sqrt{D_i}$, para que los estimadores MCO de los β 's sean insesgados y eficientes. Si es posible, realice la transformación y demuestre que el problema ha quedado solucionado.

Taller 6: Heteroscedasticidad
Econometría 06216
21-02-2011

Profesor: Carlos Giovanni González Espítia

Monitores: Adriana Caicedo – Jessica Echeverry – Samir Aristizábal

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 28 de febrero de 2011. **Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta** (no se recibirán talleres después de esa hora y fecha límite).

Instrucciones:

- Este taller debe ser escrito en computador. Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo en pareja. Por tanto el taller debe reflejar **únicamente** el trabajo de la pareja.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos **con todos ellos**.

El Dr. Barriga, jefe de la división de planeación del Departamento de mercadeo de la cadena de almacenes de ropa Tandamangapio, desea estimar el comportamiento de las ventas en cada una de sus sucursales. Para ello, ha contratado un equipo de reconocidos economistas encabezados por usted para que realice dicho estudio. Uno de los líderes de la investigación, el Dr. Jirafalles sugiere que la expresión que mejor explica el comportamiento de las ventas es la siguiente:

$$V = \beta_1 + \beta_2 \ln R + \beta_2 \frac{1}{S} + \beta_3 D \quad (1)$$

Donde V representa las ventas diarias en millones de pesos; R refleja el número de reclamos en el día i ; S corresponde al número de ofertas diarias y D representa el porcentaje de descuento ofrecido en el día i . Usted cuenta con la información contenida en el archivo adjunto T6-11-01.xls

Pregunta 1

- a. Escriba correctamente el modelo a estimar.

El modelo a estimar es:

$$V_i = \beta_1 + \beta_2 \ln R_i + \beta_3 \frac{1}{S_i} + \beta_4 D_i + \varepsilon_i \quad i = 1, 2, \dots, 100$$

b. Interprete los coeficientes a priori y justifique el signo esperado.

β_1 = No tiene interpretación económica.

Derivada para interpretar β_2 :

$$\frac{dV_i}{dLnR_i} = \beta_2$$

$$\frac{dV_i}{\frac{1}{x_{1i}} dx_{1i}} = \beta_2$$

$$\frac{dy_i}{\Delta\%x_{1i}} = \frac{\beta_2}{100}$$

β_2 = Una variación del uno por ciento en el número de reclamos en el día i, genera en promedio una variación de $\frac{\beta_2}{100}$ millones de pesos en las ventas diarias de la cadena de almacenes. Se espera signo negativo.

Derivada para interpretar β_3 :

$$\frac{dV_i}{d\frac{1}{S_i}} = \beta_3$$

$$\frac{dV_i}{-\frac{1}{S_i^2} dS_i} = \beta_3$$

$$\frac{dV_i}{\frac{dS_i}{S_i}} = -\frac{\beta_3}{S_i}$$

$$\frac{dV_i}{\Delta\%S_i} = -\frac{\beta_3}{100} \frac{1}{S_i}$$

β_3 = Una variación del uno por ciento en el número de ofertas diarias, genera en promedio, una disminución de $\frac{\beta_3}{100} \frac{1}{S_i}$ millones de pesos en las ventas diarias de la cadena de almacenes. Se espera signo positivo.

β_4 = Una variación del uno por ciento en el porcentaje de descuento ofrecido en el día i, genera en promedio, una variación de β_4 millones de pesos en las ventas diarias de la cadena de almacenes. Se espera signo positivo.

Pregunta 2

- a. Estime el modelo (1) y reporte sus resultados en una tabla.

Tabla 1

Variable dependiente: V_i

MCO

Variable	Modelo 1		Modelo 1 (corregido)	
	Parámetro	Significancia	Parámetro	Significancia
Constante	15,0246475 (20,285)	***	15,0246475 (18,116)	***
$\ln R_i$	-2,4954088 (-6,311)	***	-2,4954088 (-4,255)	***
$1 / S_i$	-24,4224684 (-6,548)	***	-24,4224684 (-3,383)	***
D_i	0,0606953 (1,695)	*	0,0606953 (1,496)	
R^2	0,7963		0,7963	
$R^2_{ajustado}$	0,7899		0,7899	
F	125,08	***	125,08	***
n	100		100	

***Nivel de significancia 1%

**Nivel de significancia 5%

*Nivel de significancia 10%

Fuente: Cálculos propios en EasyReg

- b. La Profesora Clotilde, quien ha enfatizado sus investigaciones en problemas econométricos sospecha de la existencia de un problema de heteroscedasticidad en la muestra. ¿Cuáles son los síntomas de la presencia de heteroscedasticidad? ¿Existen síntomas de este problema econométrico en este caso?

En principio, sabemos que el hecho de tener datos de corte transversal es un síntoma de la presencia de heteroscedasticidad. Además, se presentan a continuación las gráficas que reflejan el comportamiento de los errores estimados frente a las variables explicativas, la variable estimada y las observaciones con el objetivo de establecer si existe una relación entre ellas.

Gráfico 1: $\ln R_i$ vs errores estimados

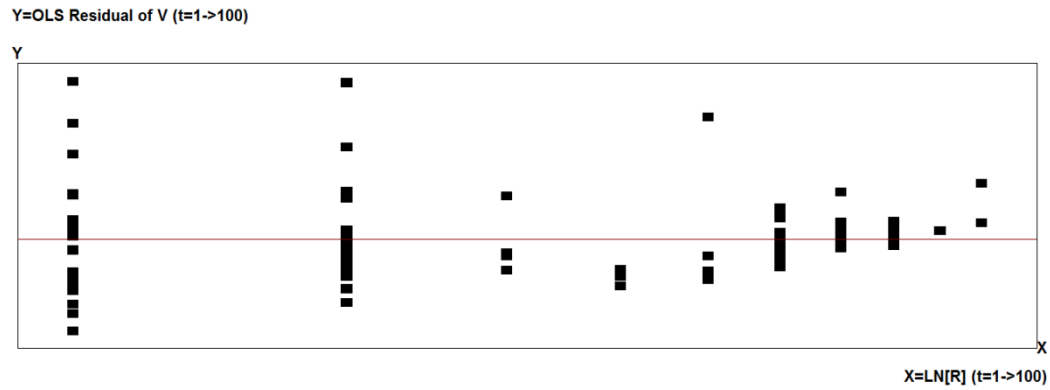


Gráfico 2: $\frac{1}{S_i}$ vs errores estimados

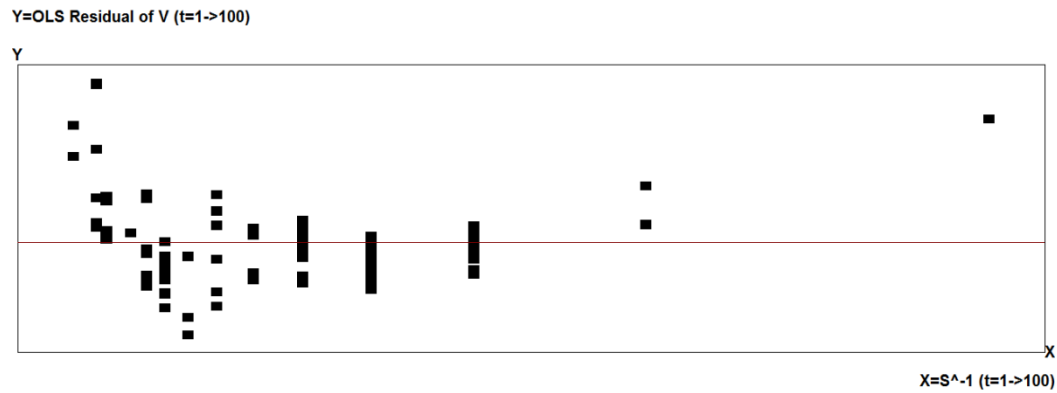


Gráfico 3: D_i vs errores estimados

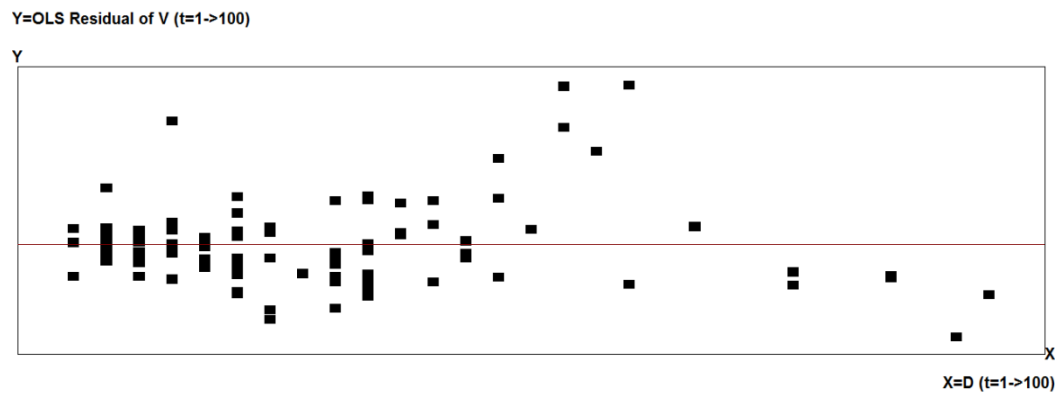


Gráfico 4: Observaciones vs errores estimados

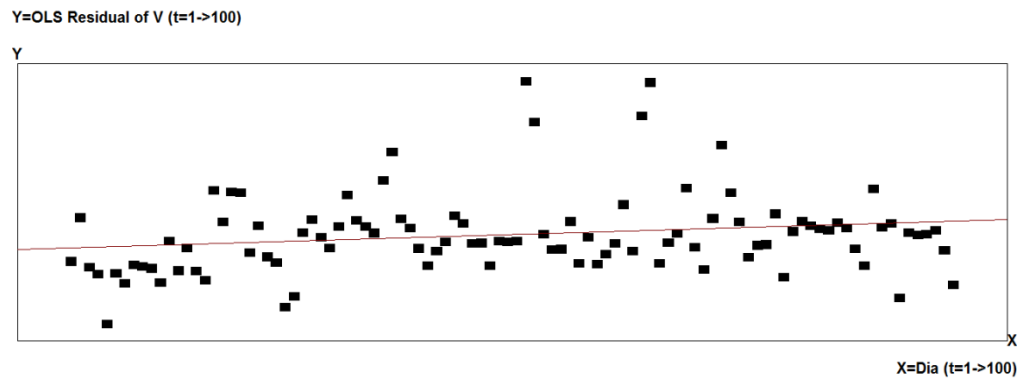


Gráfico 5: $\ln R_i$ vs errores estimados²

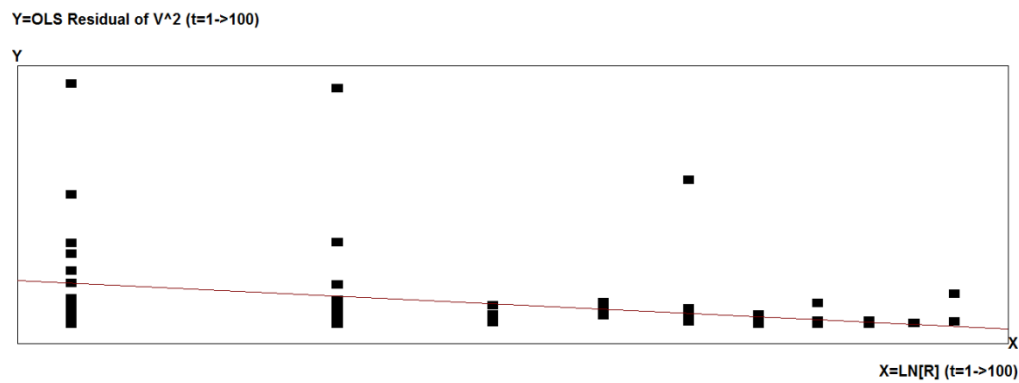


Gráfico 6: $\frac{1}{S_i}$ vs errores estimados²

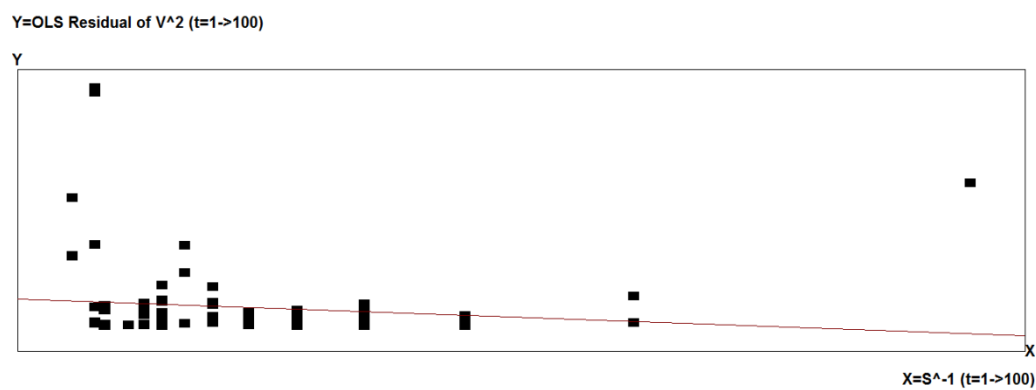


Gráfico 7: D_i vs errores estimados²

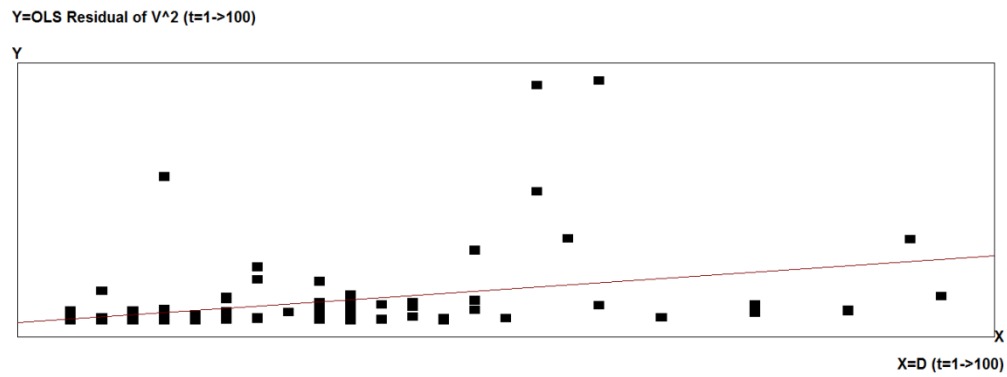


Gráfico 8: Observaciones vs errores estimados²

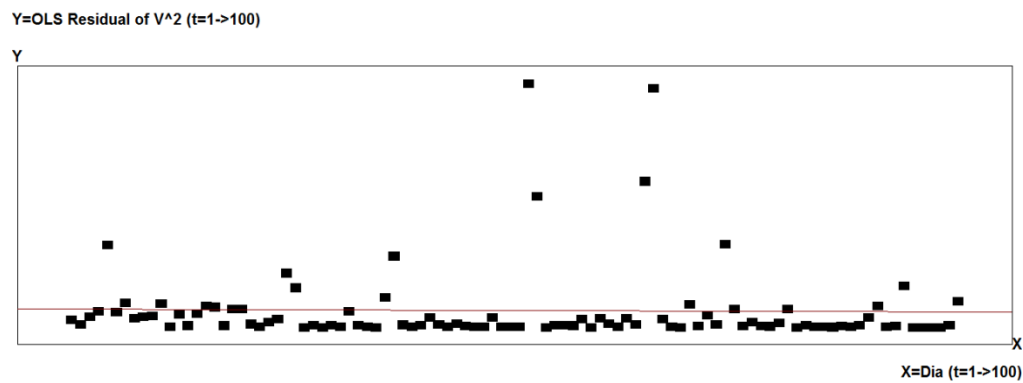


Gráfico 9: V_i vs errores estimados

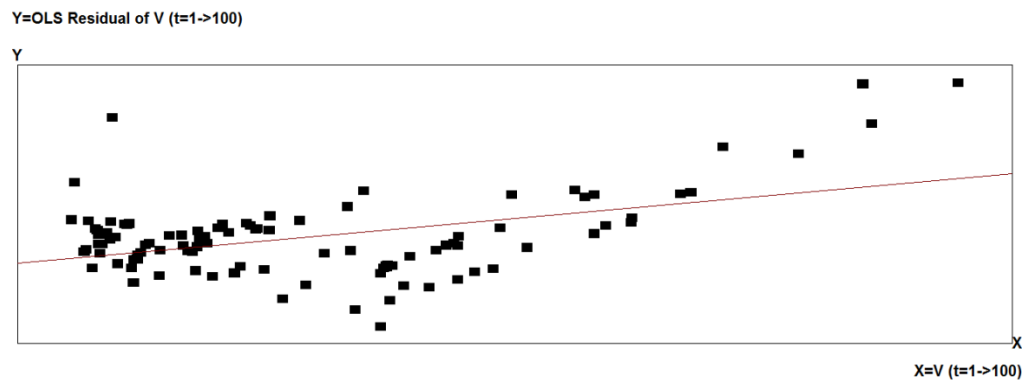
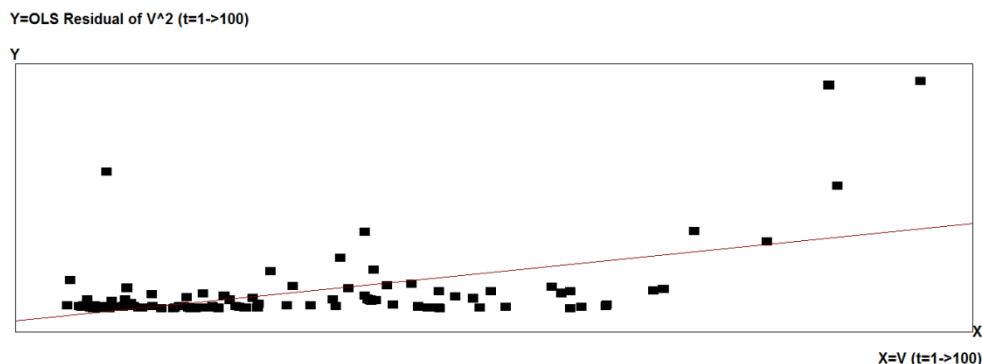


Gráfico 10: V_i vs errores estimados²



De acuerdo a los gráficos anteriores puede percibirse un comportamiento regular de las variables $\ln R_i$ y $\frac{1}{S_i}$ con respecto a los errores estimados que podrían estar causando un problema de heteroscedasticidad.

Pregunta 3

- a. El Gerente de la compañía, el Dr. Chapatí, preocupado por la posibilidad de que el modelo no sea útil en el análisis del comportamiento de las ventas, le pide que compruebe si en realidad existe un problema con el modelo; más concretamente su inquietud es: ¿Existe un problema de Heteroscedasticidad? (Realice las pruebas que considera necesarias). Sea claro en las hipótesis y reglas de decisión de cada una de las pruebas realizadas.

Para saber si existe este problema, es necesario hacer las pruebas formales:

i. Test de Goldfeld y Quandt

Los datos corresponden a series de tiempo y por tanto no tiene sentido organizarlo

La hipótesis a contrastar es:

$$H_o: \sigma_i^2 = \sigma^2$$

$$H_a: \sigma_i^2 = \sigma^2 X_i^2$$

A pesar de que hay sospecha sobre dos variables que puedan estar causando el problema de heteroscedasticidad, se procederá a efectuar la prueba con todas las variables explicativas. Así, para realizar esta prueba primero se debe ordenar de menor a mayor de acuerdo con la(s) variable(s) que se crea está(n) causando el problema. Después de esto se deben omitir los $d < \frac{1}{5}n$ (25) datos de la mitad, para después correr las dos regresiones.

Logaritmo natural del número de reclamos en el día i ($\ln R_i$)

La primera muestra contiene las observaciones 1 a 38 y la segunda, 62 a 100. Al estimar las dos regresiones a partir de las dos submuestras, tenemos que el estadístico F_{GQ} es:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{67,37680349}{13,88545418} = 4,85232983$$

$$0.05 : 1.50$$

$$0.1 :$$

Luego comparamos este valor con $F(n-d-2k, n-d-2k) \alpha = 0,01 \rightarrow F(68, 68) \alpha = 0,01 = 1,7846$. De este modo, tenemos que el estadístico Goldfeld y Quandt es mayor que el valor crítico con 68 grados de libertad en el numerador y 68 grados de libertad en el denominador, por lo cual podemos rechazar H_0 a un nivel de significancia del 1% y puede inferirse que la variable $Ln R_i$, presenta un problema de heteroscedasticidad del tipo de Goldfeld y Quandt. (Lo mismo ocurre al 95% y 90% de confianza)

Ahora, seguimos el mismo procedimiento para las otras dos variables independientes y obtenemos el siguiente cuadro que resume los resultados de esta prueba para las tres variables:

Cuadro 1: Resultados de la Prueba de Goldfeld y Quandt

	Nombre de la variable		
	$Ln R_i$	$\frac{1}{S_i}$	D_i
SSE1	13,88545418	13,1172488	200,5495829
SSE2	67,37680349	60,59268977	16,33502572
F_{GQ}	4,85232983	4,6193139	12,27727378

Recuerden que además debían hacer la prueba de Goldfeld y Quandt cuya hipótesis nula es que todas las variables causan el problema, ésta se contrasta con el estadístico que arroja EasyReg. A partir de esta prueba se puede concluir que las tres variables están presentando problemas de heteroscedasticidad del tipo de Goldfeld-Quandt.

ii. Prueba de Breush-Pagan

En esta prueba se consideran la hipótesis de una relación entre la varianza del error y un grupo de variables como medida de heteroscedasticidad (en este caso todas las variables del modelo), versus la hipótesis nula de la no existencia de esta relación como medida de homoscedasticidad, es decir:

$$H_0: \sigma_i^2 = \sigma^2$$

$$H_a: \sigma_i^2 = f(\gamma + \delta Z_i) \quad i = 1, 2, \dots, 100$$

Obtenemos la SSE del modelo original y dividimos por n para hallar σ_i^2

$$\sigma_i^2 = \frac{SSE}{n} = \frac{345,71167454}{100} = 3,4571167454$$

Con este resultado creamos la variable $\frac{\widehat{\varepsilon_i^2}}{\sigma_i^2}$ que utilizaremos como variable dependiente (VDRA_i) del modelo original.

$$VDRA_i = \beta_1 + \beta_2 Z_i + \mu_i$$

Al correr el modelo restamos SST-SSE para obtener SSR

$$SSR = SST - SSE$$

$$SSR = 561.84022631 - 449.60335167$$

$$SSR = 112,23687464$$

Calculamos ahora el estadístico BP definido como

$$BP = SSR / 2$$

$$BP = 112,23687464 / 2$$

$$BP = 56,11843732$$

Este estadístico se compara con el valor chi cuadrado con g grados de libertad y nivel de significancia α , es decir, con $\chi_{3,0,01}^2 = 11,34$. Dado que $BP > \chi_{3,0,01}^2$ entonces podemos rechazar la hipótesis nula de homoscedasticidad a favor de la presencia de una heteroscedasticidad.

iii. Prueba de White

La hipótesis a contrastar es:

$$H_0: \sigma_i^2 = \sigma^2$$

$$H_a: \text{No } H_0$$

Para realizar esta prueba, primero corremos el modelo original para encontrar la serie de residuos $\hat{\varepsilon}_i$. Luego, corremos la siguiente regresión auxiliar:

$$\varepsilon_i^2 = \beta_1 + \beta_2 \text{Ln}R_i + \beta_3 \frac{1}{S_i} + \beta_4 D_i^2 + \beta_5 \text{Ln}R_i^2 + \beta_6 \frac{1}{S_i^2} + \beta_7 D_i^2 + \beta_8 \text{Ln}R_i * \frac{1}{S_i} + \beta_9 \text{Ln}R_i * D_i + \beta_{10} \frac{1}{S_i} * D_i + \varepsilon_i \quad i = 1, 2, \dots, 100$$

El Estadístico que se debe utilizar es:

$$W_\alpha = nR^2$$

Dado que se obtuvo $R^2 = 40,78$, el estadístico de White en este caso es:

$$W_\alpha = 100 * 0.4078 = 40,78$$

El criterio de rechazo de H_0 es: $W_\alpha = \chi_{g(\alpha)}^2$

En este caso g es igual a 9 (# de parámetros de la regresión auxiliar menos 1), por ello, con un nivel de significancia del 1% el valor crítico es 21,66.

Así, dado que $W_\alpha = 40,78 > 21,66$, con un nivel de confianza del 99% se puede rechazar la hipótesis nula de que existe homoscedasticidad (lo mismo ocurre al 95% y 90% de confianza)

Finalmente, se ha comprobado que si existe el problema de heteroscedasticidad pero no se dio a conocer la forma funcional del mismo, es decir, no se pudo conocer cuál es la variable que está causando el problema.

- b.** Si encontró un problema de heteroscedasticidad, ¿cómo solucionaría el problema? Explique claramente el por qué de su respuesta. Sea lo más claro y preciso.

Para solucionar el problema de heteroscedasticidad existen dos caminos. Si conocemos la varianza de cada uno de los errores usamos el método de Mínimos Cuadrados Ponderados (MCP) donde basta dividir a ambos lados del modelo por σ_i y obtener la varianza. Así, cada observación tendrá varianza constante. Pero en este caso, no conocemos esa información ya que de acuerdo a los resultados anteriores, todas las variables están ocasionando de una u otra forma un problema de heteroscedasticidad, por lo tanto debemos utilizar la solución de White que utiliza un estimador consistente como el ilustrado a continuación para la matriz de varianzas y covarianzas de los betas estimados con el objetivo de corregir el problema

$$\text{Est. Var}(\hat{\beta}) = n(X^T X)^{-1} S_0 (X^T X)^{-1}$$

$$\text{Donde } S_0 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 x_i x_i^T$$

Pregunta 4

- a.** Uno de sus compañeros de investigación, el Dr. Lucas Tañeda, sugiere que el problema econométrico puede corregirse mediante una serie de transformaciones del modelo. ¿Tiene razón? Corrija (de ser posible) el problema de la forma como usted lo considere correcto y reporte sus resultados.

Como se mencionó en el punto anterior, se ha optado por emplear la corrección de White de la varianza de los coeficientes. Los resultados se encuentran reportados en la Tabla 1.

- b. Si la corrección del problema fue posible, muestre que ahora el modelo no tiene problemas de heteroscedasticidad.

Dado que se aplicó la corrección de White, no es posible demostrar que el problema se solucionó pues cuando volvemos a efectuar las pruebas formales realizadas en el punto 3.a, los resultados son los mismos.

Pregunta 5

Interprete los coeficientes estimados teniendo en cuenta su significancia y discuta si los signos fueron los esperados.

$\widehat{\beta}_1$: 15,0246. Este coeficiente es significativo al 99% de confianza. Carece de interpretación económica.

$\widehat{\beta}_2$: -2,495. Ante un aumento del uno por ciento en el número de reclamos en el día i , se espera una disminución de 0,0249 millones de pesos en las ventas diarias de la cadena de almacenes. Significativo al 99%. El signo fue el esperado.

$\widehat{\beta}_3$: -24,4224. Ante un aumento del uno por ciento en el número de ofertas diarias, genera en promedio, un aumento de $0,244 \cdot \frac{1}{S_i}$ millones de pesos en las ventas diarias de la cadena de almacenes. Significativo al 99%. El signo fue el esperado.

$\widehat{\beta}_4$: 0,0606. Este coeficiente no es significativo, por lo tanto, podría decirse que individualmente no influye en las ventas diarias de la cadena de almacenes.

Pregunta 6

- a. El Dr. Fredonio Gordorritúa, asesor principal del líder de la investigación, le solicita que concluya si los regresores empleados explican o no el comportamiento de las ventas de la empresa. Muestre los resultados con los que toma su decisión.

Para responder si los regresores explican las ventas diarias de la cadena de almacenes de ropa, es necesario realizar una prueba conjunta sobre todas las variables del modelo. En este caso, no es correcto emplear el F_{Global} pues depende del estimador MCO de la varianza de los errores y es este se encuentra sesgado debido al problema de Heteroscedasticidad. De este modo, lo adecuado es guiarse por el test de Wald que incluye el problema presente. En este caso, el estadístico que arroja esta prueba es 225,24, concluyendo que al menos alguna de las pendientes es estadísticamente diferente de cero al 1% de significancia. Adicional a ello, como se vio en el punto anterior, las pruebas individuales con los estadísticos t de White, arrojan que dos de las tres pendientes son significativas al 1% de significancia. De este modo, se puede concluir que los regresores empleados si explican las ventas diarias de la cadena de almacenes de ropa.

- b. Ya estando usted satisfecho con el buen trabajo que había realizado, llega María Julieta de las Nieves, una reconocida economista que le colaboró en la investigación, y le informa que la naturaleza del problema de heteroscedasticidad si podía conocerse y era: $var(\varepsilon_i) = \sigma^2 D_i$, donde D_i es una variable dummy que toma el valor de uno si la sucursal queda en el norte del país y cero en caso contrario. Ella afirma que el modelo se puede transformar dividiendo por $\sqrt{D_i}$, para que los estimadores MCO de los β 's sean insesgados y eficientes. Si es posible, realice la transformación y demuestre que el problema ha quedado solucionado.

Uno esperaría resolver el problema empleando Mínimos Cuadrados Ponderados (MCP). De tal forma que $\frac{V_i}{\sqrt{D_i}} = \frac{\beta_1}{\sqrt{D_i}} + \beta_2 \frac{\ln R_i}{\sqrt{D_i}} + \beta_3 \frac{\frac{1}{S_i}}{\sqrt{D_i}} + \beta_4 \frac{D_i}{\sqrt{D_i}} + \frac{\varepsilon_i}{\sqrt{D_i}}$. Hasta ahí parecería razonable que se puede transformar el modelo, para que los estimadores MCO de los β 's sean insesgados y eficientes. No obstante, noten que se puede dar el caso en que D_i sea cero y sería imposible dividir por cero. Por ello, para este caso esa transformación es imposible.