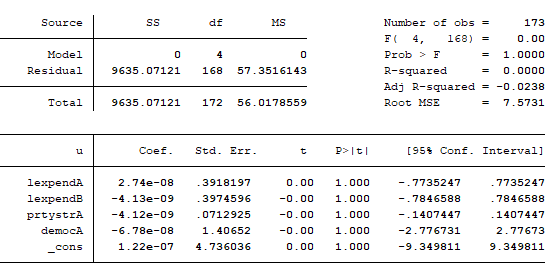
**TPº9: HETEROCEDASTICIDAD**

**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:

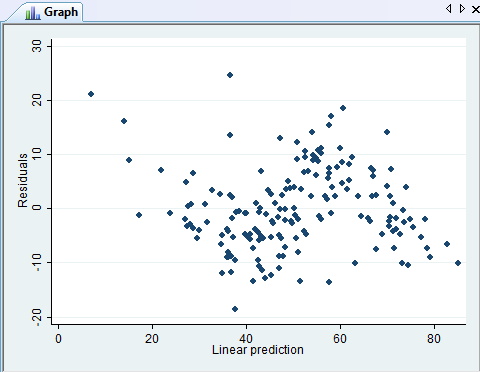
donde votosA es el porcentaje de votos recibidos por el candidato A, gastosA y gastosB son los gastos de campaña del candidato A y del candidato B, votospartA es una medida de la fortaleza del partido del candidato A (el porcentaje de votos que obtuvo el partido A en la elección presidencial más reciente) y joven es una variable dummy que es igual a 1 si ese distrito posee una proporción de votante jóvenes superior a la media nacional.

1. **Regrese los residuos de una regresión por MCO sobre todas las variables independientes y explique por qué el 𝑅2 es igual a 0 en este caso.**



El es igual a cero porque se cumple el supuesto de exogeneidad que   
Es decir, los residuos no están correlacionados con las variables explicativas.

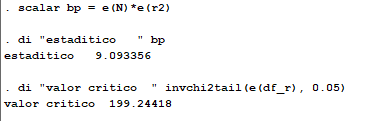
1. **Grafique los residuos de una regresión por MCO contra los valores predichos. ¿Qué se observa en este gráfico?**



Para inferir que hay existencia de heterocedasticidad en un gráfico de los errores contra la variable, debemos observar una dispersión asimétrica de los residuales a lo largo de los valores de la variable explicativa Xi. En este caso, si bien no es un caso “estéticamente” perfecto de homocedasticidad, tampoco podemos inferir a priori que haya heterocedasticidad.

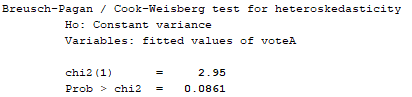
1. **Efectúe el contraste de Breusch-Pagan.**

Forma manual:



Como el estadístico es menor que el valor crítico, es decir, que cae en la zona de NO rechazo, podemos inferir que no tenemos pruebas suficientes para rechazar la hipótesis nula de homocedasticidad en el modelo.

Forma automática:

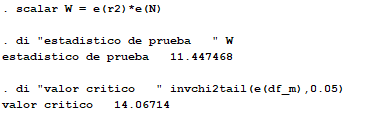


De la salida del resultado de forma automática, podemos inferir también que a un nivel de significancia del 5%, no podemos rechazar la Hipótesis Nula de que no haya heterocedasticidad. Vemos que el p-valor resultante del test es

En síntesis, fallamos al rechazar las hipotesis nulas y no podemos afirmar que estemos en presencia de heterocedasticidad en el modelo, para un nivel de significancia del 0.05

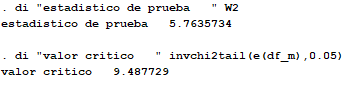
1. **Realice el caso especial del contraste de White.**

El caso especial del contraste de White consta de convertir las variables independientes en términos NO lineales y regresarlas contra los errores al cuadrado. Para esto, generamos cada variable por sí misma elevada al cuadrado y también una variable nueva por cada término de interacción (producto) entre variables. Así, conformamos una lista de variables que constituyen NO linealidades del modelo.

Armamos el estadístico de prueba LM de White, que consta de multiplicar el tamaño muestral por el R cuadrado.

Como vemos que el estadístico de prueba LM es menor que el límite o valor crítico del área , es decir, cae en zona de NO rechazo. Por ende, 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%.

1. **Efectúe el contraste de White sólo con cuadrados. Concluya: ¿Qué tan fuerte es la evidencia de heterocedasticidad?**



Empleando solo cuadrados, hay menos evidencia de que el modelo sea heteroscedástico, ya que, a diferencia de cuando usamos el modelo de White completo (incluyendo términos de interacción por producto entre variables independientes), la diferencia entre el valor del estadístico del test y el límite es mayor.

Es decir, necesitaríamos un nivel de significatividad mayor para poder rechazar la hipótesis de que el modelo sea homoscedástico si hacemos el contraste de White solo con cuadrados que si lo hacemos completo.

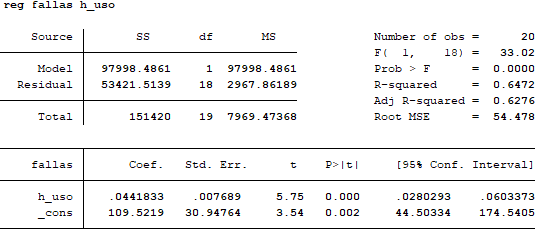
1. **Estime la regresión original con errores robustos a heterocedasticidad. ¿Existen grandes diferencias con respecto a los errores estándar originales? Compare la significatividad individual de las variables en ambos casos.**

Al corregir los errores y estadísticos de prueba t y F por robustez contra 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.

**EJERCICIO Nº 2: Heterocedasticidad conocida (MCP)**

**a) A partir de los siguientes datos, grafique los residuos al cuadrado contra los valores de 𝑋. ¿Qué relación se observa entre ambos? ¿qué supuesto acerca del comportamiento de la varianza propondría para corregir la heterocedasticidad en este caso?**

Corremos los siguientes códigos para la resolución;



Regresamos las fallas contra las horas de uso, de allí extraemos 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.

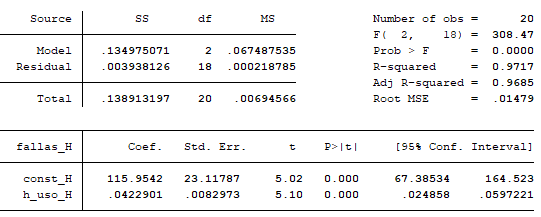
Dado que la varianza de los errores deja de ser constante para los distintos valores de X, para poder corregirla, una propuesta sería rehacer el modelo general por mínimos cuadrados ponderados (MCP) que nos permitirá hallar estimadores más eficientes y errores homoscedasticos.

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.

**b) Estime los coeficientes de regresión por el método de MCP, dado el supuesto: 𝜎𝑖2=𝜎2𝑋𝑖2.**

Generamos la constante, que no es más que B0 multiplicando el cociente de la raíz de H  
Generamos la variable explicada corregida por el término H gen   
Generamos la variable explicativa corregida por el término H

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

****

Podemos contrastar, a diferencia de la primer regresión (), un modelo con bondad de ajuste mucho mayor (), pero que no nos sirve como medida de ajuste, ya que intuitivamente las variables ajustadas no son representativas para explicar el modelo en la realidad, sino que solo se emplean para corregir el problema de heterocedasticidad. Por otro lado, la interpretación de los estimadores de mínimos cuadrados ponderados (MCP) resulta ser similar a la de los coeficientes por MCO, ya que en ambos casos son significativos, pero vemos que hay una diferencia en el valor de los beta.

**c) Los estimadores obtenidos, ¿son insesgados? ¿son eficientes?**

De los estimadores obtenidos por MCP de la regresión anterior, donde tenemos el siguiente modelo general ; 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.

**d) ¿Cómo se interpreta el 𝑅2 en este caso?**

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 , 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%.

**e) ¿Qué sucede si la función de varianza está mal especificada?**

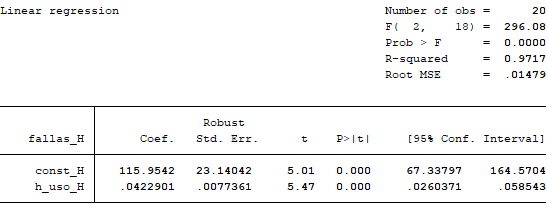
Para el caso donde la función de heterocedasticidad es incorrecta, es decir, que la función de varianza esté mal especificada, hay dos consecuencias:

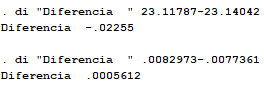
Primeramente, caemos en la problemática de que los errores estándar y los estadísticos de prueba ya no son válidos, inclusive en muestras grandes.

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

**f) Considere que la función de varianza estimada pueda estar mal especificada y calcule los errores estándar robustos a los estimadores de MCP del inciso b). ¿Varían mucho los errores estándar?**

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:





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 tienen significatividad estadística.

Pregunta: No me termino de quedar claro; ¿deberíamos darle mucha atención a la diferencia entre errores robustos y normales? Ya que, sabemos que empíricamente los errores robustos suelen ser un poquito mayores, pero la pregunta es si en una presentación, deberíamos hacer mención de eso u omitirlo, porque no tiene mucha relevancia.

**Ejercicio 5: Estimación de la función de heteroced. (MCGF)**

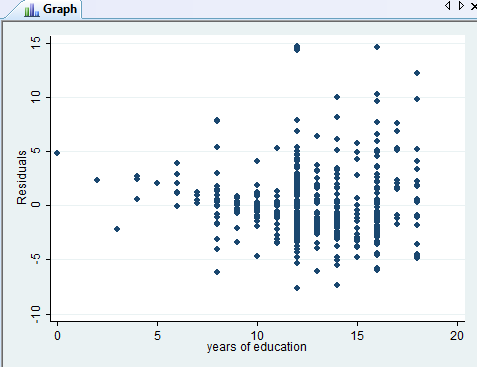
Con los datos del archivo WAGE1.dta, estime un modelo que explique la remuneración por hora de los empleados (wage) a partir de sus años de instrucción (educ), los años de experiencia (exper) y los años de antigüedad en el empleo actual (tenure).

**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.**

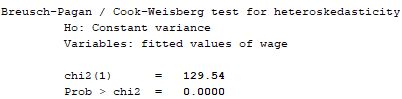
Detección gráfica:



Podemos observar que claramente hay heterocedasticidad, porque para los valores que va tomando Xi (años de educación) vemos que la varianza presenta distintas amplitudes de valores.

Detección por Test de Breusch-Pagan:

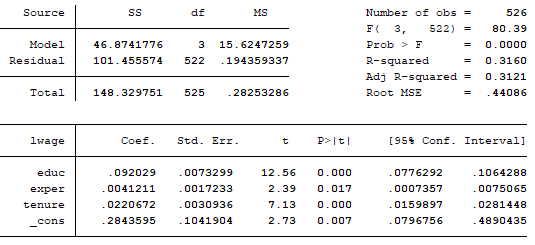
Realizando el test de manera automática, podemos llegar a los siguientes resultados:



El p-valor es menor al nivel de significancia, por lo que ante la baja probabilidad, debemos rechazar la hipótesis nula de que la varianza es constante y el modelo presenta homocedasticidad. ()

En síntesis, a un nivel de significancia de 5%, debemos inferir que estamos ante presencia de heterocedasticidad.

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



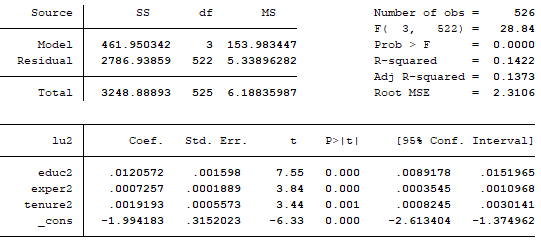
Tomando el logaritmo del salario, la variable explicada, y regresándolo contra las demás variables independientes, podemos inferir lo siguiente:

En comparación con el modelo original, la principal diferencia reside en que la variable explicativa experiencia es ahora significativa a un nivel de 5% de confianza para un pvalor de 0.017 (), cuando antes se caracterizaba por insignificancia para un pvalor de 0.064 (). El resto de coeficientes siguen siendo significativos y no presentan cambios, c excepción de la interpretación de los valores marginales, ya que ahora tenemos un modelo funcional log-nivel. Por otro lado, el es en términos prácticos igual, ya no hay diferencia significativa.

**d) Aplique el método de MCG factibles para corregir la heterocedasticidad (utilice la formulación original de la variable dependiente). Estime el ponderador a utilizar en base al siguiente supuesto acerca del comportamiento de la varianza:**

Para la resolución del inciso procedemos de la siguiente manera:

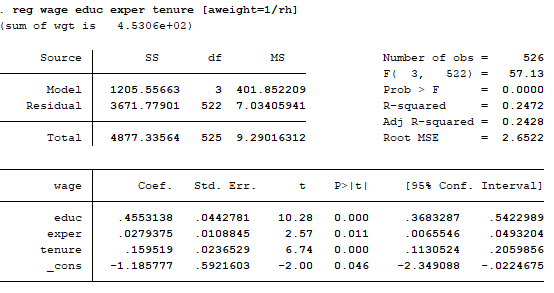
Generamos el logaritmo de los residuos al cuadrado y todas las no linealidades de las variables explicativas (cuadrados y productos), y regresaremos aquel contra todas estas.



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

Con "" podemos contstruir la función "", que es igual a su exponencial y luego la raíz cuadrada de esta:

Ahora con hMCGF debemos corregir todas las variables. Esto puede realizarse de manera manual, creando una nueva por cada variable original corregida y luego regresando el modelo nuevo o bien, utilizando el asistente automático que nos aplica la corrección sobre el comando de regresión del modelo con variables originales. Optaremos por esta última opción, para simplificar el trabajo e interpretación del código:



Nota: si hubiéramos hecho el método manual, deberíamos haber considerado corregir también la constante () e incluirla en la regresión nueva, junto con el comando para que no genere una ordenada al origen más.

**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.

Pregunta: cuando tenemos una o dos variables dummies en el modelo y debemos aplicar Minimos Cuadrados Generalizados o Ponderados, finalmente NO tiene sentido que las corrijamos o por un tema