

**Taller #5  
Econometría 06216**

**Profesor: Julio César Alonso**  
**Monitora: Valentina Gatti**

**Notas:**

- o Recuerde que tres preguntas, seleccionadas al azar, serán calificadas.
- o Este taller debe subirse en la plataforma Moodle hasta la 7:10 del 14 de Septiembre de 2009. **Sólo se recibirán talleres en formato pdf.** Cualquier otro formato no será calificado.

**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 Presidente del Estado de Morita desea conocer la situación de las finanzas públicas, del país para el periodo 1990-2007. Para esto le pide al ministro de hacienda pública que estime los ingresos del sector público, de acuerdo a la base de datos que se encuentra en el archivo T5-02-09.xls. Usted como asesor principal del ministro intuye que el modelo econométrico que explica los ingresos del sector público, es el siguiente:

$$IngPub_t = \alpha_0 + \alpha_1 PIB_t + \alpha_2 S_t + \alpha_3 NT_t + e_t \quad (1)$$

Donde  $IngPub$ ,  $PIB$ ,  $S$ ,  $Nt$ , son los ingresos públicos de Morita, el PIB del país a precios constantes de 2000, el ahorro privado, las exportaciones netas, respectivamente, todas las variables medidas a en millones de dólares.

1. Con base en la información anterior:
  - a) Estime el modelo (1) y reporte sus resultados en una tabla.
  - b) Analice la significancia individual y conjunta ¿Presenta un buen ajuste el modelo?
2. Determine si existe o no algún problema econométrico (muestre todo su trabajo y como llega a sus conclusiones)

3. Corrija el problema encontrado y reporte el modelo estimado. Explique claramente su decisión.

La firma de asesores *Microeconomics* se encuentra realizando un estudio para explicar la cantidad demandada de Café en el mercado de exportaciones, y ha encontrado que el mejor modelo para estimar la cantidad demandada de café en el exterior es:

$$Qcafe_i = \beta_1 \times Pcafe^{\beta_2} \times Pchoco^{\beta_3} \times Pte^{\beta_4} \times YD^{\beta_5} \quad (2)$$

Donde  $Qcafe$  es la cantidad demandada de café (medida en millones de costales),  $Pcafe$  es el precio de un costal de café (medido en mil dólares),  $Pchoco$  y  $Pte$  son los precios de dos bienes alternativos (chocolate y té, medido en mil dólares),  $YD$  es la renta disponible media de los consumidores de café en el exterior. La información correspondiente se encuentra en el archivo T5-02-09.xls. hoja 2.

4. De acuerdo a la información anterior:
  - a) Plantee el modelo econométrico a estimar.
  - b) Estime el modelo y presente sus resultados en una tabla.
5. Determine si existe algún problema econométrico (muestre todo su trabajo y como llega a sus conclusiones)
6. Continuando con la pregunta anterior:
  - a) Corrija el problema econométrico y estime el nuevo modelo.
  - b) Interprete los coeficientes del nuevo modelo.

**Taller #5**  
**Econometría 06216**  
**Respuestas sugeridas**

**Profesor: Julio César Alonso**  
**Monitora: Valentina Gatti**

**Notas:**

- o Recuerde que tres preguntas, seleccionadas al azar, serán calificadas.
- o Este taller debe subirse en la plataforma Moodle hasta la 7:10 del 14 de Septiembre de 2009. **Sólo se recibirán talleres en formato pdf.** Cualquier otro formato no será calificado.

**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 Presidente del Estado de Morita desea conocer la situación de las finanzas públicas, del país para el periodo 1990-2007. Para esto le pide al ministro de hacienda pública que estime los ingresos del sector público, de acuerdo a la base de datos que se encuentra en el archivo T5-02-09.xls. Usted como asesor principal del ministro intuye que el modelo econométrico que explica los ingresos del sector público, es el siguiente:

$$IngPub_t = \alpha_0 + \alpha_1 PIB_t + \alpha_2 S_t + \alpha_3 NT_t + e_t \quad (1)$$

Donde  $IngPub$ ,  $PIB$ ,  $S$ ,  $Nt$ , son los ingresos públicos de Morita, el PIB del país a precios constantes de 2000, el ahorro privado, las exportaciones netas, respectivamente, todas las variables medidas a en millones de dólares.

1. Con base en la información anterior:
  - a) Estime el modelo (1) y reporte sus resultados en una tabla. [Los resultados se presentan en la tabla 1](#)
  - b) Analice la significancia individual y conjunta ¿Presenta un buen ajuste el modelo?  
 Únicamente el coeficiente asociado a  $PIB_t$  es significativo a un 99% de confianza. El resto de los coeficiente incluido el intercepto (incluido el intercepto) no son significativos individualmente a ningún nivel. Para verificar la significancia

conjunta, se emplea el estadístico F. Este es igual a 7934,07, y un valor-p=0, lo que permite rechazar la hipótesis nula de que todos los coeficientes son conjuntamente iguales a cero a un nivel de significancia del 1%. El  $R^2$  es igual a 0,9994, es decir el 99,94% de la variabilidad en el ingreso público de Morita es explicado por las variables incluidas en el modelo. Aparentemente el ajuste del modelo es muy alto. Todo lo anterior (t bajos, F altos y  $R^2$  alto) hace pensar que puede existir grave problema de multicolinealidad. Esto puede deberse a que el comportamiento del ahorro (la parte del ahorro que no se consume) y las exportaciones netas, puede estar recogido en gran parte por el PIB.

Tabla 1: Resultados de la estimación ecuación 1

Variable dependiente: $IngPub_t$	
Estadísticos t entre paréntesis	
Ecuación (1)	
MCO	
Constante	-2,0097604 (-0,447)
$PIB_t$	3,5312066 (13,936) ***
$S_t$	0,2894518 (1,162)
$NT_t$	0,1355861 (0,894)
$R^2$	0,9994
$R^2_{ajust.}$	0,9993
F	7934,07 ***
# de Obs.	18

(\*): nivel de significancia 10%  
 (\*\*): nivel de significancia 5%  
 (\*\*\*): nivel de significancia 1%  
 MCO: Mínimos Cuadrados Ordinarios

2. Determine si existe o no algún problema econométrico (muestre todo su trabajo y como llega a sus conclusiones)

Se emplean las pruebas vistas en clase para determinar la presencia de multicolinealidad.

**Matriz de Correlaciones entre las X's.**

Se calcula el determinante de la matriz de correlación  $|R|$  empleando los valores propios obtenidos:

$$|R| = 2,09634851 \times 0,89880671 \times 0,00484477358 = 0,00912858$$

El valor que toma el determinante es cercano a cero, lo que implica que hay un problema de multicolinealidad en el modelo.

**Medida de Besley, Kuck y Welsch (1980)**

Los valores propios (eigenvalues) de la matriz son:  $\lambda_1 = 2,09634851$ ,  $\lambda_2 = 0,89880671$  y  $\lambda_3 = 0,00484477358$ . Se calcula el número de condición:

$$k(X) = \frac{\sqrt{\lambda_{max}}}{\sqrt{\lambda_{min}}} = \frac{\sqrt{2.09634851}}{\sqrt{0.00484477358}} = 20,80152$$

Como el valor de  $k(X)$  es distinto de 1, y se encuentra entre 20 y 30, entonces sí existe un problema de multicolinealidad y es grave.

**Matriz de correlación entre los coeficientes estimados:**

Se determina a partir de la matriz de varianza y covarianza estimada. La matriz es igual a:

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_0$
$\hat{\alpha}_1$	1	-0.99486	-0.35697	0.341033
$\hat{\alpha}_2$		1	0.33446	-0.33512
$\hat{\alpha}_3$			1	-0.98507
$\hat{\alpha}_0$				1

A partir de la matriz de correlación se determina que existe una fuerte correlación entre  $\hat{\alpha}_2$  y  $\hat{\alpha}_1$  cercana a -0,9, por tanto se observa un problema de multicolinealidad no perfecta grave, entre el PIB y el ahorro.

3. Corrija el problema encontrado y reporte el modelo estimado. Explique claramente su decisión.

Si bien la variable  $PIB_t$  está fuertemente correlacionada con la variable  $S_t$ , y dada la teoría económica se sabe que la el PIB depende de una serie de elementos entre ellos la inversión y el consumo, el cuál determinarán el nivel de ahorro. Sin embargo desde la teoría económica no se puede decir que el PIB y el ahorro miden el mismo concepto, ni es lo mismo, así estén fuertemente relacionadas por la dinámica de la economía. Por tanto el problema no se puede solucionar, a partir de la teoría económica.

La firma de asesores *Microeconomics* se encuentra realizando un estudio para explicar la cantidad demandada de Café en el mercado de exportaciones, y ha encontrado que el mejor modelo para estimar la cantidad demandada de café en el exterior es:

$$Qcafe_i = \beta_1 \times Pcafe^{\beta_2} \times Pchoco^{\beta_3} \times Pte^{\beta_4} \times YD^{\beta_5} \quad (2)$$

Donde  $Qcafe$  es la cantidad demandada de café (medida en millones de costales),  $Pcafe$  es el precio de un costal de café (medido en mil dólares),  $Pchoco$  y  $Pte$  son los precios de dos bienes alternativos (chocolate y té, medido en mil dólares),  $YD$  es la renta disponible

media de los consumidores de café en el exterior. La información correspondiente se encuentra en el archivo T5-02-09.xls.hoja 2.

4. De acuerdo a la información anterior:

a) Plantee el modelo econométrico a estimar.

De acuerdo con la información anterior el modelo econométrico es de la siguiente forma:

$$\ln Qcafe_i = \ln(\beta_1 \times Pcafe^{\beta_2} \times Pchoco^{\beta_3} \times Pte^{\beta_4} \times YD^{\beta_5})$$

$$\ln[Qcafe_i] = \ln \beta_1 + \beta_2 \ln[Pcafe_i] + \beta_3 \ln[Pchoco_i] + \beta_4 \ln[Pte_i] + \beta_5 \ln[YD_i]$$

Reparametrizando y agregando el término de error:

$$\ln[Qcafe_i] = \alpha + \beta_2 \ln[Pcafe_i] + \beta_3 \ln[Pchoco_i] + \beta_4 \ln[Pte_i] + \beta_5 \ln[YD_i] + \mu_i \quad (2.1)$$

Donde  $\ln \beta_1 = \alpha$

b) Estime el modelo y presente sus resultados en una tabla.

Tabla 2: Resultados de la estimación

Variable dependiente: $\ln[Qcafe_i]$	Estadísticos t entre paréntesis		
	Ecuación (2.1)	Ecuación (2.2)	
	MCO	MCO/HC	
Constante	3,1600929 (1,305)	3,1600929 (1,994)	**
$\ln[Pcafe_i]$	-0,0604796 (-1,483)	-0,0604796 (-2,718)	***
$\ln[Pchoco_i]$	0,5374623 (3,324)	0,5374623 (6,277)	***
$\ln[Pte_i]$	-0,5130241 (-2,278)	-0,5130241 (-2,333)	**
$\ln[YD_i]$	0,68662 (4,111)	0,68662 (4,286)	***
R2	0,4813	0,4813	
R2 Ajustado	0,3983	0,3983	
F	5,8	5,8	***
# de obs.	30	30	

(\*): nivel de significancia 10%

(\*\*): nivel de significancia 5%

(\*\*\*): nivel de significancia 1%

MCO: Mínimos Cuadrados Ordinarios

5. Determine si existe algún problema econométrico (muestre todo su trabajo y como llega a sus conclusiones)

En primer lugar es necesario tener en cuenta que los datos son de corte transversal lo cual hace muy probable la existencia del problema de heterocedasticidad. Por esto se realiza un análisis intuitivo a través de los siguientes gráficos.

Grafico 1: Residuos Estimados Vs. Observaciones

Y=OLS Residual of LN[Qcafe] (t=1->30)

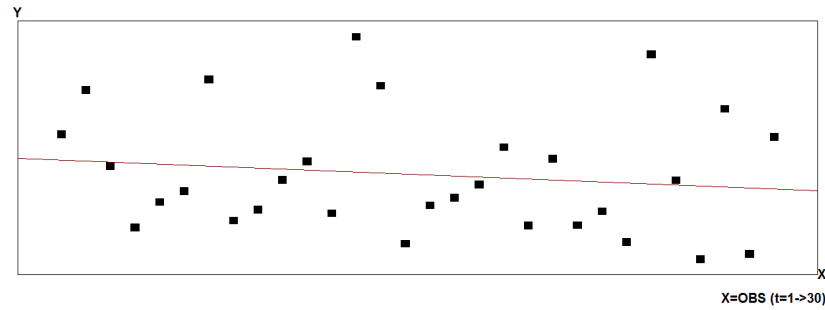


Grafico 2: Residuos Estimados Vs. Ln(Pcafe)

Y=OLS Residual of LN[Qcafe] (t=1->30)

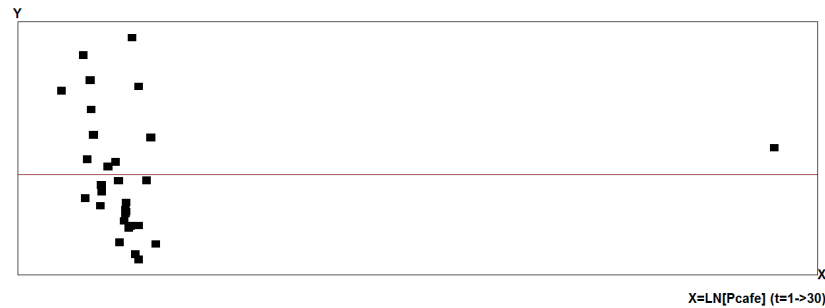


Grafico 3: Residuos Estimados Vs. Ln(Pchoco)

Y=OLS Residual of LN[Qcafe] (t=1->30)

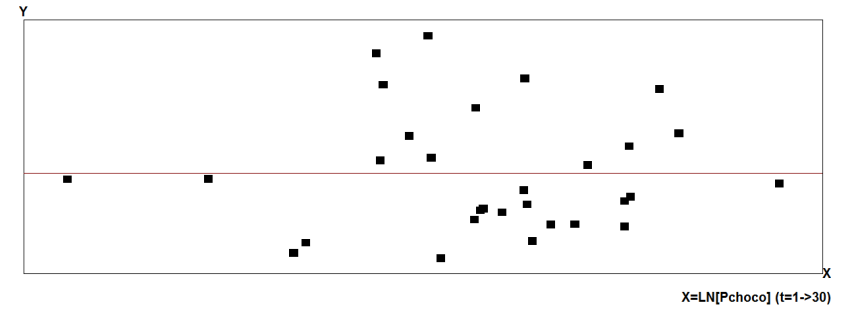


Grafico 4: Residuos Estimados Vs. Ln(Pte)

Y=OLS Residual of LN[Qcafe] (t=1->30)

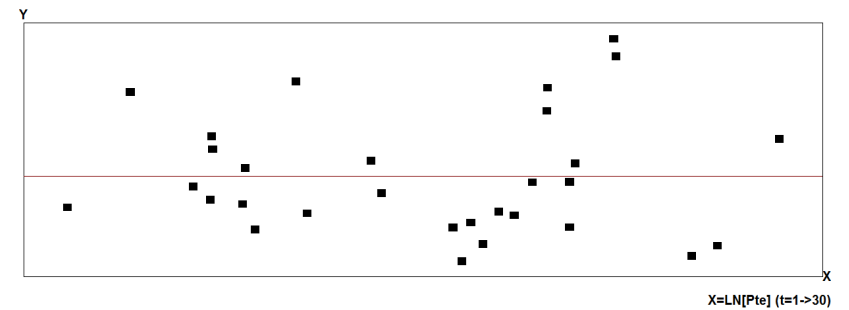
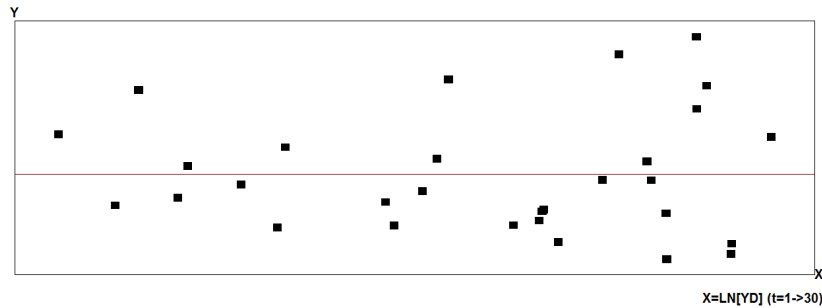


Grafico 5: Residuos Estimados Vs. Ln(YD)

Y=OLS Residual of LN[Qcafe] (t=1->30)



Como se ver en las gráficas, la serie de los residuos no parece tener una varianza constante y parece estar relacionadas con las variables explicativas.

Para estar seguros de la existencia o no de esta problema es necesario hacer las pruebas formales.

La primera prueba que se debe hacer es la de **Goldfeld y Quant**. Las hipótesis a contrastar:

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

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

Dado que no hay claridad de cual de las variables está afectando la serie de los residuos es necesario efectuar la prueba con todas las variables explicativas. Para realizar esta prueba se debe ordenar de menor a mayor de acuerdo con la variable que se cree que está causando el problema. Después de esto se deben omitir los  $d < \frac{1}{5}n$   $d < 6$  datos de la mitad, para después correr dos regresiones. Se podría pensar en retirar 5 datos de la muestra, pero en este caso no abría dos muestras iguales, por lo que el número óptimo a retirar es 4, así cada muestra quedará de 13 observaciones.

**Logaritmo Natural del Precio del Café:**

Al correr la regresión se obtiene que la suma cuadrada de los errores para los valores bajos de la variable (Donde  $i = 1, \dots, 13$ )  $\ln[Pcafe_i]$  es:

$$SSE_1 = 0.02408534$$

Por otro lado la suma de los errores al cuadrado para los datos asociados a los valores altos de la variable (Donde  $i = 17, \dots, 30$ )  $\ln[Pcafe_i]$  es:

$$SSE_2 = 0.101828$$

De lo anterior se obtiene el estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{0.101828}{0.02408534} = 4,22779998$$

Decisión: Se rechaza la hipótesis nula si  $F_{(n-d-2k, n-d-2k)\alpha} < F_{GQ}$

Por tanto, dado que  $F_{(16,16)0,01} < 4,22779998$ , entonces  $3,37204 < 4,22779998$ , por tanto se rechaza la hipótesis nula a un 99% de confianza, por tanto la variable  $\ln[Pcafe_i]$  presenta un problema de heterocedasticidad.

**Logaritmo Natural del Precio del Chocolate:**

Al correr la regresión se obtiene que la suma cuadrada de los errores para los valores bajos de la variable (Donde  $i = 1, \dots, 13$ )  $\ln[Pchoco_i]$  es:

$$SSE_1 = 0.09559955$$

Por otro lado la suma de los errores al cuadrado para los datos asociados a los valores altos de la variable (Donde  $i = 17, \dots, 30$ )  $\ln[Pchoco_i]$  es:

$$SSE_2 = 0.05493824$$

De lo anterior se obtiene el estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{0.05493824}{0.09559955} = 0,57467049$$

Decisión: Se rechaza la hipótesis nula si  $F_{(n-d-2k, n-d-2k)\alpha} < F_{GQ}$

Por tanto, dado que  $F_{(16,16)0,01} > 0,57467049$ , entonces  $3,37204 > 0,57467049$ , por tanto No se rechaza la hipótesis nula a un 99% de confianza, por tanto la variable  $\ln[Pchoco_i]$  NO presenta un problema de heterocedasticidad. Si se compara con un  $F_{(16,16)0,1} = 1,697855 > 0,57467049$ , por tanto No se rechaza la hipótesis nula a un 90% de confianza.

**Logaritmo Natural del Precio del Té:**

Al correr la regresión se obtiene que la suma cuadrada de los errores para los valores bajos de la variable (Donde  $i = 1, \dots, 13$ )  $\ln[Pte_i]$  es:

$$SSE_1 = 0.05209367$$

Por otro lado la suma de los errores al cuadrado para los datos asociados a los valores altos de la variable (Donde  $i = 17, \dots, 30$ )  $\ln[Pte_i]$  es:

$$SSE_2 = 0.09054444$$

De lo anterior se obtiene el estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{0.09054444}{0.05209367} = 1,738300335$$

Decisión: Se rechaza la hipótesis nula si  $F_{(n-d-2k, n-d-2k)\alpha} < F_{GQ}$

Por tanto, dado que  $F_{(16,16)0,01} > 1,91280572$ , entonces  $3,37204 > 1,738300335$ , por tanto No se rechaza la hipótesis nula a un 99% de confianza, por tanto la variable  $\ln[Pte]$  NO presenta un problema de heterocedasticidad tipo Goldfeld y Quant. Si se compara con un  $F_{(16,16)0,1} = 1,697855 < 1,738300335$ , por tanto se rechaza la hipótesis nula a un 90% de confianza, por tanto existe un problema de heteroscedasticidad de tipo Golfeld y Quant causado por la variable  $\ln[Pte]$ .

**Logaritmo Natural de la Renta disponible:**

Al correr la regresión se obtiene que la suma cuadrada de los errores para los valores bajos de la variable (Donde  $i = 1, \dots, 13$ )  $\ln[YD_i]$  es:

$$SSE_1 = 0.05293096$$

Por otro lado la suma de los errores al cuadrado para los datos asociados a los valores altos de la variable (Donde  $i = 17, \dots, 30$ )  $\ln[YD_i]$  es:

$$SSE_2 = 0.10735703$$

De lo anterior se obtiene el estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{0.10735703}{0.05293096} = 2,028$$

Decisión: Se rechaza la hipótesis nula si  $F_{(n-d-2k, n-d-2k)\alpha} < F_{GQ}$

Por tanto, dado que  $F_{(16,16)0,01} > 2,028$ , entonces  $3,37204 > 2,028$ , por no se rechaza la hipótesis nula a un 99% de confianza, por tanto la variable  $\ln[YD]$  NO presenta un problema de heterocedasticidad. Sin embargo si se compara con un  $F_{(16,16)0,1} = 1,697855 < 2,028$ , por tanto se rechaza la hipótesis nula a un 90% de confianza, por tanto existe un problema de heteroscedasticidad de tipo Golfeld y Quant causado por la variable  $\ln[YD]$ .

**Test de Breush-Pagan**

En este caso la hipótesis nula y la alterna son las siguientes:

$$H_0: \text{Presencia de Homocedasticidad}$$

$$H_a: \text{Relacion entre la varianza del error y un grupo de variables}$$

Lo cual se puede presentar de la siguiente manera.

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

$$H_a: \sigma_i^2 = \sigma^2(\gamma + \delta X_i)$$

Al correr la regresión del modelo planteado inicialmente (2.1) (pues se considera que el conjunto de variables que causa la heterocedasticidad son todas las variables), se obtuvo que la regresión auxiliar es:

$$\frac{\hat{\epsilon}_i^2}{\hat{\sigma}^2} = \gamma + \delta_2 \ln[Pcafe_i] + \delta_3 \ln[Pchoco_i] + \delta_4 \ln[Pte_i] + \delta_5 \ln[YD_i] + \mu_i$$

De donde se obtiene que SSR=6,91950284. Por tanto el BP=3,45975142.

Dado que la toma de la decisión se realiza de la siguiente manera Rechazar  $H_0$  cuando  $BP > x_g^2$

En este caso g es igual a 4, por tanto con un nivel de significancia del 1% el valor crítico es igual a 13,2767. Así mismo con un nivel de significancia del 5% y del 10% el valor crítico es igual a 9,487 y 7,77, respectivamente.

Por tanto dado que el  $BP < x_g^2$  con todos los niveles de significancia (10%, 5% y 1%), entonces se no existe suficiente evidencia para rechazar la hipótesis nula de que existe homocedasticidad.

**Test de White**

En este caso las hipótesis nula y alterna son las siguientes:

$$H_0: \text{Presencia de Homocedasticidad}$$

$$H_a: \text{Presencia de heterocedasticidad}$$

Lo cual se puede representar de la siguiente manera:

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

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

Para realizar esta prueba se debe correr la siguiente regresión auxiliar:

$$\hat{\epsilon}_i = \gamma + \delta_1 \ln[Pcafe_i] + \delta_2 \ln[Pchoco_i] + \delta_3 \ln[Pte_i] + \delta_4 \ln[YD_i] + \delta_5 (\ln[Pcafe_i])^2 + \delta_6 (\ln[Pchoco_i])^2 + \delta_7 (\ln[Pte_i])^2 + \delta_8 (\ln[YD_i])^2 + \delta_9 \ln[Pcafe_i] \ln[Pchoco_i] + \delta_{10} \ln[Pcafe_i] \ln[Pte_i] + \delta_{11} \ln[Pcafe_i] \ln[YD_i] + \delta_{12} \ln[Pchoco_i] \ln[Pte_i] + \delta_{13} \ln[Pchoco_i] \ln[YD_i] + \delta_{14} \ln[Pte_i] \ln[YD_i] + \mu_i$$

El estadístico que se debe utilizar es:

$$Wa = nR^2$$

Dado que se obtuvo que  $R^2 = 0.3678$ , tenemos que el estadístico de White es:

$$Wa = 30 * 0.3678 = 11,034$$

En este test la toma de decisión se realiza de la siguiente manera:

Rechazar  $H_0$  cuando  $Wa > x_{g(\alpha)}^2$ . En este caso  $g$  es igual a 14, ( $g = \#$  de parámetros de la regresión auxiliar menos 1), por tanto con un nivel de significancia del 1% el valor crítico es 29,1412.

Por tanto dado que  $Wa = 11,034 < 29,1412$ , con una confianza del 99% no se puede rechazar la hipótesis nula de que existe homocedasticidad. (lo mismo sucede con 95% y 90% de confianza).

**6.** Continuando con la pregunta anterior:

**a)** Corrija el problema econométrico y estime el nuevo modelo.

De acuerdo con los resultados del punto anterior, la prueba Goldfeld y Quant es válida para una sola variable, en este caso, simultáneamente tres variables están causando este problema, lo cual permite pensar que el problema se aproxima más a la teoría Breush-Pagan. Sin embargo esta prueba no trae nada concluyente al igual que el test de White. Por tanto como no conocemos la forma específica de la heteroscedasticidad, es decir la naturaleza del problema, se debe recurrir como último recurso a la solución de White para estimar el modelo.

Los resultados de la estimación del modelo corregido se encuentran en la tabla 2, en el literal 4.b.

**b)** Interprete los coeficientes del nuevo modelo.

$\alpha$ : Es significativo al 95% de confianza. Carece de interpretación económica.

$\beta_2 = -0,0604796$  Es significativo al 99% de confianza. Indica que por un aumento de 1% en el precio del café, la cantidad demandada de café disminuye en 0,060%.

$\beta_3 = 0,5374623$  Es significativo al 99% de confianza. Indica que por un aumento de 1% en el precio del chocolate, la cantidad demandada de café aumenta en 0,5374%. Por tanto son bienes sustitutos.

$\beta_4 = -0,5130241$  Es significativo al 95% de confianza, Indica que por un aumento de 1% en el precio del te, la cantidad demandada de café disminuye en 0,513%.

$\beta_5 = 0,68662$  Es significativo al 99% de confianza. Indica que por un aumento de 1% en la renta disponible, la cantidad demandada de café aumenta en 0,68662%.

**Taller #5**  
**Econometría 06216**  
**Respuestas sugeridas**

**Profesor: Julio César Alonso**

**Monitora: Valentina Gatti**

**Notas:**

- o Recuerde que tres preguntas, seleccionadas al azar, serán calificadas.
- o Este taller debe subirse en la plataforma Moodle hasta la 7:10 del 14 de Septiembre de 2009. **Sólo se recibirán talleres en formato pdf.** Cualquier otro formato no será calificado.

**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 Presidente del Estado de Morita desea conocer la situación de las finanzas públicas, del país para el periodo 1990-2007. Para esto le pide al ministro de hacienda pública que estime los ingresos del sector público, de acuerdo a la base de datos que se encuentra en el archivo T5-02-09.xls. Usted como asesor principal del ministro intuye que el modelo econométrico que explica los ingresos del sector público, es el siguiente:

$$IngPub_t = \alpha_0 + \alpha_1 PIB_t + \alpha_2 S_t + \alpha_3 NT_t + e_t \quad (1)$$

Donde  $IngPub$ ,  $PIB$ ,  $S$ ,  $Nt$ , son los ingresos públicos de Morita, el PIB del país a precios constantes de 2000, el ahorro privado, las exportaciones netas, respectivamente, todas las variables medidas a en millones de dólares.

1. Con base en la información anterior:

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

Los resultados se presentan en la tabla 1

b) Analice la significancia individual y conjunta ¿Presenta un buen ajuste el modelo?

Únicamente el coeficiente asociado a  $PIB_t$  es significativo a un 99% de confianza. El resto de los coeficiente incluido el intercepto (incluido el intercepto) no son significativos individualmente a ningún nivel. Para verificar la significancia

conjunta, se emplea el estadístico F. Este es igual a 7934,07, y un valor-p=0, lo que permite rechazar la hipótesis nula de que todos los coeficientes son conjuntamente iguales a cero a un nivel de significancia del 1%.

El  $R^2$  es igual a 0,9994, es decir el 99,94% de la variabilidad en el ingreso público de Morita es explicado por las variables incluidas en el modelo. Aparentemente el ajuste del modelo es muy alto. Todo lo anterior (t bajos, F altos y  $R^2$  alto) hace pensar que puede existir grave problema de multicolinealidad. Esto puede deberse a que el comportamiento del ahorro (la parte del ahorro que no se consume) y las exportaciones netas, puede estar recogido en gran parte por el PIB.

Tabla 1: Resultados de la estimación ecuación 1

Variable dependiente: $IngPub_t$	
Estadísticos t entre paréntesis	
Ecuación (1)	
MCO	
Constante	-2,0097604 (-0,447)
$PIB_t$	3,5312066 (13,936) ***
$S_t$	0,2894518 (1,162)
$NT_t$	0,1355861 (0,894)
$R^2$	0,9994
$R^2_{ajust.}$	0,9993
F	7934,07 ***
# de Obs.	18

(\*): nivel de significancia 10%

(\*\*): nivel de significancia 5%

(\*\*\*): nivel de significancia 1%

MCO: Mínimos Cuadrados Ordinarios

2. Determine si existe o no algún problema econométrico (muestre todo su trabajo y como llega a sus conclusiones)

Se emplean las pruebas vistas en clase para determinar la presencia de multicolinealidad.

**Matriz de Correlaciones entre las X's.**

Se calcula el determinante de la matriz de correlación  $|R|$  empleando los valores propios obtenidos:

$$|R| = 2,09634851 \times 0,89880671 \times 0,00484477358 = 0,00912858$$

El valor que toma el determinante es cercano a cero, lo que implica que hay un problema de multicolinealidad en el modelo.

**Medida de Besley, Kuck y Welsch (1980)**

Los valores propios (eigenvalues) de la matriz son:  $\lambda_1 = 2,09634851$ ,  $\lambda_2 = 0,89880671$  y  $\lambda_3 = 0,00484477358$ . Se calcula el número de condición:

$$k(X) = \frac{\sqrt{\lambda_{max}}}{\sqrt{\lambda_{min}}} = \frac{\sqrt{2.09634851}}{\sqrt{0.00484477358}} = 20,80152$$

Como el valor de  $k(X)$  es distinto de 1, y se encuentra entre 20 y 30, entonces sí existe un problema de multicolinealidad y es grave.

**Matriz de correlación entre los coeficientes estimados:**

Se determina a partir de la matriz de varianza y covarianza estimada. La matriz es igual a:

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_0$
$\hat{\alpha}_1$	1	-0.99486	-0.35697	0.341033
$\hat{\alpha}_2$		1	0.33446	-0.33512
$\hat{\alpha}_3$			1	-0.98507
$\hat{\alpha}_0$				1

A partir de la matriz de correlación se determina que existe una fuerte correlación entre  $\hat{\alpha}_2$  y  $\hat{\alpha}_1$  cercana a -0,9, por tanto se observa un problema de multicolinealidad no perfecta grave, entre el PIB y el ahorro.

3. Corrija el problema encontrado y reporte el modelo estimado. Explique claramente su decisión.

Si bien la variable  $PIB_t$  está fuertemente correlacionada con la variable  $S_t$ , y dada la teoría económica se sabe que la el PIB depende de una serie de elementos entre ellos la inversión y el consumo, el cuál determinarán el nivel de ahorro. Sin embargo desde la teoría económica no se puede decir que el PIB y el ahorro miden el mismo concepto, ni es lo mismo, así estén fuertemente relacionadas por la dinámica de la economía. Por tanto el problema no se puede solucionar, a partir de la teoría económica.

La firma de asesores *Microeconomics* se encuentra realizando un estudio para explicar la cantidad demandada de Café en el mercado de exportaciones, y ha encontrado que el mejor modelo para estimar la cantidad demandada de café en el exterior es:

$$Q_{cafe_i} = \beta_1 \times P_{cafe}^{\beta_2} \times P_{choco}^{\beta_3} \times P_{te}^{\beta_4} \times YD^{\beta_5} \quad (2)$$

Donde  $Q_{cafe}$  es la cantidad demandada de café (medida en millones de costales),  $P_{cafe}$  es el precio de un costal de café (medido en mil dólares),  $P_{choco}$  y  $P_{te}$  son los precios de dos bienes alternativos (chocolate y té, medido en mil dólares),  $YD$  es la renta disponible

media de los consumidores de café en el exterior. La información correspondiente se encuentra en el archivo T5-02-09.xls. hoja 2.

4. De acuerdo a la información anterior:  
a) Plantee el modelo econométrico a estimar.

De acuerdo con la información anterior el modelo econométrico es de la siguiente forma:

$$\ln Q_{cafe_i} = \ln(\beta_1 \times P_{cafe}^{\beta_2} \times P_{choco}^{\beta_3} \times P_{te}^{\beta_4} \times YD^{\beta_5})$$

$$\ln[Q_{cafe_i}] = \ln \beta_1 + \beta_2 \ln[P_{cafe_i}] + \beta_3 \ln[P_{choco_i}] + \beta_4 \ln[P_{te_i}] + \beta_5 \ln[YD_i]$$

Reparametrizando y agregando el término de error:

$$\ln[Q_{cafe_i}] = \alpha + \beta_2 \ln[P_{cafe_i}] + \beta_3 \ln[P_{choco_i}] + \beta_4 \ln[P_{te_i}] + \beta_5 \ln[YD_i] + \mu_i \quad (2.1)$$

Donde  $\ln \beta_1 = \alpha$

b) Estime el modelo y presente sus resultados en una tabla.

Tabla 2: Resultados de la estimación

	Variable dependiente: $\ln[Q_{cafe_i}]$		
	Estadísticos t entre paréntesis		
	Ecuación (2.1)	Ecuación (2.2)	
	MCO	MCO/HC	
Constante	3,1600929 (1,305)	3,1600929 (1,994)	**
$\ln[P_{cafe_i}]$	-0,0604796 (-1,483)	-0,0604796 (-2,718)	***
$\ln[P_{choco_i}]$	0,5374623 (3,324)	0,5374623 (6,277)	***
$\ln[P_{te_i}]$	-0,5130241 (-2,278)	-0,5130241 (-2,333)	**
$\ln[YD_i]$	0,68662 (4,111)	0,68662 (4,286)	***
R2	0,4813	0,4813	
R2 Ajustado	0,3983	0,3983	
F	5,8	5,8	***
# de obs.	30	30	

(\*) nivel de significancia 10%

(\*\*) nivel de significancia 5%

(\*\*\*) nivel de significancia 1%

MCO: Mínimos Cuadrados Ordinarios

5. Determine si existe algún problema econométrico (muestre todo su trabajo y como llega a sus conclusiones)

En primer lugar es necesario tener en cuenta que los datos son de corte transversal lo cual hace muy probable la existencia del problema de heterocedasticidad. Por esto se realiza un análisis intuitivo a través de los siguientes gráficos.

Grafico 1: Residuos Estimados Vs. Observaciones

Y=OLS Residual of LN[Qcafe] (t=1->30)

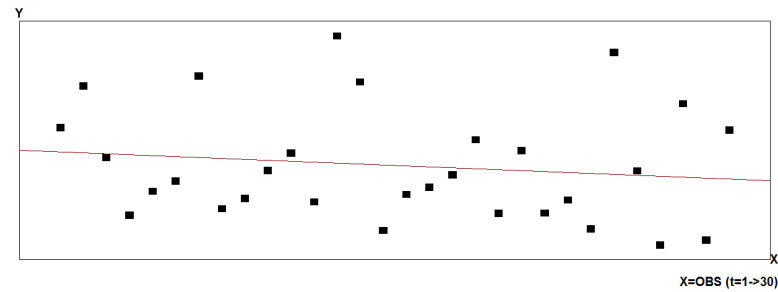


Grafico 2: Residuos Estimados Vs. Ln(Pcafe)

Y=OLS Residual of LN[Qcafe] (t=1->30)

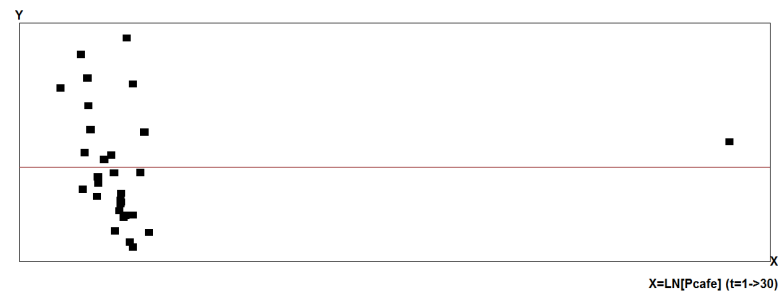


Grafico 3: Residuos Estimados Vs. Ln(Pchoco)

Y=OLS Residual of LN[Qcafe] (t=1->30)

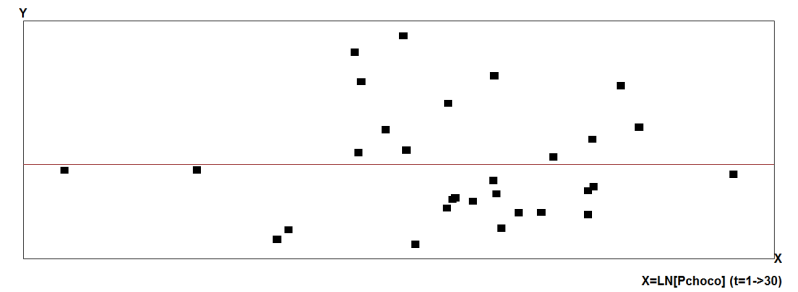


Grafico 4: Residuos Estimados Vs. Ln(Pte)

Y=OLS Residual of LN[Qcafe] (t=1->30)

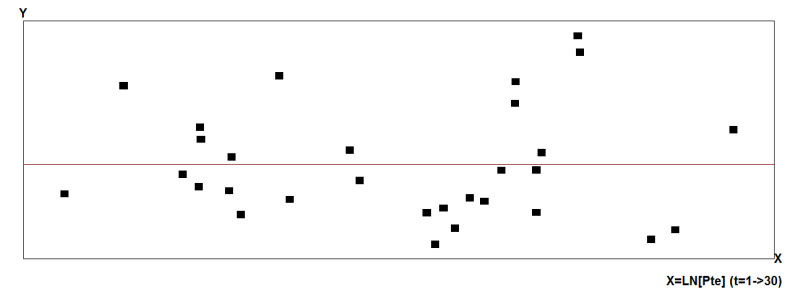
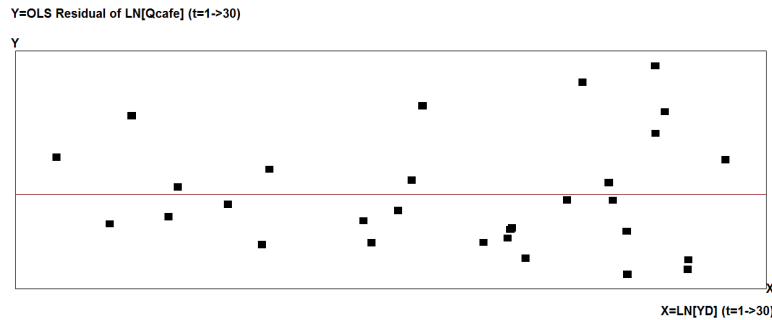


Grafico 5: Residuos Estimados Vs. Ln(YD)



Como se ve en las gráficas, la serie de los residuos no parece tener una varianza constante y parece estar relacionadas con las variables explicativas. Para estar seguros de la existencia o no de esta problema es necesario hacer las pruebas formales. La primera prueba que se debe hacer es la de **Goldfeld y Quant**. Las hipótesis a contrastar:

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

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

Dado que no hay claridad de cual de las variables está afectando la serie de los residuos es necesario efectuar la prueba con todas las variables explicativas. Para realizar esta prueba se debe ordenar de menor a mayor de acuerdo con la variable que se cree que está causando el problema. Después de esto se deben omitir los  $d < \frac{1}{5}n$   $d < 6$  datos de la mitad, para después correr dos regresiones. Se podría pensar en retirar 5 datos de la muestra, pero en este caso no habría dos muestras iguales, por lo que el número óptimo a retirar es 4, así cada muestra quedará de 13 observaciones.

**Logaritmo Natural del Precio del Café:**

Al correr la regresión se obtiene que la suma cuadrada de los errores para los valores bajos de la variable (Donde  $i = 1, \dots, 13$ )  $\ln[Pcafe_i]$  es:

$$SSE_1 = 0.02408534$$

Por otro lado la suma de los errores al cuadrado para los datos asociados a los valores altos de la variable (Donde  $i = 17, \dots, 30$ )  $\ln[Pcafe_i]$  es:

$$SSE_2 = 0.101828$$

De lo anterior se obtiene el estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{0.101828}{0.02408534} = 4,22779998$$

Decisión: Se rechaza la hipótesis nula si  $F_{(n-d-2k, n-d-2k)\alpha} < F_{GQ}$

Por tanto, dado que  $F_{(16,16)0,01} < 4,22779998$ , entonces  $3,37204 < 4,22779998$ , por tanto se rechaza la hipótesis nula a un 99% de confianza, por tanto la variable  $\ln[Pcafe_i]$  presenta un problema de heterocedasticidad.

**Logaritmo Natural del Precio del Chocolate:**

Al correr la regresión se obtiene que la suma cuadrada de los errores para los valores bajos de la variable (Donde  $i = 1, \dots, 13$ )  $\ln[Pchoco_i]$  es:

$$SSE_1 = 0.09559955$$

Por otro lado la suma de los errores al cuadrado para los datos asociados a los valores altos de la variable (Donde  $i = 17, \dots, 30$ )  $\ln[Pchoco_i]$  es:

$$SSE_2 = 0.05493824$$

De lo anterior se obtiene el estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{0.05493824}{0.09559955} = 0,57467049$$

Decisión: Se rechaza la hipótesis nula si  $F_{(n-d-2k, n-d-2k)\alpha} < F_{GQ}$

Por tanto, dado que  $F_{(16,16)0,01} > 0,57467049$ , entonces  $3,37204 > 0,57467049$ , por tanto No se rechaza la hipótesis nula a un 99% de confianza, por tanto la variable  $\ln[Pchoco_i]$  NO presenta un problema de heterocedasticidad. Si se compara con un  $F_{(16,16)0,1} = 1,697855 > 0,57467049$ , por tanto No se rechaza la hipótesis nula a un 90% de confianza.

**Logaritmo Natural del Precio del Té:**

Al correr la regresión se obtiene que la suma cuadrada de los errores para los valores bajos de la variable (Donde  $i = 1, \dots, 13$ )  $\ln[Pte_i]$  es:

$$SSE_1 = 0.05209367$$

Por otro lado la suma de los errores al cuadrado para los datos asociados a los valores altos de la variable (Donde  $i = 17, \dots, 30$ )  $\ln[Pte_i]$  es:

$$SSE_2 = 0.09054444$$

De lo anterior se obtiene el estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{0.09054444}{0.05209367} = 1,738300335$$

Decisión: Se rechaza la hipótesis nula si  $F_{(n-d-2k, n-d-2k)\alpha} < F_{GQ}$

Por tanto, dado que  $F_{(16,16)0,01} > 1,91280572$ , entonces  $3,37204 > 1,738300335$ , por tanto No se rechaza la hipótesis nula a un 99% de confianza, por tanto la variable  $\ln[Pte]$  NO presenta un problema de heterocedasticidad tipo Goldfeld y Quant. Si se compara con un  $F_{(16,16)0,1} = 1,697855 < 1,738300335$ , por tanto se rechaza la hipótesis nula a un 90% de confianza, por tanto existe un problema de heteroscedasticidad de tipo Goldfeld y Quant causado por la variable  $\ln[Pte]$ .

**Logaritmo Natural de la Renta disponible:**

Al correr la regresión se obtiene que la suma cuadrada de los errores para los valores bajos de la variable (Donde  $i = 1, \dots, 13$ )  $\ln[YD_i]$  es:

$$SSE_1 = 0.05293096$$

Por otro lado la suma de los errores al cuadrado para los datos asociados a los valores altos de la variable (Donde  $i = 17, \dots, 30$ )  $\ln[YD_i]$  es:

$$SSE_2 = 0.10735703$$

De lo anterior se obtiene el estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{0.10735703}{0.05293096} = 2,028$$

Decisión: Se rechaza la hipótesis nula si  $F_{(n-d-2k, n-d-2k)\alpha} < F_{GQ}$

Por tanto, dado que  $F_{(16,16)0,01} > 2,028$ , entonces  $3,37204 > 2,028$ , por no se rechaza la hipótesis nula a un 99% de confianza, por tanto la variable  $\ln[YD]$  NO presenta un problema de heterocedasticidad. Sin embargo si se compara con un  $F_{(16,16)0,1} = 1,697855 < 2,028$ , por tanto se rechaza la hipótesis nula a un 90% de confianza, por tanto existe un problema de heteroscedasticidad de tipo Goldfeld y Quant causado por la variable  $\ln[YD]$ .

**Test de Breush-Pagan**

Para descartar cualquier tipo de problema que pueda causar la variable  $\ln[Pcafe_i]$ , realizamos un prueba breusch-Pagan para esta variable.

En este caso las hipótesis a contrastar son las siguientes:

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

$$H_a: \sigma_i^2 = \sigma^2(\gamma + \delta \ln[Pcafe_i])$$

Para lo que se debe correr la siguiente regresión:

$$\frac{\hat{\epsilon}_i^2}{\hat{\sigma}^2} = \gamma + \delta \ln[Pcafe_i] + \mu_i$$

Después de correr la regresión auxiliar se calcula el estadístico Breusch-Pagan, donde:  $BP = 0,35457249$ . Dado que el  $BP < x_g^2$  (con un nivel de significancia de 1% y 1 grado de libertad), no hay suficiente evidencia para rechazar la hipótesis nula con un 99% de confiabilidad.

Por tanto no es posible asegurar que la variable  $\ln[Pcafe_i]$  sea la que esté causando la heterocedasticidad. Por tal razón se debe hacer la prueba Breusch-Pagan con todas las variables, ya que no se conoce con certeza la naturaleza de la heterocedasticidad.

En este caso la hipótesis nula y la alterna son las siguientes:

$$H_0: \text{Presencia de Homocedasticidad}$$

$$H_a: \text{Relacion entre la varianza del error y un grupo de variables}$$

Lo cual se puede presentar de la siguiente manera.

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

$$H_a: \sigma_i^2 = \sigma^2(\gamma + \delta X_i)$$

Al correr la regresión del modelo planteado inicialmente (2.1) (pues se considera que el conjunto de variables que causa la heterocedasticidad son todas las variables), se obtuvo que la regresión auxiliar es:

$$\frac{\hat{\epsilon}_i^2}{\hat{\sigma}^2} = \gamma + \delta_2 \ln[Pcafe_i] + \delta_3 \ln[Pchoco_i] + \delta_4 \ln[Pte_i] + \delta_5 \ln[YD_i] + \mu_i$$

De donde se obtiene que  $SSR=6,91950284$ . Por tanto el  $BP=3,45975142$ .

Dado que la toma de la decisión se realiza de la siguiente manera

Rechazar  $H_0$  cuando  $BP > x_g^2$

En este caso  $g$  es igual a 4, por tanto con un nivel de significancia del 1% el valor crítico es igual a 13,2767. Así mismo con un nivel de significancia del 5% y del 10% el valor crítico es igual a 9,487 y 7,77, respectivamente.

Por tanto dado que el  $BP < x_g^2$  con todos los niveles de significancia (10%, 5% y 1%), entonces se no existe suficiente evidencia para rechazar la hipótesis nula de que existe homocedasticidad.

**Test de White**

En este caso las hipótesis nula y alterna son las siguientes:

$H_0$ : Presencia de Homocedasticidad

$H_a$ : Presencia de heterocedasticidad

Lo cual se puede representar de la siguiente manera:

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

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

Para realizar esta prueba se debe correr la siguiente regresión auxiliar:

$$\begin{aligned} \hat{\varepsilon}_i = & \gamma + \delta_1 \ln[Pcafe_i] + \delta_2 \ln[Pchoco_i] + \delta_3 \ln[Pte_i] + \delta_4 \ln[YD_i] + \delta_5 (\ln[Pcafe_i])^2 \\ & + \delta_6 (\ln[Pchoco_i])^2 + \delta_7 (\ln[Pte_i])^2 + \delta_8 (\ln[YD_i])^2 \\ & + \delta_9 \ln[Pcafe_i] \ln[Pchoco_i] + \delta_{10} \ln[Pcafe_i] \ln[Pte_i] \\ & + \delta_{11} \ln[Pcafe_i] \ln[YD_i] + \delta_{12} \ln[Pchoco_i] \ln[Pte_i] \\ & + \delta_{13} \ln[Pchoco_i] \ln[YD_i] + \delta_{14} \ln[Pte_i] \ln[YD_i] + \mu_i \end{aligned}$$

El estadístico que se debe utilizar es:

$$W\alpha = nR^2$$

Dado que se obtuvo que  $R^2 = 0.3678$ , tenemos que el estadístico de White es:

$$W\alpha = 30 * 0.3678 = 11,034$$

En este test la toma de decisión se realiza de la siguiente manera:

Rechazar  $H_0$  cuando  $W\alpha > x_{g(\alpha)}^2$ . En este caso  $g$  es igual a 14, ( $g = \#$  de parámetros de la regresión auxiliar menos 1), por tanto con un nivel de significancia del 1% el valor crítico es 29,1412.

Por tanto dado que  $W\alpha = 11,034 < 29,1412$ , con una confianza del 99% no se puede rechazar la hipótesis nula de que existe homocedasticidad. (lo mismo sucede con 95% y 90% de confianza).

Finalmente, no se conoce realmente cuál es la variable que está causando la hetero, lo cual se comprueba con las pruebas realizadas anteriormente.

**6. Continuando con la pregunta anterior:**

**a)** Corrija el problema econométrico y estime el nuevo modelo.

De las pruebas del punto anterior se puede concluir lo siguiente:

En la prueba Goldfeld y Quant se encontró evidencia de un posible problema de heteroscedasticidad causado por tres variables. En un principio la variable que mayor posibilidad tenía de ser la causante del problema era  $\ln[Pcafe_i]$ . Para

corroborar el resultado anterior fue necesario realizar la prueba de Breush-Pagan para esta variable. El resultado de la prueba no permitió confirmar la existencia de heteroscedasticidad. Al ser el resultado de la prueba individual no concluyente se realizó la prueba conjunta que tampoco arrojó evidencia sobre un problema de heteroscedasticidad.

Con las dos pruebas anteriores no se pudo identificar que una variable estuviese causando heteroscedasticidad en el modelo. Igualmente, el test de White acepta la hipótesis de homocedasticidad por lo que se puede concluir que no conocemos la forma específica de la heteroscedasticidad.

Así, como no conocemos la forma específica de la heteroscedasticidad, es decir la naturaleza del problema, se debe recurrir como último recurso a la solución de White para estimar el modelo.

Los resultados de la estimación del modelo corregido se encuentran en la tabla 2, en el literal 4.b.

**b)** Interprete los coeficientes del nuevo modelo.

$\alpha$ : Es significativo al 95% de confianza. Carece de interpretación económica.

$\beta_2 = -0,0604796$  Es significativo al 99% de confianza. Indica que por un aumento de 1% en el precio del café, la cantidad demandada de café disminuye en 0,060%.

$\beta_3 = 0,5374623$  Es significativo al 99% de confianza. Indica que por un aumento de 1% en el precio del chocolate, la cantidad demandada de café aumenta en 0,5374%.

$\beta_4 = -0,5130241$  Es significativo al 95% de confianza. Indica que por un aumento de 1% en el precio del te, la cantidad demandada de café disminuye en 0,513%.

$\beta_5 = 0,68662$  Es significativo al 99% de confianza. Indica que por un aumento de 1% en la renta disponible, la cantidad demandada de café aumenta en 0,68662%.