MÍNIMOS CUADRADOS PONDERADOS APLICADOS AL ESTUDIO DE LA PRESIÓN SANGUÍNEA DIASTÓLICA SEGÚN LA EDAD EN MUJERES

ISADORE NABI

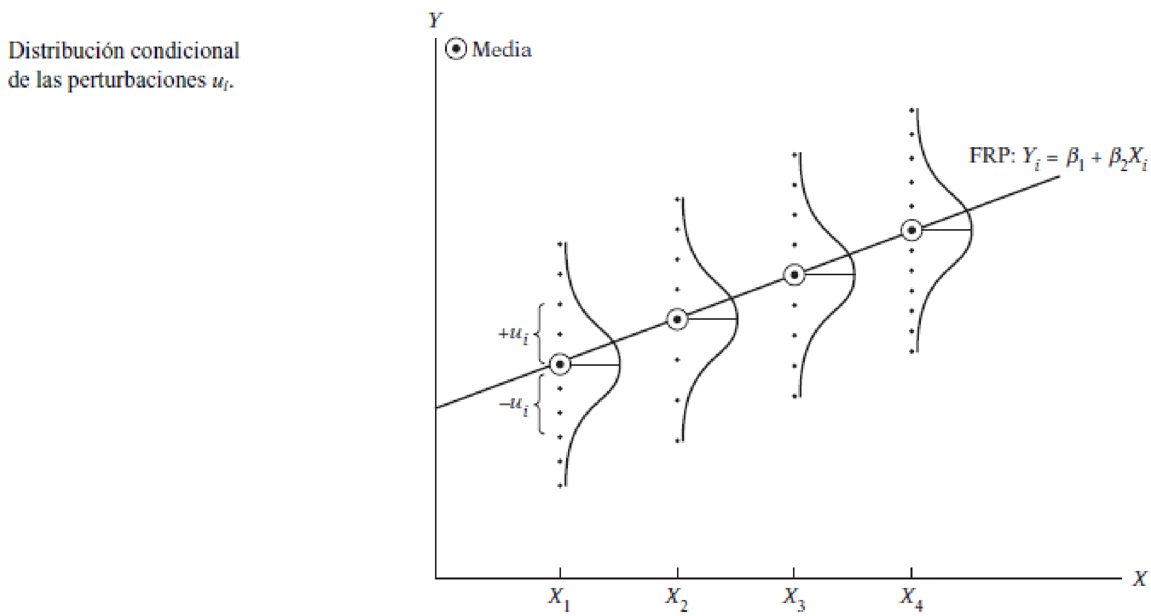
I. FUNDAMENTOS TEÓRICOS GENERALES DE LOS MÍNIMOS CUADRADOS PONDERADOS

I.I. Heteroscedasticidad

I.I. I. Conceptualización

El modelo clásico de regresión lineal (MCRL) está fundamentado en el teorema de Gauss-Markov, que implica una serie de supuestos sobre el comportamiento de las variables explicativas o predictores, así como del término de error (Nabi, 2021, págs. 8-22). El supuesto relativo a que la esperanza matemática o valor esperado del término de perturbación estocástica es igual a cero, dados los valores de las variables regresoras, es el supuesto de homocedasticidad (Gujarati & Porter, 2010, pág. 63); lo anterior puede enunciarse, de manera equivalente, afirmando que la varianza del término de error debe ser la misma (o, en su defecto, ser tendencialmente la misma) sin importar los valores de las variables regresoras (Gujarati & Porter, 2010, pág. 365). El incumplimiento de este supuesto es el escenario de *heterocedasticidad* en el conjunto de datos.

Figura 1. Nulidad del Valor Esperado del Término de Perturbación Estocástica



Fuente: (Gujarati & Porter, 2010, pág. 63).

Figura 2. Diferencias entre las Distribuciones Teóricas de Perturbaciones Homoscedásticas y Heteroscedásticas

FIGURA 11.1
Perturbaciones homosce-dásticas.

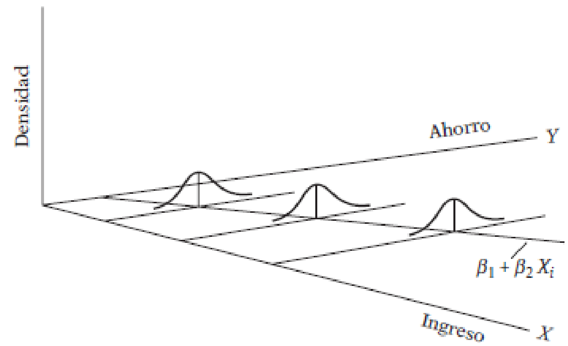
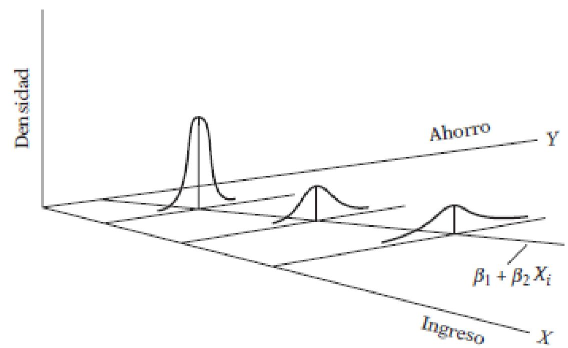


FIGURA 11.2
Perturbaciones heterosce-dásticas.



Fuente: (Gujarati & Porter, 2010, pág. 366).

I.I. II. Causas de Heteroscedasticidad

La naturaleza de la heteroscedasticidad (ausencia de homoscedasticidad) obedece a una o más de las siguientes causas (Gujarati & Porter, 2010, págs. 366-368):

1. Con base en los modelos de aprendizaje de los errores, a medida que la gente aprende disminuyen sus errores en el tiempo, por ejemplo, la relación entre horas de práctica de mecanografía y los errores mecanográficos es heterocedástica.
2. Cambios naturales en determinada clase de patrón. Por ejemplo, a nivel de los sistemas económicos, a medida que aumentan los ingresos, las personas poseen más opciones de compra, por lo que en función de las opciones de compra que les proporcione cada nivel de ingreso (y dado una estructura psíquica de gustos y preferencias, generada por su base genética y su proceso de socialización) pueden variar (o no) su nivel de ahorro.

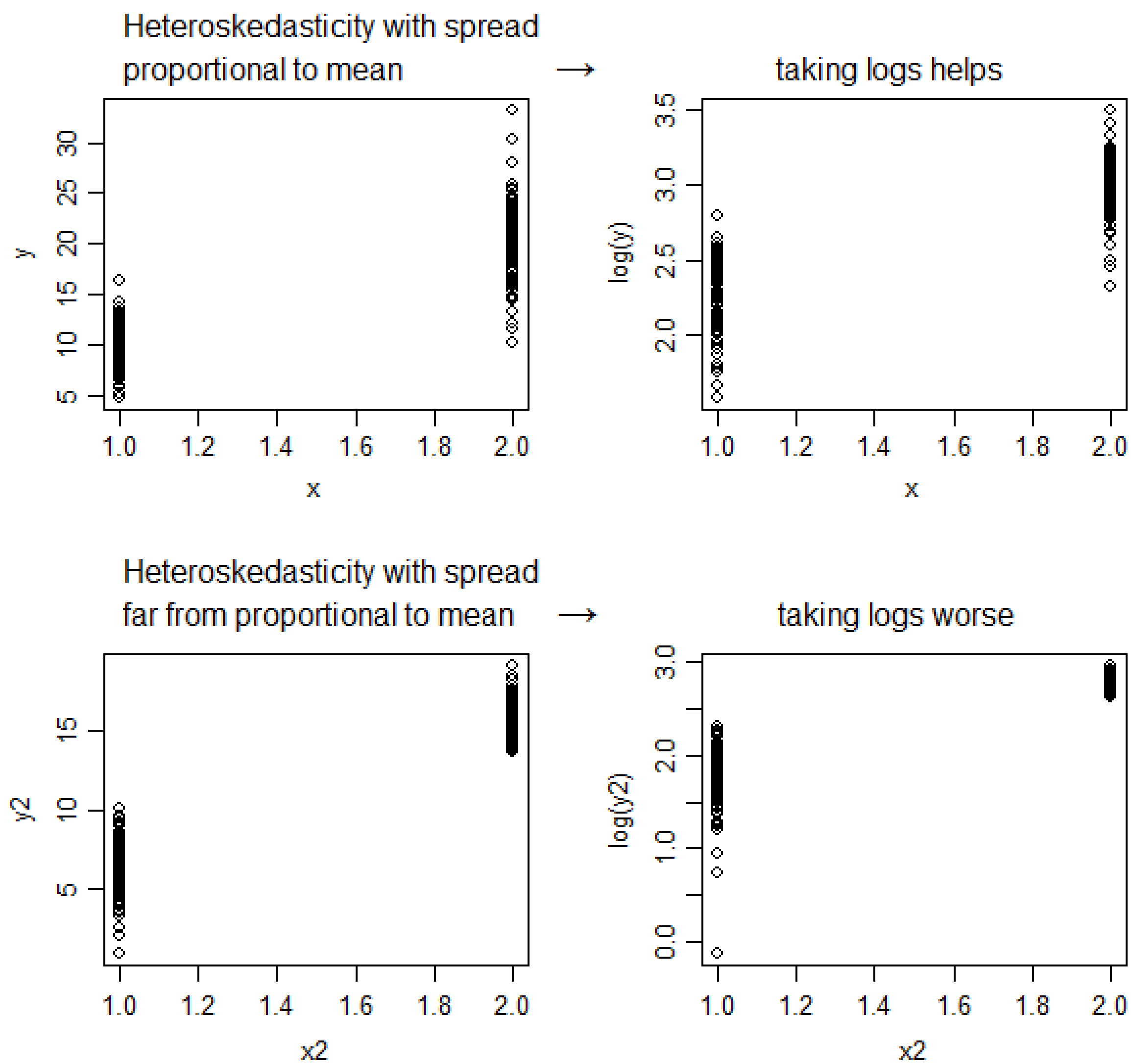
3. A medida que mejoran las técnicas de recolección de datos, es probable que la varianza se reduzca. Así, por ejemplo, es probable que los bancos con equipos complejos de procesamiento de información cometan menos errores en los informes mensuales o trimestrales de sus clientes que bancos que no los posean.
4. La existencia de datos atípicos, los cuales son observaciones que son muy grandes o pequeñas en comparación con la abrumadora generalidad de observaciones. Gráficamente esto se expresa como un aislamiento de estas pocas observaciones del resto del conjunto de datos. De forma más precisa, un dato atípico es una observación dentro de la muestra que proviene (estadísticamente hablando) de una población distinta a la que genera las demás observaciones de la muestra.
5. Problemas de especificación en el modelo de regresión (omisión de variables estadísticamente relevantes).
6. Asimetría en la distribución de una o más regresoras incluidas en el modelo. Por ejemplo, en el contexto de los sistemas económicos, al analizar la distribución del ingreso, a causa de diversos factores como el proceso de acumulación de capital que centraliza progresivamente la riqueza en cada vez menos individuos.
7. Incorrecta transformación de los datos. Esto ocurre porque la heteroscedasticidad es, geométricamente hablando (que no debe confundirse con su representación gráfica), una dispersión irregular[1] de los datos y, por ello, algunas transformaciones sobre las variables pueden corregirla (al menos parcialmente); sin embargo, también puede empeorarla. Como ejemplo tómese la transformación logarítmica que, como se señala en (Cross Validated, 2018), siempre que la dispersión del error sea aproximadamente proporcional al valor esperado condicional de la variable

[1] Con esto los autores buscan implicar que es una distribución no-gaussiana.

de respuesta, la heterocedasticidad tenderá a mejorar, sin embargo, si lo anterior no se cumple, entonces la empeorará; geométricamente hablando, esto ocurre porque la transformación logarítmica redistribuye al conjunto de datos de tal forma que hace que los valores más extremos (los valores más altos) tiendan a situarse más a la derecha del primer cuadrante del plano, mientras que hace que los valores más bajos tiendan a situarse más a la izquierda de dicho cuadrante.

8. Forma funcional incorrecta. Esto ocurre, por ejemplo, cuando se quiere modelar un fenómeno no-lineal con un modelo lineal o viceversa y no debe confundirse con un problema de especificación, que es relativo a la omisión de variables.

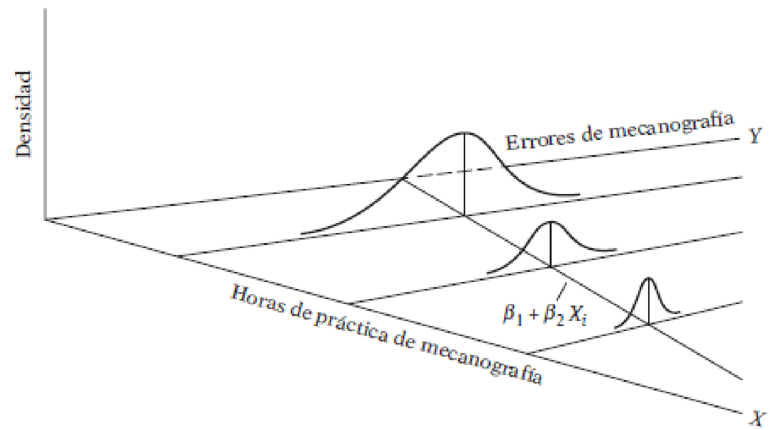
Figura 3. Posibles Resultados de la Corrección de Heteroscedasticidad Vía Transformaciones Sobre la Variable Dependiente



Fuente: (Cross Validated, 2018).

Figura 4. Caso Aplicado de Error Heterocedástico

FIGURA 11.3
Ilustración de la heteroscedasticidad.



Fuente: (Gujarati & Porter, 2010, pág. 367).

I.I. III. Consecuencias de la Heteroscedasticidad en la Estimación MCO

Es posible realizar la estimación MCO empleando la siguiente ecuación de la varianza del estimador[2] (Gujarati & Porter, 2010, pág. 370), asumiendo que se conocen las verdaderas σ_i^2:

$$var(\hat{\beta}_2) = \frac{\sum x_i \sigma_i^2}{(\sum x_i^2)^2}$$

Sin embargo, a pesar de que la ecuación anterior considera explícitamente la presencia de heteroscedasticidad, el estimador así calculado tendrá una mayor varianza que la del estimador real, lo que implica que los intervalos de confianza construidos con base en tales estimadores serán innecesariamente[3] grandes con relación a lo que en realidad son (Gujarati & Porter, 2010, pág. 374).

Por tanto, las pruebas t y F arrojarán resultados imprecisos. Por ejemplo, lo que parece un coeficiente estadísticamente no significativo (puesto que el valor t es

[2] Se presenta el caso más simple en que sólo existe una regresora.

[3] Sin que este hecho venga justificado por alguna razón con implicaciones prácticas.

más bajo de lo apropiado) puede resultar significativo si se estableciesen intervalos de confianza correctos con base en el procedimiento de MCG.

La validez estadística de las inferencias se debilita aún más si ni siquiera se utiliza la ecuación anterior para la estimación de los parámetros y la estimación por MCO se realiza en su forma tradicional, puesto que conllevará a un sesgo importante en la varianza del estimador calculado, es decir, que esta varianza del estimador calculado, en promedio, sobreestimará o subestimará la varianza del estimador real, lo cual resulta aún más problemático si se considera que, en general, no es posible determinar si el sesgo es positivo (sobreestimación) o negativo (subestimación)[4] (Gujarati & Porter, 2010, págs. 373-374). Esto usualmente conduce a inferencias muy equivocadas.

I.II. Mínimos Cuadrados Generalizados (MCG)

Como señalan (Gujarati & Porter, 2010, pág. 371), los estimadores usuales de mínimos cuadrados ordinarios (MCO) de algún coeficiente β_i pueden en ocasiones no ser los mejores estimadores, aunque sean insesgados. Ello ocurre en el escenario en que la relación entre la respuesta y el predictor (o los predictores) presenta gran variabilidad, *i.e.*, en aquellos donde el supuesto de homogeneidad de varianza es claramente incumplido.

En tales casos, al efectuar una regresión, es recomendable aprovechar el conocimiento que se posee sobre dicha variabilidad. Idealmente es deseable diseñar un esquema de estimación de manera tal que las observaciones que surgen de poblaciones con mayor variabilidad reciban menos peso que las que provienen de poblaciones con menor variabilidad, pues las primeras están más concentradas alrededor de sus valores medios que las últimas, lo que permite estimar la función

---

[4] No es posible puesto que el sentido de tal sesgo depende de la naturaleza de la relación entre σ_i^2 y los valores tomados por la variable explicativa X.

de regresión poblacional (FRP) en forma más precisa. Sin embargo, el método de MCO usual no sigue esta estrategia y, por consiguiente, no aprovecha la "información" contenida en la variabilidad desigual de la variable dependiente Y, por cuanto este método asigna igual peso o importancia a cada observación.

Sin embargo, existe un método de estimación, conocido como mínimos cuadrados generalizados (MCG), que toma en cuenta esa información explícitamente y, por consiguiente, es capaz de producir estimadores que son MELI.

Como señala (Greene, 2015, pág. 7), en términos geométricos y de poder determinar unívocamente cuáles son los mejores estimadores entre la clase de estimadores insesgados, el hecho de que existan $m + 1$ hiperplanos MCO para modelar un solo conjunto de datos en $m + 1$ variables es problemático[5]. Se desea tener un modelo lineal único para los datos, para el cual es válido resolver para cualquiera de las variables con fines de predicción.

Los mínimos cuadrados generalizados multivariados resuelven este problema buscando minimizar la media generalizada promedio[6] de las desviaciones cuadradas entre los datos y el hiperplano en todas las variables simultáneamente. Para el hiperplano de regresión resultante, es válido resolver cualquiera de las variables con fines de predicción.

Como señalan (Gujarati & Porter, 2010, pág. 373), el término de error estocástico de los MCO se transforma en el escenario MCG en una suma ponderada de residuos

[5] Fundamentalmente por dos motivos. El primero es por los criterios de selección que se requerirían para escoger cuál es el modelo óptimo dado el conjunto de datos. El segundo es por el costo computacional implicado.

[6] La media generalizada ("generalized mean" en la literatura en inglés) es una generalización de todos los tipos de media, conocida también como media de Hölder. Su siguiente y máxima generalización es la media cuasi-aritmética o media de Kolmogórov, la cual es equivalente conceptualmente a la media aritmética en este contexto; al hablar de una media generalizada "promedio" se hace referencia a la división del estadístico de prueba entre el número de observaciones durante el proceso de estimación.

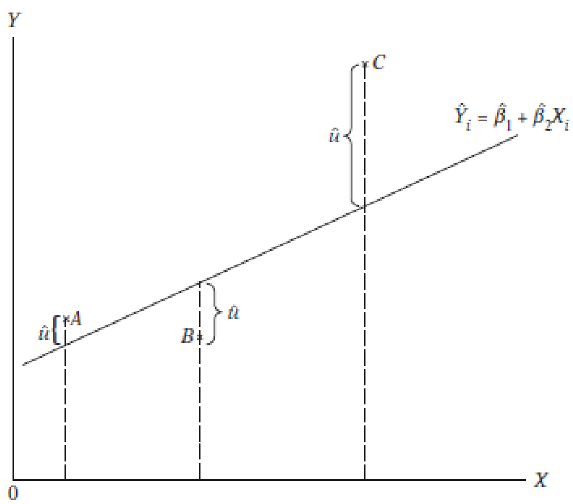
al cuadrado, donde $w_i = 1/\sigma_i^2$ actúan como ponderación, tal como se presenta a continuación:

Figura 5. Pesos Asignados a los Componentes del Término de Error en MCG

FIGURA 11.7
Diagrama de dispersión hipotético.

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i)^2$$



Fuente: (Gujarati & Porter, 2010, pág. 374).

Como puede observarse, en el método de MCO la suma de cuadrados residual (SCR) es equivalente a la del método MCG cuando los cuadrados de los residuos no se ponderan o, lo que es lo mismo, todas las ponderaciones son iguales. Como muestra la figura 1, en MCG el peso asignado a cada observación es inversamente proporcional a su σ_i, es decir, las observaciones que provienen de una población con una σ_i más grande tendrán una ponderación relativamente menor, y las de una población con un σ_i menor tendrán una ponderación proporcionalmente mayor al reducir la SCR.

Como señalan (Gujarati & Porter, 2010, pág. 376), aunque ya se estableció que, en caso de heteroscedasticidad, son los MCG y no los MCO los que son MELI, existen ejemplos en los que los MCO pueden ser MELI a pesar de la heteroscedasticidad. No obstante, dichos casos son poco frecuentes en la práctica y su existencia obedece a que el teorema de Gauss-Markov proporciona la condición suficiente (pero no necesaria) para que los MCO sean eficientes. La condición suficiente y necesaria para que los MCO sean MELI la establece el teorema de Kruskal[7].

I.III. Mínimos Cuadrados Ponderados (MCP)

Como señalan (Gujarati & Porter, 2010, pág. 389), la heterocedasticidad no destruye las propiedades de insesgamiento y consistencia de los estimadores MCO; sin embargo, éstos ya no son eficientes, ni siquiera asintóticamente. Esta falta de eficiencia resta credibilidad a los procedimientos habituales de pruebas de hipótesis. Existen dos tipos de medidas correctivas que generan dos tipos de modelos, en función de que conocer o no la covarianza de los miembros del término de error. El método más directo de corrección de heterocedasticidad tiene lugar cuando la verdadera matriz de covarianza del error[8] es conocida y los estimadores así obtenidos son los mejores estimadores lineales insesgados (MELI). Este método es el de los Mínimos Cuadrados Ponderados (MCP).

Como se señala en (Wikipedia, 2022), el método de MCP es un caso especial del método de MCG, el cual ocurre cuando todas las entradas fuera de la diagonal de la matriz de covarianzas Ω son 0. Esta situación surge cuando las varianzas de los valores observados son desiguales (es decir, hay heterocedasticidad presente), pero donde no existen correlaciones entre las variaciones observadas[9]. El peso de la

---

[7] Para mayores detalles, véase la investigación de Denzil G. Fiebig y Michael McAleer y Robert Bartels titulada "Properties of Ordinary Least Squares Estimators in Regression Models with Nonspherical Disturbances", publicada en el Journal of Econometrics, vol. 54, núm. 1-3, octubre-diciembre de 1992, p. 321-334.

[8] Los autores hacen referencia aquí a los valores poblacionales.

[9] Esto implica que la corrección por MCP es inútil si las regresoras están significativamente correlacionadas, es decir, si existe un nivel de multicolinealidad (que puede considerarse así, por ejemplo, cuando el factor de inflación de varianza es superior a 10).

unidad i es proporcional al recíproco[10] de la varianza de la respuesta de la unidad i.

I.IV. Mínimos Cuadrados Generalizados Factibles (MCGF)

Si se desconoce la covarianza de los errores Ω, que es el caso ampliamente más usual, se puede obtener una estimación coherente Ω, digamos $\hat{\Omega}$, utilizando una versión implementable de MCG conocida como estimador de mínimos cuadrados generalizados factibles (MCGF). En MCGF, el modelado procede en dos etapas:

1. El modelo es estimado por MCO u otro estimador consistente (pero ineficiente), y los residuales se usan para construir un estimador consistente de la matriz de covarianza de errores (para hacerlo, a menudo se necesita para examinar el modelo agregando restricciones adicionales, por ejemplo, si los errores siguen un proceso de series de tiempo, un estadístico generalmente necesita algunos supuestos teóricos sobre este proceso para asegurarse de que esté disponible un estimador consistente).

2. Utilizando el estimador consistente de la matriz de covarianza de los errores, se pueden implementar las ideas MCG.

Mientras que MCG es más eficiente que MCO bajo heterocedasticidad o autocorrelación, esto no es cierto para MCGF. El estimador factible es, siempre que la matriz de covarianza de errores se calcule de manera consistente, asintóticamente más eficiente, pero para una muestra de tamaño pequeño o mediano, en realidad puede ser menos eficiente que MCO. Por eso, algunos autores prefieren utilizar MCO y reformular sus inferencias simplemente considerando un estimador alternativo para la varianza del estimador robusto a la heteroscedasticidad o autocorrelación en serie. Pero para muestras grandes se

---

[10] El recíproco de un objeto matemático A es otro objeto matemático B tal que el producto de A por B es igual a la unidad. Por ejemplo, el recíproco de un escalar k que pertenece a los reales es $\frac{1}{k}$.

prefiere FGLS sobre MCO bajo heterocedasticidad o correlación serial. Una nota de advertencia es que el estimador MCGF no siempre es consistente. Un caso en el que MCGF podría ser inconsistente es si hay efectos fijos específicos individuales.

En general, este estimador (el MCGF) tiene propiedades diferentes a MCG. Para muestras grandes (es decir, asintóticamente) todas las propiedades son (en condiciones apropiadas) comunes con respecto a MCG, pero para muestras finitas se desconocen las propiedades de los estimadores MCGF: varían dramáticamente con cada modelo en particular y, como regla general, sus distribuciones exactas no se pueden derivar analíticamente[11]. Para muestras finitas, MCGF puede ser incluso menos eficiente que MCO en algunos casos.

Por lo tanto, aunque GLS puede ser factible, no siempre es prudente aplicar este método cuando la muestra es pequeña. Un método que a veces se usa para mejorar la precisión de los estimadores en muestras finitas es iterar, es decir, tomar los residuales de MCGF para actualizar el estimador de covarianza de errores y luego actualizar la estimación de MCGF, aplicando la misma idea de manera iterativa hasta que los estimadores varíen menos que algún nivel tolerado[12]. Pero este método no necesariamente mejora mucho la eficiencia del estimador si la muestra original era pequeña. Una opción razonable cuando las muestras no son demasiado grandes es aplicar MCO, pero descartando el estimador de varianza clásico $\sigma^2 * (X'X)^{(-1)}$ (que es inconsistente en este marco) y usando un estimador HAC (Heteroscedasticidad y Autocorrelación Consistente).

Como señalan (Gujarati & Porter, 2010, pág. 447), es importante observar que siempre que se estimen los parámetros del modelo transformado con un método MCGF o un MCGE, los coeficientes estimados no necesariamente tendrán las

---

[1111] Mediante inferencia formal, que por su naturaleza es de carácter apriorística.

[12] Lo cual, evidentemente, depende de las necesidades objetivas y concretas de la investigación.

propiedades óptimas usuales del modelo clásico, como ser MELI, sobre todo en muestras pequeñas. Sin profundizar en complejidades técnicas, es posible enunciar como principio general que, en régimen asintótico, siempre que se utilice un estimador en lugar de su verdadero valor, los coeficientes de MCO estimados quizás presenten las propiedades óptimas usuales. Asimismo, los procedimientos convencionales para pruebas de hipótesis son, en estricto sentido, válidos de modo asintótico. En consecuencia, para muestras pequeñas, se debe tener cuidado al interpretar los resultados estimados.

Además, al utilizar MCGF, si no se incluye la primera observación (como se hace, por ejemplo, en el método Cochrane-Orcutt[13]), se pueden ver afectados de modo adverso no sólo los valores numéricos, sino también la eficiencia de los estimadores, sobre todo si el tamaño de la muestra es pequeño y las regresoras no son, estrictamente hablando, no estocásticas (esto es especialmente cierto si las regresoras muestran una tendencia, lo cual es muy común en los datos económicos).

I.V. Consideraciones sobre el Patrón de Heterocedasticidad para la Investigación Aplicada con MCP

Para que aplicar MCP a un determinado conjunto de datos sea necesario, se requiere primero detectar si en efecto existe heteroscedasticidad en los mismos. Esta detección puede llevarse a cabo mediante diferentes metodologías, tanto gráficas/exploratorias como formales. Los métodos gráficos son útiles no solo como primera aproximación, sino por otros fines que se verán más adelante.

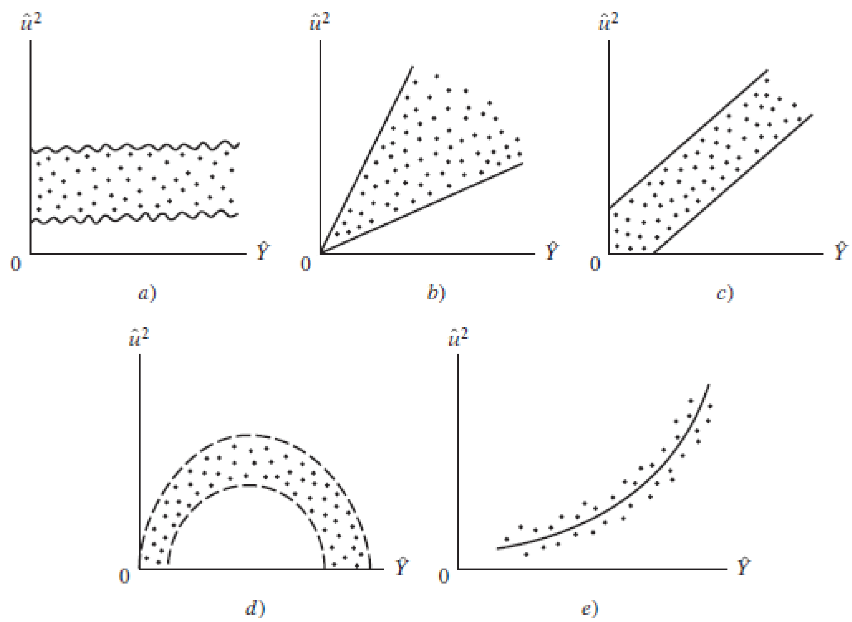
---

[13] Procedimiento econométrico de estimación que ajusta el modelo de regresión lineal clásico ante autocorrelación (correlación serial en el término de error).

Para emplear métodos gráficos en la detección de heteroscedasticidad, debe llevarse a cabo una regresión MCO sin modificación alguna. Posteriormente, deben graficarse los residuos de la regresión para realizar el examen exploratorio.

Una opción para la detección gráfica de heteroscedasticidad es graficar los residuos al cuadrado, sea contra la variable dependiente o contra las regresoras, tal como se muestra a continuación[14].

Figura 6. Gráfica de Patrones Hipotéticos de los Residuos de la Regresión MCO Elevados al Cuadrado vs Variable Dependiente

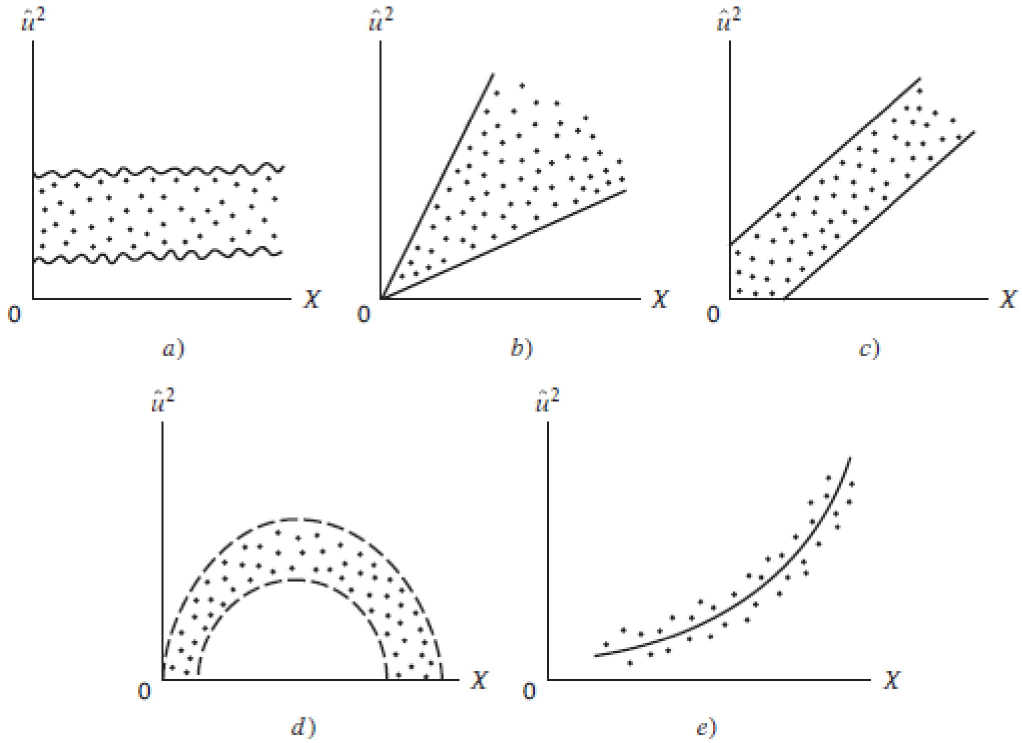


Fuente: (Gujarati & Porter, 2010, pág. 377).

---

[14] En el caso del modelo con solamente una variable explicativa o regresora, el gráfico de los residuos al cuadrado frente a los componentes de la variable dependiente es equivalente a graficar los residuos antes referidos frente a la variable regresora, por ello las figuras 6 y 7 son equivalentes. Sin embargo, ésta no es la situación cuando se considera un modelo con dos o más variables regresoras. En tal caso, los residuos al cuadrado pueden graficarse frente a cualquier variable incluida en el modelo.

Figura 7. Gráfica de Patrones Hipotéticos de los Residuos de la Regresión MCO Elevados al Cuadrado vs Variable Explicativa



Fuente: (Gujarati & Porter, 2010, pág. 378).

Un patrón como el de la figura 7.c, por ejemplo, sugiere que la varianza del término de perturbación se relaciona linealmente con la variable X, implicando con ello que es lógico esperar que la varianza heteroscedástica del término de error (perturbación estocástica en el contexto teórico) sea proporcional al valor de la variable explicativa. Conocer esto puede contribuir a transformar los datos de manera que, en la regresión sobre los datos transformados, la varianza de las perturbaciones sea homoscedástica. Para este caso concreto, se deberán dividir la variable dependiente, el intercepto, la variable explicativa y el error entre $\sqrt{X}$.

Por otro lado, si por criterio experto, métodos gráficos, la prueba de Park y/o la prueba de Glejser se considera que la varianza del error es proporcional al cuadrado de la variable explicativa, se deben dividir la variable dependiente, el intercepto, la variable explicativa y el error entre X.

En la misma línea, si por alguna razón se considera que la varianza del error es proporcional al cuadrado del valor medio de la variable dependiente Y, deben dividirse la variable dependiente, el intercepto, la variable explicativa y el error entre la esperanza matemática, valor medio o valor esperado de Y. Si la proporcionalidad es respecto al valor medio de Y y no al cuadrado de dicho valor medio, la división antes mencionada se efectuará con $\sqrt{Y}$ como denominador.

En (Olmo Chica, 2011) se exploran los escenarios, mediante regresiones auxiliares[15], en que la varianza del error tiene relación lineal con la inversa de la variable explicativa. Esto mismo puede aplicarse implementando otras transformaciones (por ejemplo, logarítmicas o exponenciales) sobre la variable explicativa o la variable dependiente en las regresiones auxiliares antes descritas.

II. CASO DE APLICACIÓN: PREDICCIÓN DE LA PRESIÓN SANGUÍNEA

Se toman 54 mujeres adultas entre 20 y 60 años y se les mide la presión sanguínea diastólica para ver su relación con la edad, el único predictor.

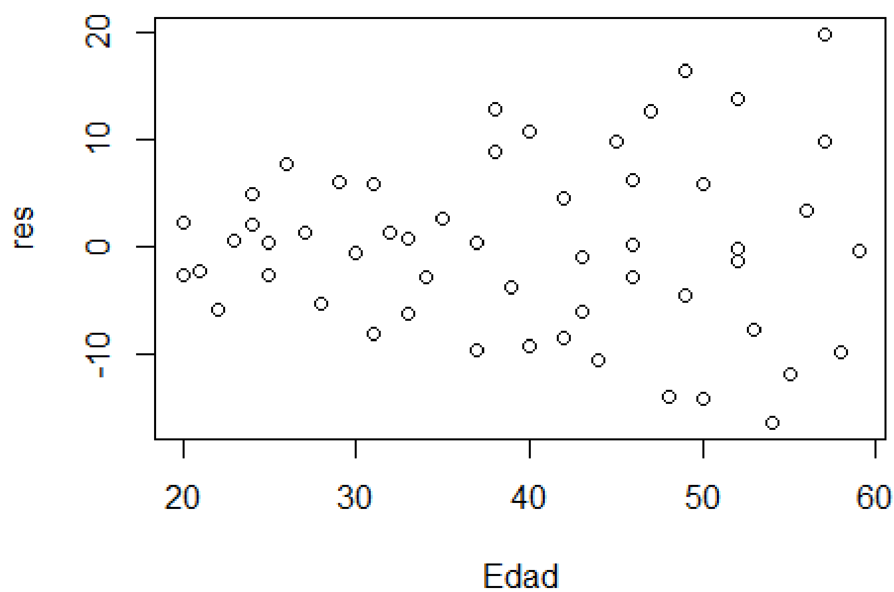
Con ellos puede estimarse un modelo de regresión que tenga como respuesta la presión y como predictor la edad. Es recomendable, por ejemplo, elaborar gráficos de los residuales y de los residuales al cuadrado contra edad, tal como se vio

---

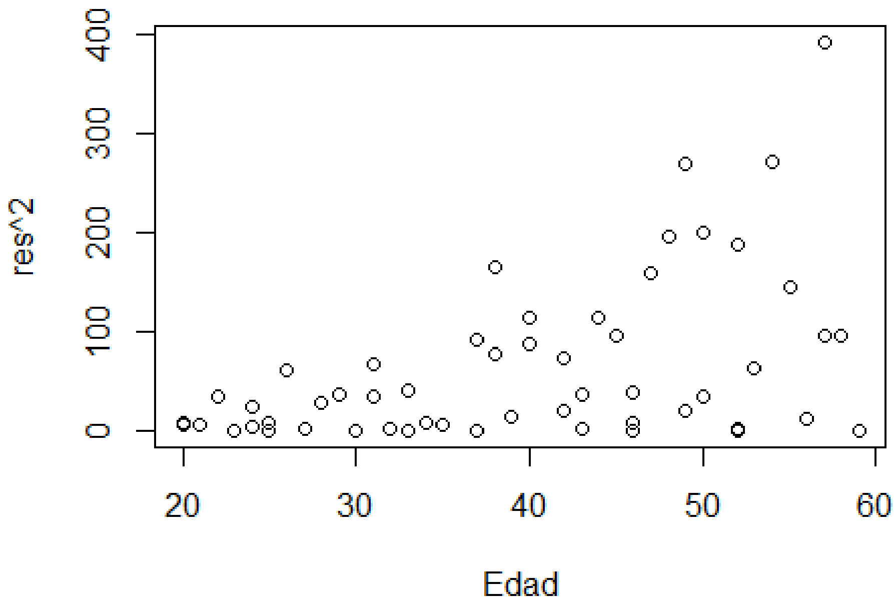
[15] En tales regresiones, la variable dependiente es el valor absoluto de los residuos y la variable explicativa para cada modelo son las transformaciones realizadas sobre la variable explicativa original. Lo que determina en estas regresiones auxiliares si la relación lineal entre el valor absoluto de los residuos y cada variable explicativa usada en cada modelo es el valor p del coeficiente (no se debe confundir con el intercepto, aunque este es también un coeficiente) y el coeficiente de correlación de Pearson R^2 resultante de la regresión.

anteriormente, con la finalidad de estudiar el comportamiento de la varianza residual.

```
load("presang.Rdata")
attach(base)

mod=lm(Presion~Edad,base)
res=mod$res
plot(Edad,res)
```




```
plot(Edad,res^2)
```



Puesto que en el gráfico de los residuales contra la edad tiene forma de megáfono, se sugiere encontrar la función de desviación estándar al ajustar el modelo de residuales absolutos contra edad. El segundo gráfico muestra una tendencia creciente por lo que se puede encontrar la función de variancia al ajustar el modelo de residuales cuadráticos contra edad.

En términos de lo antes expuesto, debe estimarse un modelo de desviación estándar poniendo como respuesta los residuales absolutos y como predictor la edad, tras lo cual deben encontrarse los valores estimados en este modelo y tomarse como las desviaciones estándar en la construcción de los pesos. El vector de ponderaciones w_i se construye tal que

$$w_i = \frac{1}{\hat{s}_i^2}$$

.

```
modres1=lm(abs(res)~Edad)
s=predict(modres1)
w=1/(s^2)
```

Para realizar el ajuste vía mínimos cuadrados ponderados es necesario construir la matriz W. Ello puede lograrse almacenando los pesos en una matriz diagonal (una matriz cuyos elementos fuera de la diagonal son ceros). Si w es el vector de ponderaciones o pesos, entonces W=diag(w) es la matriz diagonal.

```
W=diag(w)
```

Habiendo construido la matriz W puede realizarse la estimación de los coeficientes de regresión mediante:

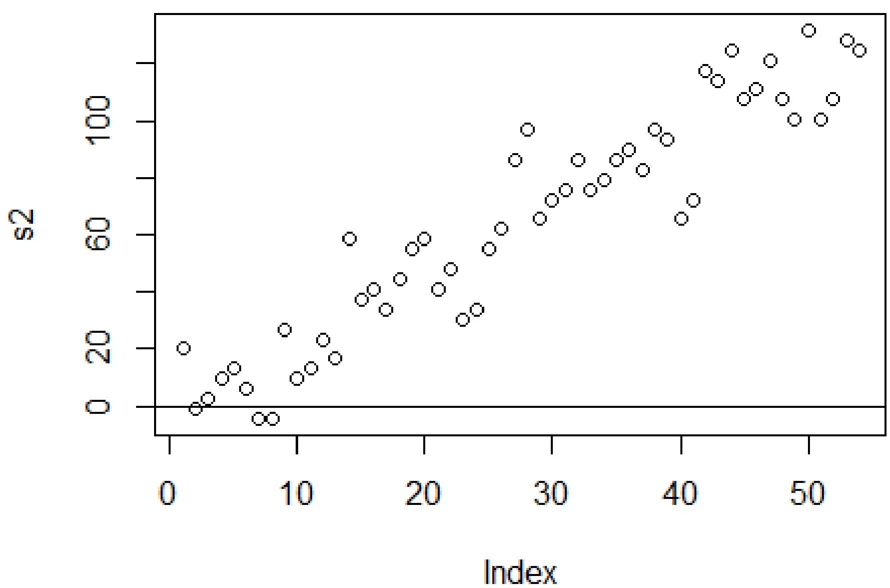
$$\hat{\beta}^{(w)} = (X^T W X)^{-1}(X^T W Y)$$

```
y=Presion
x=cbind(1,Edad)
bw=solve(t(x) %*% W %*% x) %*% (t(x) %*% W %*% y)
round(bw,2)

##       [,1]
##      55.57
## Edad  0.60
```

Adicionalmente, puede estimarse un modelo de la varianza en el que se establezca como respuesta a los residuales cuadráticos y como predictor la edad. Tras encontrarse los valores estimados en el modelo antes descrito, estos deben tomarse como las varianzas en la construcción de los pesos.

```
res2=res^2
modres2=lm(res2~Edad)
s2=predict(modres2)
```

```
plot(s2)
abline(h=0)
```



El lector puede observar que algunas estimaciones son negativas, por lo que no son adecuadas como estimaciones de las varianzas (puesto que esta es por definición positiva, dado que no existen distancias negativas y la varianza no es más que la distancia cuadrática que existe entre una observación y la media), por lo que no debe continuarse con este camino.

Además, pueden compararse los coeficientes obtenidos en el modelo sin ponderar y en el modelo que considera las ponderaciones.

```
round(cbind(mod$coef,bw),2)

##              [,1]  [,2]
## (Intercept) 56.16 55.57
## Edad         0.58  0.60
```

Es de especial interés la comparación entre el CMR (el cuadrado medio residual, que es el resultado de dividir la suma de cuadrados totales residuales entre los grados de libertad n-p) del modelo original sin poderar con el del modelo ponderado, en donde en el segundo caso se encuentra mediante

$$CMR^{(w)} = \frac{\sum w_i (y_i - \hat{y}_i)^2}{n - p}$$

.

```
cmres=anova(mod)[2,3]
n=nrow(base)
p=2
fitw=x%*%bw
resw=y-fitw
scresw=sum(w*resw^2)
cmresw=scresw/(n-p)
c(cmres,cmresw)

## [1] 66.353173  1.471414
```

La matriz de covarianzas de los coeficientes de regresión se determina mediante la ecuación

$$\hat{V}\left(\hat{\beta}^{(w)}\right) = CMR^{(w)}(X^T W X)^{-1}$$

```
vcovw=cmresw*solve(t(x) %*% W %*% x)
vcovw

##                     Edad
##       6.3550256 -0.189363636
## Edad -0.1893636  0.006278666
```

Podría ser de interés para el investigador comparar el error estándar del coeficiente para *edad* obtenido sin ponderar con el que se obtiene en el escenario ponderado.

```
sqrt(vcov(mod)[2,2])

## [1] 0.09695116

sqrt(vcovw[2,2])

## [1] 0.07923803
```

La respuesta media ponderada (el valor esperado ponderado de la respuesta) puede encontrarse mediante la ecuación

$$\bar{y}^{(w)} = \frac{\sum w_i y_i}{\sum w_i}$$

.

```
medw=sum(w*y)/sum(w)
medw

## [1] 73.55134
```

También puede calcularse la suma de cuadrados total ponderada mediante la siguiente ecuación:

$$SCT^{(w)} = \sum w_i \left(y_i - \bar{y}^{(w)}\right)^2$$

```
sctw=sum(w*(y-medw)^2)
sctw

## [1] 159.8543
```

Tras esto, puede también obtenerse el R^2 ponderado mediante

$$R^2_{(w)} = 1 - \frac{SCRes^{(w)}}{SCTot^{(w)}}$$

.

```
r2w=1-scresw/sctw
r2w

## [1] 0.5213548
```

Es de especial interés la comparación entre el R^2 obtenido sin ponderar y el ponderado.

```
r2=summary(mod)$r.sq
round(c(r2,r2w),2)

## [1] 0.41 0.52
```

Finalmente, pueden obtenerse todos los resultados anteriores de forma automatizada mediante la sintaxis lm indicando el vector de ponderaciones en el argumento weigths=w.

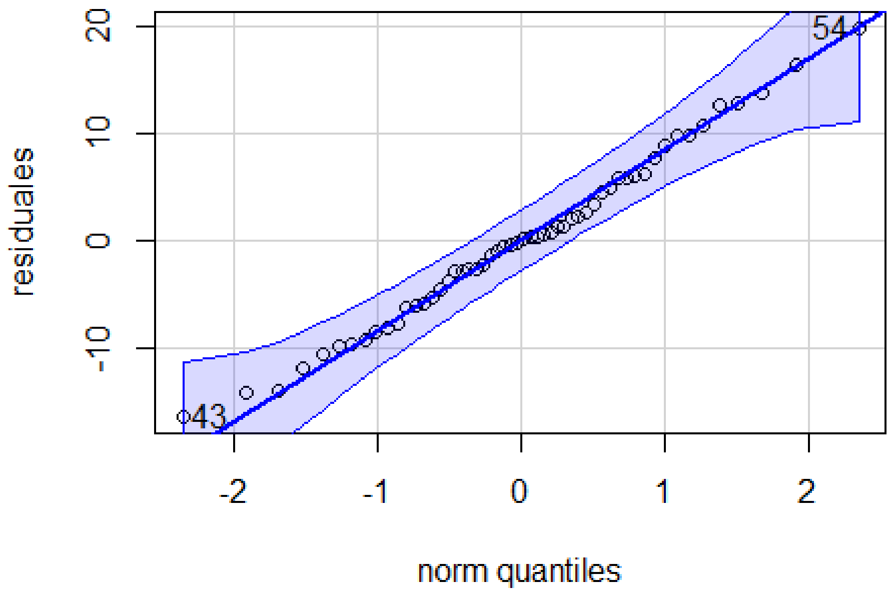
```
modw=lm(Presion~Edad, weights = w)
summary(modw)

##
## Call:
## lm(formula = Presion ~ Edad, weights = w)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0230 -0.9939 -0.0327  0.9250  2.2008
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 55.56577    2.52092  22.042  < 2e-16 ***
```

```
## Edad         0.59634    0.07924   7.526 7.19e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.213 on 52 degrees of freedom
## Multiple R-squared:  0.5214, Adjusted R-squared:  0.5122
## F-statistic: 56.64 on 1 and 52 DF,  p-value: 7.187e-10
```

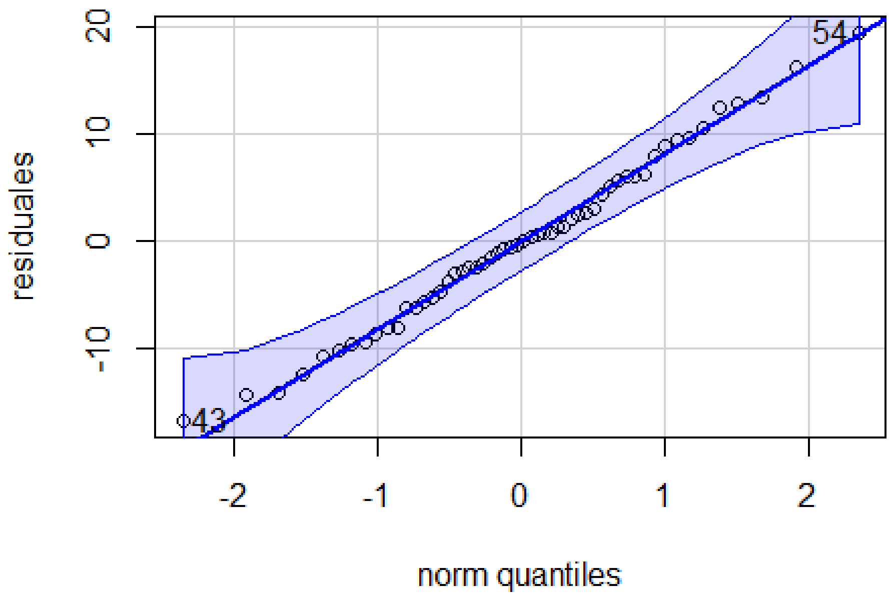
Puede estudiarse la homogeneidad de varianza de los residuos formalmente y la normalidad de dichos residuos de forma gráfica, con la finalidad de comparar los resultados para ambos modelos.

```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

bptest(mod,studentize=T)

##
##  studentized Breusch-Pagan test
##
## data:  mod
## BP = 12.541, df = 1, p-value = 0.0003981

library(car)

## Loading required package: carData
```

```
qqPlot(mod$res,ylab="residuales")
```




```
## [1] 54 43

library(lmtest)

bptest(modw,studentize=F)

##
##  Breusch-Pagan test
##
## data:  modw
## BP = 1159.8, df = 1, p-value < 2.2e-16

library(car)

qqPlot(modw$res,ylab="residuales")
```




```
## [1] 54 43
```

Nótese que a pesar de que en ambos modelos los residuos parecen distribuirse normalmente, tal apariencia es invalidada por el hecho de que el contraste de White indica que tales residuos no tienen una varianza homogénea, por lo que no pueden ser normales.

Si las estimaciones de los coeficientes cuando se usaron ponderaciones difieren sustancialmente de los primeros coeficientes cuando no se ponderó, se sugiere iterar el proceso de ponderación usando los residuales de los mínimos cuadrados ponderados para volver a estimar la función de varianza o desviación estándar y luego obtener nuevos pesos. A menudo bastan una o dos iteraciones para estabilizar los coeficientes de regresión.

Cuando hay poca diferencia en las varianzas de los errores el método de mínimos cuadrados ponderados no es muy útil. En tal escenario, el coeficiente de determinación R^2 no tiene un significado exacto, por lo que debe usarse con cautela.

III. REFERENCIAS

Cross Validated. (23 de Marzo de 2018). *Will log transformation always mitigate heteroskedasticity?* Obtenido de StackExchange: https://stats.stackexchange.com/questions/336315/will-log-transformation-always-mitigate-heteroskedasticity

Greene, N. (2015). Generalized Least-Squares Regressions V: Multiple Variables. *City University of New York Academic Works*, 17-25. Obtenido de https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1079&context=kb_pubs

Gujarati, D., & Porter, D. (2010). *Econometría.* México, D.F.: Fondo de Cultura Económica.

Nabi, I. (24 de Septiembre de 2021). *SUPUESTOS DEL MODELO CLÁSICO DE REGRESIÓN LINEAL Y DE LOS MODELOS LINEALES GENERALIZADOS.* Obtenido de Marxist Statistics: https://marxianstatistics.files.wordpress.com/2021/09/supuestos-del-modelo-clasico-de-regresion-lineal-y-de-los-modelos-lineales-generalizados-isadore-nabi.pdf

Olmo Chica, J. (2011). *Guía multimedia para la elaboración de un modelo econométrico (GUIME 2.0)*. Obtenido de Universidad de Granada: https://www.ugr.es/~jchica/Pagina2/Modelo/Ejercicios%20gretl4.htm

Wikipedia. (20 de Junio de 2022). *Generalized least squares*. Obtenido de Least squares: https://en.wikipedia.org/wiki/Generalized_least_squares