TP N°9: Heterocedasticidad

Junghanss – Freysselinard – Marco – Werner

**Actividad 1**

1] FALSO. Los EMC resultan insesgados ante la existencia de heterocedasticidad.

2] VERDADERO. Los EMC son consistentes ante la presencia de heterocedasticidad.

3] VERDADERO.

**Actividad 2**

1] Existen errores en los cálculos del estimador de la matriz de varianzas y covarianzas de los estimadores de mínimos cuadrados.

Se suele perder eficiencia sobre el estimador cuadrático mínimo.

2] El patrón que se debe exhibir para sospechar de la existencia de heterocedasticidad, es que a medida que aumenta el valor predicho, los residuos se vayan alejando de la media.

3] El ejemplo mas simple de heterocedasticidad son los modelos de predicción utilizados en los activos financieros. Esto se debe a que la volatilidad de estos activos puede variar enormemente de un periodo a otro.

**Ejercicio N°1: Detección de Heteroscedasticidad**

**El archivo "VOTE1.dta" contiene datos de 173 distritos electorales. A partir de ellos es posible estudiar si los gastos de campaña afectan los resultados de las elecciones a partir del modelo:**

**VA = B0 + B1\*logGA + B2\*logGB + B3\*VPA + B4\*J + u**

**donde VA es el porcentaje de votos recibidos por el candidato A, GA y GB son los gastos de campaña del candidato A y del B, VPA es una medida de la fortaleza del partido del candidato A y J es una dummy que es igual a 1 si este distrito posee una proporción de votantes jóvenes superior a la media nacional.**

. rename voteA VA

. rename prtystrA VPA

. rename lexpendA logGA

. rename lexpendB logGB

. rename democA J

**a) Regrese los residuos de una regresión por MCO sobre todas las variables independientes y explique por qué el R^2 es igual a 0 en este caso.**

. reg VA logGA logGB VPA J

Source | SS df MS Number of obs = 173

-------------+------------------------------ F( 4, 168) = 169.23

Model | 38822.1774 4 9705.54436 Prob > F = 0.0000

Residual | 9635.07113 168 57.3516139 R-squared = 0.8012

-------------+------------------------------ Adj R-squared = 0.7964

Total | 48457.2486 172 281.728189 Root MSE = 7.5731

------------------------------------------------------------------------------

VA | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

logGA | 5.779294 .3918197 14.75 0.000 5.00577 6.552819

logGB | -6.237836 .3974596 -15.69 0.000 -7.022495 -5.453178

VPA | .2519176 .0712925 3.53 0.001 .1111729 .3926623

J | 3.792943 1.40652 2.70 0.008 1.016213 6.569674

\_cons | 37.66142 4.736036 7.95 0.000 28.3116 47.01123

-----------------------------------------------------------------------------

. predict u, r

. reg u logGA logGB VPA J

Source | SS df MS Number of obs = 173

-------------+------------------------------ F( 4, 168) = 0.00

Model | 0 4 0 Prob > F = 1.0000

Residual | 9635.07121 168 57.3516143 R-squared = 0.0000

-------------+------------------------------ Adj R-squared = -0.0238

Total | 9635.07121 172 56.0178559 Root MSE = 7.5731

------------------------------------------------------------------------------

u | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

logGA | 2.74e-08 .3918197 0.00 1.000 -.7735247 .7735247

logGB | -4.13e-09 .3974596 -0.00 1.000 -.7846588 .7846588

VPA | -4.12e-09 .0712925 -0.00 1.000 -.1407447 .1407447

J | -6.78e-08 1.40652 -0.00 1.000 -2.776731 2.77673

\_cons | 1.22e-07 4.736036 0.00 1.000 -9.349811 9.349811

-----------------------------------------------------------------------------

. di e(r2)

0

El R^2 es igual a cero porque se cumple el supuesto de exogeneidad: E(u|X)=0: es decir, los residuos no están correlacionados con las variables explicativas.

**b) Grafique los residuos de una regresión por MCO contra los valores predichos. ¿Qué se observa en este gráfico?**

. reg VA logGA logGB VPA J

Source | SS df MS Number of obs = 173

-------------+------------------------------ F( 4, 168) = 169.23

Model | 38822.1774 4 9705.54436 Prob > F = 0.0000

Residual | 9635.07113 168 57.3516139 R-squared = 0.8012

-------------+------------------------------ Adj R-squared = 0.7964

Total | 48457.2486 172 281.728189 Root MSE = 7.5731

------------------------------------------------------------------------------

VA | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

logGA | 5.779294 .3918197 14.75 0.000 5.00577 6.552819

logGB | -6.237836 .3974596 -15.69 0.000 -7.022495 -5.453178

VPA | .2519176 .0712925 3.53 0.001 .1111729 .3926623

J | 3.792943 1.40652 2.70 0.008 1.016213 6.569674

\_cons | 37.66142 4.736036 7.95 0.000 28.3116 47.01123

------------------------------------------------------------------------------

. predict Yhat, xb

. scatter u Yhat

. 

La verdad que no hay una relación evidente entre los residuos y la predicción, por lo que no podemos decir con ningún tipo de seguridad que la varianza efectivamente varía a medida que cambian los valores que toman las variables explicativas.

**c) Efectúe el Contraste de Breusch-Pagan**

. reg VA logGA logGB VPA J

Source | SS df MS Number of obs = 173

-------------+------------------------------ F( 4, 168) = 169.23

Model | 38822.1774 4 9705.54436 Prob > F = 0.0000

Residual | 9635.07113 168 57.3516139 R-squared = 0.8012

-------------+------------------------------ Adj R-squared = 0.7964

Total | 48457.2486 172 281.728189 Root MSE = 7.5731

------------------------------------------------------------------------------

VA | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

logGA | 5.779294 .3918197 14.75 0.000 5.00577 6.552819

logGB | -6.237836 .3974596 -15.69 0.000 -7.022495 -5.453178

VPA | .2519176 .0712925 3.53 0.001 .1111729 .3926623

J | 3.792943 1.40652 2.70 0.008 1.016213 6.569674

\_cons | 37.66142 4.736036 7.95 0.000 28.3116 47.01123

------------------------------------------------------------------------------

. estat hettes

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of VA

chi2(1) = 2.95

Prob > chi2 = 0.0861

Como el valor del estadístico de prueba es 0,0861, que es mayor a 0,05, entonces, a un nivel de significatividad del 5%, no podemos rechazar la hipótesis nula: no podemos decir que no hay homoscedasticidad.

**d) Realice el Caso Especial del Contraste de White.**

. reg VA logGA logGB VPA J

Source | SS df MS Number of obs = 173

-------------+------------------------------ F( 4, 168) = 169.23

Model | 38822.1774 4 9705.54436 Prob > F = 0.0000

Residual | 9635.07113 168 57.3516139 R-squared = 0.8012

-------------+------------------------------ Adj R-squared = 0.7964

Total | 48457.2486 172 281.728189 Root MSE = 7.5731

------------------------------------------------------------------------------

VA | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

logGA | 5.779294 .3918197 14.75 0.000 5.00577 6.552819

logGB | -6.237836 .3974596 -15.69 0.000 -7.022495 -5.453178

VPA | .2519176 .0712925 3.53 0.001 .1111729 .3926623

J | 3.792943 1.40652 2.70 0.008 1.016213 6.569674

\_cons | 37.66142 4.736036 7.95 0.000 28.3116 47.01123

------------------------------------------------------------------------------

Generamos la predicción de Y y la predicción de Y al cuadrado.

De b), ya tenemos al la predicción de Y: Yhat.

Predicción de Y al cuadrado (Yhat2):

. gen Yhat2 = Yhat^2

Residuos de la regresión original al cuadrado (u2):

. gen u2 = u^2

Ahora, debemos regresar u2 contra Yhat e Yhat2:

. reg u2 Yhat Yhat2

Source | SS df MS Number of obs = 173

-------------+------------------------------ F( 2, 170) = 2.79

Model | 37152.5553 2 18576.2776 Prob > F = 0.0645

Residual | 1133583.18 170 6668.13637 R-squared = 0.0317

-------------+------------------------------ Adj R-squared = 0.0203

Total | 1170735.74 172 6806.60313 Root MSE = 81.659

------------------------------------------------------------------------------

u2 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

Yhat | -4.263681 2.166534 -1.97 0.051 -8.540455 .0130928

Yhat2 | .0357354 .0212419 1.68 0.094 -.0061964 .0776672

\_cons | 171.8584 53.14213 3.23 0.001 66.95494 276.7619

------------------------------------------------------------------------------

Haremos el contraste mediante LM: El LM del test es el producto entre el R^2 (e(r2)) de esa regresión y el tamaño muestral (e(N)):

. scalar LMWEtest = e(N)\*e(r2)

. di " Valor LM White Especial del Test " LMWEtest

Valor LM White Especial del Test 5.4900452

. di " Valor LM White Especial Límite " invchi2tail(e(df\_m),0.05)

Valor LM White Especial Límite 5.9914645

Como LMtest<LMlímite, entonces no rechazamos la hipótesis nula que plantea que el modelo es homoscedástico: al igual que concluimos por el Contraste de Breusch-Pagan, no podemos rechazar la hipótesis nula que plantea que la varianza es homoscedástica, a un nivel de significatividad del 5%.

**e) Efectúe el Contraste de White solo con cuadrados. Concluya: ¿Qué tan fuerte es la evidencia de heteroscedasticidad?**

Debemos, primeramente, definir todos los cuadrados de las explicativas:

. gen logGA2 = logGA^2

. gen logGB2 = logGB^2

. gen VPA2 = VPA^2

. gen J2 = J^2

Generados todos los cuadrados, debemos regresar los residuos al cuadrado del modelo original (u2) contra todos ellos:

. reg u2 logGA2 logGB2 VPA2 J2

Source | SS df MS Number of obs = 173

-------------+------------------------------ F( 4, 168) = 1.45

Model | 39003.5916 4 9750.89789 Prob > F = 0.2206

Residual | 1131732.15 168 6736.50087 R-squared = 0.0333

-------------+------------------------------ Adj R-squared = 0.0103

Total | 1170735.74 172 6806.60313 Root MSE = 82.076

------------------------------------------------------------------------------

u2 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

logGA2 | -.7396229 .5227247 -1.41 0.159 -1.771578 .2923325

logGB2 | -.2062947 .505968 -0.41 0.684 -1.205169 .7925798

VPA2 | -.0070299 .0078073 -0.90 0.369 -.0224429 .0083832

J2 | 8.636997 15.3613 0.56 0.575 -21.68906 38.96305

\_cons | 95.11818 31.35934 3.03 0.003 33.20903 157.0273

Empleamos LM para contrastar, sabiendo que el estadístico LM del test en White es igual a n\*(R^2):

. scalar LMWtest = e(N)\*e(r2)

. di " Valor LM del Test = " LMWtest

Valor LM del Test = 5.7635734

. di " Valor LM Límite = " invchi2tail(e(df\_m),0.05)

Valor LM Límite = 9.487729

Empleando solo cuadrados, hay menos evidencia de que el modelo sea heteroscedástico, ya que, a diferencia de cuando usamos el modelo de White completo (no solo con cuadrados, como acá) la diferencia entre el valor del estadístico del test y el límite es muy grande (5,7 y 9,4, mientras que antes eran 5,49 y 5,99). Es decir, necesitaríamos un nivel de significatividad más grande para poder rechazar que el modelo sea homoscedástico si hacemos el contraste de White solo con cuadrados que si lo hacemos completo.

**f) Estime la regresión original con errores robustos a la heterocedasticidad. ¿Existen grandes diferencias con respecto a los errores estándar originales? Compare la significatividad individual de las variables en ambos casos.**

Empleamos el comando "reg ..., r":

. reg VA logGA logGB VPA J, r

Linear regression Number of obs = 173

F( 4, 168) = 164.20

Prob > F = 0.0000

R-squared = 0.8012

Root MSE = 7.5731

------------------------------------------------------------------------------

| Robust

VA | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

logGA | 5.779294 .533142 10.84 0.000 4.726773 6.831815

logGB | -6.237836 .3561583 -17.51 0.000 -6.940959 -5.534714

VPA | .2519176 .0660631 3.81 0.000 .1214969 .3823383

J | 3.792943 1.452168 2.61 0.010 .9260957 6.659791

\_cons | 37.66142 4.41894 8.52 0.000 28.93761 46.38522

------------------------------------------------------------------------------

Y también corremos el modelo sin corregir para compararlos:

.

. reg VA logGA logGB VPA J

Source | SS df MS Number of obs = 173

-------------+------------------------------ F( 4, 168) = 169.23

Model | 38822.1774 4 9705.54436 Prob > F = 0.0000

Residual | 9635.07113 168 57.3516139 R-squared = 0.8012

-------------+------------------------------ Adj R-squared = 0.7964

Total | 48457.2486 172 281.728189 Root MSE = 7.5731

------------------------------------------------------------------------------

VA | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

logGA | 5.779294 .3918197 14.75 0.000 5.00577 6.552819

logGB | -6.237836 .3974596 -15.69 0.000 -7.022495 -5.453178

VPA | .2519176 .0712925 3.53 0.001 .1111729 .3926623

J | 3.792943 1.40652 2.70 0.008 1.016213 6.569674

\_cons | 37.66142 4.736036 7.95 0.000 28.3116 47.01123

Al corregir por heteroscedasticidad, aumenta la significatividad individual de todas las variables, excepto de la variable J. No obstante, no cambia en absoluto la conclusión de ninguno de los tests individuales. Los estadísticos t de prueba varían (aunque muy poco) en todos los casos, ya que dependen del error estándar de los coeficientes, los cuales cambian al ajustar heteroscedasticidad.

**h) ¿Vale la pena calcular los errores robustos a heteroscedasticidad en este caso? Explique.**

A priori, dado que los test de detección de heteroscedasticidad no rechazaron la hipótesis nula de que hay homoscedasticidad, y debido también a que los tests de significatividad individual no cambiaron sus conclusiones en absoluto, entonces podríamos tender a afirmar que no hace falta corregir empleando errores robustos.

Sin embargo, por definición, corregir por errores estándar robustos a la heterocedasticidad sirve tanto para cuando efectivamente hay como para cuando en realidad no hay heteroscedasticidad. De este modo, entonces podemos decir que, antes desconocimiento de si verdaderamente hay o no heteroscedasticidad, sí vale la pena calcular los errores robustos a la heterocedasticidad.

**Ejercicio 2: Heterocedasticidad conocida (MCP)**

La variable “fallas” representa la cantidad de fallas en la producción y “h\_uso” es la variable que almacena los datos de horas de utilización de la máquina de producción.

**Punto (A):**

Corremos los siguientes códigos para la resolución;

reg fallas h\_uso predict u, r gen u2 = u^2 gen h\_uso2 = h\_uso^2 reg u2 fallas h\_uso scatter u2 h\_uso

\* scatter u2 h\_uso2 (No son necesarios, pero tenemos la posibilidad también de graficarlos) \* scatter u2 fallas

Regresamos las fallas contra las horas de uso, de allí extramoes los residuales de la regresión, los elevamos al cuadrado y hacemos una nueva variable. Esta es la que utilizaremos para graficar los errores al cuadrado para cada valor que toma la variable explicativa Xi.

Entonces, cuando graficamos los residuos al cuadrado contra los valores de la variable explicativa X (las horas de uso) podemos observar que hay poca dispersión de los residuos en los valores de Xi al principio y luego aumenta cuando cambia Xi, lo que nos indica presencia de heterocedasticidad.

Dado que la varianza de los errores deja de ser constante para los distintos valores de X, para poder corregirla, una propuesta sería rehacer el modelo general por mínimos cuadrados ponderados (MCP) que nos permitirá hallar estimadores más eficientes y errores homoscedasticos.

Conociendo la forma que toman los residuos al cuadrado respecto X, podemos inferir cómo será el término que h(x) que utilizaremos para corregir la ecuación: utilizaremos la raíz cuadra del término H, que dividirá cada elemento de la regresión.

**Punto (B):** Corremos los siguientes códigos para la resolución;

gen H = h\_uso2 gen rH = H^(12)

gen const\_H = 1rH Generamos la constante, que no es más que B0 multiplicando el cociente de la raíz de H gen fallas\_H = fallasrH Generamos la variable explicada corregida por el término H gen h\_uso\_H = h\_usorH Generamos la variable explicativa corregida por el término H

reg fallas\_H const\_H h\_uso\_H, noconst

Corremos la regresión nueva que contiene la ecuación del modelo original pero corregida por el término H

**Punto (C):**

De los estimadores obtenidos por MCP de la regresión anterior, donde tenemos el siguiente modelo general: Fallas = (115.95) + (0.042)X ; podemos inferir que:

- Los estimadores son, por propiedad estadística, mucho más eficientes que los del modelo original. - En relación a su insesgadez, son estimadores insesgados, ya que esa propiedad GM la cumplían en su modelo original y con la transformación, los supuestos de Gauss-Markov que ya se cumplían, siguen cumpliéndose.

Solo cambia que ahora también se cumple el supuesto de errores homocedasticos.

**Punto (D):**

Si bien tenemos un R^2 de 0.971 para esta regresión, raramente nos serviría para utilizar como medida de bondad de ajuste, ya que las variables transformadas no explican lo que literalmente estamos estudiando detrás.

Para este caso, deberíamos seguir utilizando como medida de ajuste el R^2 del modelo original.

Cabe destacar como recordatorio, que la presencia de heteroscedasticidad no afecta nuestro R^2, por lo que podemos seguir empleando el original, que es de 0.6472 y representa un ajuste de la X sobre Y de casi un 65%.

**Punto (E):**

Para el caso donde la función de heterocedasticidad es incorrecta, es decir, que la función de varianza esté mal especificada, hay dos consecuencias:

Primeramente, caemos en la problemática de que los errores estándar y los estadísticos de prueba ya no son válidos, inclusive en muestras grandes.

Por otro lado, la otra consecuencia es que no sabremos con ninguna certeza que los estimadores MCP sean más eficientes que los de MCO.

**Punto (F):** Asumiendo hipotéticamente que la función H(x) que hemos empleado en el punto (B) está mal especificada, podemos proceder convirtiendo los errores estándar en robustos de la siguiente manera:

reg fallas\_H const\_H h\_uso\_H, noconst r di "Diferencia " 23.11787-23.14042 di "Diferencia " .0082973-.0077361

Cuando convertimos los errores en robustos, podemos evaluar el tipo de diferencia que presentan frente a los errores estándar del modelo inicial por MCP.

La diferencia de los errores de la constante es negativa pero cercana a cero, además, la diferencia de los errores estándar de la variable X (horas de uso) es prácticamente cero, por lo que no tiene significatividad estadística.

**Ejercicio 5: Estimación de la función de heterocedasticidad (MCGF)**

. use "WAGE1.DTA"

. reg wage educ exper tenure

Source | SS df MS Number of obs = 526

-------------+------------------------------ F( 3, 522) = 76.87

Model | 2194.1116 3 731.370532 Prob > F = 0.0000

Residual | 4966.30269 522 9.51398984 R-squared = 0.3064

-------------+------------------------------ Adj R-squared = 0.3024

Total | 7160.41429 525 13.6388844 Root MSE = 3.0845

------------------------------------------------------------------------------

wage | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

educ | .5989651 .0512835 11.68 0.000 .4982176 .6997126

exper | .0223395 .0120568 1.85 0.064 -.0013464 .0460254

tenure | .1692687 .0216446 7.82 0.000 .1267474 .2117899

\_cons | -2.872735 .7289643 -3.94 0.000 -4.304799 -1.440671

------------------------------------------------------------------------------

**a) ¿Es de esperar que exista heterocedasticidad en este modelo? ¿Por qué?**

Sí, porque cuanto mayor es el nivel de educación, más son las oportunidades laborales, y por ende aumenta la varianza de los empleos obtenidos y por ende de los salarios remunerados. A menor educación, menor es la varianza porque los salarios están muy en torno al salario mínimo.

**b) Realice las pruebas que considere necesarias para evaluar la existencia de heterocedasticidad.**

Detección Gráfica:

. predict u, r

. gen u2 = u^2

. scatter u2 educ



Como vemos, a mayor nivel educativo y de experiencia, mayor es la varianza, lo cual sustenta lo que explicábamos en el punto a)

.

Contraste de Breusch-Pagan:

Paso a paso, metódicamente:

. reg u2 educ exper tenure

Source | SS df MS Number of obs = 526

-------------+------------------------------ F( 3, 522) = 15.53

Model | 24875.0022 3 8291.66739 Prob > F = 0.0000

Residual | 278734.78 522 533.974674 R-squared = 0.0819

-------------+------------------------------ Adj R-squared = 0.0767

Total | 303609.782 525 578.304347 Root MSE = 23.108

------------------------------------------------------------------------------

u2 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

educ | 1.570524 .3841997 4.09 0.000 .8157566 2.325292

exper | .1104628 .090326 1.22 0.222 -.0669843 .28791

tenure | .6931533 .1621544 4.27 0.000 .3745979 1.011709

\_cons | -15.70645 5.461164 -2.88 0.004 -26.43501 -4.977886

------------------------------------------------------------------------------

. scalar bp = e(N)\*e(r2)

. di "estadistico de prueba " bp

estadistico de prueba 43.095618

. di "valor critico " invchi2tail(e(df\_r), 0.05)

valor critico 576.25936

. scalar F = (e(r2)/e(df\_m) / (1-e(r2)) / e(df\_r))

. di F

.00005699

. di invFtail(e(df\_m),e(df\_r),0.05)

2.6219805

Por comandos:

. estat hettes

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of u2

chi2(1) = 181.58

Prob > chi2 = 0.0000

Como vemos, tanto paso a paso como por comando el p valor es menor a 0,05 e incluso 0,01: se rechaza la hipótesis nula de que no hay heteroscedasticidad.

**c) Estime el modelo tomando como variable dependiente al logaritmo del salario. ¿Qué conclusiones puede obtener?**

. reg lwage educ exper tenure

Source | SS df MS Number of obs = 526

-------------+------------------------------ F( 3, 522) = 80.39

Model | 46.8741776 3 15.6247259 Prob > F = 0.0000

Residual | 101.455574 522 .194359337 R-squared = 0.3160

-------------+------------------------------ Adj R-squared = 0.3121

Total | 148.329751 525 .28253286 Root MSE = .44086

------------------------------------------------------------------------------

lwage | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

educ | .092029 .0073299 12.56 0.000 .0776292 .1064288

exper | .0041211 .0017233 2.39 0.017 .0007357 .0075065

tenure | .0220672 .0030936 7.13 0.000 .0159897 .0281448

\_cons | .2843595 .1041904 2.73 0.007 .0796756 .4890435

------------------------------------------------------------------------------

.

.

.

**d) Aplique el método de MCGF para corregir la heteroscedasticidad (utilice la formulación original de la variable dependiente).**

Generamos el logaritmo de los residuos al cuadrado y todas las no linealidades de las variables explicativas, y regresaremos aquel contra todas estas:

Generamos las variables:.

. gen lu2 = log(u2)

. gen educ2 = educ^2

. gen exper2 = exper^2

. gen tenure2 = tenure^2

. gen educ\_exper = educ\*exper

. gen educ\_tenure = educ\*tenure

. gen exper\_tenure = exper\*tenure

. Y regresamos:

. reg lu2 educ2 exper2 tenure2 educ\_tenure educ\_exper exper\_tenure

Source | SS df MS Number of obs = 526

-------------+------------------------------ F( 6, 519) = 14.64

Model | 470.298109 6 78.3830182 Prob > F = 0.0000

Residual | 2778.59082 519 5.35373955 R-squared = 0.1448

-------------+------------------------------ Adj R-squared = 0.1349

Total | 3248.88893 525 6.18835987 Root MSE = 2.3138

------------------------------------------------------------------------------

lu2 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

educ2 | .011105 .0019352 5.74 0.000 .0073032 .0149068

exper2 | .000757 .0004452 1.70 0.090 -.0001176 .0016316

tenure2 | .0013523 .0017204 0.79 0.432 -.0020274 .0047321

educ\_tenure | .0032405 .0029287 1.11 0.269 -.0025131 .0089942

educ\_exper | -.000014 .0015543 -0.01 0.993 -.0030675 .0030394

exper\_tenure | -.0006249 .0014895 -0.42 0.675 -.0035511 .0023013

\_cons | -1.923647 .3217903 -5.98 0.000 -2.555819 -1.291476

------------------------------------------------------------------------------

De esta regresión, tomamos la predicción del logaritmo de los residuos al cuadrado, que denominamos "g". Como logu2 es la variable explicada en este último modelito, su predicción es simplemente la predicción lineal de éste:

predict g, xb

Con "g" podemos contstruir la función "hMCGF", que es igual a su exponencial:

. gen hMCGF = exp(g)

Y aplicamos raíz cuadrada:

. gen raízhMCGF = hMCGF^(1/2)

Y con hMCGF, corregimos todas las variables, dividiéndolas por la raíz de hMCGF:

Corrección/transformación de la constante (la tomamos como 1 original):

. gen constransMCGF = 1 / raízhMCGF

.

Corrección/transformación de wage:

. gen wagetransMCGF = wage / raízhMCGF

Corrección/transformación de educ:

. gen eductransMCGF = educ / raízhMCGF

.

Corrección/transformación de exper:

. gen expertransMCGF = exper / raízhMCGF

.

Corrección/transformación de tenure:

. gen tenuretransMCGF = tenure / raízhMCGF

.

.

Y, finalmente, regresamos el modelo transformado, sacándole la constante:

. reg wagetransMCGF eductransMCGF expertransMCGF tenuretransMCGF constransMCGF, noconst

Source | SS df MS Number of obs = 526

-------------+------------------------------ F( 4, 522) = 561.62

Model | 10750.5944 4 2687.6486 Prob > F = 0.0000

Residual | 2498.03185 522 4.78550162 R-squared = 0.8114

-------------+------------------------------ Adj R-squared = 0.8100

Total | 13248.6263 526 25.1875024 Root MSE = 2.1876

---------------------------------------------------------------------------------

wagetransMCGF | Coef. Std. Err. t P>|t| [95% Conf. Interval]

----------------+----------------------------------------------------------------

eductransMCGF | .34131 .0387957 8.80 0.000 .265095 .417525

expertransMCGF | .0325364 .0103307 3.15 0.002 .0122415 .0528313

tenuretransMCGF | .1473849 .0250565 5.88 0.000 .0981608 .1966089

constransMCGF | .08199 .4941906 0.17 0.868 -.8888568 1.052837

---------------------------------------------------------------------------------

Más rápidamente sería haberlo hecho con comando:

. reg wage educ exper tenure [aweight = 1/hMCGF]

(sum of wgt is 4.5620e+02)

Source | SS df MS Number of obs = 526

-------------+------------------------------ F( 3, 522) = 43.17

Model | 714.649778 3 238.216593 Prob > F = 0.0000

Residual | 2880.2352 522 5.51769195 R-squared = 0.1988

-------------+------------------------------ Adj R-squared = 0.1942

Total | 3594.88498 525 6.84739996 Root MSE = 2.349

------------------------------------------------------------------------------

wage | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

educ | .34131 .0387957 8.80 0.000 .265095 .417525

exper | .0325364 .0103307 3.15 0.002 .0122415 .0528313

tenure | .1473849 .0250565 5.88 0.000 .0981608 .1966089

\_cons | .0819902 .4941906 0.17 0.868 -.8888566 1.052837

------------------------------------------------------------------------------

.

**e) Los estimadores MCGF obtenidos, ¿son insesgados?, ¿son consistentes?, ¿son eficientes?**

Los estimadores MCGF obtenidos no son insesgados, ya que el ponderador se estima con los mismos datos.

No obstante, siguen siendo consistentes y asintóticamente más eficiente que el MCO.