



**UNIVERSIDAD
DE GRANADA**

ANÁLISIS DE UNA SERIE TEMPORAL

MARÍA ISABEL SERRANO DELGADO

ÍNDICE

1 Descripción de la Serie	2
2 Identificación de Modelos	3
3 Validación	27
3.1 Sobreajuste	27
3.2 Estudio de los residuos	29
3.2.1 Estudio de la Normalidad	29
3.2.2 Estudio de la Media	31
3.2.3 Estudio de la Homocedasticidad	31
3.2.4 Estudio de la Incorrelación	32
3.2.5 Estudio de la Aleatoriedad.....	32
4 Predicción	33

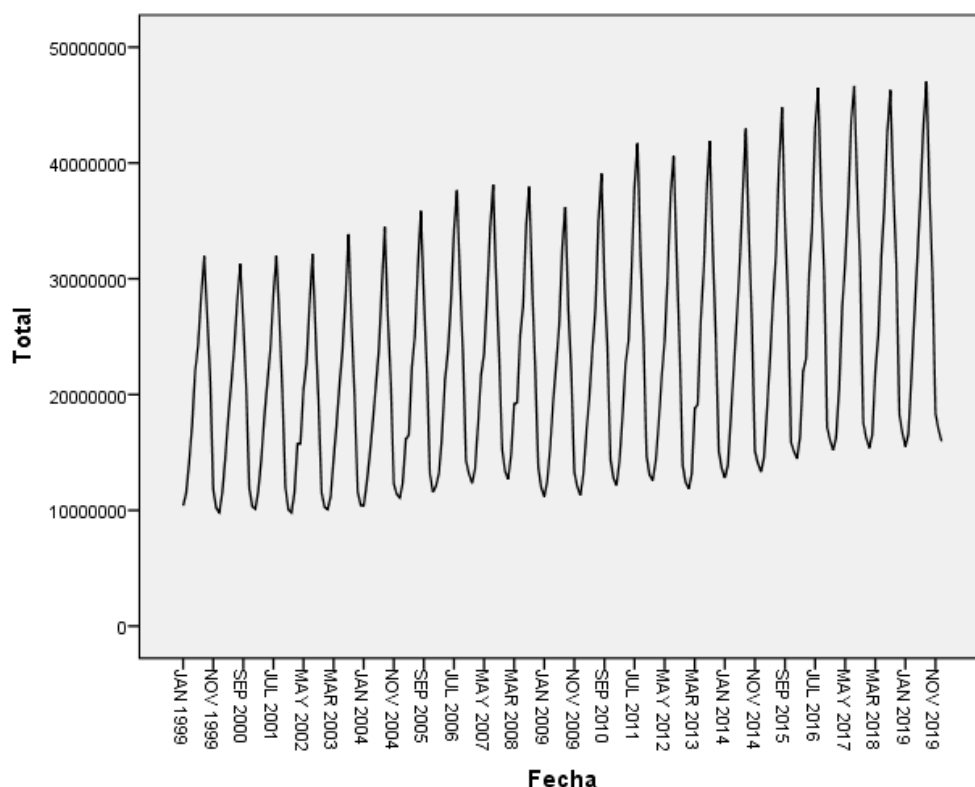
1. Descripción de la Serie

Para la realización de este estudio, hemos extraído los datos oficiales del Instituto Nacional de Estadística (INE) a través del portal datos.gob.es. Concretamente, hemos seleccionado la serie temporal correspondiente a la Encuesta de Ocupación Hotelera (EOH), utilizando la variable de Pernoctaciones Totales a nivel nacional (Identificador API: 67190).

[Viajeros y pernoctaciones por comunidades autónomas y provincias. EOH \(Identificador API: 67190\) - Conjunto de datos - Datos.gob.es](https://datos.gob.es/api/67190)

El periodo de estudio abarca desde enero de 1999 hasta enero de 2020, realizado por meses. Hemos decidido finalizar la serie en esta fecha para excluir el impacto de la pandemia de COVID-19, la cual provocó un cierre total del sector turístico en abril de 2020, registrando 0 pernoctaciones. Este suceso constituye un valor atípico extremo que distorsionaría las predicciones del modelo estacional.

Una vez depurada la base de datos y definidas las fechas con frecuencia mensual, procedemos a analizar la estacionariedad y estacionalidad de la serie. Para ello, en primer lugar, realizamos el gráfico de secuencia:



Analizando el gráfico de secuencia, observamos un patrón claramente estacional con picos recurrentes cada 12 meses. Además, se aprecia que la oscilación de estos picos aumenta a medida que crece la serie, lo que indica que la varianza no es constante, heterocedasticidad; esto nos sugiere que será necesario aplicar una Transformación Logarítmica para estabilizarla.

Respecto a la estacionariedad, la serie no es estacionaria en media, ya que se observa una clara tendencia creciente a lo largo de los años.

2. Identificación de Modelos

Para poder proponer modelos a analizar, debemos dibujar en primer lugar las gráficas de autocorrelación y autocorrelación parcial, con un número de retardos entre $T/4$ y $T/3$, siendo T el número de observaciones, en este caso $T = 253$.

Tenemos que establecer un número de retardos entre 63.25 y 84.33. En este caso, estableceremos el número de retornos como 72. Para obtener las gráficas de estas funciones lo hacemos, en una primera toma de contacto, mediante la aproximación de Bartlett, teniendo en cuenta que realizaremos la transformación de logaritmos ya.

▸ ACF

[Conjunto_de_datos1] C:\Users\Isabela\OneDrive\Documentos

Descripción del modelo

Nombre del modelo	MOD_1
Nombre de la serie	1
Transformación	Logaritmo natural
Diferenciación no estacional	0
Diferenciación estacional	0
Longitud del período estacional	12
Número máximo de retardos	72
Proceso asumido para el cálculo de los errores típicos de las autocorrelaciones	MA de orden igual al número del retardo menos uno (se usa la aproximación de Bartlett) ^a
Mostrar y representar	Todos los retardos

Aplicando las especificaciones del modelo de MOD_1

a. No aplicable para el cálculo de los errores típicos de las autocorrelaciones parciales.

Resumen del procesamiento de los casos

	Total
Longitud de la serie	253
Número de valores perdidos	0
Negativo o cero antes de la transformación Log	
Perdidos definidos por el usuario	0
Perdidos del sistema	0
Número de valores válidos	253
Número de primeros retardos computables	252

Total

Autocorrelaciones

Serie: Total

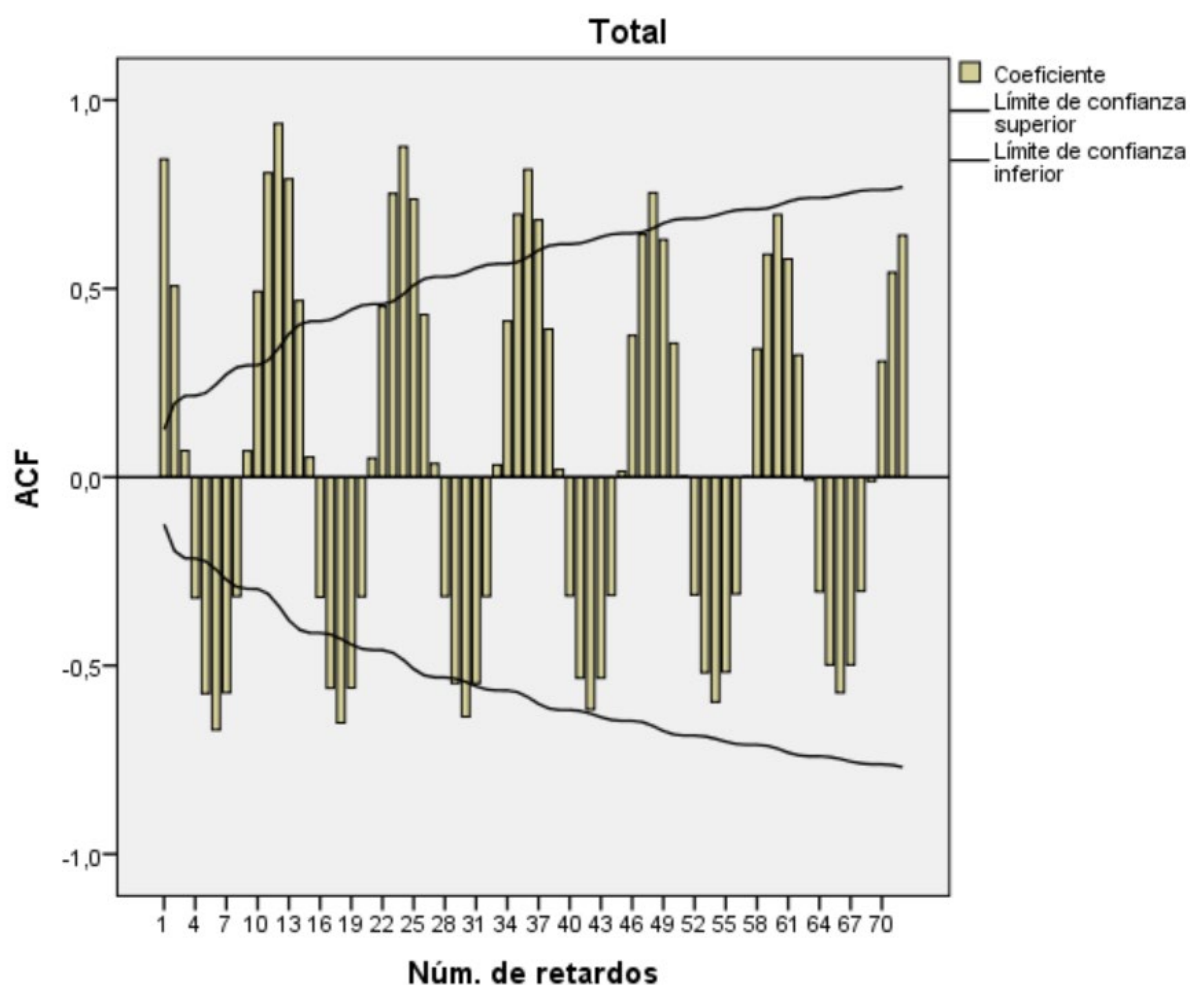
Retardo	Autocorrelación	Típ. Error ^a	Estadístico de Box-Ljung		
			Valor	gl	Sig. ^b
1	,843	,063	182,033	1	,000
2	,507	,098	247,991	2	,000
3	,070	,108	249,245	3	,000
4	-,321	,108	275,905	4	,000
5	-,574	,112	361,673	5	,000
6	-,670	,123	478,996	6	,000
7	-,572	,136	564,815	7	,000
8	-,316	,146	591,117	8	,000
9	,069	,148	592,385	9	,000
10	,492	,148	656,696	10	,000
11	,807	,155	830,336	11	,000
12	,938	,171	1065,643	12	,000
13	,790	,190	1233,616	13	,000
14	,468	,202	1292,688	14	,000
15	,053	,207	1293,441	15	,000
16	-,318	,207	1321,015	16	,000
17	-,560	,209	1406,672	17	,000
18	-,652	,214	1523,458	18	,000
19	-,559	,222	1609,654	19	,000
20	-,317	,228	1637,442	20	,000
21	,050	,229	1638,145	21	,000
22	,452	,229	1695,159	22	,000
23	,752	,233	1853,837	23	,000
24	,877	,242	2070,369	24	,000
25	,737	,255	2223,998	25	,000
26	,431	,263	2276,725	26	,000
27	,035	,266	2277,084	27	,000
28	-,317	,266	2305,825	28	,000
29	-,547	,267	2391,885	29	,000

30	-,635	,272	2508,702	30	,000
31	-,547	,277	2595,629	31	,000
32	-,316	,282	2624,850	32	,000
33	,032	,283	2625,151	33	,000
34	,414	,283	2675,634	34	,000
35	,697	,285	2819,201	35	,000
36	,816	,292	3017,103	36	,000
37	,681	,301	3155,821	37	,000
38	,392	,307	3202,014	38	,000
39	,020	,309	3202,134	39	,000
40	-,315	,309	3232,203	40	,000
41	-,533	,310	3318,735	41	,000
42	-,617	,314	3435,057	42	,000
43	-,533	,319	3522,277	43	,000
44	-,314	,322	3552,700	44	,000
45	,015	,323	3552,770	45	,000
46	,375	,323	3596,674	46	,000
47	,644	,325	3726,438	47	,000
48	,754	,330	3905,403	48	,000
49	,630	,337	4030,887	49	,000
50	,355	,341	4070,867	50	,000
51	,002	,343	4070,868	51	,000
52	-,313	,343	4102,227	52	,000
53	-,519	,344	4189,097	53	,000
54	-,596	,347	4304,391	54	,000
55	-,517	,351	4391,651	55	,000
56	-,310	,354	4423,168	56	,000
57	,001	,355	4423,168	57	,000
58	,340	,355	4461,403	58	,000
59	,591	,356	4577,484	59	,000
60	,696	,360	4739,477	60	,000
61	,579	,366	4852,007	61	,000
62	,323	,369	4887,337	62	,000
63	-,008	,370	4887,358	63	,000
64	-,304	,370	4918,933	64	,000
65	-,498	,371	5004,124	65	,000

66	-,572	,374	5117,017	66	,000
67	-,498	,377	5203,088	67	,000
68	-,303	,380	5235,095	68	,000
69	-,011	,381	5235,137	69	,000
70	,307	,381	5268,426	70	,000
71	,543	,382	5372,852	71	,000
72	,641	,385	5519,392	72	,000

a. El proceso subyacente asumido es MA con el orden igual al número de retardo menos uno. Se ha utilizado la aproximación de Bartlett.

b. Basado en la aproximación chi cuadrado asintótica.



Las bandas de confianza se calculan al 95%, y en este caso lo que se compara es el nivel de autocorrelación con $2 \times$ desviación típica, de tal manera que, si este último es más pequeño, se sale de la barra de confianza, lo que significaría que esa autocorrelación es significativamente distinta de 0.

Como observamos en el gráfico un claro patrón en los datos por tanto diferenciamos en ciclo para eliminarlo. Esto ya lo suponíamos ya que en el gráfico de secuencias se veía una clara estacionalidad.

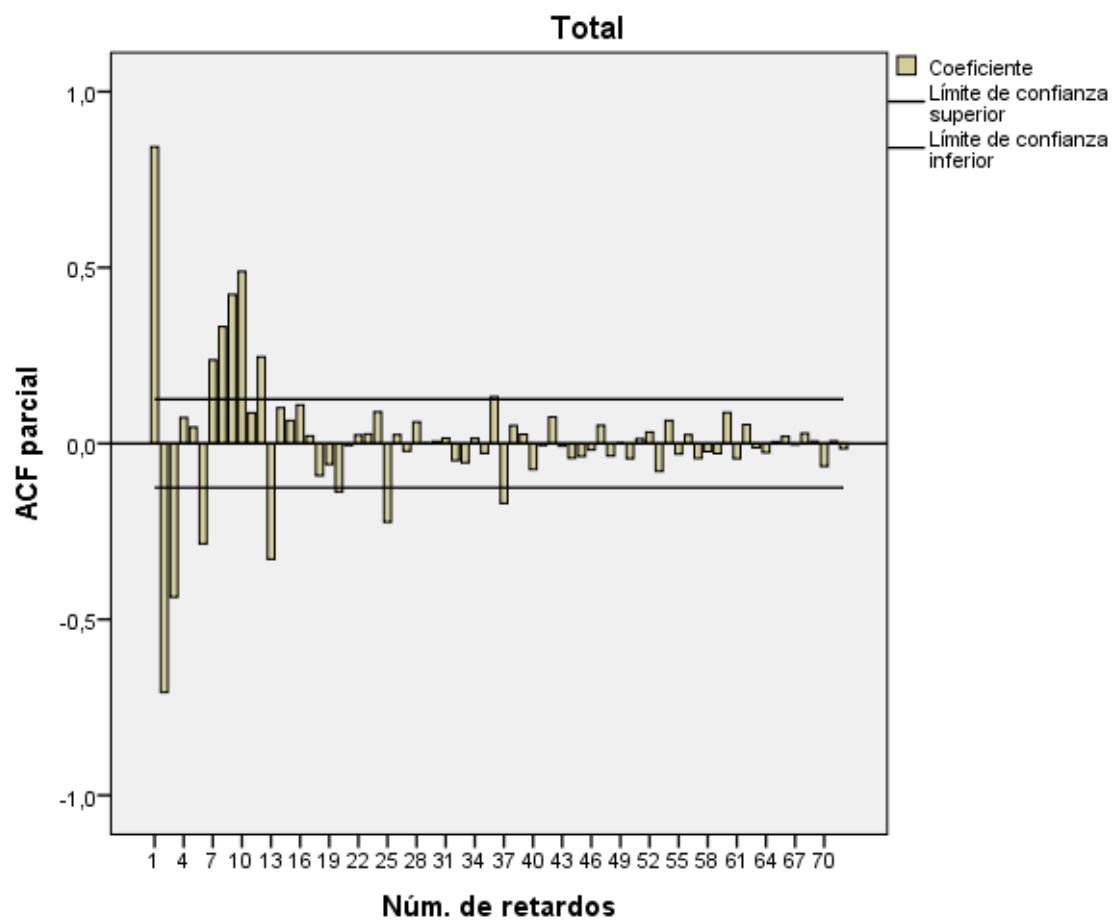
Autocorrelaciones parciales

Serie: Total

Retardo	Autocorrelación parcial	Típ. Error
1	,843	,063
2	-,708	,063
3	-,437	,063
4	,073	,063
5	,046	,063
6	-,285	,063
7	,237	,063
8	,331	,063
9	,423	,063
10	,489	,063
11	,087	,063
12	,247	,063
13	-,329	,063
14	,101	,063
15	,065	,063
16	,110	,063
17	,022	,063
18	-,092	,063
19	-,059	,063
20	-,137	,063
21	-,006	,063
22	,024	,063
23	,027	,063
24	,090	,063
25	-,224	,063
26	,025	,063
27	-,022	,063
28	,061	,063
29	,000	,063
30	,006	,063
31	,015	,063
32	-,048	,063
33	-,054	,063

34	,014	,063
35	-,028	,063
36	,133	,063
37	-,170	,063
38	,051	,063
39	,026	,063
40	-,073	,063
41	-,005	,063
42	,075	,063
43	-,007	,063
44	-,041	,063
45	-,037	,063
46	-,018	,063
47	,051	,063
48	-,035	,063
49	,001	,063
50	-,043	,063
51	,013	,063
52	,032	,063
53	-,079	,063
54	,065	,063
55	-,030	,063
56	,025	,063
57	-,042	,063
58	-,023	,063
59	-,028	,063
60	,088	,063
61	-,043	,063
62	,053	,063
63	-,012	,063
64	-,026	,063
65	,005	,063

66	,020	,063
67	-,003	,063
68	,029	,063
69	,006	,063
70	-,065	,063
71	,007	,063
72	-,015	,063



Para asegurarnos de que tenemos que hacer la diferenciación estacional, vamos a ver las FAC y FACP de los retardos múltiplos de 12.

ACF

[Conjunto_de_datos1] C:\Users\Isabela\OneDrive\Documentos

Descripción del modelo

Nombre del modelo	MOD_2
Nombre de la serie	1
Transformación	Total
	Logaritmo natural
Diferenciación no estacional	0
Diferenciación estacional	0
Longitud del período estacional	12
Número máximo de retardos	72
Proceso asumido para el cálculo de los errores típicos de las autocorrelaciones	MA de orden igual al número del retardo menos uno (se usa la aproximación de Bartlett) ^a
Mostrar y representar	Sólo para los intervalos periódicos

Aplicando las especificaciones del modelo de MOD_2

- a. No aplicable para el cálculo de los errores típicos de las autocorrelaciones parciales.

Total

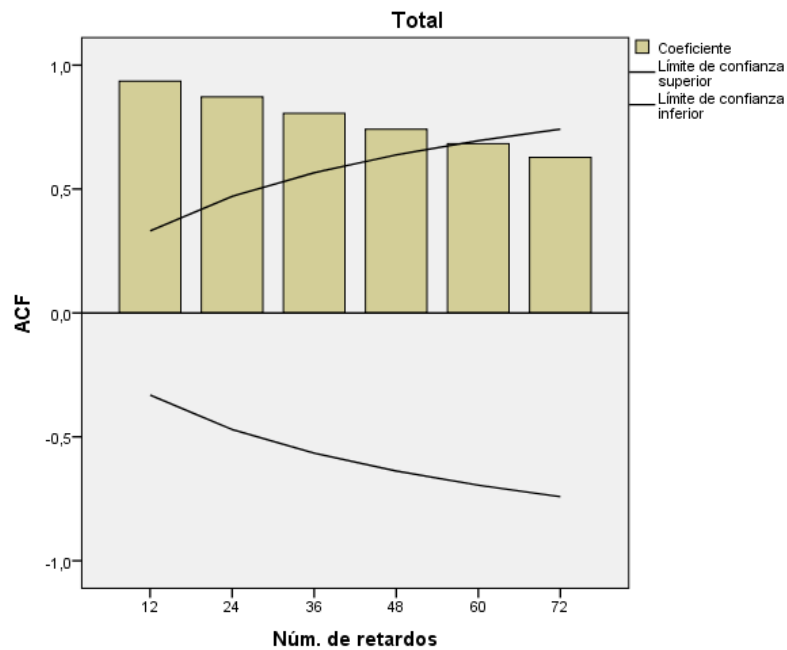
Autocorrelaciones

Serie: Total

Retardo	Autocorrelación	Típ. Error ^a	Estadístico de Box-Ljung		
			Valor	gl	Sig. ^b
12	,938	,171	1065,643	12	,000
24	,877	,242	2070,369	24	,000
36	,816	,292	3017,103	36	,000
48	,754	,330	3905,403	48	,000
60	,696	,360	4739,477	60	,000
72	,641	,385	5519,392	72	,000

- a. El proceso subyacente asumido es MA con el orden igual al número de retardo menos uno. Se ha utilizado la aproximación de Bartlett.

- b. Basado en la aproximación chi cuadrado asintótica.



Al analizar los retardos estacionales en la Función de Autocorrelación Simple, observamos un patrón de decaimiento lento. Los coeficientes en los retardos 12, 24, 36 y 48 son significativamente positivos y van disminuyendo su magnitud progresivamente, llegando a rozar el límite de significación en el retardo 60 y perdiendo significación estadística en el retardo 72.

Este comportamiento en los ciclos confirma que la serie no es estacionaria estacionalmente, lo que justifica la necesidad de aplicar una diferenciación estacional de orden 1 para eliminar esta tendencia cíclica.

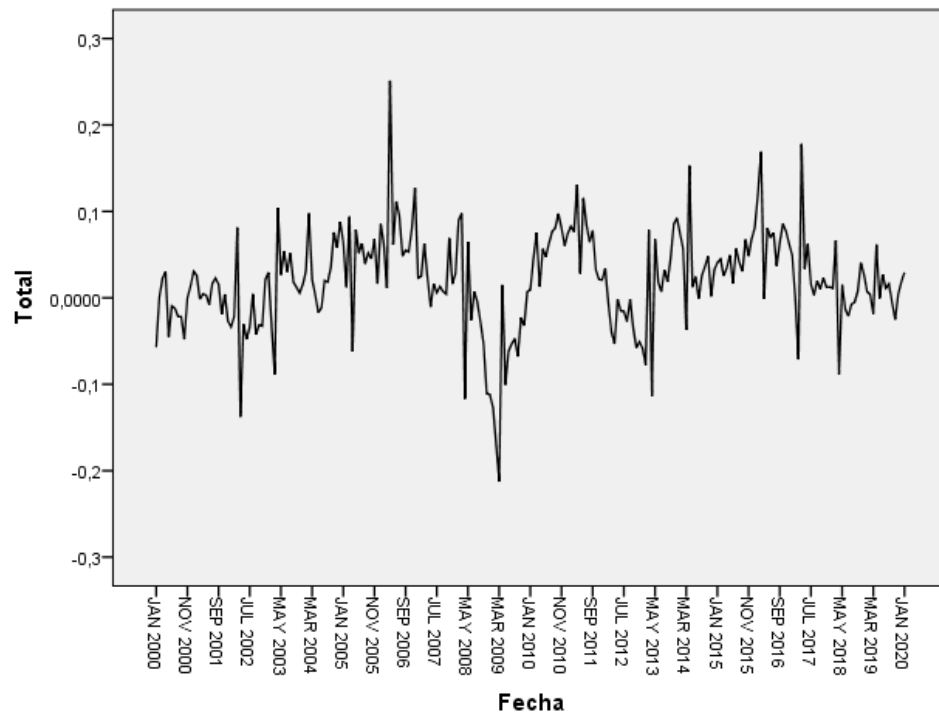
Gráfico de secuencia

[Conjunto_de_datos1] C:\Users\Isabela\OneDr

Descripción del modelo

Nombre del modelo	MOD_4
Serie o secuencia	1
Transformación	Total
	Logaritmo natural
Diferenciación no estacional	0
Diferenciación estacional	1
Longitud del periodo estacional	12
Etiquetas del eje horizontal	Fecha_
Mostrar intervenciones	Ninguna
Líneas de referencia	Ninguna
Área bajo la curva	No rellenado

Aplicando las especificaciones del modelo de MOD_4



Transformaciones: log natural, diferencia estacional(1, periodo 12)

Al observar el gráfico de secuencia de la serie transformada, vemos que la serie oscila de forma aleatoria alrededor de una media constante cero y con una varianza estabilizada.

La desaparición de la tendencia creciente y de los patrones cíclicos repetitivos nos permite concluir que la serie transformada sí es estacionaria y ya no presenta componente estacional, cumpliendo así los requisitos necesarios para proceder a la estimación de modelos ARMA. Estudiaremos sus FAP y FAS

ACF

[Conjunto_de_datos1] C:\Users\Isabela\OneDrive\Documentos

Descripción del modelo

Nombre del modelo	MOD_5
Nombre de la serie	1
Transformación	Total
	Logaritmo natural
Diferenciación no estacional	0
Diferenciación estacional	1
Longitud del período estacional	12
Número máximo de retardos	72
Proceso asumido para el cálculo de los errores típicos de las autocorrelaciones	MA de orden igual al número del retardo menos uno (se usa la aproximación de Bartlett) ^a
Mostrar y representar	Todos los retardos

Aplicando las especificaciones del modelo de MOD_5

- a. No aplicable para el cálculo de los errores típicos de las autocorrelaciones parciales.

Resumen del procesamiento de los casos

	Total
Longitud de la serie	253
Número de valores perdidos	0
Negativo o cero antes de la transformación Log	
Perdidos definidos por el usuario	0
Perdidos del sistema	0
Número de valores válidos	253
Número de valores perdidos debido a la diferenciación	12
Número de primeros retardos computables después de la diferenciación	240

Total

Autocorrelaciones

Serie: Total

Retardo	Autocorrelación	Típ. Error ^a	Estadístico de Box-Ljung		
			Valor	gl	Sig. ^b
1	,396	,064	38,232	1	,000
2	,517	,074	103,733	2	,000
3	,453	,088	154,136	3	,000
4	,398	,097	193,341	4	,000
5	,377	,103	228,698	5	,000
6	,335	,109	256,667	6	,000
7	,273	,113	275,277	7	,000
8	,284	,116	295,517	8	,000
9	,180	,119	303,652	9	,000
10	,143	,120	308,873	10	,000
11	,170	,121	316,272	11	,000
12	-,182	,122	324,720	12	,000
13	,085	,123	326,584	13	,000
14	-,045	,123	327,108	14	,000
15	-,029	,123	327,326	15	,000
16	-,077	,123	328,878	16	,000
17	-,062	,123	329,876	17	,000
18	-,087	,123	331,856	18	,000
19	-,118	,124	335,526	19	,000
20	-,167	,124	342,878	20	,000
21	-,139	,125	348,045	21	,000
22	-,201	,126	358,893	22	,000
23	-,109	,127	362,061	23	,000
24	-,209	,127	373,902	24	,000
25	-,164	,129	381,165	25	,000
26	-,172	,130	389,250	26	,000
27	-,190	,131	399,129	27	,000

28	-,164	,132	406,507	28	,000
29	-,180	,133	415,483	29	,000
30	-,206	,134	427,268	30	,000
31	-,170	,135	435,334	31	,000
32	-,165	,136	443,001	32	,000
33	-,140	,137	448,481	33	,000
34	-,121	,137	452,589	34	,000
35	-,229	,138	467,527	35	,000
36	,049	,139	468,216	36	,000
37	-,146	,139	474,297	37	,000
38	-,035	,140	474,641	38	,000
39	-,059	,140	475,661	39	,000
40	-,055	,140	476,528	40	,000
41	-,031	,140	476,815	41	,000
42	-,040	,140	477,289	42	,000
43	-,051	,140	478,061	43	,000
44	-,016	,140	478,138	44	,000
45	-,062	,140	479,292	45	,000
46	-,014	,140	479,348	46	,000
47	,080	,140	481,293	47	,000
48	-,180	,141	491,116	48	,000
49	,054	,142	492,007	49	,000
50	-,094	,142	494,703	50	,000
51	-,056	,142	495,685	51	,000
52	-,022	,142	495,838	52	,000
53	-,057	,142	496,837	53	,000
54	,005	,142	496,844	54	,000
55	,021	,142	496,980	55	,000
56	-,018	,142	497,087	56	,000
57	,015	,142	497,162	57	,000
58	,015	,142	497,238	58	,000
59	-,060	,142	498,384	59	,000
60	,071	,142	500,021	60	,000
61	-,044	,142	500,662	61	,000
62	,051	,142	501,511	62	,000
63	,069	,142	503,083	63	,000
64	,049	,143	503,863	64	,000
65	,052	,143	504,780	65	,000
66	,014	,143	504,845	66	,000
67	,004	,143	504,852	67	,000
68	,011	,143	504,895	68	,000
69	,019	,143	505,015	69	,000
70	-,009	,143	505,041	70	,000
71	-,010	,143	505,078	71	,000
72	,039	,143	505,597	72	,000

a. El proceso subyacente asumido es MA con el orden igual al número de retardo menos uno. Se ha utilizado la aproximación de Bartlett.

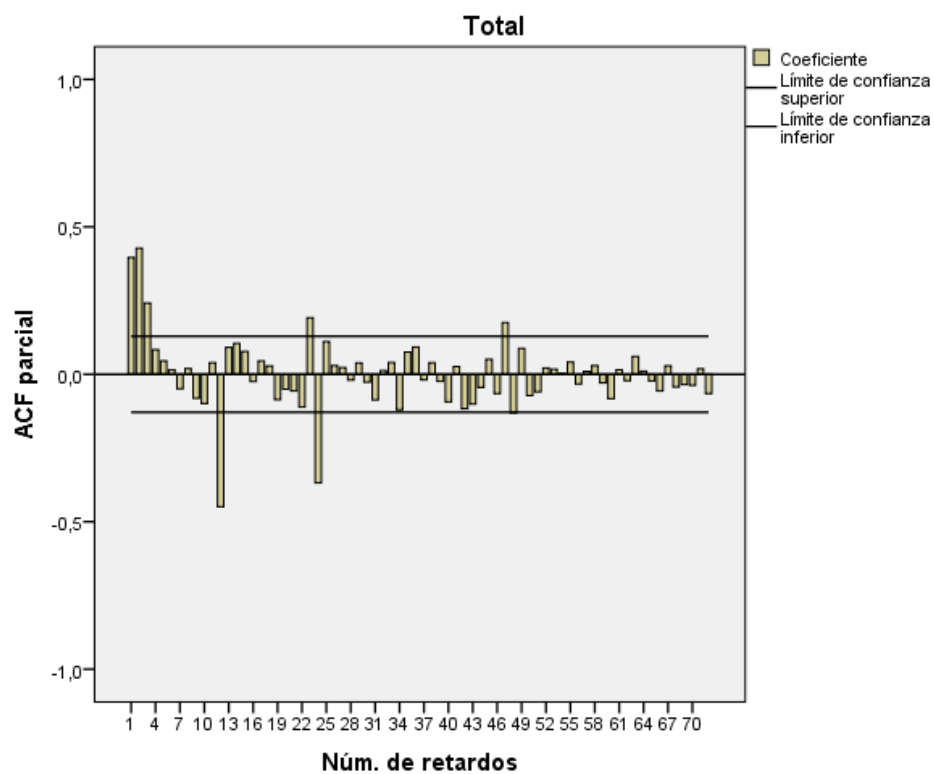
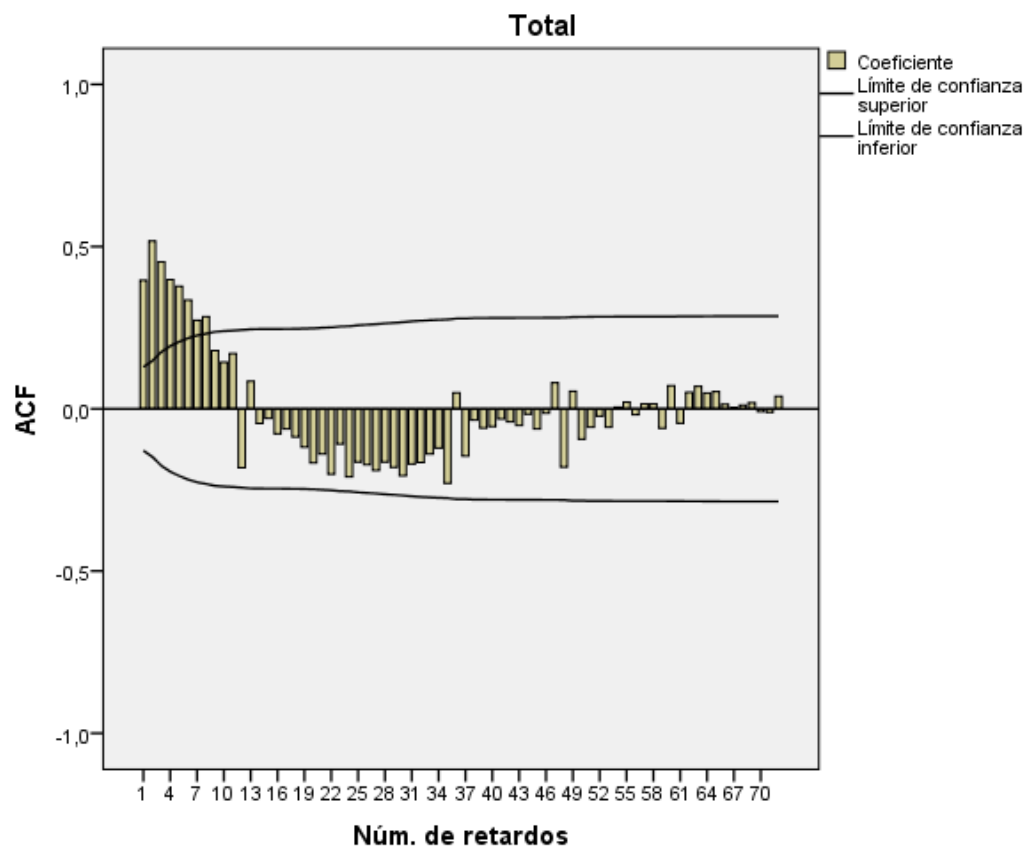
b. Basado en la aproximación chi cuadrado asintótica.

Autocorrelaciones parciales

Serie: Total

Retardo	Autocorrelación parcial	Típ. Error
1	,396	,064
2	,427	,064
3	,241	,064
4	,084	,064
5	,045	,064
6	,015	,064
7	-,049	,064
8	,019	,064
9	-,081	,064
10	-,099	,064
11	,039	,064
12	-,449	,064
13	,091	,064
14	,104	,064
15	,078	,064
16	-,025	,064
17	,046	,064
18	,028	,064
19	-,086	,064
20	-,050	,064
21	-,057	,064
22	-,111	,064
23	,191	,064
24	-,368	,064
25	,111	,064
26	,029	,064
27	,023	,064
28	-,020	,064
29	,038	,064
30	-,027	,064
31	-,087	,064
32	,012	,064

33	,040	,064
34	-,120	,064
35	,075	,064
36	,093	,064
37	-,019	,064
38	,039	,064
39	-,023	,064
40	-,093	,064
41	,026	,064
42	-,117	,064
43	-,101	,064
44	-,044	,064
45	,051	,064
46	-,066	,064
47	,175	,064
48	-,131	,064
49	,088	,064
50	-,073	,064
51	-,059	,064
52	,022	,064
53	,017	,064
54	,002	,064
55	,041	,064
56	-,033	,064
57	,010	,064
58	,030	,064
59	-,030	,064
60	-,082	,064
61	,015	,064
62	-,022	,064
63	,061	,064
64	,010	,064
65	-,022	,064
66	-,057	,064
67	,029	,064
68	-,043	,064
69	-,034	,064
70	-,038	,064
71	,019	,064
72	-,065	,064



Analizando el correlograma de la serie estacionaria, observamos que la gran mayoría de los coeficientes han quedado dentro de las bandas de confianza. Sin embargo, observamos un pico significativo en el primer retardo tanto de la ACF como de la ACF Parcial. Esto indica que existe una dependencia a corto plazo que debemos capturar. Según la metodología Box-Jenkins, esto sugiere la inclusión de un parámetro regular de orden 1, ya sea de Media Móvil o Autoregresivo.

Realizaremos lo mismo sólo para los retardos periódicos para observarlo mejor.

ACF

[Conjunto_de_datos1] C:\Users\Isabela\OneDrive\Documentos

Descripción del modelo

Nombre del modelo	MOD_6
Nombre de la serie	1
Transformación	Total
	Logaritmo natural
Diferenciación no estacional	0
Diferenciación estacional	1
Longitud del período estacional	12
Número máximo de retardos	72
Proceso asumido para el cálculo de los errores típicos de las autocorrelaciones	MA de orden igual al número del retardo menos uno (se usa la aproximación de Bartlett) ^a
Mostrar y representar	Sólo para los intervalos periódicos

Aplicando las especificaciones del modelo de MOD_6

a. No aplicable para el cálculo de los errores típicos de las autocorrelaciones parciales.

Resumen del procesamiento de los casos

	Total
Longitud de la serie	253
Número de valores perdidos	0
Negativo o cero antes de la transformación Log	
Perdidos definidos por el usuario	0
Perdidos del sistema	0
Número de valores válidos	253
Número de valores perdidos debido a la diferenciación	12
Número de primeros retardos computables después de la diferenciación	240

Total

Autocorrelaciones

Serie: Total

Retardo	Autocorrelación	Típ. Error ^a	Estadístico de Box-Ljung		
			Valor	gl	Sig. ^b
12	-,182	,122	324,720	12	,000
24	-,209	,127	373,902	24	,000
36	,049	,139	468,216	36	,000
48	-,180	,141	491,116	48	,000
60	,071	,142	500,021	60	,000
72	,039	,143	505,597	72	,000

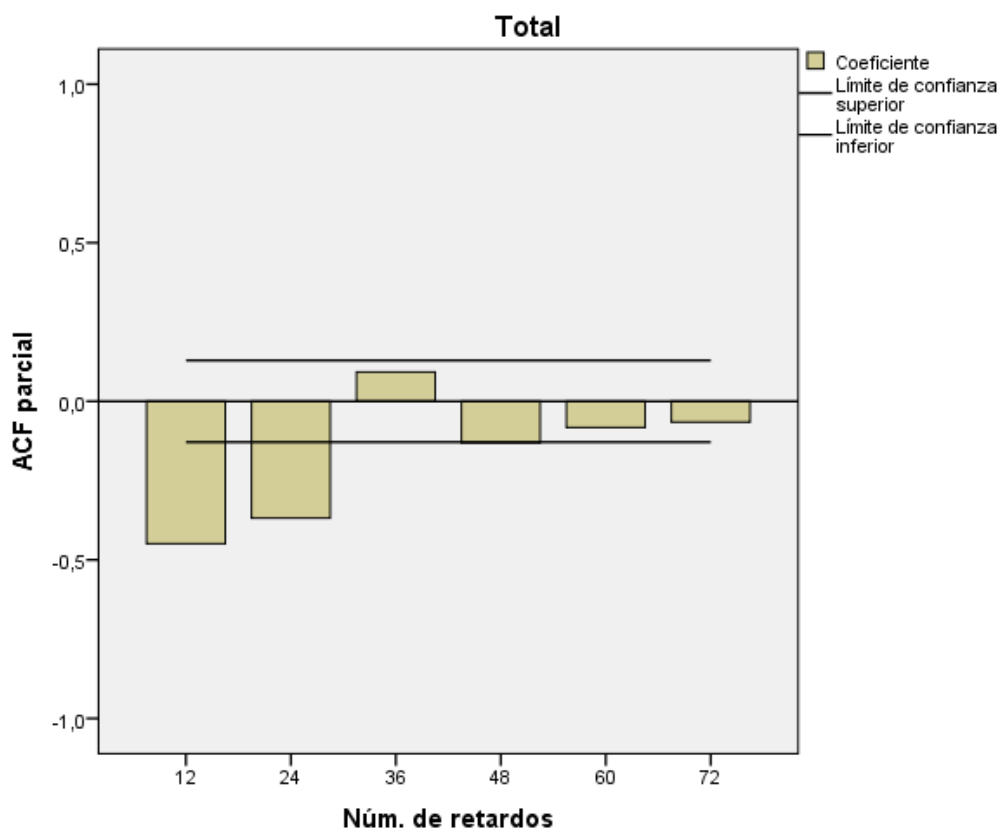
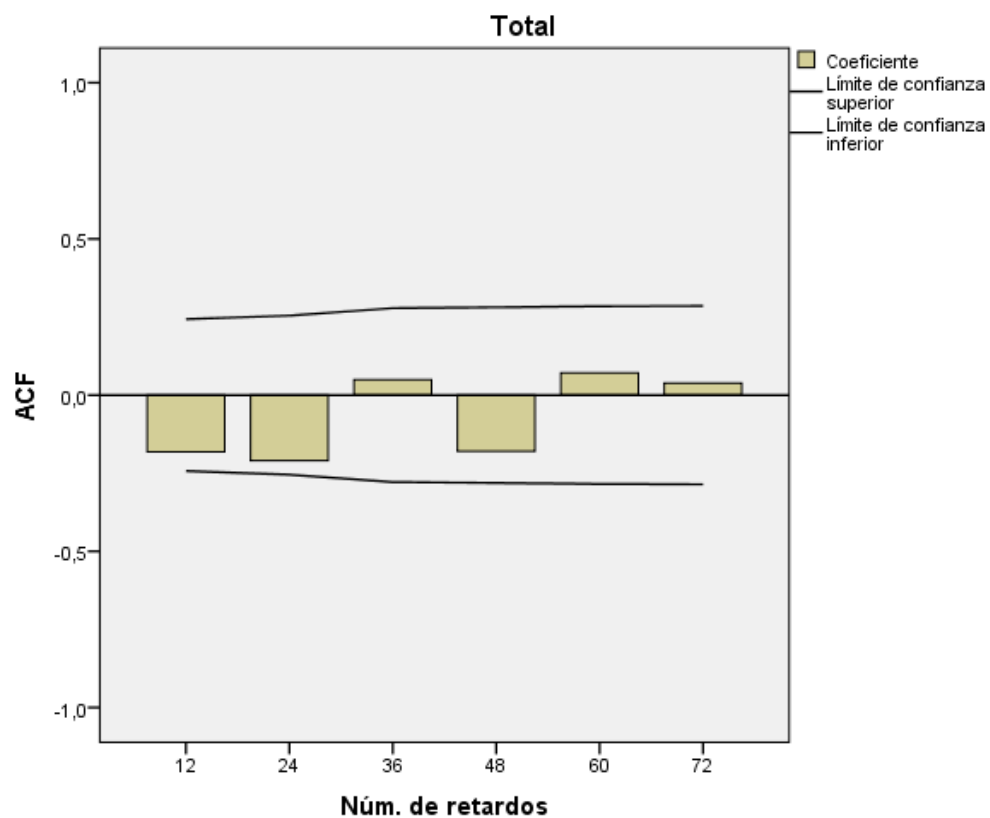
a. El proceso subyacente asumido es MA con el orden igual al número de retardo menos uno. Se ha utilizado la aproximación de Bartlett.

b. Basado en la aproximación chi cuadrado asintótica.

Autocorrelaciones parciales

Serie: Total

Retardo	Autocorrelación parcial	Típ. Error
12	-,449	,064
24	-,368	,064
36	,093	,064
48	-,131	,064
60	-,082	,064
72	-,065	,064



Al analizar el correlograma de la serie transformada estacionaria diferenciada, observamos un comportamiento interesante en la estructura estacional:

En la ACF Estacional, se observa que no hay picos significativos, todos están dentro de las bandas de confianza.

Sin embargo, en la ACF Parcial Estacional, hay dos picos significativos en el 12 y 24, lo que indica que decrece.

Como la ACF estacional se corta, es decir, es nula y la ACF Parcial tiene estructura, el componente es Autoregresivo.

Observando en la ACF de todos los retardos tanto Estacional como Parcial, se tiene que en el retardo 1 tienen un pico muy significativo y luego caen rápidamente. Esto suele indicar una estructura simple de orden 1. Puede modelarse como AR(1) o MA(1).

Dado que la ACF muere relativamente rápido después del retardo 1, un AR(1) es una opción sólida.

El modelo que sugerimos será el siguiente: SARIMA (1, 0, 0) (2, 1, 0) transformado log.

En la parte Regular identificamos un componente Autorregresivo de orden 1, como explicamos anteriormente por el comportamiento de la ACF. La diferencia es 0, dado que trabajamos sobre la serie que ya ha sido diferenciada previamente. Finalmente, el orden de Media Móvil es 0, ya que la estructura se explica suficientemente con el componente autorregresivo.

En la parte Estacional: Asignamos un orden 2 al componente Autorregresivo Estacional, ya que en la ACF Parcial observamos dos picos significativos, en los retardos 12 y 24.

Establecemos otra diferencia en 1 para capturar y limpiar la estacionalidad de los datos.

Por último, el componente de Media Móvil Estacional es 0, dado que la ACF estacional es nula y no presenta picos significativos en los retardos estacionales.

Modelizador para series temporales

[Conjunto_de_datos1] C:\Users\Isabela\OneD:

Descripción del modelo

			Tipo de modelo
ID del modelo	Total	Modelo_1	ARIMA(1,0,0) (2,1,0)

Resumen del modelo

Ajuste del modelo

Estadístico de ajuste	Media	ET	Mínimo	Máximo	Percentil						
					5	10	25	50	75	90	95
R-cuadrado estacionaria	,449	.	,449	,449	,449	,449	,449	,449	,449	,449	,449
R-cuadrado	,990	.	,990	,990	,990	,990	,990	,990	,990	,990	,990
RMSE	953703,329	.	953703,329	953703,329	953703,329	953703,329	953703,329	953703,329	953703,329	953703,329	953703,329
MAPE	3,086	.	3,086	3,086	3,086	3,086	3,086	3,086	3,086	3,086	3,086
MaxAPE	22,144	.	22,144	22,144	22,144	22,144	22,144	22,144	22,144	22,144	22,144
MAE	686031,326	.	686031,326	686031,326	686031,326	686031,326	686031,326	686031,326	686031,326	686031,326	686031,326
MaxAE	3489247,460	.	3489247,460	3489247,460	3489247,460	3489247,460	3489247,460	3489247,460	3489247,460	3489247,460	3489247,460
BIC normalizado	27,627	.	27,627	27,627	27,627	27,627	27,627	27,627	27,627	27,627	27,627

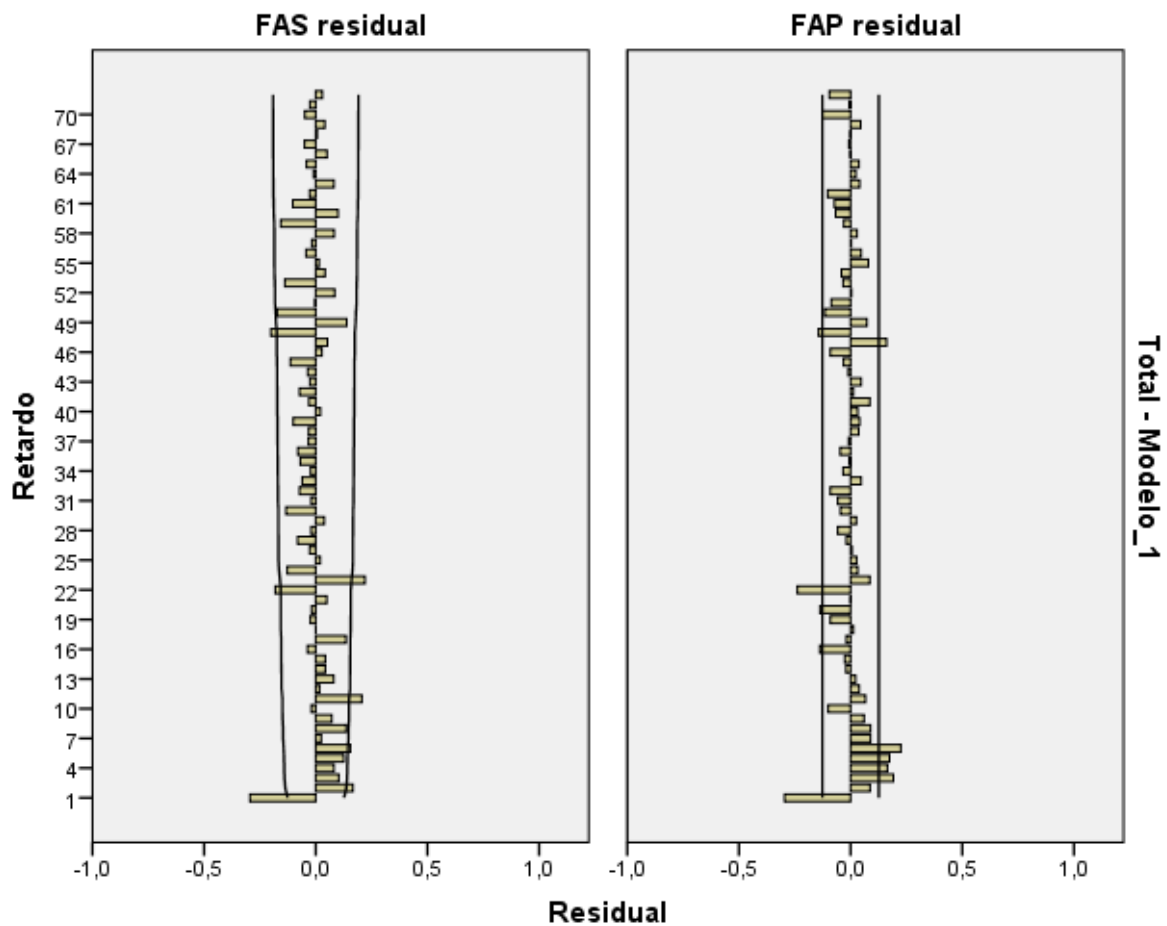
Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo	Ljung-Box Q(18)			Número de valores atípicos
		R-cuadrado estacionaria	Estadísticos	GL	Sig.	
Total-Modelo_1	0	,449	67,060	15	,000	0

Dado que no pasa el test de Ljung-Box, ya que su p-valor es 0, y para que se acepte el modelo $p\text{-valor} > 0.05$, concluimos que este modelo no va a ser bueno.

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
Total-Modelo_1	Total	Log natural	Constante		,021	,005	4,638	,000
			AR	Retardo 1	,723	,047	15,483	,000
			AR, estacional	Retardo 1	-,718	,059	-12,213	,000
				Retardo 2	-,472	,059	-7,998	,000
			Diferenciación estacional	1				



Al analizar la FAS y FAP de los residuos, observamos que el retardo 1 es significativamente distinto de 0. Esto sugiere que el modelo AR(1) es insuficiente para explicar la dependencia a corto plazo. Por otra parte, vemos que los retardos estacionales, como el 12 y 24 son también significativos

Por tanto, para corregir el modelo, proponemos añadir un término de Media Móvil regular (MA 1) para capturar esa estructura restante. El nuevo modelo a estimar será un SARIMA(1,0,1)(2,1,0)

Modelizador para series temporales

[Conjunto_de_datos1] C:\Users\Isabela\OneD:

Descripción del modelo

			Tipo de modelo
ID del modelo	Total	Modelo_1	ARIMA(1,0,1) (2,1,0)

Resumen del modelo

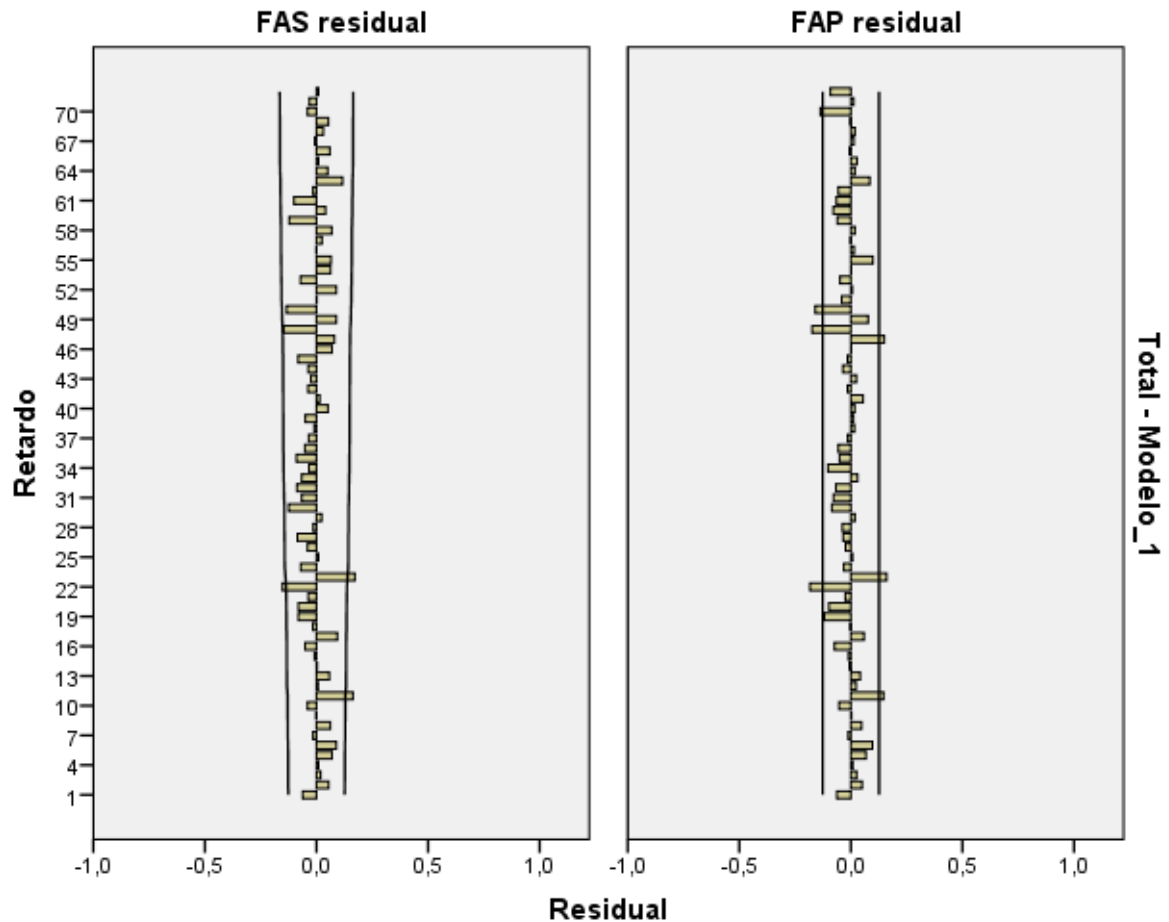
Estadístico de ajuste	Media	ET	Mínimo	Máximo	Ajuste del modelo						
					Percentil						
					5	10	25	50	75	90	95
R-cuadrado estacionaria	,566	.	,566	,566	,566	,566	,566	,566	,566	,566	,566
R-cuadrado	,992	.	,992	,992	,992	,992	,992	,992	,992	,992	,992
RMSE	843898,621	.	843898,621	843898,621	843898,621	843898,621	843898,621	843898,621	843898,621	843898,621	843898,621
MAPE	2,837	.	2,837	2,837	2,837	2,837	2,837	2,837	2,837	2,837	2,837
MaxAPE	16,322	.	16,322	16,322	16,322	16,322	16,322	16,322	16,322	16,322	16,322
MAE	619875,105	.	619875,105	619875,105	619875,105	619875,105	619875,105	619875,105	619875,105	619875,105	619875,105
MaxAE	2779356,739	.	2779356,739	2779356,739	2779356,739	2779356,739	2779356,739	2779356,739	2779356,739	2779356,739	2779356,739
BIC normalizado	27,405	.	27,405	27,405	27,405	27,405	27,405	27,405	27,405	27,405	27,405

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo	Ljung-Box Q(18)			Número de valores atípicos
		R-cuadrado estacionaria	Estadísticos	GL	Sig.	
Total-Modelo_1	0	,566	17,051	14	,253	0

Parámetros del modelo ARIMA

				Estimación	ET	t	Sig.
Total-Modelo_1	Total	Log natural	Constante	,019	,011	1,748	,082
			AR Retardo 1	,960	,021	44,782	,000
			MA Retardo 1	,604	,061	9,831	,000
			AR, estacional Retardo 1	-,686	,061	-11,318	,000
			Retardo 2	-,423	,061	-6,914	,000
			Diferenciación estacional	1			



Observamos que este modelo si pasa el Test de Ljung-Box, ya que su $p\text{-valor}=0.253>0.05$, sin embargo, fijándonos en los parámetros del Modelo, la constante no es significativa, así que la quitaremos y realizamos el mismo modelo de SARIMA(1,0,1)(2,1,0) Transformada Log, sin Constante.

Modelizador para series temporales

[Conjunto_de_datos1] C:\Users\Isabela\OneD:

Descripción del modelo

			Tipo de modelo
ID del modelo	Total	Modelo_1	ARIMA(1,0,1) (2,1,0)

Resumen del modelo

Ajuste del modelo											
Estadístico de ajuste	Media	ET	Mínimo	Máximo	Percentil						
					5	10	25	50	75	90	95
R-cuadrado estacionaria	,564	.	,564	,564	,564	,564	,564	,564	,564	,564	,564
R-cuadrado	,992	.	,992	,992	,992	,992	,992	,992	,992	,992	,992
RMSE	846091,738	.	846091,738	846091,738	846091,738	846091,738	846091,738	846091,738	846091,738	846091,738	846091,738
MAPE	2,859	.	2,859	2,859	2,859	2,859	2,859	2,859	2,859	2,859	2,859
MaxAPE	15,797	.	15,797	15,797	15,797	15,797	15,797	15,797	15,797	15,797	15,797
MAE	626633,243	.	626633,243	626633,243	626633,243	626633,243	626633,243	626633,243	626633,243	626633,243	626633,243
MaxAE	2774181,519	.	2774181,519	2774181,519	2774181,519	2774181,519	2774181,519	2774181,519	2774181,519	2774181,519	2774181,519
BIC normalizado	27,388	.	27,388	27,388	27,388	27,388	27,388	27,388	27,388	27,388	27,388

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo	Ljung-Box Q(18)			Número de valores atípicos
		R-cuadrado estacionaria	Estadísticos	GL	Sig.	
Total-Modelo_1	0	,564	15,821	14	,324	0

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
Total-Modelo_1	Total	Log natural	AR	Retardo 1	,976	,015	63,689	,000
			MA	Retardo 1	,614	,057	10,724	,000
			AR, estacional	Retardo 1	-,687	,060	-11,419	,000
				Retardo 2	-,424	,061	-6,964	,000
			Diferenciación estacional		1			

Nos quedamos con este Modelo, ya que sigue siendo significativo en el test de Ljung-Box.

El Modelo Final será: SARIMA(1,0,1)(2,1,0) Sin Constante, Transformación Logarítmica.

Tiene como ecuación:

$$(1 - 0,976B)(1 + 0,687B^{12} + 0,424B^{24})(1 - B^{12})\ln(Y_t) = (1 - 0,614B)\varepsilon_t$$

3. Validación

En este apartado comprobaremos la validez del modelo elegido, si el modelo adecuado o no. En caso de que no fuera adecuado, deberíamos volver a la fase de identificación y proponer un modelo alternativo.

3.1 SOBREAJUSTE

En este apartado añadiremos más parámetros a cada una de las partes del modelo. Si este fuera bueno, debería rechazar los parámetros extra.

Los modelos que comprobaremos serán:

- SARIMA (2, 0, 1)(2, 1, 0)
- SARIMA (1, 0, 2)(2, 1, 0)

SARIMA (2, 0, 1) (2, 1, 0)

Modelizador para series temporales

[Conjunto_de_datos1] C:\Users\Isabela\OneD:

Descripción del modelo

			Tipo de modelo
ID del modelo	Total	Modelo_1	ARIMA(2,0,1) (2,1,0)

Resumen del modelo

Ajuste del modelo

Figura 10. Métricas de ajuste											
Estadístico de ajuste	Media	ET	Mínimo	Máximo	Percentil						
					5	10	25	50	75	90	95
R-cuadrado estacionaria	,567	.	,567	,567	,567	,567	,567	,567	,567	,567	,567
R-cuadrado	,992	.	,992	,992	,992	,992	,992	,992	,992	,992	,992
RMSE	839764,088	.	839764,088	839764,088	839764,088	839764,088	839764,088	839764,088	839764,088	839764,088	839764,088
MAPE	2,873	.	2,873	2,873	2,873	2,873	2,873	2,873	2,873	2,873	2,873
MaxAPE	15,073	.	15,073	15,073	15,073	15,073	15,073	15,073	15,073	15,073	15,073
MAE	625609,450	.	625609,450	625609,450	625609,450	625609,450	625609,450	625609,450	625609,450	625609,450	625609,450
MaxAE	2789340,321	.	2789340,321	2789340,321	2789340,321	2789340,321	2789340,321	2789340,321	2789340,321	2789340,321	2789340,321
BIC normalizado	27,396	.	27,396	27,396	27,396	27,396	27,396	27,396	27,396	27,396	27,396

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo	Ljung-Box Q(18)			Número de valores atípicos
		R-cuadrado estacionaria	Estadísticos	GL	Sig.	
Total-Modelo_1	0	,567	13,416	13	,416	0

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
Total-Modelo_1	Total	Log natural	AR	Retardo 1	,819	,111	7,366	,000
				Retardo 2	,149	,103	1,453	,147
			MA	Retardo 1	,525	,097	5,429	,000
				AR, estacional	Retardo 1	-,672	,062	-10,918
			Diferenciación estacional	Retardo 2	-,401	,063	-6,361	,000
					1			

Al analizar los resultados, observamos que, aunque el modelo sigue cumpliendo con el test de Ljung-Box, el nuevo parámetro introducido, AR 2 no es estadísticamente significativo con un p-valor > 0,05. Esto indica que la inclusión de este retardo es redundante y no aporta capacidad explicativa adicional volviendo al Modelo SARIMA(1,0,1)(2,1,0), al que estamos sobreajustando.

SARIMA (1, 0, 2)(2, 1, 0)

Modelizador para series temporales

[Conjunto_de_datos1] C:\Users\Isabela\OneD:

Descripción del modelo

			Tipo de modelo
ID del modelo	Total	Modelo_1	ARIMA(1,0,2) (2,1,0)

Resumen del modelo

Ajuste del modelo

Estadístico de ajuste	Media	ET	Mínimo	Máximo	Percentil						
					5	10	25	50	75	90	95
R-cuadrado estacionaria	,567	.	,567	,567	,567	,567	,567	,567	,567	,567	,567
R-cuadrado	,992	.	,992	,992	,992	,992	,992	,992	,992	,992	,992
RMSE	839680,799	.	839680,799	839680,799	839680,799	839680,799	839680,799	839680,799	839680,799	839680,799	839680,799
MAPE	2,874	.	2,874	2,874	2,874	2,874	2,874	2,874	2,874	2,874	2,874
MaxAPE	15,025	.	15,025	15,025	15,025	15,025	15,025	15,025	15,025	15,025	15,025
MAE	625862,724	.	625862,724	625862,724	625862,724	625862,724	625862,724	625862,724	625862,724	625862,724	625862,724
MaxAE	2800479,533	.	2800479,533	2800479,533	2800479,533	2800479,533	2800479,533	2800479,533	2800479,533	2800479,533	2800479,533
BIC normalizado	27,395	.	27,395	27,395	27,395	27,395	27,395	27,395	27,395	27,395	27,395

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo	Ljung-Box Q(18)			Número de valores atípicos
		R-cuadrado estacionaria	Estadísticos	GL	Sig.	
Total-Modelo_1	0	,567	13,384	13	,419	0

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
Total-Modelo_1	Total	Log natural	AR	Retardo 1	,972	,017	56,926	,000
			MA	Retardo 1	,679	,068	9,961	,000
				Retardo 2	-,095	,068	-1,400	,163
			AR, estacional	Retardo 1	-,672	,062	-10,911	,000
				Retardo 2	-,401	,063	-6,352	,000
			Diferenciación estacional		1			

Nuevamente, el coeficiente asociado al segundo retardo de media móvil MA 2 resulta no significativo.

Conclusión : Al rechazarse todos estos modelos, se aprueba la hipótesis nula y nuestro modelo será válido.

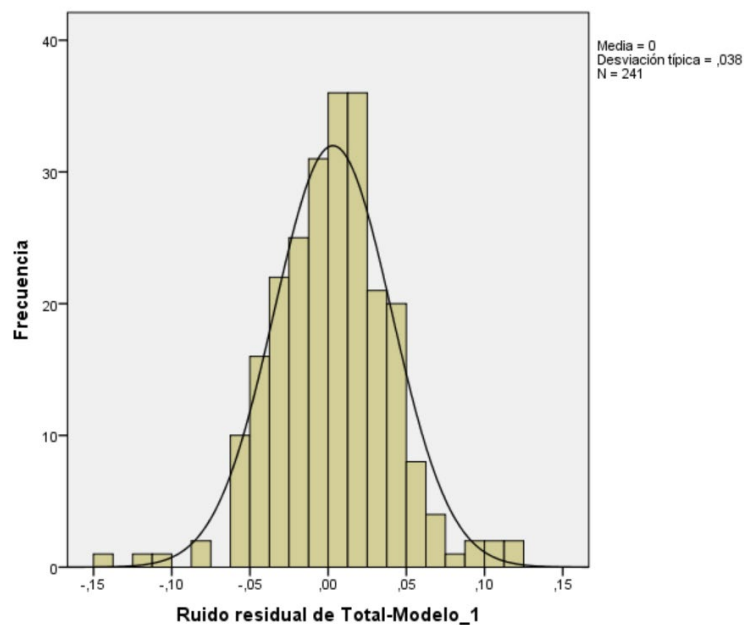
3.2 ESTUDIO DE LOS RESIDUOS

3.2.1. NORMALIDAD

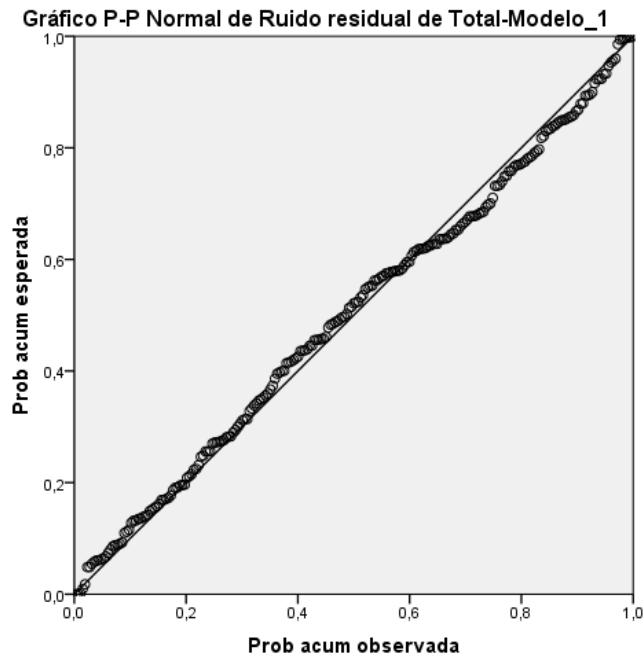
Realizamos el gráfico de la Normal para ver si los datos se ajustan correctamente, más adelante se realizará el Gráfico P-P también.

Gráfico

[Conjunto_de_datos1] C:\Users\Isabela\OneDrive\Documentos\spss isa series datos.sav



Observamos que no tiene mala pinta, los datos se ajustan muy bien. Realizaremos el Gráfico P-P para ver si los datos se ajustan a la recta.



Como habíamos comprobado antes, los datos se ajustan muy bien a la recta teórica. Aún así, para asegurarnos, realizaremos el Test de Kolmogorov-Smirnov.

Pruebas no paramétricas

[Conjunto_de_datos1] C:\Users\Isabela\OneDrive\

Prueba de Kolmogorov-Smirnov para una muestra

		Ruido residual de Total-Modelo_1
N		241
Parámetros normales ^{a,b}	Media	,00
	Desviación típica	,038
Diferencias más extremas	Absoluta	,049
	Positiva	,049
	Negativa	-,037
Z de Kolmogorov-Smirnov		,759
Sig. asintót. (bilateral)		,612

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

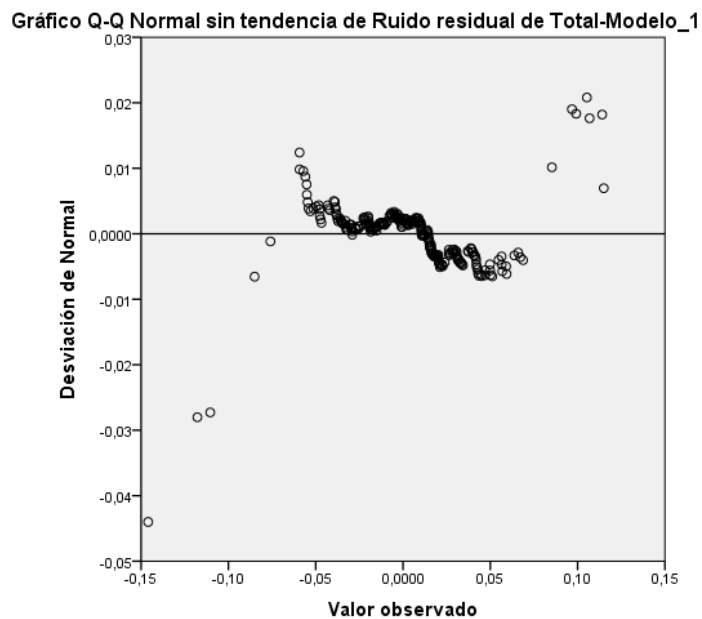
El test de Kolmogorov-Smirnov verifica a un nivel de significación de $0,612 > 0.05$ que los residuos del modelo provienen de una distribución normal.

$\begin{cases} H_0: \text{los residuos son normales} \\ H_1: \text{los residuos no son normales} \end{cases}$

Es decir, no rechazamos H_0 , la hipótesis nula, los residuos de este modelo pueden ser normales.

3.2.2. ESTUDIO DE LA MEDIA

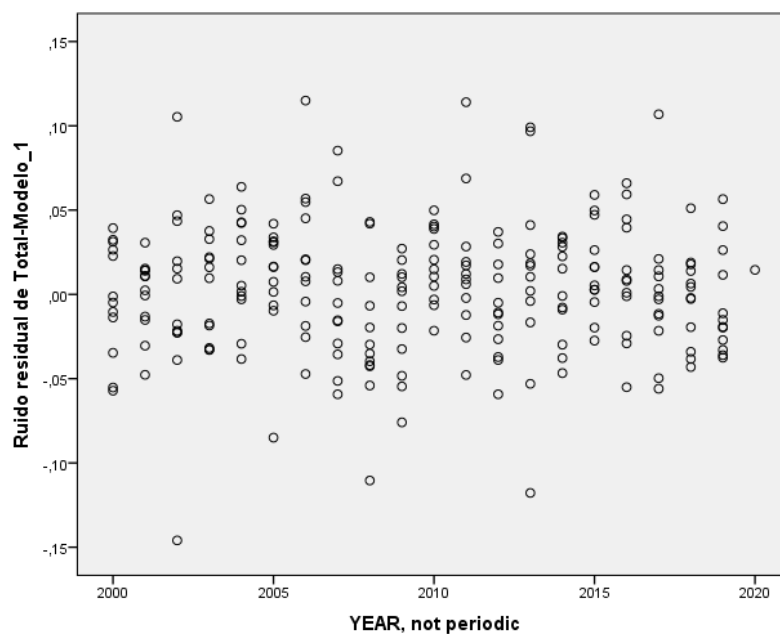
Una vez verificado que los residuos son normales con media 0 debemos de ver cómo se distribuyen, de forma que representamos la nube de puntos.



Se puede ver como los primeros y últimos valores se encuentran más alejados del 0, lo cual es lógico ya que al representar el gráfico normal estos valores también se encontraban alejados de la recta.

3.2.3. ESTUDIO DE LA HOMOCEDASTICIDAD

Para estudiar la existencia de homocedasticidad en los residuos del modelo, obtendremos un diagrama de dispersión en el que comprobaremos, de forma visual, la existencia de la misma.



Viendo cómo se distribuyen los puntos podemos decir que no existe heterocedasticidad en los residuos del modelo, es decir, tienen varianza constante.

3.2.4. ESTUDIO DE LA INCORRELACIÓN

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo	Ljung-Box Q(18)			Número de valores atípicos
		R-cuadrado estacionaria	Estadísticos	GL	Sig.	
Total-Modelo_1	0	,564	15,821	14	,324	0

Para el modelo aceptado, al 95% de significación, tenemos que $sig = 0.324 > 0.05 = \alpha$. Esto nos indicaría que, para el contraste

$$\begin{cases} H_0: \text{los residuos están incorrelados} \\ H_1: \text{los residuos no están incorrelados} \end{cases}$$

no rechazaríamos la hipótesis nula. Es decir, los residuos están incorrelados.

3.2.5. ESTUDIO DE LA ALEATORIEDAD

Probaremos también su aleatoriedad, usaremos para esto el test de rachas. La hipótesis alternativa en este caso será que los valores de los residuos

Prueba de rachas

	Ruido residual de Total-Modelo_1
Valor de prueba ^a	0
Casos < Valor de prueba	120
Casos >= Valor de prueba	121
Casos en total	241
Número de rachas	102
Z	-2,517
Sig. asintót. (bilateral)	,012

a. Mediana

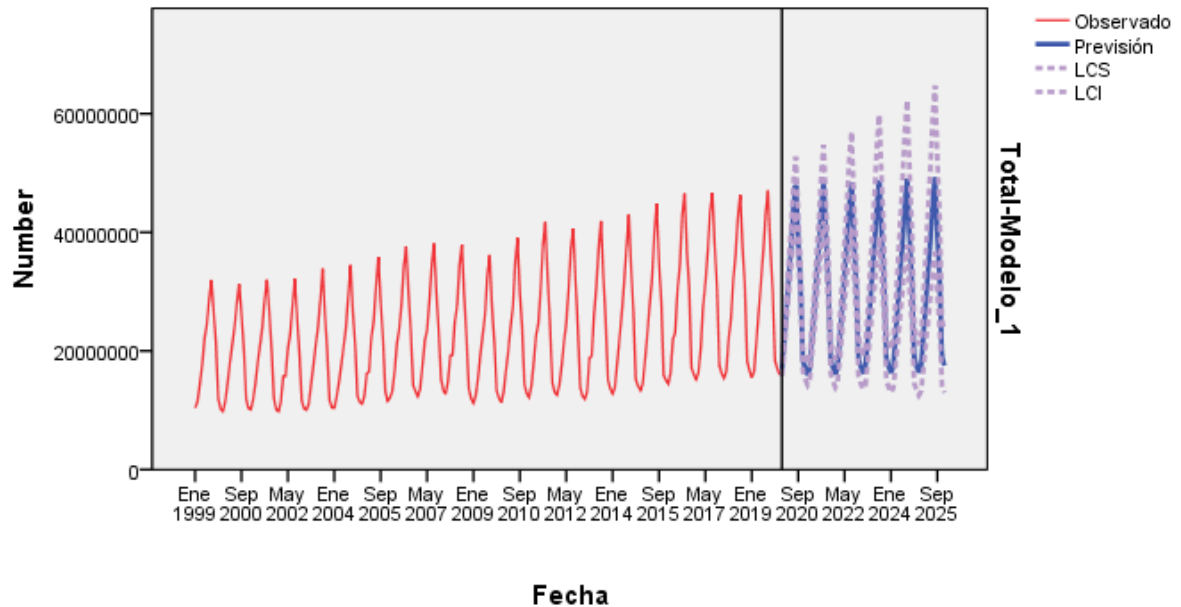
Para este caso, tenemos el siguiente contraste:

$$\begin{cases} H_0: \text{los residuos son aleatorios} \\ H_1: \text{los residuos no son aleatorios} \end{cases}$$

Al un nivel de confianza del 95%, tenemos que $sig = 0.012 < 0.05 = \alpha$, por lo que se habría evidencias para rechazar la hipótesis nula. En este caso, tenemos que los datos no son aleatorios.

4. Predicción

Por último, obtenemos una predicción hasta el día de hoy, es decir, 5 años después del último dato obtenido en la serie.



Tras realizar la predicción podemos observar en los intervalos de confianza están próximos a la estimación por tanto será una buena predicción, además vemos como la varianza es constante y la media también.