

Universidad Panamericana
Maestría en Ciencia de Datos
Econometría

Actividad RLM

Enrique Ulises Báez Gómez Tagle

20 de septiembre de 2025

Índice

1	Pregunta 1	3
2	Pregunta 2	3
2.1	Grafique los datos en un diagrama de dispersión y analice.	3
2.2	Efectúe la regresión de Y sobre X y examine los residuos de esta regresión (realice la gráfica y pruebas correspondientes). ¿Qué observa?	4
2.3	Como los datos de Chile parecen atípicos, repita la regresión en b) sin la información sobre Chile. Ahora examine los residuos de esta regresión. ¿Qué observa? (realice la gráfica y pruebas correspondientes).	5
2.4	Si, con base en los resultados de b), concluye que hubo heteroscedasticidad en la varianza del error, pero con base en los resultados de c) modifica este resultado, ¿qué conclusiones generales obtiene?	6
3	Pregunta 3	7
3.1	Haga la regresión del gasto alimentario sobre el gasto total y examine los residuos obtenidos en dicha regresión.	7
3.2	Grafique los residuos al cuadrado obtenidos en el inciso a) contra el gasto total y verifique si existe algún patrón sistemático.	8
3.3	Si la gráfica del inciso b) sugiere heteroscedasticidad, aplique las pruebas de Park, correlación de orden de Spearman y White para determinar si la sensibilidad respecto de la heteroscedasticidad observada en b) se sustenta con estas pruebas.	9
3.4	Efectúe la regresión del logaritmo del gasto alimentario sobre el logaritmo del gasto total. Si observa heteroscedasticidad en el modelo lineal del inciso a) pero no en el modelo log-log, ¿a qué conclusión llega? Muestre todas las pruebas necesarias.	9
4	Pregunta 4	11
4.1	Cuando hay presencia de autocorrelación, los estimadores de MCO son sesgados e ineficientes.	11
4.2	La prueba DW de Durbin–Watson supone que la varianza del término de error u_t es homoscedástica.	11
4.3	La transformación de primeras diferencias para eliminar la autocorrelación supone que el coeficiente de autocorrelación ρ es -1.	11
4.4	Los valores R^2 de dos modelos, de los cuales uno corresponde a una regresión en primeras diferencias y el otro a una regresión en niveles, no son directamente comparables.	11

4.5	Un DW de Durbin–Watson significativo no necesariamente denota autocorrelación de primer orden.	11
4.6	En presencia de autocorrelación, las varianzas calculadas convencionalmente y los errores estándar de los valores pronosticados son ineficientes.	12
4.7	La exclusión de una o varias variables importantes de un modelo de regresión puede producir un valor DW significativo.	12
5	Pregunta 5	12
5.1	a) DW=1.05	12
5.2	b) DW=1.40	12
5.3	c) DW=2.50	12
5.4	d) DW=3.97	12
6	Pregunta 6	12
6.1	a) Regresión de primeras diferencias	12
6.2	b) Regresión de promedios móviles	13
7	Link al repositorio con código fuente	13

1. Pregunta 1

Tiene la siguiente información: SCR_1 basado en las primeras 30 observaciones 55, gl 25; SCR_2 basado en las últimas 30 observaciones 140, gl 25. Realice la prueba de heteroscedasticidad de Goldfeld-Quandt en el nivel de significancia de 5 %.

Hipótesis

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad (\text{homoscedasticidad})$$

$$H_1 : \sigma_2^2 > \sigma_1^2 \quad (\text{varianza mayor en el 2° subgrupo; prueba de una cola})$$

Estadístico de prueba

$$F = \frac{SCR_2/gl_2}{SCR_1/gl_1} = \frac{140/25}{55/25} = \frac{140}{55} = 2.545455.$$

Bajo H_0 , $F \sim F(gl_2, gl_1) = F(25, 25)$.

Regla de decisión (una cola, $\alpha = 0.05$)

Rechazar H_0 si $F > F_{1-\alpha}(25, 25) = 1.955447$.

Resultados

$$F_{\text{obs}} = 2.545455,$$

$$F_{\text{crítico}}(\text{una cola}, \alpha = 0.05, 25, 25) = 1.955447,$$

$$p\text{-valor (una cola)} = 0.011518,$$

$$p\text{-valor (dos colas)} = 0.023036.$$

Conclusión

Como $F_{\text{obs}} = 2.545455 > 1.955447$ y el p-valor de una cola es $0.011518 < 0.05$, **rechazamos H_0 de homoscedasticidad** al 5 % y concluimos que existe evidencia de **heteroscedasticidad**, con varianza del error *mayor en el segundo subgrupo* (últimas 30 observaciones).

Si planteamos la prueba bilateral, el p-valor = $0.023036 < 0.05$ también lleva a rechazar H_0 .

2. Pregunta 2

La tabla 11.9 presenta información acerca de los precios de acciones (Y) y los precios al consumidor (X) expresados en cambios porcentuales anuales para un corte transversal de 20 países.

2.1. Grafique los datos en un diagrama de dispersión y analice.

La gráfica de dispersión (Figura 1) muestra una relación positiva entre X y Y , pero se observa un valor atípico correspondiente a Chile, con valores muy superiores al resto. Este outlier puede distorsionar la regresión y afectar las pruebas de supuestos.

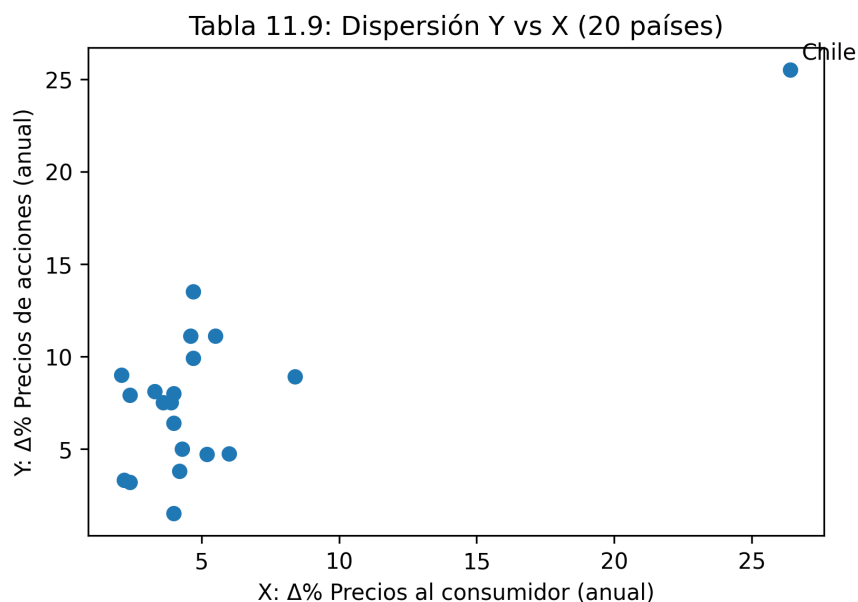


Figura 1: Tabla 11.9: Dispersión Y vs X (20 países)

2.2. Efectúe la regresión de Y sobre X y examine los residuos de esta regresión (realice la gráfica y pruebas correspondientes). ¿Qué observa?

La regresión de Y sobre X muestra un coeficiente positivo (Cuadro 1), aunque influido por el outlier.

Cuadro 1: Pregunta 2(b): Coeficientes OLS para Y X (todos los países)

Parámetro	Coef	EE	t	p-valor	IC 2.5 %	IC 97.5 %
Intercept	3.7781	0.9993	3.7807	0.0014	1.6786	5.8775
X	0.8033	0.1366	5.8801	0.0000	0.5163	1.0903

El gráfico de residuos vs. ajustados (Figura 2) sugiere heteroscedasticidad y presencia de valores influyentes.

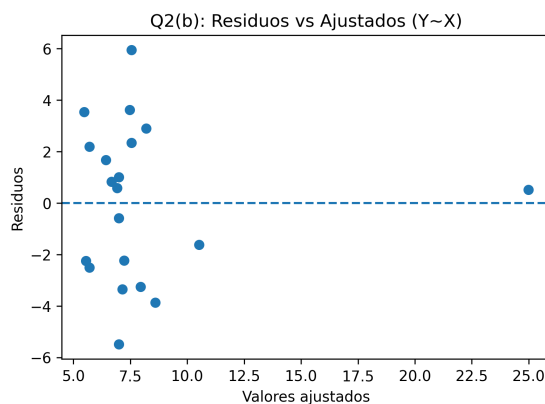


Figura 2: Q2(b): Residuos vs Ajustados ($Y \sim X$)

El QQ-plot (Figura 3) revela desviaciones de normalidad en colas.

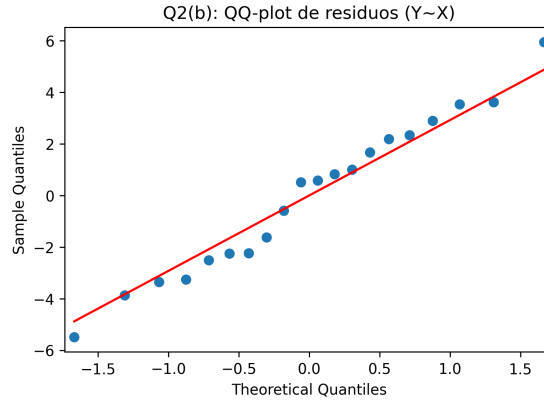


Figura 3: Q2(b): QQ-plot de residuos ($Y \sim X$)

Las pruebas de heteroscedasticidad (Cuadro 2) confirman rechazo de homoscedasticidad.

Cuadro 2: Pregunta 2(b): Pruebas de heterocedasticidad para $Y \sim X$ (todos)

Prueba	Estadístico	p-valor
Breusch-Pagan	0.5949	0.4405
White	1.1586	0.5603

- 2.3. Como los datos de Chile parecen atípicos, repita la regresión en b) sin la información sobre Chile. Ahora examine los residuos de esta regresión. ¿Qué observa? (realice la gráfica y pruebas correspondientes).

Al excluir a Chile, los coeficientes se estabilizan (Cuadro 3).

Cuadro 3: Pregunta 2(c): Coeficientes OLS para $Y \sim X$ (sin Chile)

Parámetro	Coef	EE	t	p-valor	IC 2.5 %	IC 97.5 %
Intercept	4.9327	2.1864	2.2561	0.0375	0.3198	9.5456
X	0.5209	0.4934	1.0557	0.3059	-0.5200	1.5618

Los residuos vs. ajustados (Figura 4) muestran mayor homogeneidad.

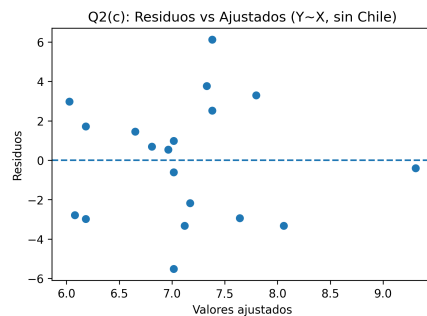


Figura 4: Q2(c): Residuos vs Ajustados ($Y \sim X$, sin Chile)

El QQ-plot (Figura 5) se ajusta mejor a la normal.

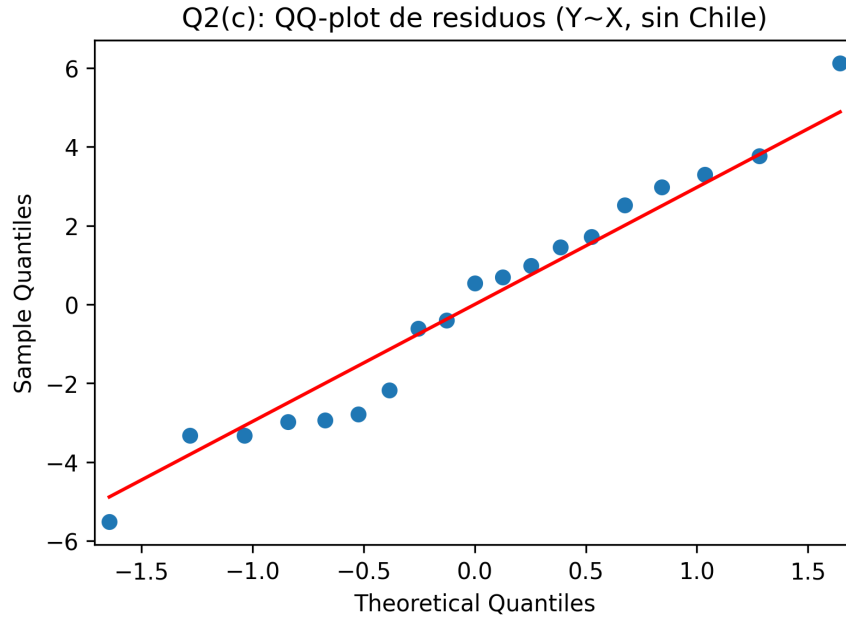


Figura 5: Q2(c): QQ-plot de residuos ($Y \sim X$, sin Chile)

Las pruebas de heteroscedasticidad (Cuadro 4) ya no son significativas, sugiriendo homoscedasticidad sin Chile.

Cuadro 4: Pregunta 2(c): Pruebas de heterocedasticidad para $Y \sim X$ (sin Chile)

Prueba	Estadístico	p-valor
Breusch-Pagan	0.0275	0.8684
White	1.5742	0.4552

2.4. Si, con base en los resultados de b), concluye que hubo heteroscedasticidad en la varianza del error, pero con base en los resultados de c) modifica este resultado, ¿qué conclusiones generales obtiene?

Comparando (b) y (c): la heteroscedasticidad detectada se debe al valor atípico de Chile. Al removerlo, los supuestos de homoscedasticidad y normalidad se cumplen mejor.

Conclusión: el modelo no presenta problemas salvo por la influencia de un outlier.

Cuadro 5: Pregunta 2(d): Comparación de OLS y pruebas de heterocedasticidad (todos vs. sin Chile)

Modelo	β_0	β_1	R^2	BP p-valor	White p-valor
Todos	3.7781	0.8033	0.6576	0.4405	0.5603
Sin Chile	4.9327	0.5209	0.0615	0.8684	0.4552

3. Pregunta 3

Gasto alimentario en India. En la tabla 2.8 se proporcionaron datos sobre el gasto en alimentos y el gasto total de 55 familias de India.

3.1. Haga la regresión del gasto alimentario sobre el gasto total y examine los residuos obtenidos en dicha regresión.

Modelo lineal:

$$\text{Food}_i = \beta_0 + \beta_1 \text{Total}_i + \varepsilon_i, \quad i = 1, \dots, 55.$$

Estimación OLS y significancia (ver Cuadro 6): el coeficiente de Total es $\hat{\beta}_1 = 0.4368$ ($t = 5.577$, $p < 0.001$), con $R^2 = 0.37$.

Cuadro 6: Pregunta 3(a): Coeficientes OLS de Food sobre Total

Parámetro	Coef	EE	t	p-valor	IC 2.5 %	IC 97.5 %
Intercept	94.2088	50.8563	1.8524	0.0695	-7.7961	196.2137
Total	0.4368	0.0783	5.5770	0.0000	0.2797	0.5939

Diagnósticos de residuos.

El gráfico de residuos vs. ajustados sugiere varianza creciente (patrón en “abanico”), indicio de heteroscedasticidad.

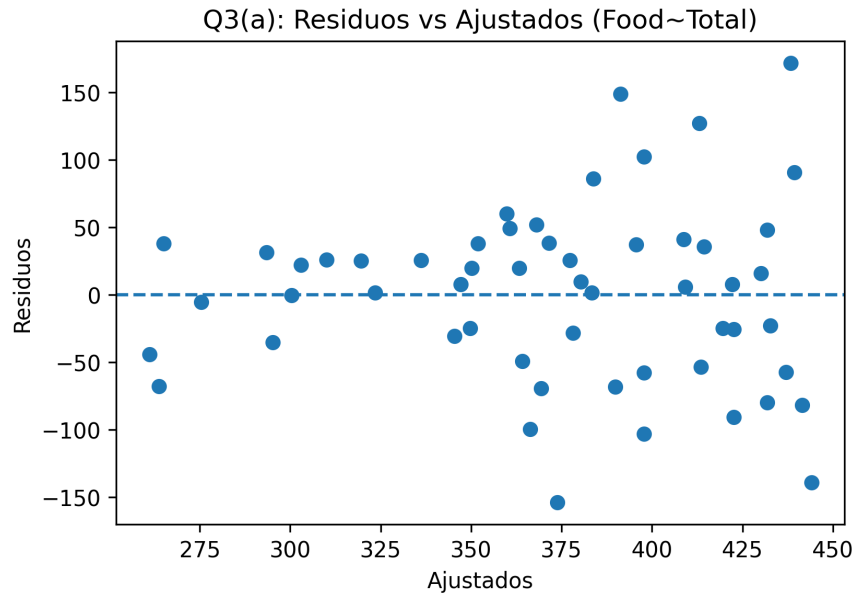


Figura 6: Q3(a): Residuos vs Ajustados (Food Total)

La normalidad de residuos es razonable (Jarque–Bera $p = 0.8792$), según el QQ-plot:

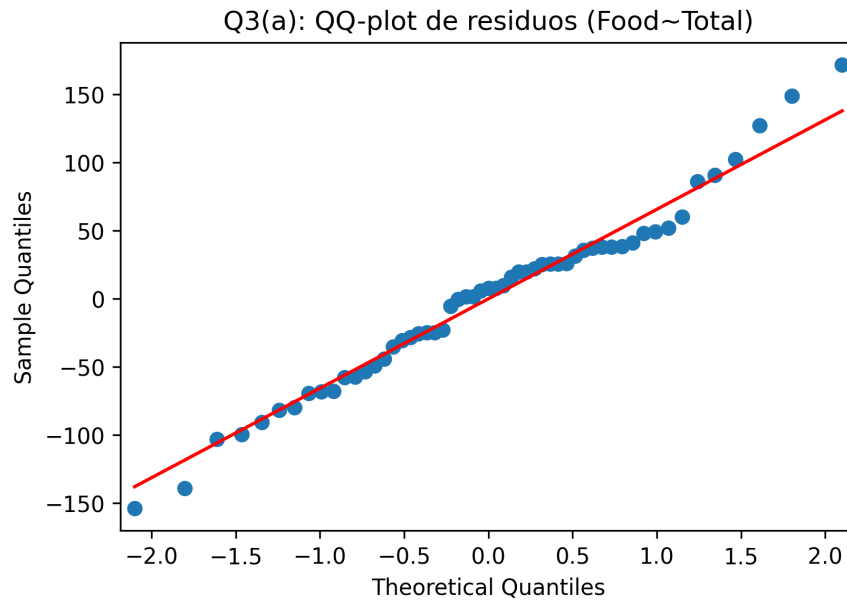


Figura 7: Q3(a): QQ-plot de residuos (Food Total)

3.2. Grafique los residuos al cuadrado obtenidos en el inciso a) contra el gasto total y verifique si existe algún patrón sistemático.

La relación positiva entre $\hat{\varepsilon}_i^2$ y $Total_i$ es evidente, con mayor dispersión a niveles altos de gasto total.

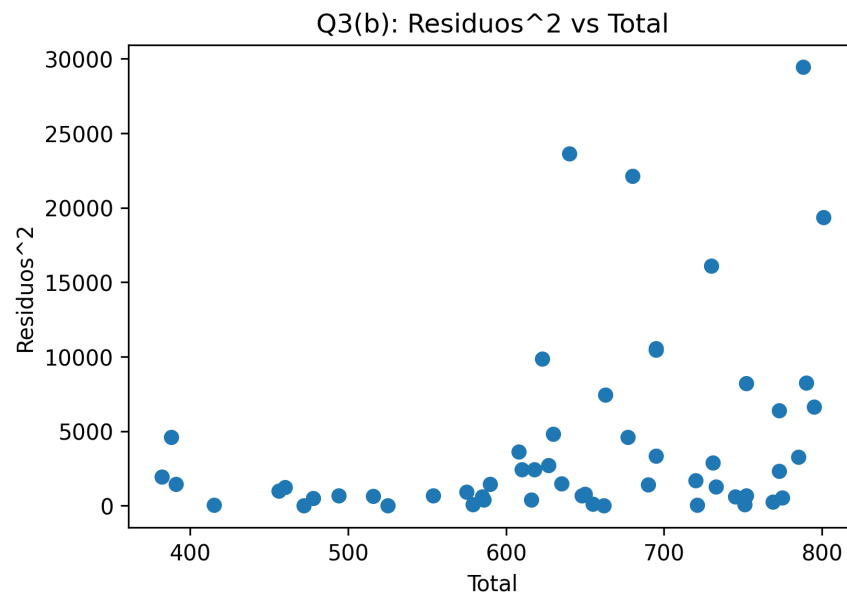


Figura 8: Q3(b): $\hat{\varepsilon}^2$ vs. Total

La gráfica sugiere un patrón sistemático de varianza creciente.

- 3.3. Si la gráfica del inciso b) sugiere heteroscedasticidad, aplique las pruebas de Park, correlación de orden de Spearman y White para determinar si la sensación respecto de la heteroscedasticidad observada en b) se sustenta con estas pruebas.

Prueba de Park.

Modelo probado:

$$\ln(\hat{\varepsilon}_i^2) = \alpha + \gamma \ln(\text{Total}_i) + u_i.$$

Hipótesis: $H_0: \gamma = 0$ (homoscedasticidad) vs. $H_1: \gamma \neq 0$.

Resultado (Cuadro 7): $\hat{\gamma} = 3.7032$, $t = 2.386$, $p = 0.0206 \Rightarrow$ **rechazo de H_0** .

Cuadro 7: Pregunta 3(c): Prueba de Park ($\ln(\text{resid}^2)$ $\ln(\text{Total})$)

Parámetro	Coef	EE	t	p-valor	IC 2.5 %	IC 97.5 %
Intercept	-16.8629	10.0014	-1.6861	0.0977	-36.9231	3.1974
ln_Total	3.7032	1.5519	2.3863	0.0206	0.5906	6.8159

Correlación de Spearman.

Se contrasta la monotonía entre $|\hat{\varepsilon}_i|$ (o $\hat{\varepsilon}_i^2$) y Total_i . Resultado: $\rho = 0.3627$, $p = 0.0065 \Rightarrow$ **evidencia de heteroscedasticidad**.

Cuadro 8: Pregunta 3(c): Correlación de Spearman para heteroscedasticidad

Métrica	rho	p-valor
—resid— vs Total	0.3627	0.0065
resid^2 vs Total	0.3627	0.0065

Prueba de White.

Se reporta el estadístico y su p -valor en el Cuadro 9. El contraste compara contra una χ^2 con g.l. correspondientes. Los resultados son consistentes con la presencia de heteroscedasticidad.

Cuadro 9: Pregunta 3(c): Prueba de White (modelo lineal)

Prueba	Estadístico	p-valor
White	7.3745	0.0250

Las tres pruebas respaldan la sospecha de heteroscedasticidad detectada en (a)-(b).

- 3.4. Efectúe la regresión del logaritmo del gasto alimentario sobre el logaritmo del gasto total. Si observa heteroscedasticidad en el modelo lineal del inciso a) pero no en el modelo log-log, ¿a qué conclusión llega? Muestre todas las pruebas necesarias.

Modelo log-log:

$$\ln(\text{Food}_i) = \beta_0 + \beta_1 \ln(\text{Total}_i) + \eta_i.$$

Estimación OLS (Cuadro 10): $\hat{\beta}_1 = 0.7363$ ($t = 6.100$, $p < 0.001$), con $R^2 = 0.412$.

Cuadro 10: Pregunta 3(d): Coeficientes OLS (log-log)

Parámetro	Coef	EE	t	p-valor	IC 2.5 %	IC 97.5 %
Intercept	1.1543	0.7780	1.4838	0.1438	-0.4061	2.7147
ln_Total	0.7363	0.1207	6.0998	0.0000	0.4942	0.9784

Los residuos del modelo log-log muestran varianza más estable y mejor normalidad:

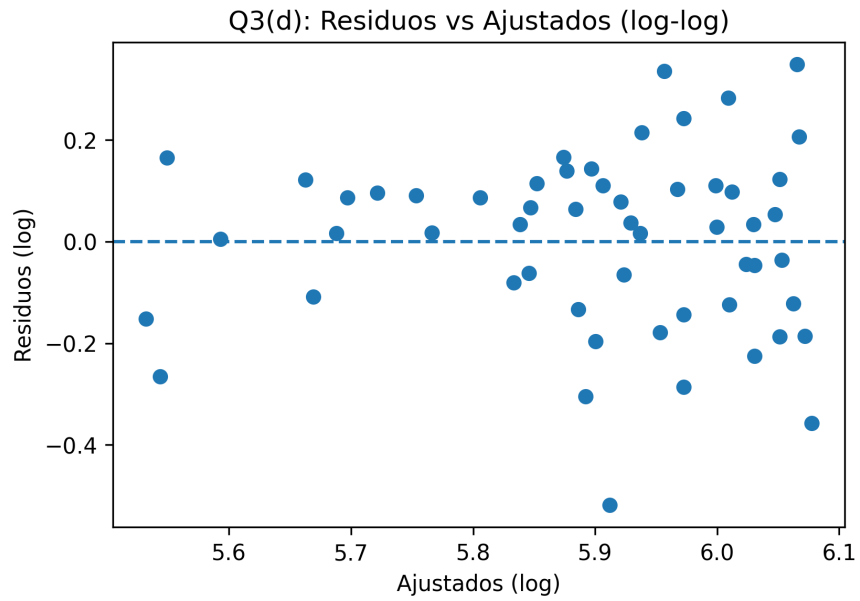


Figura 9: Q3(d): Residuos vs Ajustados (log-log)

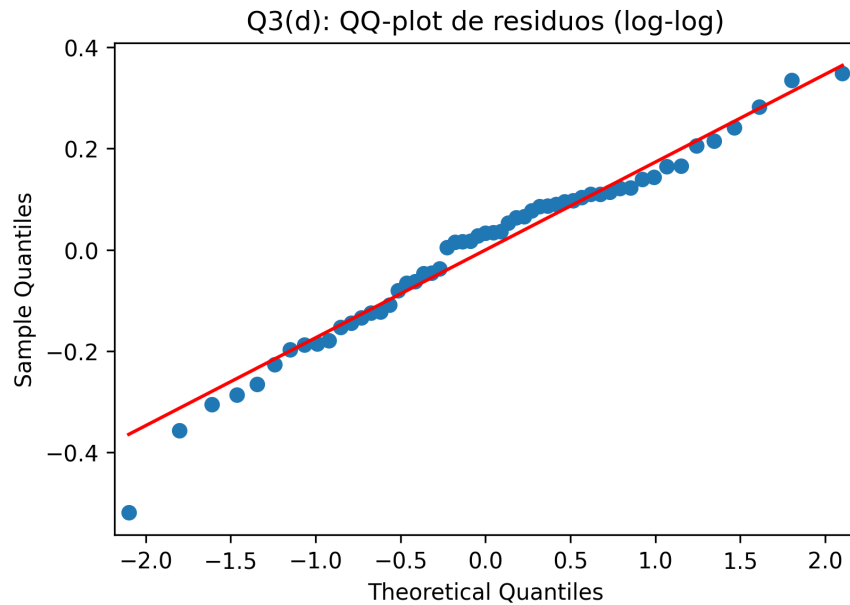


Figura 10: Q3(d): QQ-plot de residuos (log-log)

Pruebas (Cuadro 11) indican no rechazar homoscedasticidad en el modelo transformado.

Cuadro 11: Pregunta 3(d): Pruebas de heterocedasticidad (log-log)

Prueba	p-valor
White	0.3600
Breusch-Pagan	0.2355

La heteroscedasticidad en el modelo lineal parece deberse a varianza proporcional al nivel de gasto. La transformación logarítmica estabiliza la varianza y mejora los supuestos clásicos, por lo que el modelo log-log es preferible para hacer inferencia.

4. Pregunta 4

Establezca si las siguientes afirmaciones son verdaderas o falsas. Justifique brevemente.

4.1. Cuando hay presencia de autocorrelación, los estimadores de MCO son sesgados e ineficientes.

Falso. Con $E(u_t | X) = 0$, los estimadores MCO siguen siendo *insesgados* y consistentes; lo que falla es la eficiencia y la inferencia porque $\text{Var}(\hat{\beta})$ ya no es $\sigma^2(X'X)^{-1}$ sino $(X'X)^{-1}X'\Sigma X(X'X)^{-1}$ con $\Sigma \neq \sigma^2 I$. Los errores estándar “convencionales” quedan mal calculados; usar GLS o correcciones tipo Newey–West.

4.2. La prueba DW de Durbin–Watson supone que la varianza del término de error u.t es homoscedástica.

Verdadero. El contraste DW se deriva bajo $H_0: \rho = 0$ suponiendo errores homoscedásticos e independientes en ausencia de autocorrelación y sin y_{t-1} como regresor. Heteroscedasticidad o términos dependientes afectan su validez.

4.3. La transformación de primeras diferencias para eliminar la autocorrelación supone que el coeficiente de autocorrelación rho es -1.

Falso. La *primeras diferencias* corresponde al caso especial de cuasi-diferenciación cuando $\rho = 1$ (no -1). Para $u_t = \rho u_{t-1} + \varepsilon_t$, la transformación correcta es $y_t - \rho y_{t-1}$ y $x_t - \rho x_{t-1}$; si $\rho = 1$ se reduce a diferenciar.

4.4. Los valores R^2 de dos modelos, de los cuales uno corresponde a una regresión en primeras diferencias y el otro a una regresión en niveles, no son directamente comparables.

Verdadero. En primeras diferencias la variable dependiente cambia de escala y varianza, por lo que TSS y, en consecuencia, $R^2 = 1 - SSR/TSS$ se definen sobre magnitudes distintas al modelo en niveles; no son comparables entre especificaciones con dependientes diferentes.

4.5. Un DW de Durbin–Watson significativo no necesariamente denota autocorrelación de primer orden.

Verdadero. Un DW extremo rechaza $H_0: \rho = 0$, pero la causa puede ser autocorrelación de orden superior, estacionalidad, omisiones o mala especificación. El test no identifica que el proceso sea estrictamente AR(1).

4.6. En presencia de autocorrelación, las varianzas calculadas convencionalmente y los errores estándar de los valores pronosticados son ineficientes.

Verdadero. Con $\Sigma \neq \sigma^2 I$, los estimadores de varianza “convencionales” son inválidos o inconsistentes y tienden a subestimar la incertidumbre; por tanto los errores estándar y bandas de predicción basadas en ellos no son confiables. Use GLS o correcciones robustas (HAC/Newey–West).

4.7. La exclusión de una o varias variables importantes de un modelo de regresión puede producir un valor DW significativo.

Verdadero. Omitir variables relevantes con persistencia temporal hace que el término de error incluya un componente serialmente correlacionado (w_t), induciendo $\rho \neq 0$ y, por tanto, un estadístico DW significativo aunque el verdadero error estructural fuese no correlacionado. %%

5. Pregunta 5

Con una muestra de 50 observaciones y 4 variables explicativas, ¿qué puede decir sobre autocorrelación si

5.1. a) DW=1.05

Evidencia de autocorrelación positiva. Regla aproximada: $\hat{\rho} \approx 1 - \frac{DW}{2} = 1 - 0.525 = 0.475$. Con tablas DW (n=50, k=4), valores tan bajos suelen caer por debajo de $d_L \rightarrow$ rechazar $H_0 : \rho = 0$ a favor de $\rho > 0$.

5.2. b) DW=1.40

Autocorrelación positiva moderada (posible) o zona indeterminada. Aproximación: $\hat{\rho} \approx 1 - 0.70 = 0.30$. Sin tablas exactas, es probable que DW esté entre d_L y $d_U \rightarrow$ conclusión formal: *indeterminada* al 5%; indicios de $\rho > 0$.

5.3. c) DW=2.50

Indicios de autocorrelación negativa. Aproximación: $\hat{\rho} \approx 1 - 1.25 = -0.25$. En términos de reglas DW, valores por encima de 2 sugieren $\rho < 0$; confirmar con comparación a $4 - d_U$ (n=50, k=4).

5.4. d) DW=3.97

Autocorrelación negativa muy fuerte. $\hat{\rho} \approx 1 - \frac{3.97}{2} = 1 - 1.985 = -0.985$ (cercano a -1). Normalmente $DW > 4 - d_L \rightarrow$ rechazar $H_0 : \rho = 0$ a favor de $\rho < 0$. %%

6. Pregunta 6

Establezca las circunstancias en que sería adecuado cada uno de los siguientes métodos de estimación del coeficiente de autocorrelación de primer orden ρ :

6.1. a) Regresión de primeras diferencias

Adecuada cuando la autocorrelación es de primer orden *positiva* y elevada ($0 < \rho < 1$, típicamente $\rho \approx 1$) y se desea eliminar la persistencia. La cuasi-diferenciación $y_t - \rho y_{t-1}$ se reduce a *primeras diferencias* si $\rho = 1$. Útil con tendencias suaves o nivel casi integrado, sin y_{t-1} como regresor y con T moderado-grande (la pérdida de una observación no es crítica). No es recomendable si ρ es negativo o pequeño.

6.2. b) Regresión de promedios móviles

Adecuada cuando la autocorrelación de primer orden es *negativa* ($\rho \approx -1$) y los errores alternan de signo entre periodos (serrucho). La transformación de *promedios móviles* (p.ej. usar $\frac{1}{2}(y_t + y_{t-1})$ y $\frac{1}{2}(x_t + x_{t-1})$) suaviza esa alternancia y elimina gran parte de la correlación serial. Es preferible con ruido de alta frecuencia y sin tendencia marcada. No es la opción indicada para $\rho > 0$.

7. Link al repositorio con código fuente

<https://github.com/enriquegomeztagle/MCD-Econometria/tree/main/HWs/RegressionAssumptions>