

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	Pregunta 7	13
7.1	a) Interprete los resultados.	13
7.2	b) Obtenga los residuos y los residuos estandarizados de la regresión anterior y gráfíquelos. ¿Qué opina sobre la presencia de autocorrelación en estos residuos?	14
7.3	c) Estime el estadístico DW de Durbin–Watson y comente sobre la naturaleza de la autocorrelación presente en los datos.	15
7.4	d) Realice las pruebas gráficas de autocorrelación de los errores poblacionales.	16
8	Pregunta 8	17
8.1	a) Estime la regresión anterior.	17
8.2	Realice las pruebas gráficas de normalidad y de no autocorrelación.	18
8.3	Con los residuos estimados, investigue si hay autocorrelación positiva mediante i) la prueba de Durbin–Watson y ii) las dos pruebas analíticas de normalidad para los errores poblacionales.	21
8.4	Si sospecha que la estructura autorregresiva del error es de orden $p = 1$, verifíquelo con la prueba de Breusch–Godfrey.	21
8.5	Con base en los resultados de esta prueba, ¿cómo transformaría los datos para eliminar la autocorrelación?	22
9	Link al repositorio con código fuente	22

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.

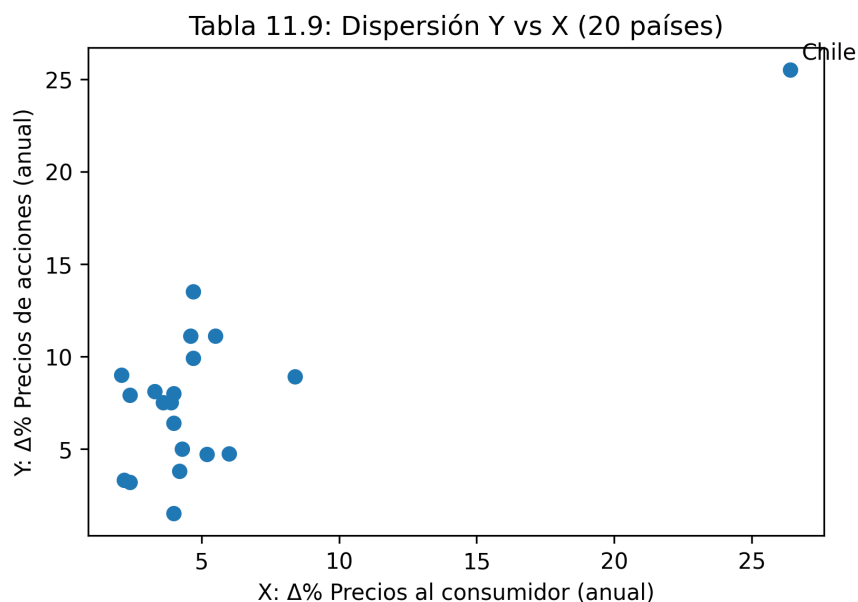


Figura 1: Tabla 11.9: Dispersión Y vs X (20 países)

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.

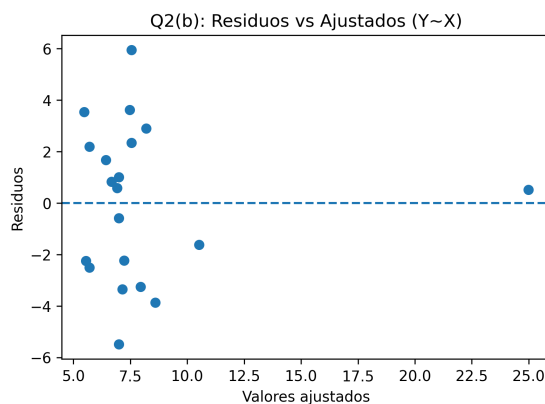


Figura 2: Q2(b): Residuos vs Ajustados ($Y \sim X$)

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

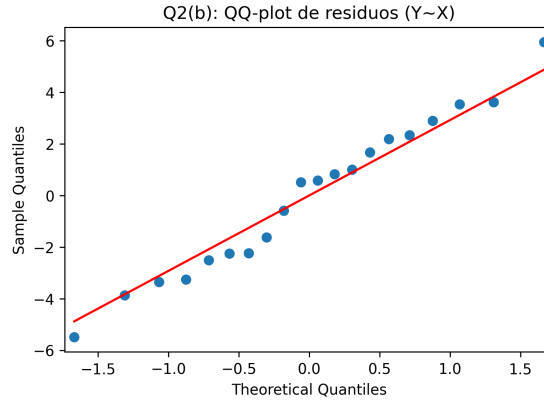


Figura 3: Q2(b): QQ-plot de residuos ($Y \sim X$)

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.

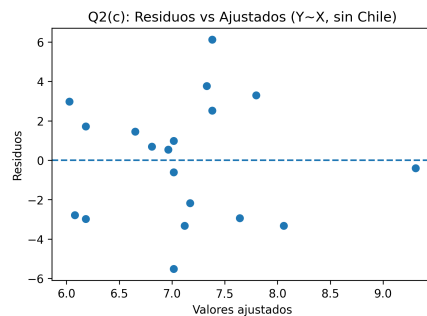


Figura 4: Q2(c): Residuos vs Ajustados ($Y \sim X$, sin Chile)

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

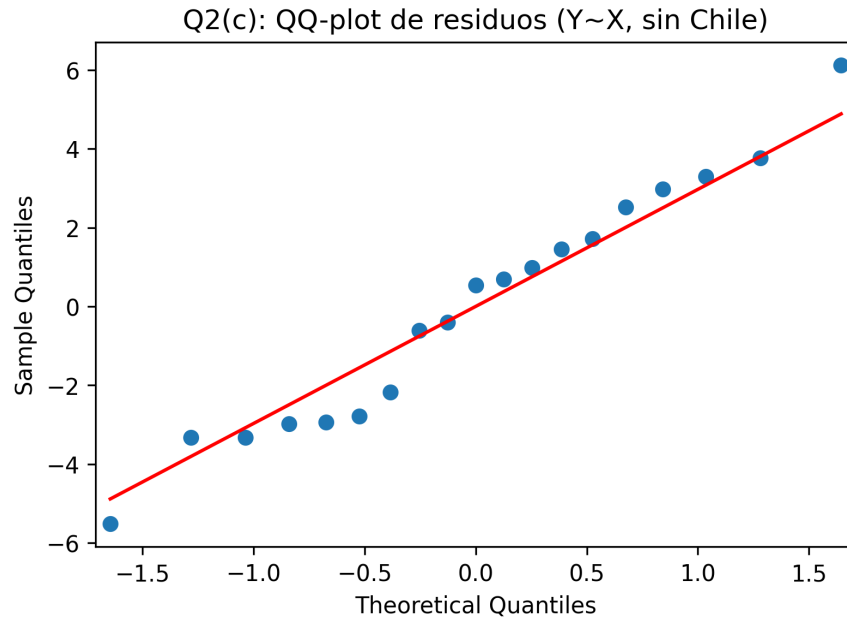


Figura 5: Q2(c): QQ-plot de residuos ($Y \sim X$, sin Chile)

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.

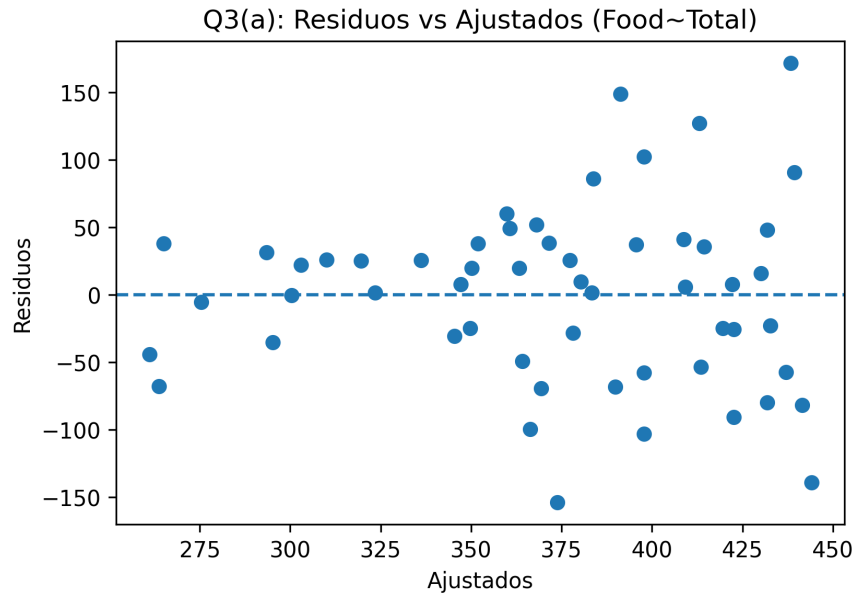


Figura 6: Q3(a): Residuos vs Ajustados (Food Total)

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

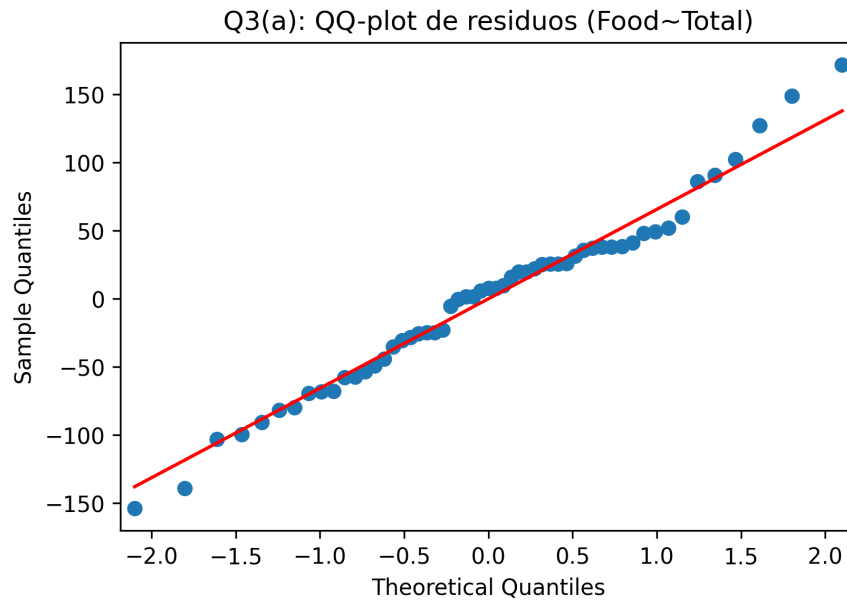


Figura 7: Q3(a): QQ-plot de residuos (Food Total)

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.

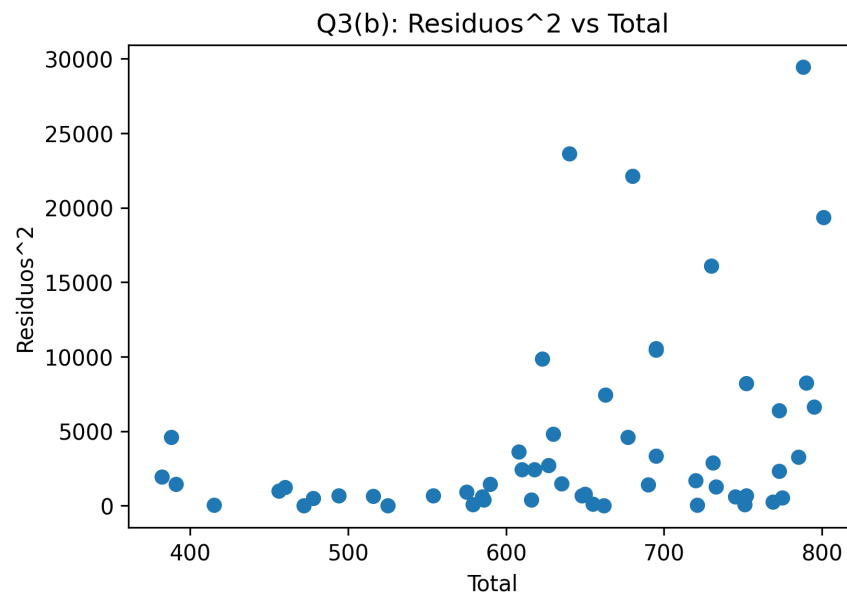


Figura 8: Q3(b): $\hat{\varepsilon}^2$ vs. Total

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:

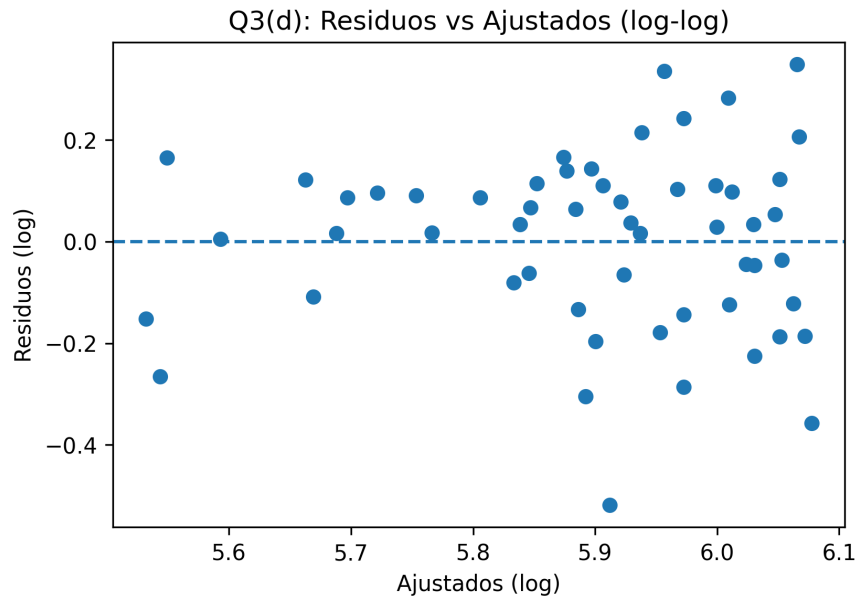


Figura 9: Q3(d): Residuos vs Ajustados (log-log)

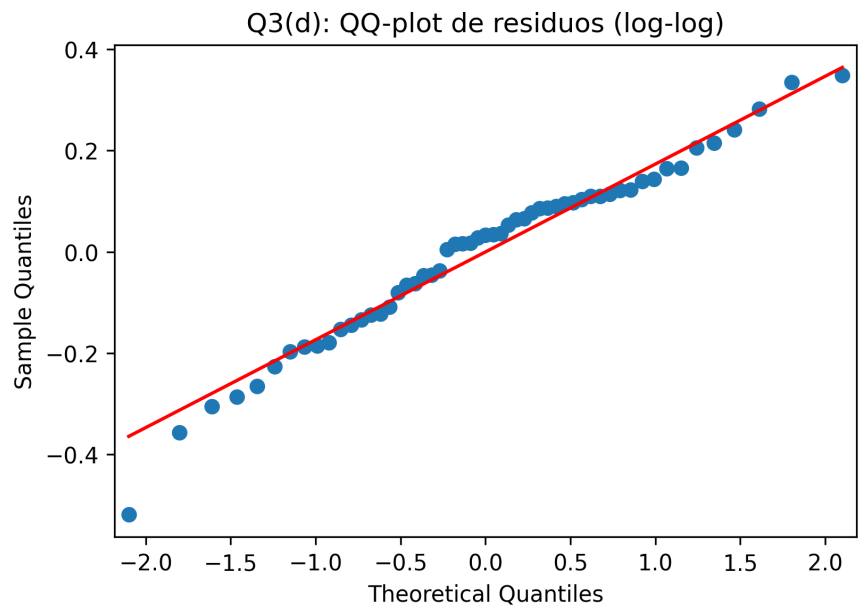


Figura 10: Q3(d): QQ-plot de residuos (log-log)

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

Conclusión indeterminada al 5 % (posible $\rho < 0$, no concluyente). Aproximación: $\hat{\rho} \approx 1 - 1.25 = -0.25$. Para probarr $4 - DW$ con d_L y d_U : aquí $4 - 2.50 = 1.50$ y, con $n = 50, k = 4$, suele cumplirse $d_L < 1.50 < d_U \Rightarrow$ *zona indeterminada*. No se rechaza ni se acepta $H_0 : \rho = 0$.

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.

Cuadro 12: Resultados OLS para $\ln C_t$ con regresores $\ln I_t, \ln L_t, \ln H_t, \ln A_t$. Errores estándar entre paréntesis.

Parámetro	Coef.	EE	t	p-valor
const	-1.5004	1.0030	-1.4959	0.1472
$\ln I$	0.4675	0.1660	2.8165	0.0093
$\ln L$	0.2794	0.1147	2.4357	0.0223
$\ln H$	-0.0052	0.1429	-0.0360	0.9715
$\ln A$	0.4414	0.1065	4.1447	0.0003
R^2	0.9361	NaN	NaN	NaN
R^2 ajustado	0.9259	NaN	NaN	NaN
N	30.0000	NaN	NaN	NaN

6.2. b) Regresión de promedios móviles

Útil como **heurística cuando** la autocorrelación de primer orden es *negativa* ($\rho \approx -1$) y los errores alternan de signo ("serrucho"). Promediar observaciones sucesivas (p.ej. $\frac{1}{2}(y_t + y_{t-1})$ y $\frac{1}{2}(x_t + x_{t-1})$) puede atenuar esa alternancia; sin embargo, *no es el método estándar para estimar ρ* . En aplicaciones formales se prefiere **GLS** tipo *Cochrane–Orcutt* o *Prais–Winsten* estimando $\rho < 0$, o bien revisar si hay *sobre-diferenciación* que induzca ρ negativo. No es adecuado para $\rho > 0$ ni para tendencias marcadas.

7. Pregunta 7

Consulte los datos sobre la industria del cobre de la tabla 12.7.

Con base en esta información, estime el siguiente modelo de regresión:

$$\ln C_t = \beta_1 + \beta_2 \ln I_t + \beta_3 \ln L_t + \beta_4 \ln H_t + \beta_5 \ln A_t + u_t.$$

7.1. a) Interprete los resultados.

Modelo log–log (elasticidades). Los coeficientes se interpretan como elasticidades de C respecto a cada insumo:

- $\hat{\beta}_2 = 0.468$ (*Inversión*): un 1 % más de I se asocia con ≈ 0.47 % más de C ($t = 2.82$, $p = 0.009$).
- $\hat{\beta}_3 = 0.279$ (*Trabajo L*): efecto positivo y significativo ($t = 2.44$, $p = 0.022$).
- $\hat{\beta}_4 = -0.005$ (*Horas H*): no significativo ($p = 0.972$); no hay evidencia de efecto marginal de H ceteris paribus.
- $\hat{\beta}_5 = 0.441$ (A): efecto positivo y significativo ($t = 4.15$, $p < 0.001$).

Ajuste global alto: $R^2 = 0.936$ ($R^2_{aj} = 0.926$) y $F = 91.5$ ($p < 10^{-13}$). Normalidad de residuos consistente con Jarque–Bera ($p = 0.736$).

- 7.2. b) Obtenga los residuos y los residuos estandarizados de la regresión anterior y gráfíquelos. ¿Qué opina sobre la presencia de autocorrelación en estos residuos?

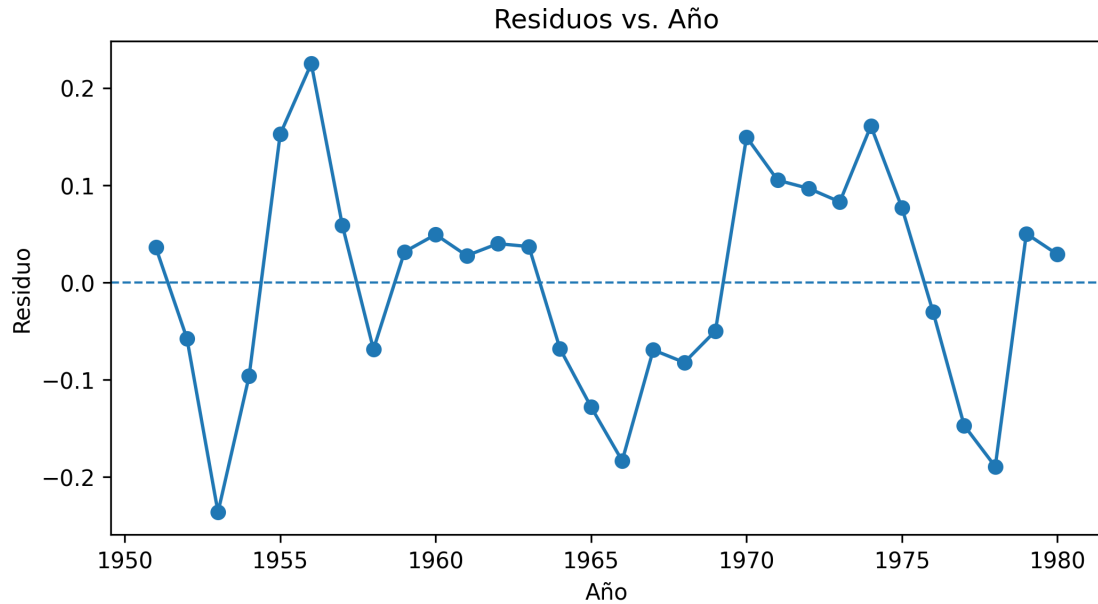


Figura 11: Residuos vs. tiempo

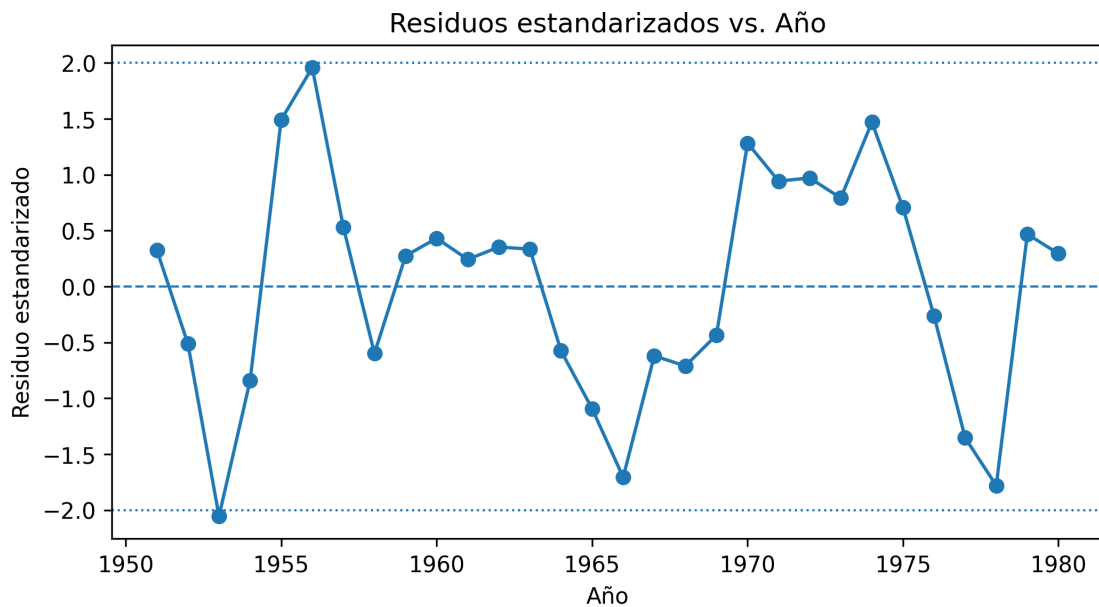


Figura 12: Residuos estandarizados vs. tiempo

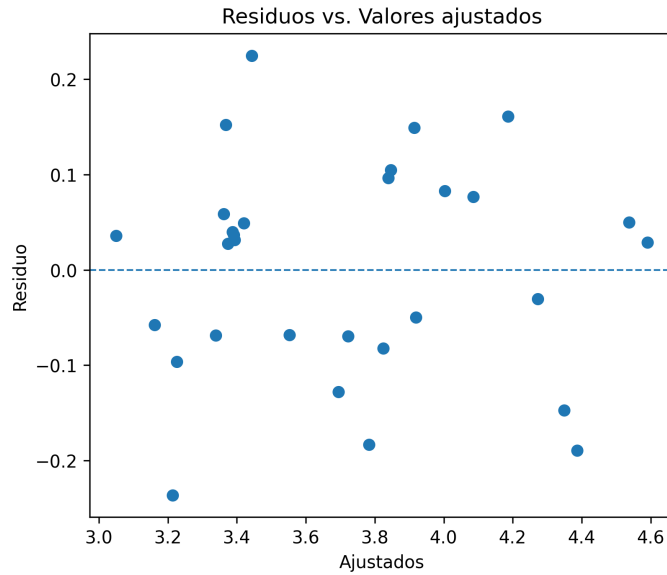


Figura 13: Residuos vs. ajustados

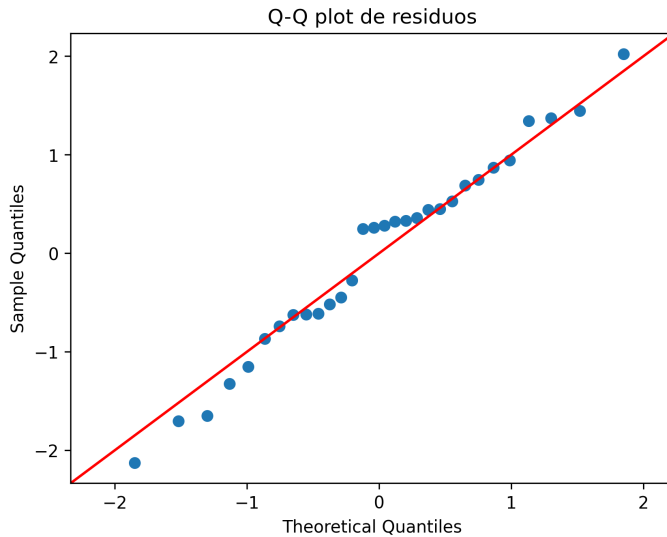


Figura 14: QQ-plot de residuos

Los residuos muestran rachas persistentes (bloques positivos/negativos), lo que sugiere **autocorrelación positiva**. La nube residuos-ajustados no muestra patrón de abanico marcado (heteroscedasticidad no evidente). El QQ-plot es cercano a la diagonal (normalidad aceptable).

7.3. c) Estime el estadístico DW de Durbin–Watson y comente sobre la naturaleza de la autocorrelación presente en los datos.

Durbin–Watson. $DW = 0.9549$. Con $n = 30$ y $k = 4$, un valor tan bajo cae típicamente por debajo de $d_L \Rightarrow$ **rechazo de $H_0 : \rho = 0$ a favor de autocorrelación positiva de primer orden**. Esto concuerda con: (i) el diagrama u_t vs. u_{t-1} con pendiente positiva y (ii) la ACF con barra significativa en el rezago 1.

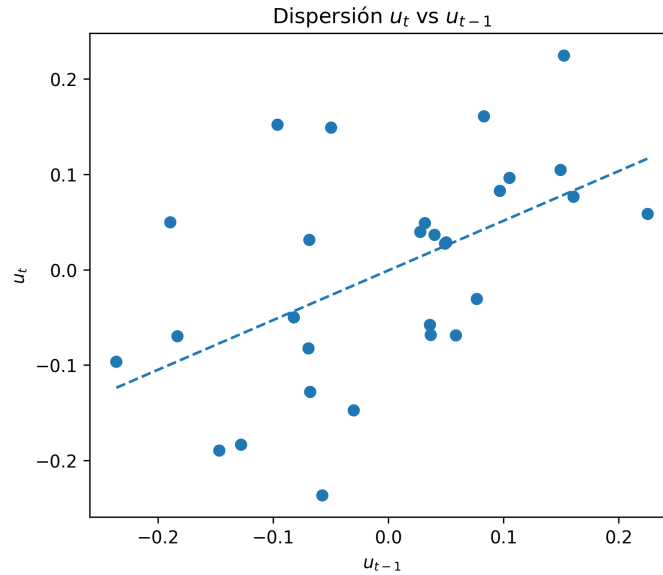


Figura 15: Dispersión u_t vs. u_{t-1}

7.4. d) Realice las pruebas gráficas de autocorrelación de los errores poblacionales.

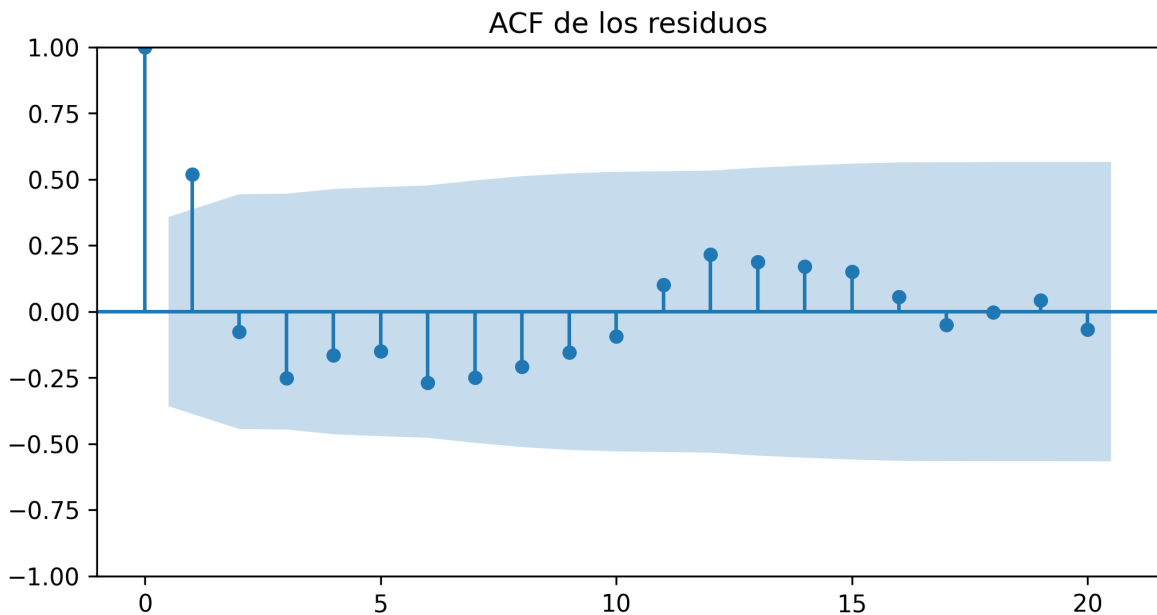


Figura 16: Función de autocorrelación (ACF) de los residuos

La ACF exhibe una barra positiva y significativa en $\ell = 1$ y decaimiento posterior, consistente con un proceso **AR(1)** con $\rho > 0$. En conjunto con $DW < 2$, la evidencia apunta a autocorrelación positiva en los errores.

8. Pregunta 8

Con los datos proporcionados en la tabla 12.9, estime el modelo

$$Y_t = \beta_1 + \beta_2 X_t + u_t,$$

donde Y son inventarios y X ventas, ambas medidas en miles de millones de dólares.

8.1. a) Estime la regresión anterior.

Estimamos

$$Y_t = \beta_1 + \beta_2 X_t + u_t, \quad t = 1950, \dots, 1990.$$

Cuadro 13: Estadísticos descriptivos de ventas, inventarios y razón inventarios/ventas.

Variable	N	Media	Desv.Est.	Mínimo	P25	P50	P75	Máximo
sales	41.0	208038.7	107874.9	46486.0	113201.0	206326.0	299766.0	411663.0
inventories	41.0	312958.1	131513.5	84646.0	188378.0	339516.0	423082.0	509902.0
ratio	41.0	1.6009	0.2086	1.2141	1.4362	1.6576	1.7676	1.9240

Cuadro 14: Matriz de correlaciones.

Variable	Ventas	Inventarios	Razón
sales	1.0000	0.9711	-0.9156
inventories	0.9711	1.0000	-0.8094
ratio	-0.9156	-0.8094	1.0000

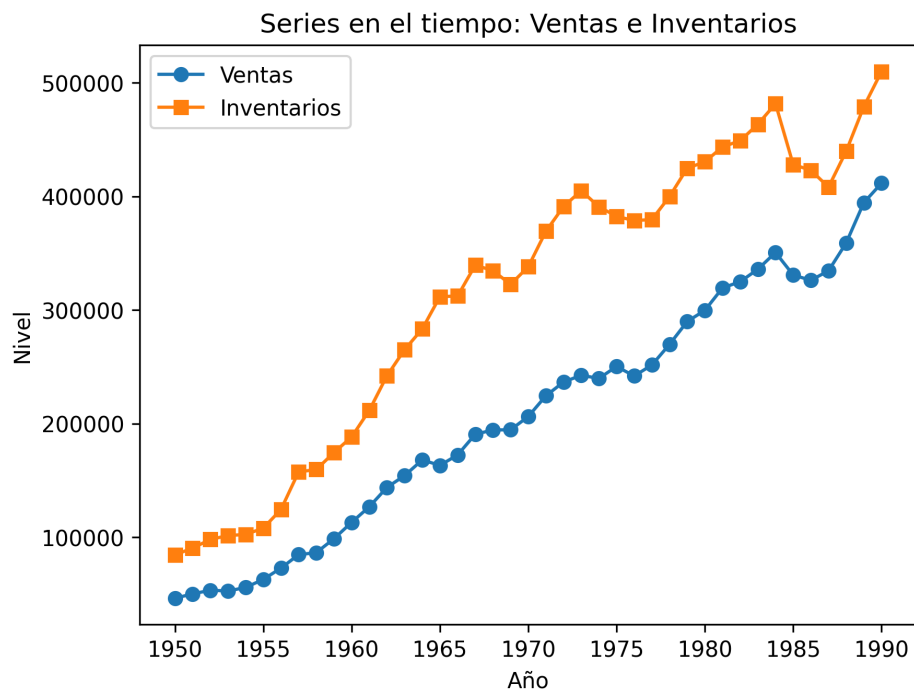


Figura 17: Series en el tiempo: ventas e inventarios

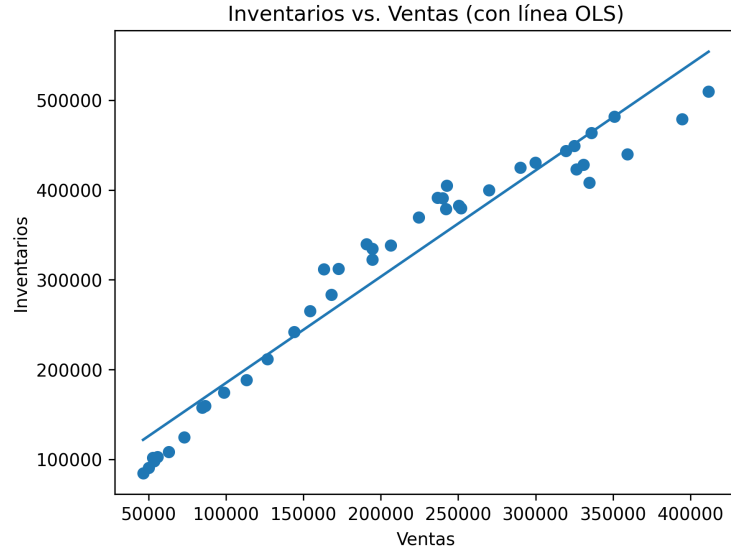


Figura 18: Inventarios vs. ventas con recta OLS

OLS (Modelo 1).

Cuadro 15: Resultados OLS: Inventarios sobre Ventas.

Parámetro	Coef	EE	t	p-valor	IC 2.5 %	IC 97.5 %
const	66669.0	10895.7	6.12	0.000	44630.3	88707.7
sales	1.184	0.047	25.40	0.000	1.090	1.278

Lectura: $\hat{\beta}_2 = 1.184$ ($p < 0.001$), $R^2 = 0.943$. Sin embargo, veremos abajo que los residuos presentan fuerte dependencia temporal.

OLS (Modelo 2, con tendencia):

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 \text{Year}_t + u_t.$$

Cuadro 16: Resultados OLS: Inventarios sobre Ventas y Año.

Parámetro	Coef	EE	t	p-valor	IC 2.5 %	IC 97.5 %
const	6057807.3	7324071.9	0.83	0.413	-8769001.2	20884615.8
sales	1.523	0.418	3.65	0.001	0.678	2.369
year	-3077.0	3761.6	-0.82	0.419	-10692.1	4538.0

El término de tendencia no resulta significativo y no corrige la dependencia temporal.

8.2. Realice las pruebas gráficas de normalidad y de no autocorrelación.

Gráficas de residuos (M1 y M2).

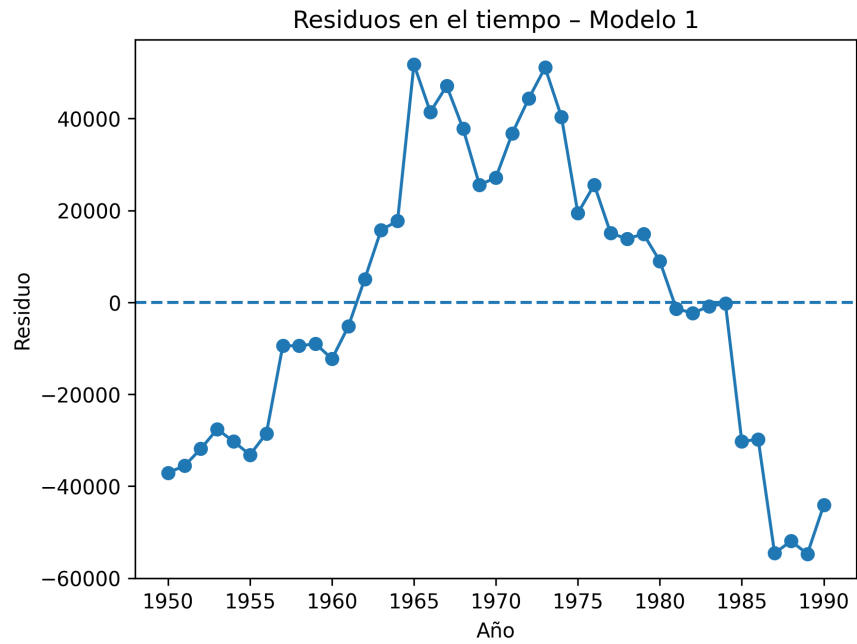


Figura 19: Residuos en el tiempo — Modelo 1

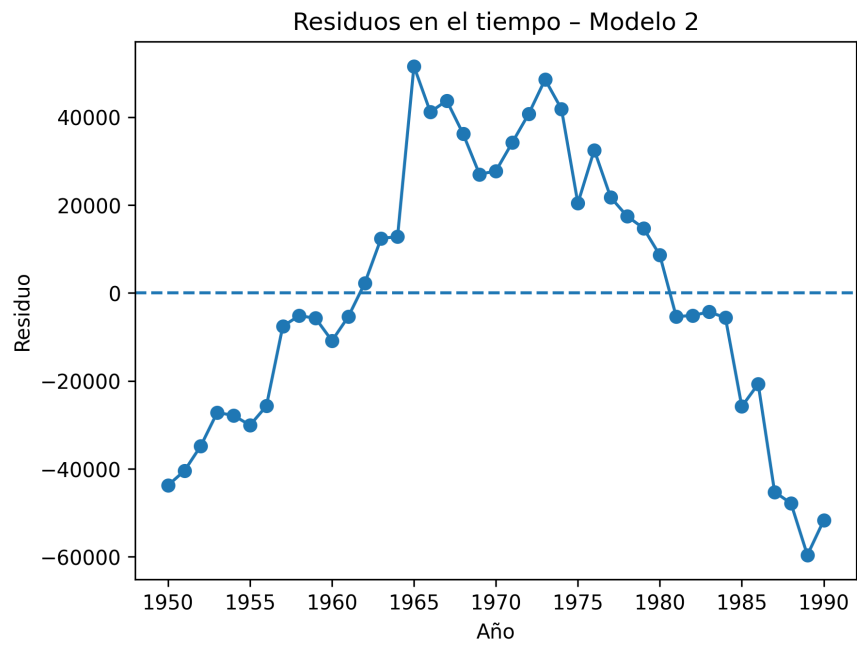


Figura 20: Residuos en el tiempo — Modelo 2

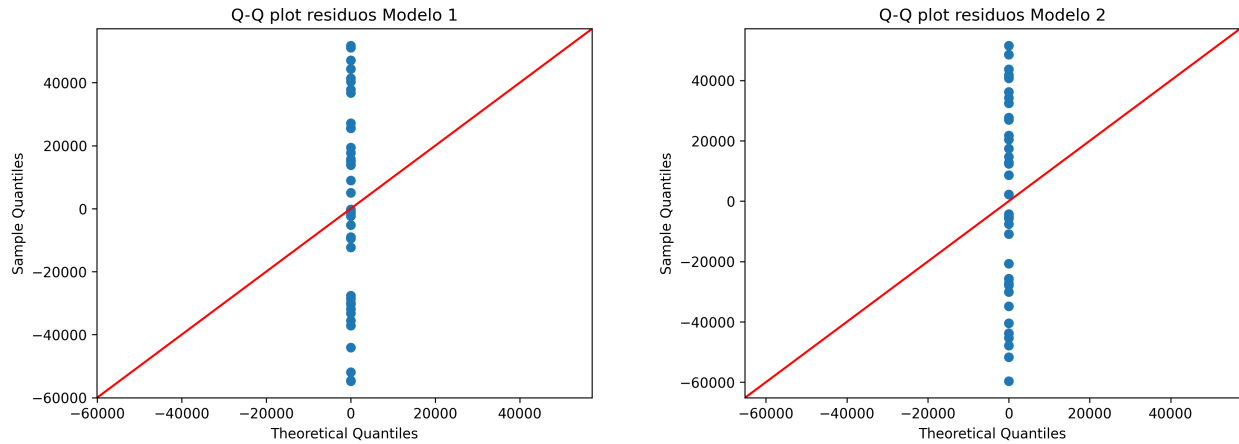


Figura 21: Q-Q-plot de residuos — M1 (izq.) y M2 (der.)

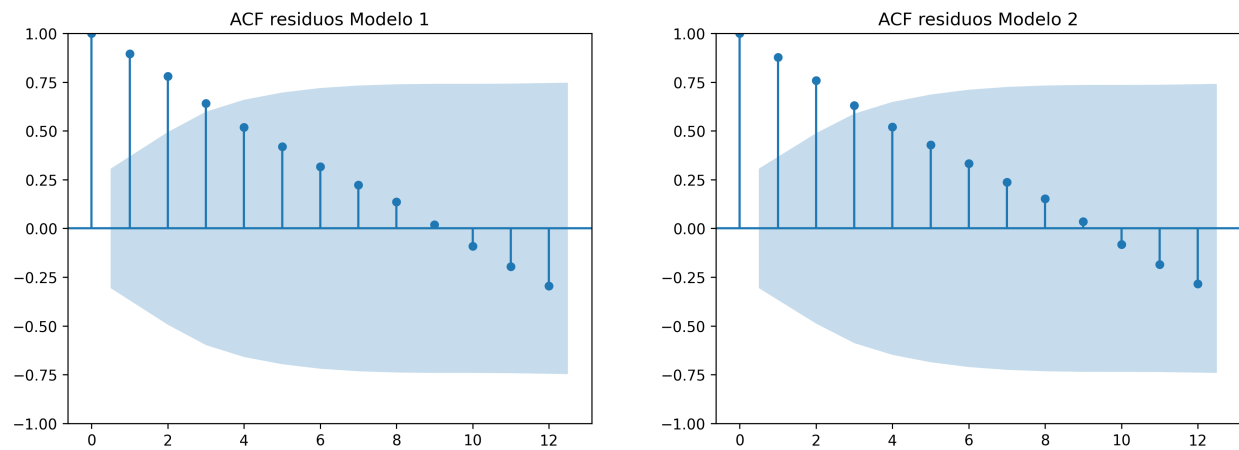


Figura 22: ACF de residuos — M1 (izq.) y M2 (der.)

Hay rachas largas en el tiempo y barras significativas en la ACF para rezagos bajos \Rightarrow **autocorrelación positiva**. Los Q-Q-plots no muestran colas extremas — **normalidad aceptable**. También es útil observar la evolución de la razón Inventarios/Ventas:

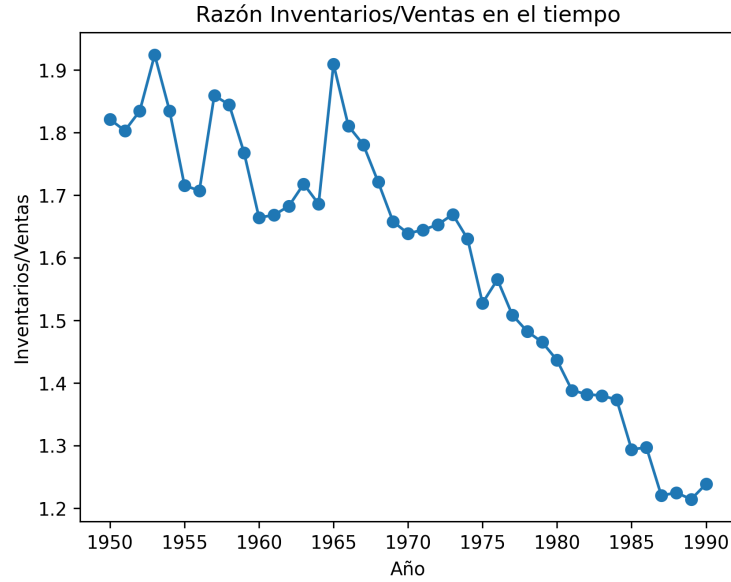


Figura 23: Razón Inventarios/Ventas

8.3. Con los residuos estimados, investigue si hay autocorrelación positiva mediante i) la prueba de Durbin–Watson y ii) las dos pruebas analíticas de normalidad para los errores poblacionales.

El estadístico de Durbin–Watson (DW) reportado por el script es:

$$DW_{M1} = 0.1256,$$

$$DW_{M2} = 0.1254.$$

Como aproximación, $\hat{\rho} \approx 1 - \frac{DW}{2} \Rightarrow \hat{\rho} \approx 0.94$; valores tan bajos de DW sugieren **autocorrelación positiva muy fuerte**.

Para la normalidad, el script reporta Jarque–Bera (JB): $p=0.360$ (M1) y $p=0.369$ (M2), por lo que **no se rechaza la normalidad** de residuos.

Además, el contraste RESET rechaza la especificación (ver cuadro de diagnósticos): hay evidencia de dinámica/curvatura omitida.

Cuadro 17: Pruebas de supuestos y métricas de ajuste para los modelos estimados.

Modelo	Durbin–Watson	Breusch–Pagan p	White p	Jarque–Bera p	RESET p	R^2	R^2 Ajustado	N
M1: Inv Ventas	0.1256	0.1819	0.2371	0.3596	0.0000	0.9430	0.9415	41
M2: Inv Ventas + Año	0.1254	0.2126	0.1524	0.3690	0.0000	0.9440	0.9410	41

8.4. Si sospecha que la estructura autorregresiva del error es de orden $p = 1$, verifíquelo con la prueba de Breusch–Godfrey.

Breusch–Godfrey (BG, $p = 1$). El procedimiento consiste en estimar la regresión auxiliar sobre los residuos OLS

$$\hat{u}_t = \alpha + \gamma_1 X_t + \cdots + \gamma_k Z_{kt} + \phi \hat{u}_{t-1} + v_t,$$

(Z_{kt} son los regresores originales) y contrastar $H_0: \phi = 0$ vía $LM \sim \chi^2_1$.

En la tabla 17, usar las columnas *BG LM p-valor* y *BG F p-valor* para decidir: p-valores pequeños \Rightarrow rechazo de H_0 (evidencia de AR(1) positiva), coherente con $DW \approx 0.126$ y la ACF.

8.5. Con base en los resultados de esta prueba, ¿cómo transformaría los datos para eliminar la autocorrelación?

Dada la evidencia de AR(1) en los errores, se recomienda estimar por *GLS* tipo Prais–Winsten/Cochrane–Orcutt, aplicando la cuasi-diferenciación

$$Y_t - \rho Y_{t-1} = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t,$$

con ρ estimado (p.ej., por MCO sobre \hat{u}_t vs. \hat{u}_{t-1} o por máxima verosimilitud). Como alternativa robusta en presencia de tendencia común, trabajar en diferencias/elasticidades:

$$\begin{aligned}\Delta Y_t &= \beta_2 \Delta X_t + e_t, \\ \ln Y_t &= \alpha + \beta \ln X_t + u_t.\end{aligned}$$

En cualquiera de los casos, re-evaluar ACF/PACF y DW, y reportar errores estándar *HAC* (*Newey–West*) si persiste dependencia débil.

9. Link al repositorio con código fuente

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