

Taller 5: Multicolinealidad y Heteroscedasticidad

Econometría 06216

08-02-2010

Profesores: Julio César Alonso.

Monitoras: Sasha Magyaroff - Carolina Restrepo.

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 AM del 13 de Septiembre de 2010.
- Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta.

Instrucciones:

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

La secretaría de turismo de Caralibro, un municipio relativamente grande de una república caribeña desea estudiar los determinantes de la distancia que viajan los hogares en sus vacaciones. Usted ha sido contratado como consultor para realizar esta investigación. Usted cuenta con una muestra de 200 hogares del municipio de Caralibro. La muestra fue seleccionada aleatoriamente con el fin de investigar qué tan lejos tienden a viajar en sus vacaciones los hogares de dicha ciudad. Empleando esta muestra responda las siguientes preguntas.

Pregunta 1

Tras una revisión bibliográfica, usted llegó a la conclusión que el mejor modelo para realizar esta investigación es:

$$Km_i = \beta_1 + \beta_2 Ingreso_i + \beta_3 NumHijos_i + \beta_4 edad_i + \varepsilon_i \quad (1)$$

A. Interprete los coeficientes a priori y explique el signo esperado de cada coeficiente.

B. Estime el modelo (1) y reportelo en una tabla.

Pregunta 2

- A. Discuta si existen o no síntomas de multicolinealidad en el modelo estimado en la pregunta anterior.
- B. ¿Existe algún problema de multicolinealidad? Sea lo más claro posible.

Pregunta 3

Ahora, independientemente de su respuesta en la pregunta anterior, siga trabajando con el modelo (1) y responda las siguientes preguntas.

- A. ¿Existen síntomas de un problema de heteroscedasticidad? Sea lo más claro posible.

Pregunta 4

- A. ¿Existe un problema de heteroscedasticidad? Muestre claramente cómo llega a su conclusión.
- 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 5

Solucione el problema de heteroscedasticidad y si es posible demuestre que el problema se solucionó. Explique claramente su resultado y muéstrelo en una tabla. Muestre todos los pasos que emplee para llegar a su conclusión.

Pregunta 6

Tras una segunda revisión bibliográfica se encontró que es común encontrar en la literatura el siguiente modelo:

$$Km_i = \beta_1 + \beta_2 Ingreso_i + \beta_3 Ingreso_i^2 + \beta_4 NumHijos_i + \beta_5 edad_i + \varepsilon_i \quad (2)$$

- A. Estime el modelo 2. Reporte en una tabla el modelo que no tenga problemas de heteroscedasticidad.
- B. ¿Cuál de los dos modelos es mejor el (1) o (2)? Explique claramente su respuesta y emplee un método que permita concluir con un nivel de confianza específico.

Taller 5: Multilinealidad y Heteroscedasticidad

Econometría 06216

08-02-2010

Profesores: Julio César Alonso.

Monitoras: Sasha Magyaroff - Carolina Restrepo.

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 AM del 09 de Agosto de 2010.
- Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta.

Instrucciones:

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

La secretaría de turismo de Caralibro, un municipio relativamente grande de una república caribeña desea estudiar los determinantes de la distancia que viajan los hogares en sus vacaciones. Usted ha sido contratado como consultor para realizar esta investigación. usted cuenta con una muestra de 200 hogares del municipio de Caralibro. La muestra fue seleccionada aleatoriamente con el fin de investigar qué tan lejos tienden a viajar en sus vacaciones los hogares de dicha ciudad. Empleando esta muestra responda las siguientes preguntas.

Pregunta 1

Tras una revisión bibliográfica, usted llegó a la conclusión que el mejor modelo para realizar esta investigación es:

$$Km_i = \beta_1 + \beta_2 Ingreso_i + \beta_3 NumHijos_i + \beta_4 edad_i + \varepsilon_i \quad (1)$$

A. Interprete los coeficientes a priori y explique el signo esperado de cada coeficiente.

Respuesta Sugerida

β_1 : Es la parte de la distancia recorrida en las últimas vacaciones del hogar en km que no dependen ni del ingreso, la edad ni del número de hijos.

β_2 : Ante un aumento de mil dólares al año en el ingreso del jefe del hogar, se espera que la distancia recorrida en las últimas vacaciones del hogar varíe en β_2 kilómetros. Se espera que tenga signo positivo.

β_3 : Ante un aumento de un hijo en la familia, se espera que la distancia recorrida en las últimas vacaciones del hogar varíe en β_3 kilómetros. Se espera que tenga signo negativo, ya que entre más miembros tenga la familia, más difícil será recorrer grandes distancias.

β_4 : Ante un aumento de un año en la edad del jefe del hogar, se espera que la distancia recorrida en las últimas vacaciones del hogar varíe en β_4 kilómetros. Se espera que tenga signo positivo.

B. Estime el modelo (1) y repórtelo en una tabla.

Respuesta Sugerida

El resultado se reporta en la siguiente tabla.

Table 1: Estimación del modelo (1)

	(1)+	(2)++	(3)+++
	Km	Km	Km
(Intercept)	-391.548** [-2.306]	-424.996*** [-3.5]	-391.548*** [-2.773]
Ingreso	14.201*** [7.889]	13.947*** [9.42]	14.201*** [7.399]
NumHijos	-81.826*** [-3.016]	-76.806*** [-3.515]	-81.826*** [-2.835]
edad	15.741*** [4.189]	16.717*** [5.527]	15.741*** [4.01]
R^2	0.341	0.457	0.341
$adj.R^2$	0.331	0.449	0.331
N	200	200	200

t-values in brackets

* (p ≤ 0.1), ** (p ≤ 0.05), *** (p ≤ 0.01)

+ La primera columna corresponde al método de mínimos cuadrados ordinarios (MCO).

++ La segunda columna corresponde al método de mínimos cuadrados ponderados (MCP) empleando $Ingreso_i$ como ponderador.

+++ La tercera columna corresponde al método de mínimos cuadrados ordinarios (MCO) para los coeficientes y los t calculados se basan en la corrección de White.

Pregunta 2

A. Discuta si existen o no síntomas de multicolinealidad en el modelo estimado en la pregunta anterior.

Respuesta Sugerida

Parece no haber síntomas de multicolinealidad en el modelo estimado, ya que para ello debería generarse una contradicción entre la bondad de ajuste del modelo (R^2 alto), la significancia conjunta (F alto) y la significancia individual (t bajos). Sin embargo para este caso, se encuentra que el R^2 es bajo y el F es alto, y los t son altos, lo que hace que no se genere una inconsistencia entre la significancia individual y la significancia global.

B. ¿Existe algún problema de multicolinealidad? Sea lo más claro posible.

Respuesta Sugerida

Las pruebas no brindan evidencia en favor de la presencia de heteroscedasticidad. Esta conclusión se puede derivar de:

- $\kappa = 1.546$.
- $|X| = 0.829$
- Y las correlaciones de los coeficientes estimados no superan 0.8 en valor absoluto. La matriz de correlaciones de los coeficientes estimados es:
 - 1.

Pregunta 3

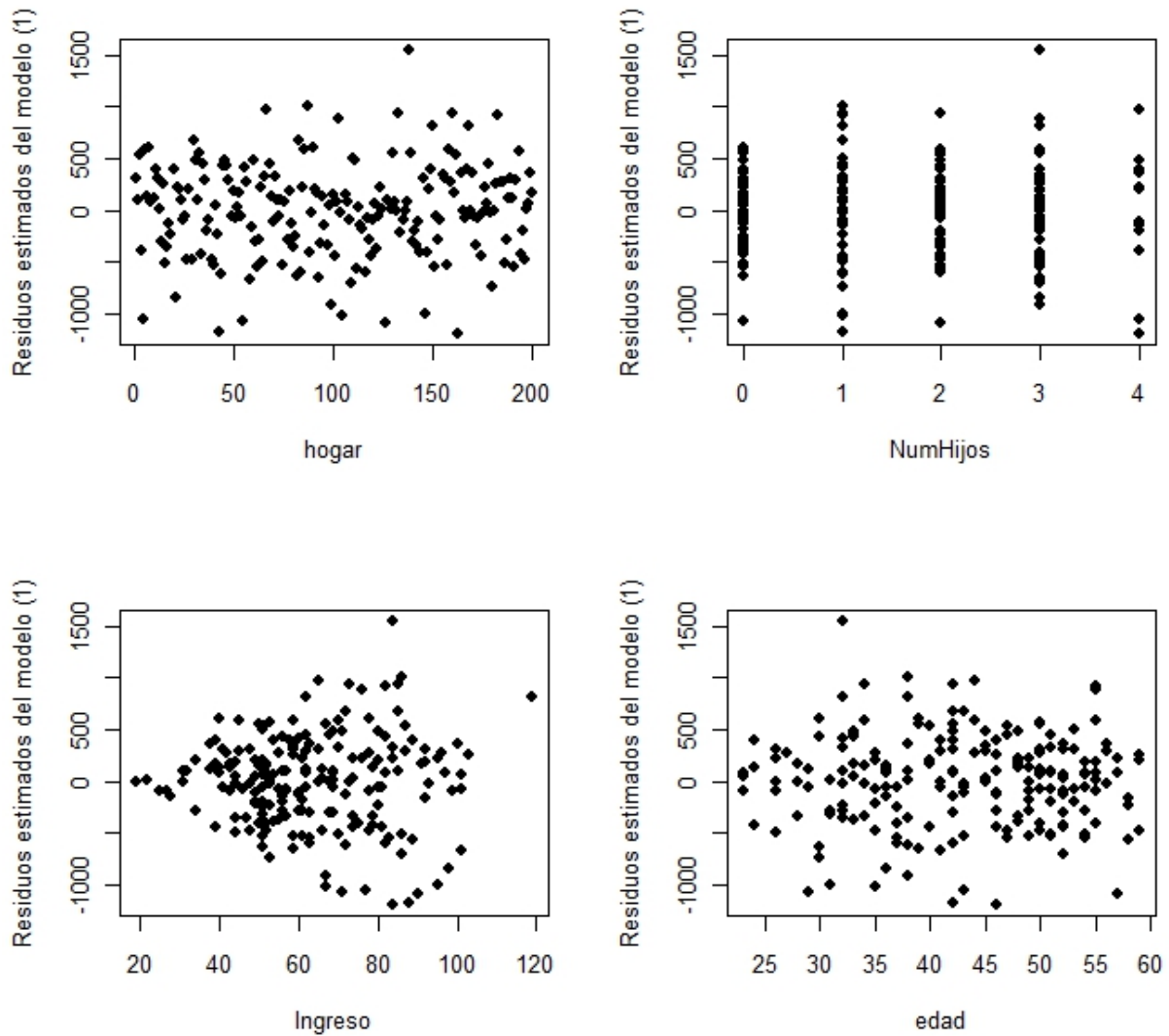
Ahora, independientemente de su respuesta en la pregunta anterior, siga trabajando con el modelo (1) y responda las siguientes preguntas.

A. ¿Existen síntomas de un problema de heteroscedasticidad? Sea lo más claro posible.

Respuesta Sugerida

Los gráficos de los residuos permiten intuir un problema de heteroscedasticidad. En especial parece que el ingreso está generando dicho problema. (Ver gráfico)

Figure 1: Gráfico de los residuos del modelo (1)



Pregunta 4

A. ¿Exite un problema de heteroscedasticidad? Muestre claramente cómo llega a su conclusión.

Respuesta Sugerida

Ustedes deberían realizar las pruebas de Goldfeld y Quandt, Breusch-Pagan y White. A continuación se presentan los resultados de cada una de esas pruebas.

Nota: es necesario mostrar cómo se llega a cada resultado y conclusión.

Prueba de Goldfeld y Quandt.

Table 2: Resultados de la prueba de Goldfeld y Quandt

H_A	Estadístico	p-valor
$\sigma_i^2 = (hogar_i)^2 \cdot \sigma^2$	1.1721	0.2438
$\sigma_i^2 = (Ingreso_i)^2 \cdot \sigma^2$	3.4123	0
$\sigma_i^2 = (NumHijos_i)^2 \cdot \sigma^2$	1.2053	0.2073
$\sigma_i^2 = \left(\frac{1}{edad_i}\right)^2 \cdot \sigma^2$	1.6926	0.011

Como se reporta en la tabla 2 en la única prueba que se puede rechazar la hipótesis nula, con un nivel de confianza del 99%, es aquella cuya alterna es que $\sigma_i^2 = Ingreso_i \cdot \sigma^2$. Es decir, con esta prueba parece existir evidencia que la heteroscedasticidad está relacionada con la variable ingreso y que la forma funcional de la heteroscedasticidad es cuadrática.

Prueba de Breusch y Pagan.

Table 3: Resultados de la prueba de Breusch y Pagan

H_A	Estadístico	p-valor
$\sigma_i^2 = f(hogar_i)$	0.022	0.8822
$\sigma_i^2 = f(Ingreso_i)$	24.7869	0
$\sigma_i^2 = f(NumHijos_i)$	2.2425	0.1343
$\sigma_i^2 = f(edad_i)$	1.8065	0.1789
$\sigma_i^2 = f(Ingreso_i, NumHijos_i, edad_i)$	36.3258	0

Como se reporta en la tabla 3 dos pruebas permiten rechazar la hipótesis nula. En esos dos casos la varianza es función de la variable $Ingreso_i$. Es decir, con esta prueba parece existir evidencia que la heteroscedasticidad está relacionada con el ingreso.

Prueba de White.

El estadístico de la prueba de White es en este caso 39.9928 y el correspondiente valor p es 0. Así se puede rechazar la hipótesis nula de homoscedasticidad. **Conclusión final.**

Las tres pruebas brindan evidencia en favor de la existencia de heteroscedasticidad. De hecho las pruebas de Breusch y Pagan y Goldfeld y Quandt permiten concluir que la variable que genera el problema es el ingreso. Es más, gracias a la prueba de Goldfeld y Quandt podemos concluir que en este caso se tiene que $\sigma_i^2 = Ingreso_i \cdot \sigma^2$.

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.

Respuesta Sugerida

De acuerdo al resultado del apartado anterior, podemos concluir que la mejor forma de resolver el problema es emplear los mínimos cuadrados ponderados. Es posible utilizar esta solución debido a que se conoce tanto la variable que causa el problema, como su forma funcional. Dicha información permite conocer los pesos de ponderación, en este caso corresponde a $Ingreso_i$. Por tanto, se debe correr el siguiente modelo, que corrige la muestra haciendo que desaparezca el problema de heteroscedasticidad.

$$\frac{Km_i}{Ingreso_i} = \beta_1 \frac{1}{Ingreso_i} + \beta_2 + \frac{NumHijos_i}{Ingreso_i} + \frac{edad_i}{Ingreso_i} + \frac{\varepsilon_i}{Ingreso_i} \tag{2}$$

Ustedes debían mostrar que el nuevo término de error debe ser homoscedástico si es verdad que $\sigma_i^2 = Ingreso_i \cdot \sigma^2$.

Pregunta 5

Solucione el problema de heteroscedasticidad y si es posible demuestre que el problema se solucionó. Explique claramente su resultado y muéstrello en una tabla. Muestre todos los pasos que emplee para llegar a su conclusión.

Respuesta Sugerida

NOTA: Ustedes deben presentar todos los pasos para llegar a la conclusión.

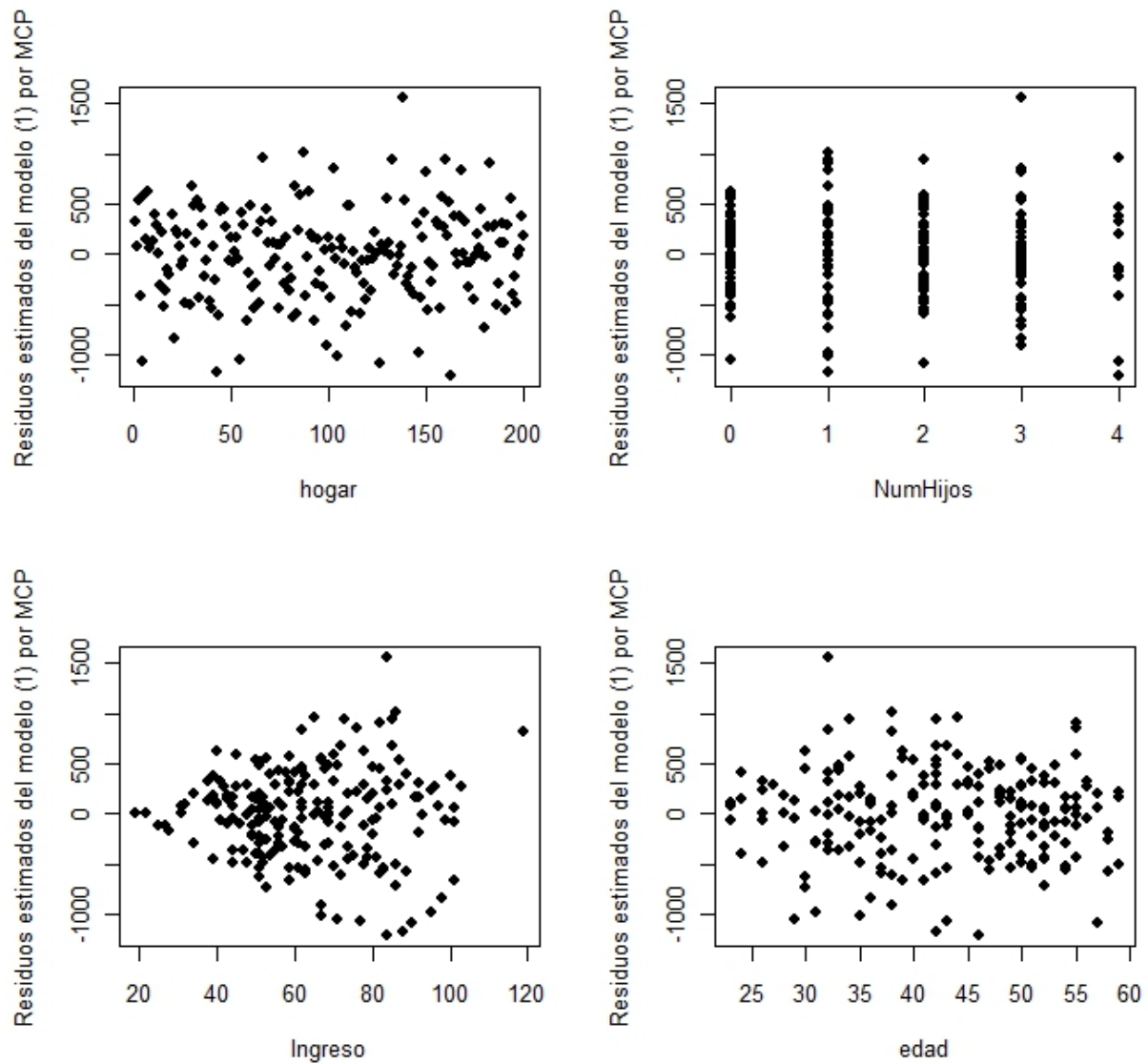
Para resolver el problema, debemos estimar el modelo (1) por el método de mínimos cuadrados ponderados, empleando como ponderador el ingreso. Es decir, el modelo a estimar será:

$$\frac{Km_i}{Ingreso_i} = \beta_1 \frac{1}{Ingreso_i} + \beta_2 + \frac{NumHijos_i}{Ingreso_i} + \frac{edad_i}{Ingreso_i} + \frac{\varepsilon_i}{Ingreso_i} \quad (3)$$

Los resultados se presentan en la segunda columna de la tabla 1. Para determinar si el problema ha sido corregido debemos realizar las pruebas de nuevo sobre el modelo corregido. A continuación se presentan los resultados.

Los gráficos de los residuos permiten intuir que el problema de heteroscedasticidad aún persiste. En especial, el ingreso sigue generando el problema de heteroscedasticidad. (Ver gráfico)

Figure 2: Gráfico de los residuos del modelo (1) por MCP



Prueba de Goldfeld y Quandt.

Table 4: Resultados de la prueba de Goldfeld y Quandt del modelo (1) estimado por MCP

H_A	Estadístico	p-valor
$\sigma_i^2 = (hogar_i)^2 \cdot \sigma^2$	1.1721	0.2438
$\sigma_i^2 = (Ingreso_i)^2 \cdot \sigma^2$	3.4123	0
$\sigma_i^2 = (NumHijos_i)^2 \cdot \sigma^2$	1.2053	0.2073
$\sigma_i^2 = \left(\frac{1}{edad_i}\right)^2 \cdot \sigma^2$	1.6926	0.011

Como se reporta en la tabla 4, nuevamente la única prueba que puede rechazar la hipótesis nula, con un nivel de confianza del 99%, es aquella cuya alterna es que $\sigma_i^2 = Ingreso_i \cdot \sigma^2$. Es decir, con esta prueba parece aún existir evidencia que la heteroscedasticidad está relacionada con la variable ingreso y que la forma funcional de la heteroscedasticidad es cuadrática.

Prueba de Breusch y Pagan.

Table 5: Resultados de la prueba de Breusch y Pagan del modelo (1) estimado por MCP

H_A	Estadístico	p-valor
$\sigma_i^2 = f(hogar_i)$	0.022	0.8822
$\sigma_i^2 = f(Ingreso_i)$	24.7869	0
$\sigma_i^2 = f(NumHijos_i)$	2.2425	0.1343
$\sigma_i^2 = f(edad_i)$	1.8065	0.1789
$\sigma_i^2 = f(Ingreso_i, NumHijos_i, edad_i)$	36.3258	0

Como se reporta en la tabla 5, nuevamente dos pruebas permiten rechazar la hipótesis nula. En esos dos casos la varianza es función de la variable $Ingreso_i$. Es decir, con esta prueba parece existir evidencia que la heteroscedasticidad está relacionada con el ingreso.

Prueba de White.

El estadístico de la prueba de White es en este caso 39.9928 y el correspondiente valor p es 0. Así se puede rechazar la hipótesis nula de homoscedasticidad.

Conclusión final.

Las tres pruebas brindan evidencia en favor de la existencia aún de heteroscedasticidad en el modelo estimado por MCP. Así, la mejor solución será emplear los estimadores MCO para los coeficientes y la matriz de varianzas y covarianzas de White para estimar los t calculados. Esta solución se presenta en la última columna de la tabla 1.

Pregunta 6

Tras una segunda revisión bibliográfica se encontró que es común encontrar en la literatura el siguiente modelo:

$$Km_i = \beta_1 + \beta_2 Ingreso_i + \beta_3 Ingreso_i^2 + \beta_4 NumHijos_i + \beta_5 edad_i + \varepsilon_i \quad (4)$$

A. Estime el modelo 4. Reporte en una tabla el modelo que no tenga problemas de heteroscedasticidad.

Respuesta Sugerida

Primero que todo, es importante reconocer que al estimar el modelo (4) por MCO existirá un problema de heteroscedasticidad (a continuación se presentan los gráficos y las pruebas que soportan esta afirmación). Es más, la corrección de MCP no resuelve el problema y por eso se debe emplear los coeficientes estimados por MCO y los t calculados por medio de la corrección de White. Este resultado se reporta en la tabla 2, en la cual también se reporta el modelo (1) con el mismo método.

+ La primera y cuarta columna corresponde al método de mínimos cuadrados ordinarios (MCO).

++ La segunda columna corresponde al método de mínimos cuadrados ponderados (MCP) empleando $Ingreso_i$ como ponderador.

+++ La tercera y quinta columna corresponde al método de mínimos cuadrados ordinarios (MCO) para los coeficientes y los t calculados se basan en la corrección de White.

Table 6: Estimación del modelo (1) y (4) empleando la corrección de White.

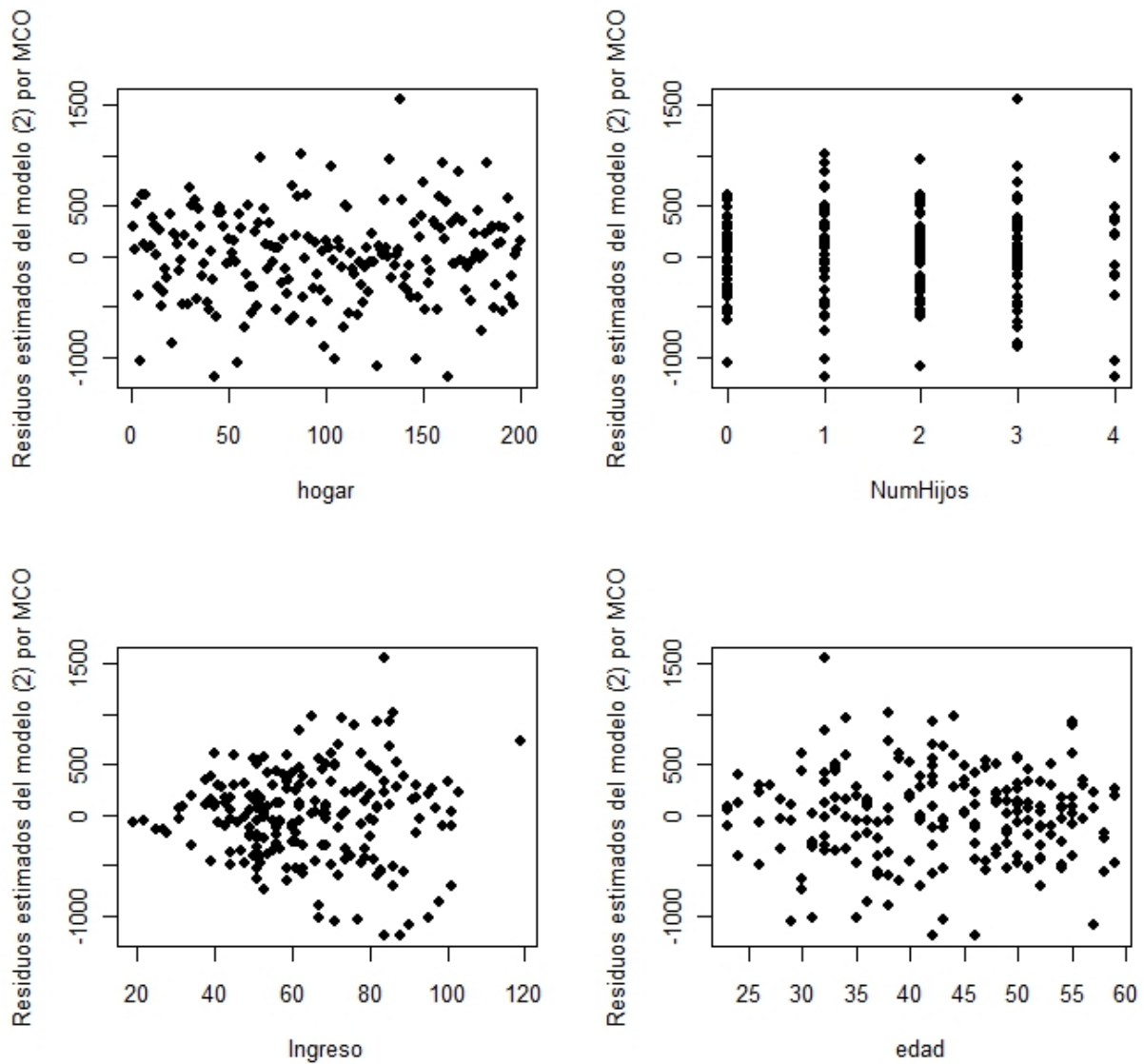
	(1)+	(2)++	(3)++	(4)+	(5)++
	Km	Km	Km	Km	Km
(Intercept)	-391.548**	-424.996***	-391.548***	-258.155	-258.155
	[-2.306]	[-3.5]	[-2.773]	[-0.773]	[-0.98]
Ingreso	14.201***	13.947***	14.201***	9.575	9.575
	[7.889]	[9.42]	[7.399]	[0.945]	[1.074]
NumHijos	-81.826***	-76.806***	-81.826***	-82.807***	-82.807***
	[-3.016]	[-3.515]	[-2.835]	[-3.037]	[-2.872]
edad	15.741***	16.717***	15.741***	15.946***	15.946***
	[4.189]	[5.527]	[4.01]	[4.207]	[4.044]
I(Ingreso ²)				0.035	0.035
				[0.464]	[0.492]
R^2	0.341	0.457	0.341	0.341	0.341
$adj.R^2$	0.331	0.449	0.331	0.328	0.328
N	200	200	200	200	200

t-values in brackets

* ($p \leq 0.1$), ** ($p \leq 0.05$), *** ($p \leq 0.01$)

Los gráficos de los residuos permiten intuir que el problema de heteroscedasticidad aún persiste. En especial, el ingreso sigue generando el problema de heteroscedasticidad. (Ver gráfico)

Figure 3: Gráfico de los residuos del modelo (2) por MCO



Prueba de Goldfeld y Quandt.

Table 7: Resultados de la prueba de Goldfeld y Quandt del modelo (4) por MCO

H_A	Estadístico	p-valor
$\sigma_i^2 = (hogar_i)^2 \cdot \sigma^2$	1.189	0.2261
$\sigma_i^2 = (Ingreso_i)^2 \cdot \sigma^2$	3.4017	0
$\sigma_i^2 = (NumHijos_i)^2 \cdot \sigma^2$	1.2072	0.2068
$\sigma_i^2 = \left(\frac{1}{edad_i}\right)^2 \cdot \sigma^2$	1.6822	0.0123

Como se reporta en la tabla 7, nuevamente la única prueba que puede rechazar la hipótesis nula, con un nivel de confianza

del 99 %, es aquella cuya alterna es que $\sigma_i^2 = Ingreso_i \cdot \sigma^2$. Es decir, con esta prueba parece aún existir evidencia que la heteroscedasticidad está relacionada con la variable ingreso y que la forma funcional de la heteroscedasticidad es cuadrática.

Prueba de Breusch y Pagan.

Table 8: Resultados de la prueba de Breusch y Pagan del modelo (4) estimado por MCO

H_A	Estadístico	p-valor
$\sigma_i^2 = f(hogar_i)$	0.0112	0.9158
$\sigma_i^2 = f(Ingreso_i)$	24.21	0
$\sigma_i^2 = f(NumHijos_i)$	2.2785	0.1312
$\sigma_i^2 = f(edad_i)$	1.8167	0.1777
$\sigma_i^2 = f(Ingreso_i, NumHijos_i, edad_i)$	35.7532	0

Como se reporta en la tabla 8, nuevamente dos pruebas permiten rechazar la hipótesis nula. En esos dos casos la varianza es función de la variable *Ingreso_i*. Es decir, con esta prueba parece existir evidencia que la heteroscedasticidad está relacionada con el ingreso.

Prueba de White.

El estadístico de la prueba de White es en este caso 39.3898 y el correspondiente valor p es 0. Así se puede rechazar la hipótesis nula de homoscedasticidad. **Conclusión final.**

Las tres pruebas brindan evidencia en favor de la existencia aún de heteroscedasticidad en el modelo estimado por MCP. Así, la mejor solución será emplear los estimadores MCO para los coeficientes y la matriz de varianzas y covarianzas de White para estimar los t calculados. Esta solución se presenta en la última columna de la tabla 6.

B. ¿Cuál de los dos modelos es mejor el (1) y (4)? Explique claramente su respuesta y emplee un método que permita concluir con un nivel de confianza específico.

Respuesta Sugerida

Para responder esta pregunta, podemos emplear una prueba de Wald, cuya hipótesis nula implica que el modelo es el (1) (modelo restringido) y el modelo no restringido es (4). En este caso el estadístico de Wald es 0.2424, y su respectivo valor p es 0.623. Por lo tanto no existe suficiente evidencia para rechazar la hipótesis de que el modelo (1) es mejor que el modelo (2). Así deberíamos emplear el modelo (1) y no el (2).

Nota 1: Emplear el R^2 ajustado no es tan buena opción, pues no permite hacer una prueba de hipótesis que permita responder con un nivel de confianza específico.

Nota 2: Otra posibilidad de probar si el modelo restringido es mejor que el sin restringir, es emplear la prueba de significancia individual, pero con la corrección de White.