1º
semestre
2022/23

ANÁLISIS DE UNA SERIE TEMPORAL

SERIES TEMPORALES

CARMEN PLATA FERNÁNDEZ

GRADO EN ESTADÍSTICA | UNIVERSIDAD DE GRANADA

ÍNDICE

INTRODUCCIÓN A LA SERIE

DATOS

Usaremos los datos: "Tipo de interés medio al inicio de las hipotecas constituidas en las viviendas." Cogidos por el INE como indicadores nacionales mensuales sobre las hipotecas.

Estos datos, como hemos dicho anteriormente, son mensuales desde octubre de 2012 hasta septiembre de 2022, un total de 10 años.

Por tanto son 120 datos.

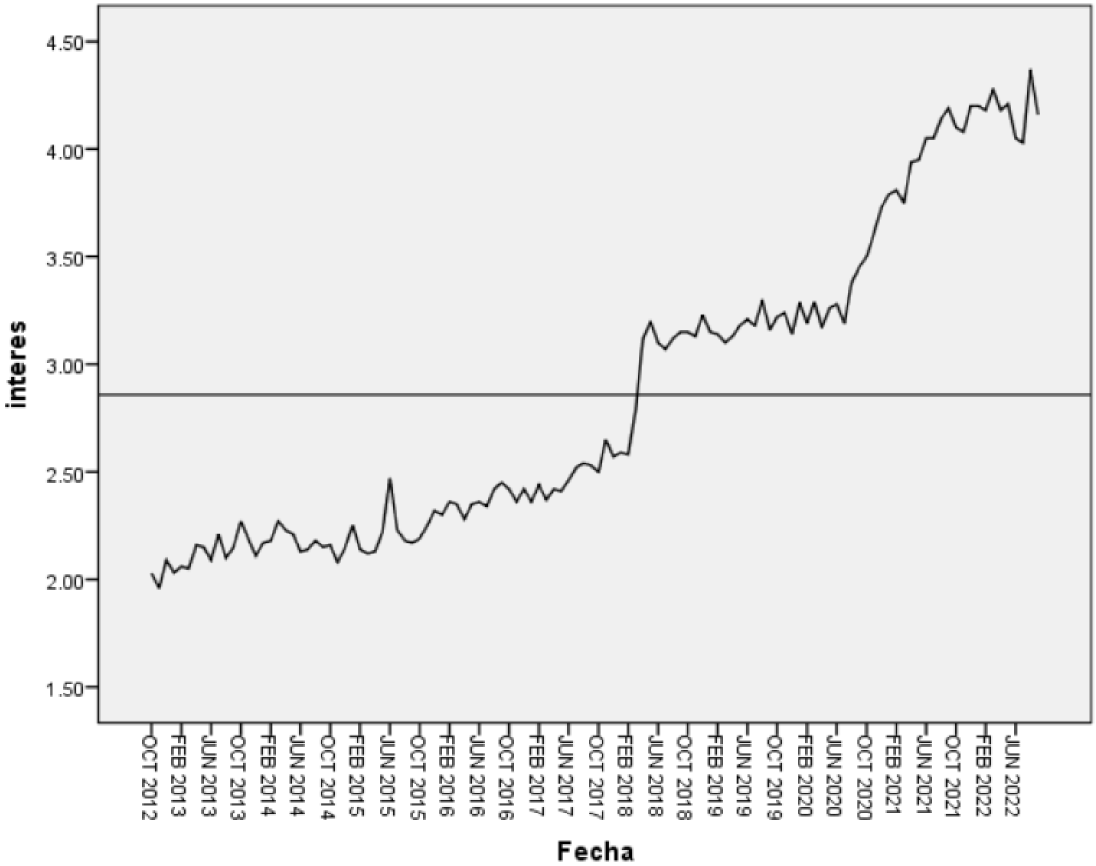
Podemos verlos en este link: https://www.ine.es/jaxiT3/Datos.htm?t=24457#!tabs-grafico.

OBJETIVO

Nuestro objetivo es hacer un análisis de esta serie tal que, podamos obtener un modelo del que extraer conclusiones y predecir los datos con el menor error posible.

GRÁFICO

Veamos el gráfico de secuencia de la serie, el cual obtenemos en SPSS en Analizar > Predicciones > Gráficos de secuencia. Aquí introducimos la variable "interés" en "Variables". En Formato, añadimos la "Línea de referencia en la media de la serie" para poder verificar bien que haya claramente diferencia de medias, como explico después.



En el gráfico de secuencias veremos si es estacionaria y si existe estacionalidad.

Vemos que no es estacionaria, ya que no sigue ninguna tendencia y tiene claramente diferencia de medias; es decir, existen tramos (entre los años 2012 y 2017 la media será sobre 2.25; entre 2018 y 2020 una media de 3.25 y entre 2020 y 2022 una media de 3.75).

No existe estacionalidad ya que, no hay ningún crecimiento fuerte ni cambios bruscos de nivel (picos).

Ya prevemos que hará falta diferenciarla por no ser estacionaria.

Como hemos visto que la serie no cumple la propiedad de estacionariedad, la idea es que los retardos en las autocorrelaciones vayan descendiendo lentamente, esto lo comprobaremos más adelante.

Realizaremos el análisis de Box-Jenkins.

Como hemos visto en clase, este se divide en cuatro fases:

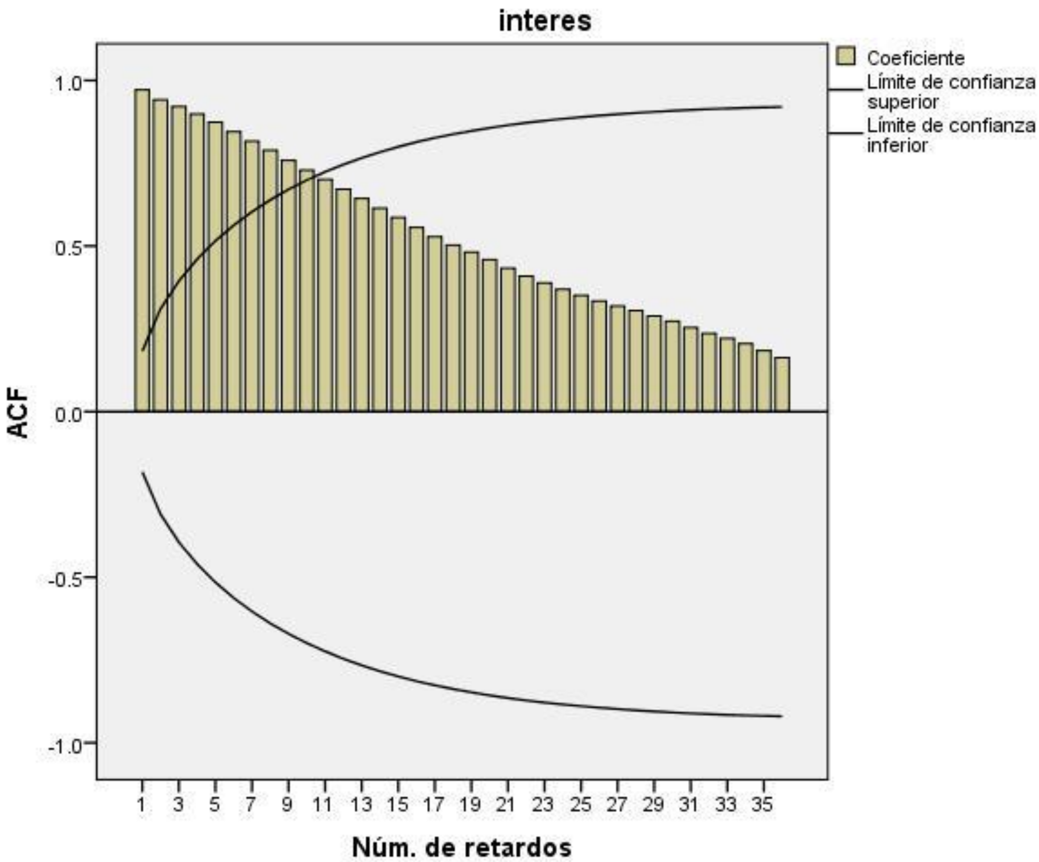
- IDENTIFICACIÓN
- ESTIMACIÓN
- VALIDACIÓN
- PREDICCIÓN

MODELO DE BOX-JENKINS

1. IDENTIFICACIÓN

Para poder identificarla debemos dibujar las funciones de autocorrelación y autocorrelación parcial.

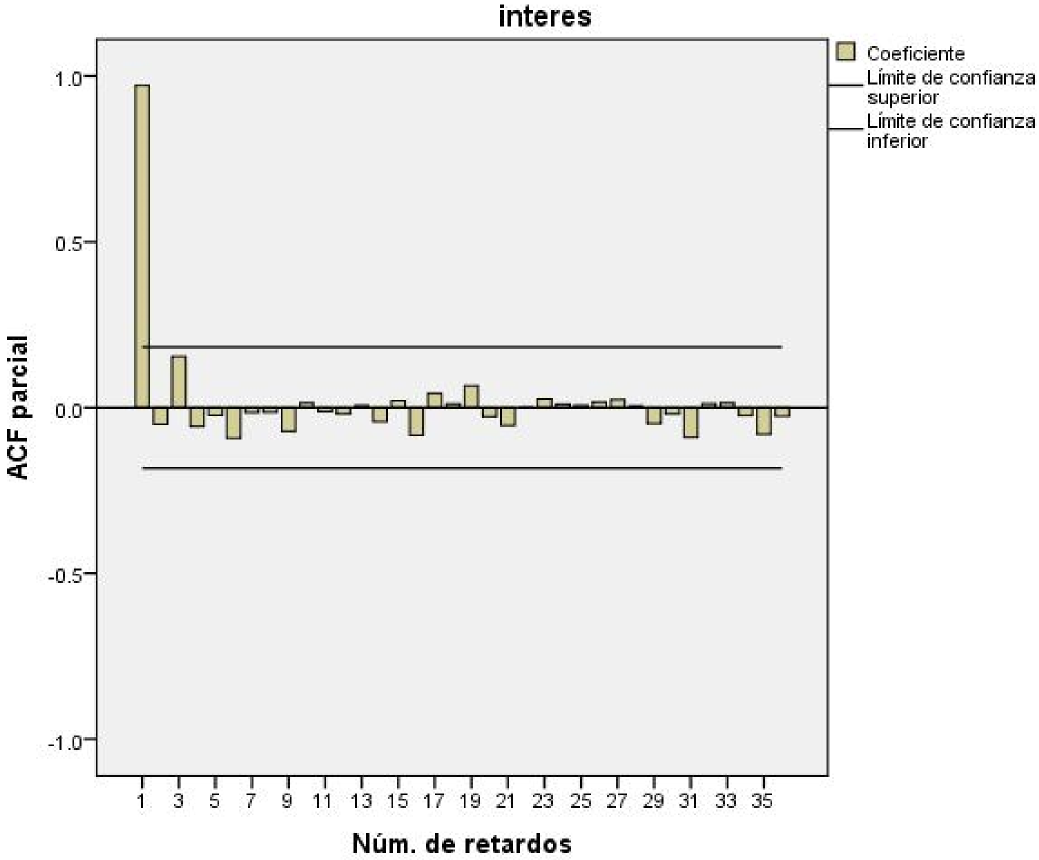
Para poder analizarla, usaremos un número de retardos que se encuentre entre T/3 y T/4, y que sea múltiplo de 12 (ya que los datos son mensuales). T=120, por lo que usaremos 36 retardos. Para calcular la variabilidad, utilizaremos la aproximación de Bartlett.



Vemos que en la función de autocorrelación, hay un decrecimiento lento, lo que también es signo de que no sea estacionaria.

Son significativamente distintos de cero los 10 primeros retardos. El hecho de que haya muchos retardos significativos, también es signo de que no es estacionaria.

Al haber un decaimiento, necesitaremos incluir términos AR en el modelo.



Vemos que en la función de autocorrelación parcial hay un corte en el primer retardo.

Solo es significativamente distintos de cero este valor.

Como aparece un corte, deberemos incluir un AR(p).

Si decreciese sin tener aparentemente un corte, entonces incluiríamos un término MA, pero no es el caso.

Teniendo esto en cuenta, probaremos un modelo AR(1).

2. ESTIMACIÓN

Modelo AR (1)

En SPSS le damos a Analizar > Predicciones > CrearModelosTradicionales, e introducimos la variable "interés" como "variable dependiente".

Cambiamos el método a modelizador ARIMA, y en criterio, como queremos ajustar un modelo AR(1) ponemos (1,0,0).

Cambiamos a la ventana "Estadísticos" y seleccionamos las tres opciones de

“Estadísticos de modelos individuales”. En la pestaña de “Gráficos”, quitamos el gráfico de la serie, y añadimos la “Función de autocorrelación simple (FAS) residual” y la “Función de autocorrelación parcial (FAP) residual”

En la pestaña Opciones vamos a poner, en número de retardos, 36 con un intervalo de confianza al 95%.

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
interes-Modelo_1	interes	Sin transformación	Constante		3.089	1.393	2.218	.028
			AR	Retardo 1	.995	.012	85.129	.000

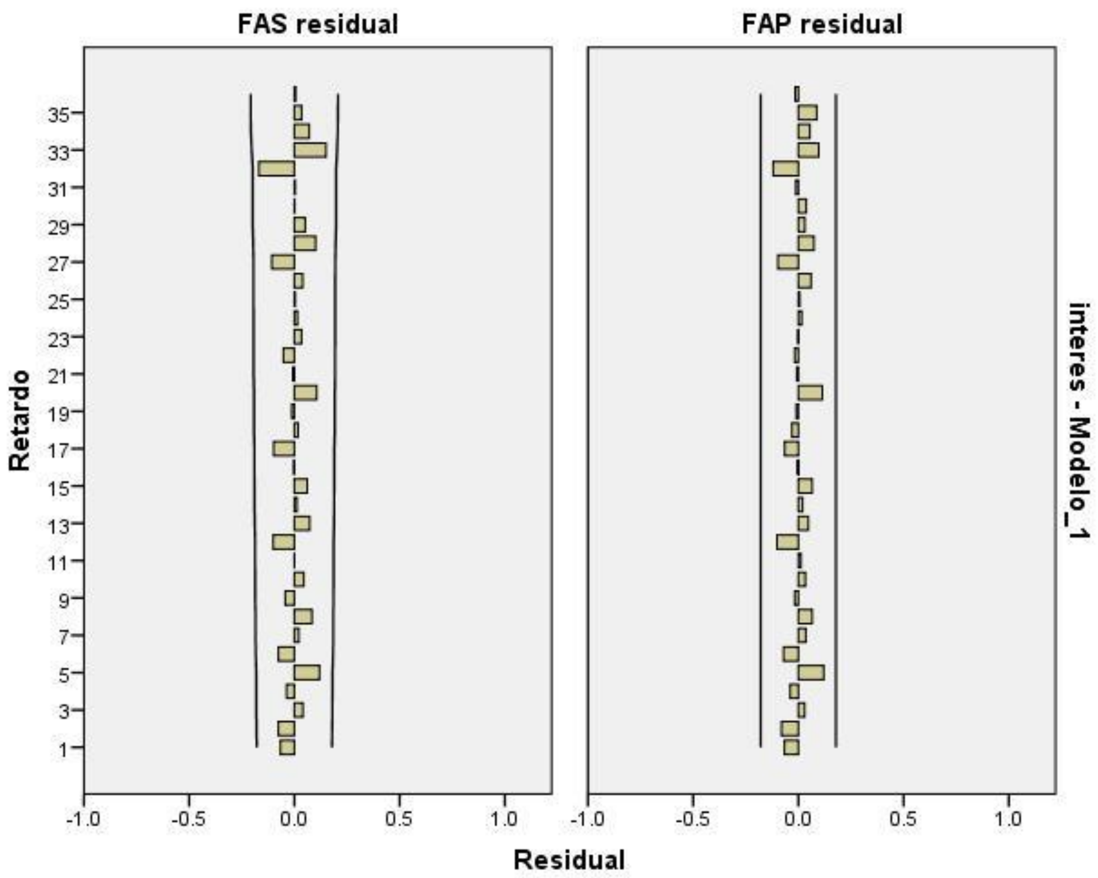
Ahora debemos estudiar si las estimaciones son significativas o no. Si resulta que no lo son, se podrían eliminar del modelo y el modelo sería un modelo con menos parámetros, más parsimonioso, que es preferible a un modelo más complejo, que introduce más distorsiones.

Nuestro test de hipótesis será:

$$\begin{cases} H_0, & parámetro = 0 \\ H_1, & parámetro \neq 0 \end{cases}$$

Para ello miramos la significación de cada uno de los parámetros. Como la significación es <0.05, se rechaza H0, y el parámetro se considera no nulo. Los dos parámetros, al ser menor que 0.05, son significativamente distintos de 0. Si alguno fuera 0, se quitaría el modelo y se volvería a hacer la estimación.

Miremos las funciones de autocorrelación y la parcial residuales:



Los parámetros se encuentran dentro de las bandas de confianza.

Los residuos son incorrelados.

Miramos el test de Ljung-Box para ver si el modelo es válido.

El contraste de hipótesis será:

$$\begin{cases} H_0: \rho_1 = \rho_2 = \cdots = \rho_{18} = 0 \\ H_1: \rho_i \neq 0, \qquad i\epsilon(1{,}18) \end{cases}$$

Estamos comparando los 18 primeros.

Al ser la significación 0.915, se rechaza la hipótesis nula.

Por tanto, este modelo es válido.

Comprobamos si los parámetros pueden ser 1. Si no admitiera el valor 1, sería estacionaria y por tanto, tendríamos un modelo válido.

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
interes-Modelo_1	interes	Sin transformación	Constante		3.089	1.393	2.218	.028
			AR	Retardo 1	.995	.012	85.129	.000

Esto lo estudiamos viendo el intervalo de confianza al 95% de los parámetros.

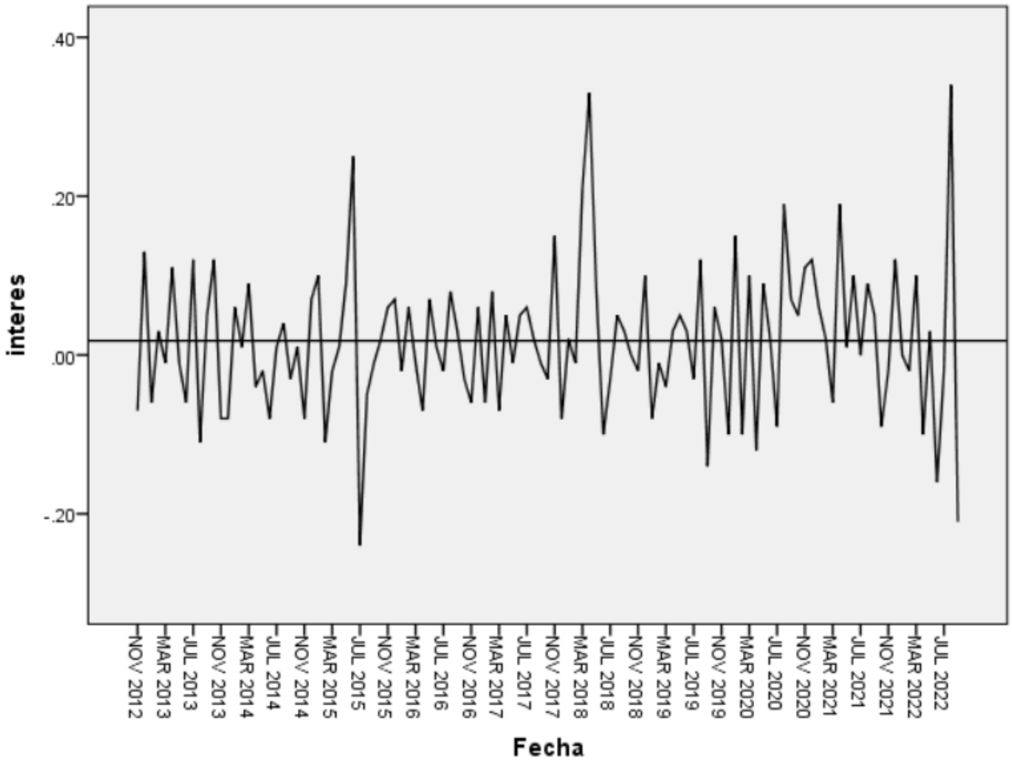
En la constante sería: $[3.089 \pm (1{,}96 \times 1.393)]$

Lo que nos da como resultado: (0.35872,5.81928)

Admite el 1, por lo que deberíamos realizar la diferenciación.

DIFERENCIACIÓN

Al realizar una diferenciación, realizamos el gráfico de secuencias como hemos explicado anteriormente, pero le añadiremos marcar la casilla "Diferencia" con un 1. Nos quedaría un gráfico de secuencia de la serie:

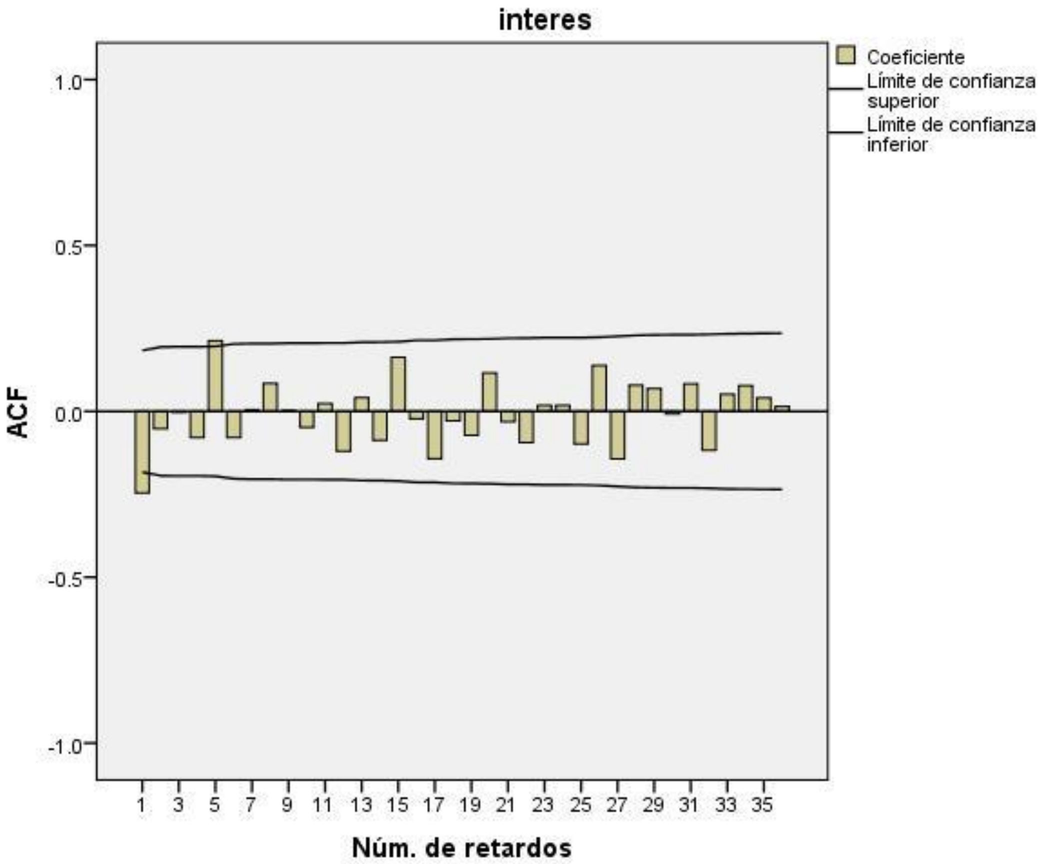


Transformaciones: diferencia(1)

Al hacer una diferencia, ya si se percibe cierta homogeneidad con la media cerca de 0 es decir, podríamos decir que se ha vuelto estacionaria.

Además, se observan ciertos picos que serían cambios bruscos de nivel los cuales son señal de que en nuestra serie existe estacionalidad.

Veamos sus autocorrelaciones:

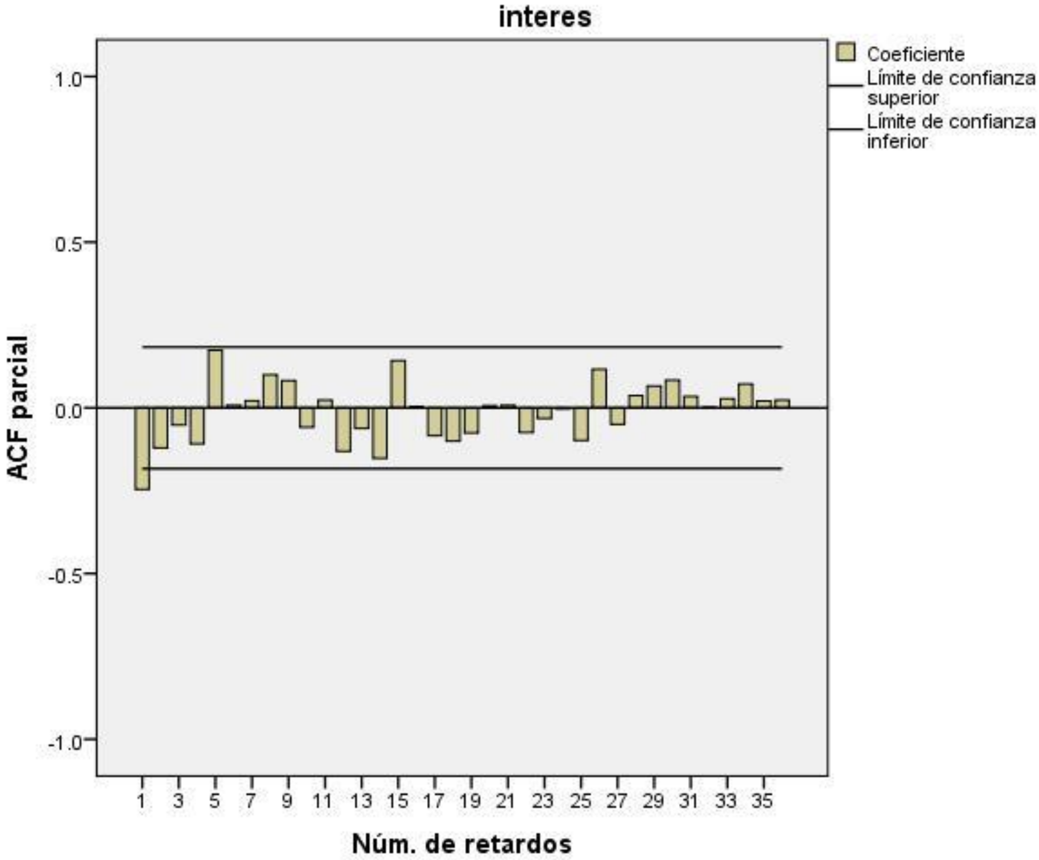


En la función de autocorrelación el primer valor se encuentra fuera de del límite de confianza, por lo que será significativamente distinto de cero.

Hay un corte en el primer retardo.

Como hemos diferenciado, al no haber ya un descenso en los retardos, es una señal de que ya podrá ser estacionaria.

Veamos la función de autocorrelación parcial:



Vemos que en el primer valor es significativamente distinto de 0. En este caso también hay un corte en el primer retardo, por tanto probaremos tanto un modelo ARIMA (1,1,0) como una ARIMA (0,1,1).

MODELO ARIMA (0,1,1)

Obtenemos los datos en “Analizar” > “Predicciones” > “CrearModelosTradicionales”.

Ponemos el modelo ARIMA (0,1,1) y obtenemos los siguientes datos:

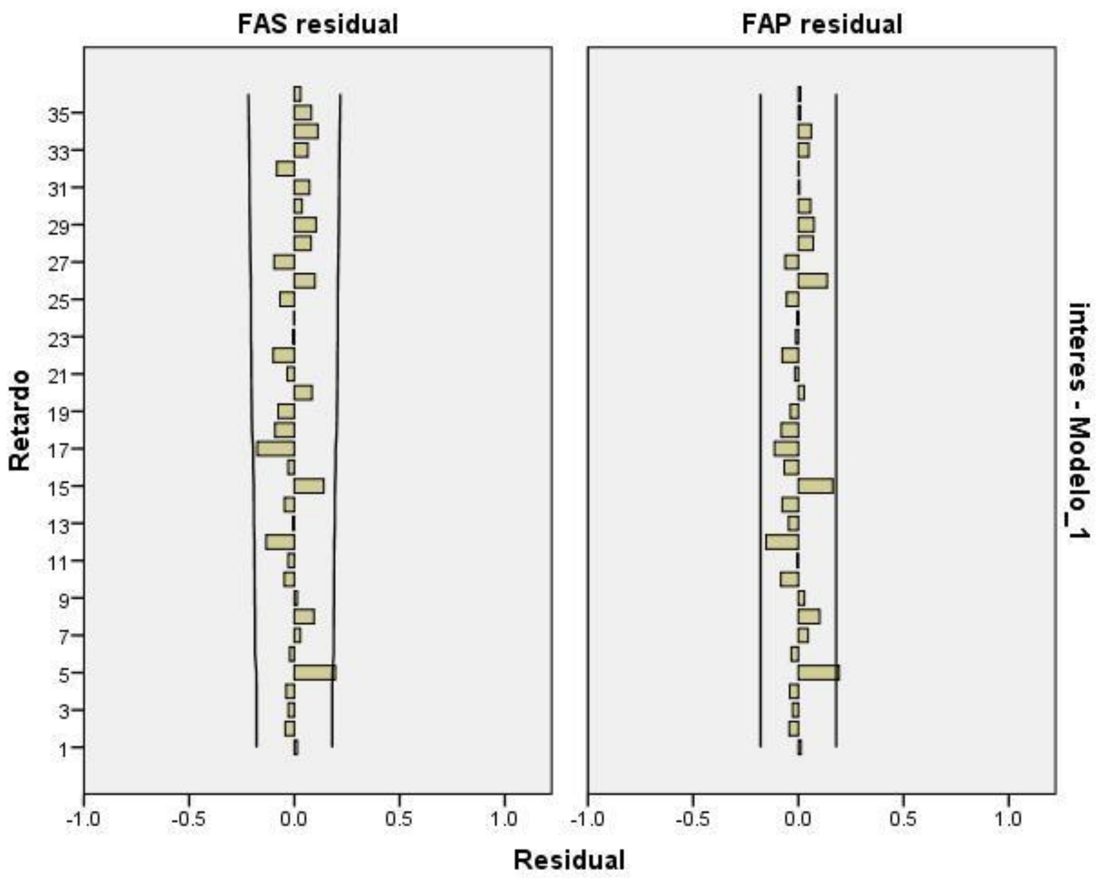
Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
interes-Modelo_1	interes	Sin transformación	Constante		.018	.006	3.201	.002
			Diferencia		1			
			MA	Retardo 1	.305	.088	3.448	.001

Veamos si los coeficientes son significativos.

Igual que antes, ambos coeficientes son significativamente distintos de cero.

Miremos si los parámetros de las funciones de autocorrelación y autocorrelación parcial se encuentran dentro de las bandas de confianza.



Se cumple.

Por ahora, podríamos decir que el modelo es válido, pero debemos comprobar el test de Ljung-Box:

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.	
interes-Modelo_1	0	.076	18.163	17	.379	0

El nivel de significación es 0.379, por tanto mayor que 0.05, por lo que es un modelo válido.

Veamos si hay que diferenciar otra vez, o es estacionaria:

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
interes-Modelo_1	interes	Sin transformación	Constante		.018	.006	3.201	.002
			Diferencia		1			
			MA	Retardo 1	.305	.088	3.448	.001

Constante: $[0.018 \pm (1{,}96 \times 0.006)]$ = (0.00624,0.02976)

Aquí se cumple.

Veamos que en el otro parámetro tampoco se encuentra el valor 1: $[0.305 \pm (1{,}96 \times 0.088)]$ = (0.13252,0.47748)

Hemos encontrado un modelo válido y estacionario.

MODELO ARIMA (1,1,0)

Realizamos los pasos hechos anteriormente, pero poniendo el modelo ARIMA (1,1,0) y obtenemos los siguientes datos:

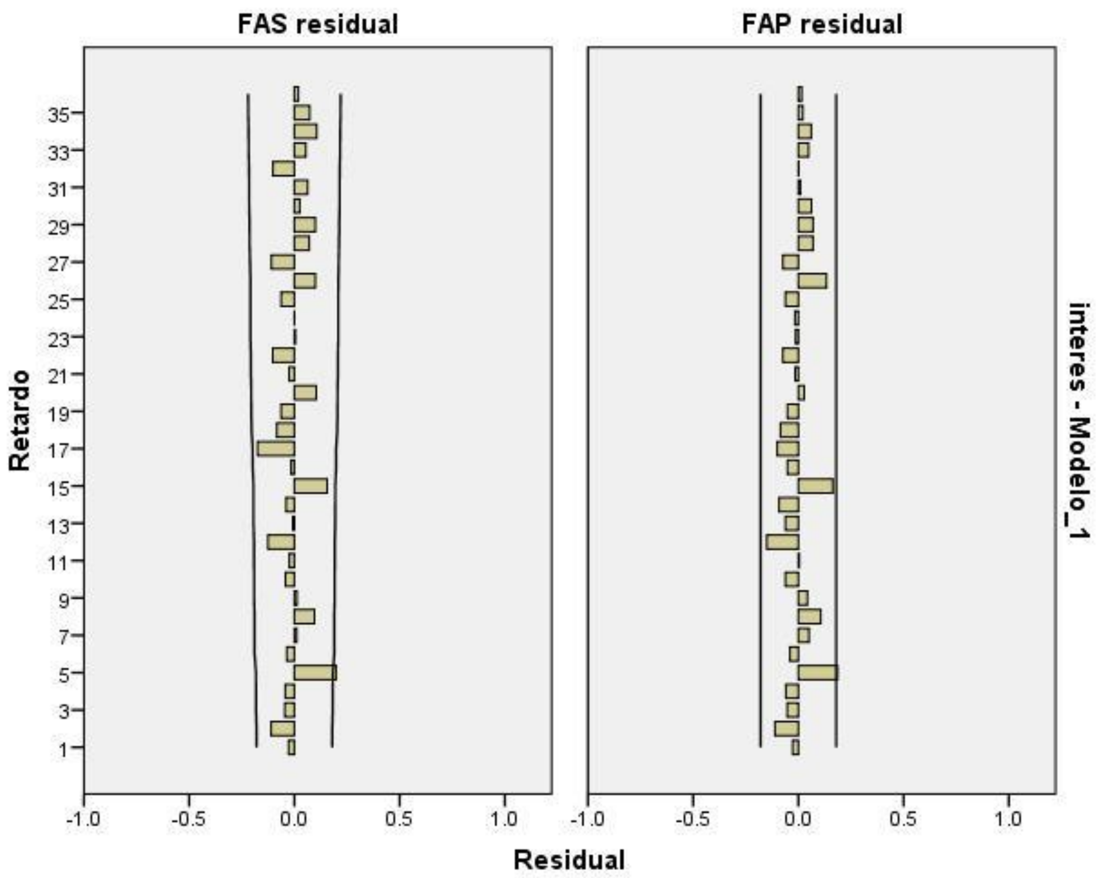
Estudiamos si los coeficientes son significativamente distintos a 0.

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
interes-Modelo_1	interes	Sin transformación	Constante		.018	.007	2.795	.006
			AR	Retardo 1	-.259	.092	-2.826	.006
			Diferencia		1			

Igual que antes ambos coeficientes lo son.

Veamos si los parámetros de las funciones de autocorrelación y autocorrelación parcial se encuentran dentro de las bandas de confianza.



Se cumple.

Por ahora, podríamos decir que el modelo es válido, pero debemos mirar el test de Ljung-Box:

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.	
interes-Modelo_1	0	.064	19.703	17	.290	0

El nivel de significación es 0.290, mayor que 0.05, por lo que es un modelo válido.

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
interes-Modelo_1	interes	Sin transformación	Constante		.018	.007	2.795	.006
			AR	Retardo 1	-.259	.092	-2.826	.006
			Diferencia		1			

Observemos que el valor 1 no se encuentra dentro de los parámetros del modelo:

La constantes estaría en el intervalo: ($[0.018 \pm (1{,}96 \times 0.007)]$) = (0.00428,0.03172)

Aquí se cumple. Veamos que en el otro parámetro tampoco se encuentra: ($[-0.259 \pm (1{,}96 \times 0.092)]$) = (-0.43932,-0.07868)

Hemos encontrado un modelo válido y estacionario.

3. VALIDACIÓN

La fase de validación nos servirá para saber si el modelo que hemos escogido es adecuado. Si lo es, utilizamos el modelo; si no lo es, volvemos a la fase de identificación.

Tenemos el problema que después del análisis realizado, hemos llegado a la conclusión de que tenemos más de un modelo válido:

- ARIMA (1,1,0)
- ARIMA (0,1,1)

Para elegir entre un modelo u otro, debemos usar algún criterio del estadístico de ajuste, generalmente el criterio BIC.

El criterio BIC es un compromiso entre la verosimilitud y el número de parámetros.

Elegimos aquel que tenga un BIC más pequeño con al menos tres unidades de diferencia.

No podemos elegir mediante el BIC porque no hay tres unidades de diferencia.

Nos quedaremos con el modelo ARIMA (0,1,1); ya que este tiene un valor de R^2= 0.076, mayor que el del modelo ARIMA (1,1,0), el cual es 0.064.

Llevamos a cabo un estudio de los residuos.

ESTUDIO DE LOS RESIDUOS

ESTUDIO DE LA INCORRELACIÓN

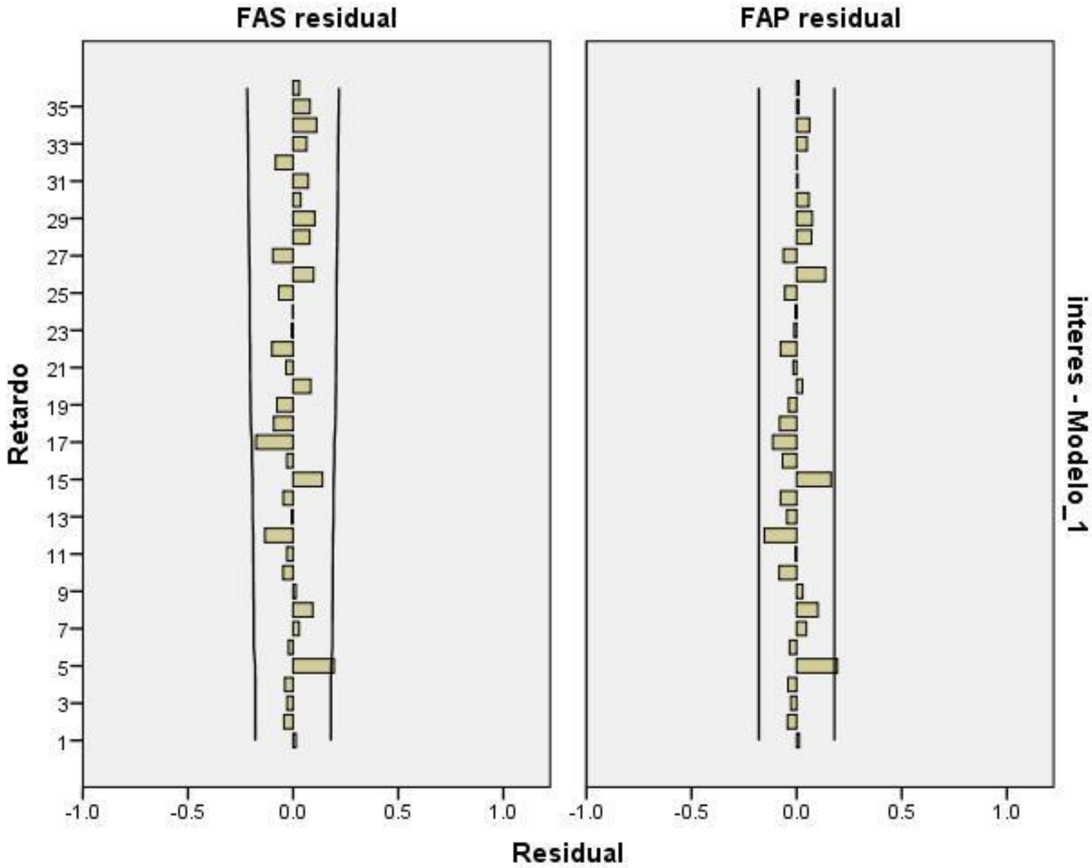
Observamos que de los cinco primeros valores, estos están dentro de las bandas de confianza.

Individualmente, cada coeficiente de autocorrelación se puede considerar nulo.

Se puede estudiar a través de distintos tests: Test Boc-Pierre, Test Ljung-Box, Test McLeod, Test McLeod-Li….

Los tests sirven para rechazar, pero no rechazan modelos “pobres”, es decir, que rechazan los buenos. Es un método para eliminar modelos, más que para decidir.

En este caso hemos aplicado anteriormente el Test Ljung-Box y nuestro modelo lo ha rechazado, por lo que es válido.



ESTUDIO DE LA ALEATORIEDAD

Veamos si los residuos son aleatorios o no.

$$\begin{cases} H_0: La\ muestra\ es\ aleatoria \\ H_1: La\ muesta\ no\ es\ aleatoria \end{cases}$$

Prueba de rachas

	Ruido residual de interes-Modelo_1
Valor de prueba[a]	.00
Casos < Valor de prueba	59
Casos >= Valor de prueba	60
Casos en total	119
Número de rachas	67
Z	1.198
Sig. asintót. (bilateral)	.231

a. Mediana

Vemos que no existen evidencias para rechazar el test de aleatoriedad, ya que sig. Es > 0.05 (0.231) por lo que suponemos que es aleatoria.

Podría seguir una normal, vamos a estudiarlo.

ESTUDIO DE LA NORMALIDAD

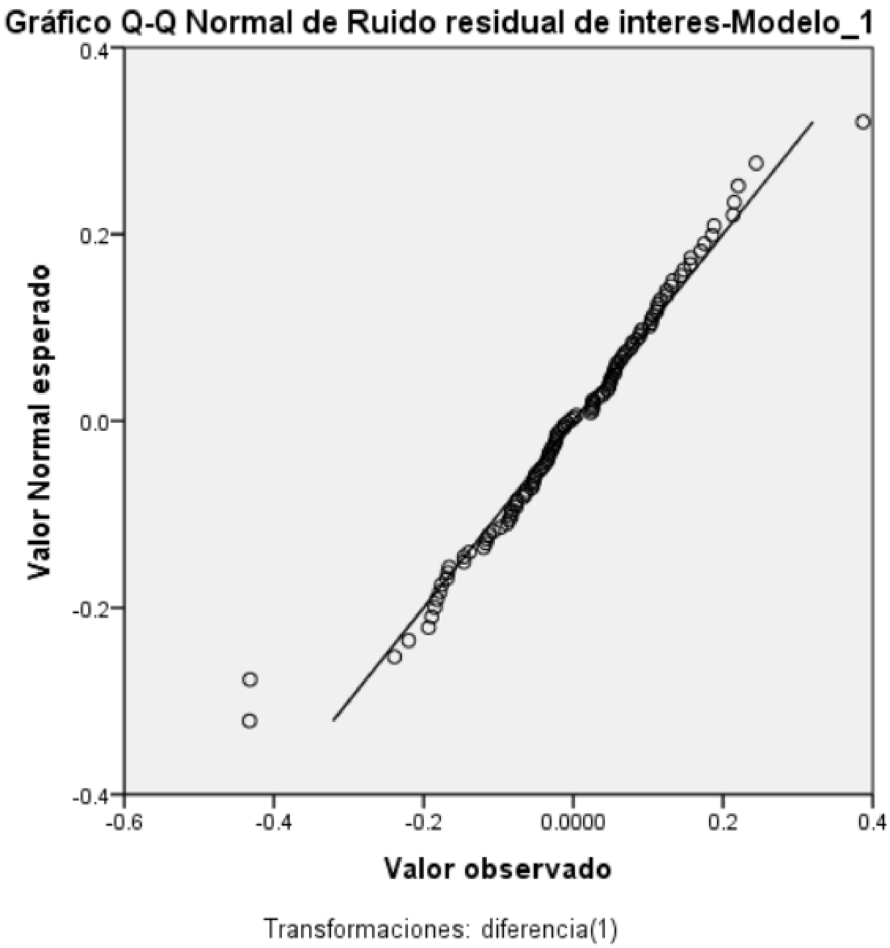
El estudio de la normalidad nos sirve para saber si los residuos siguen una normal.

Si esta hipótesis es violada, posiblemente debamos realizar una transformación Box-Cox.

La principal herramienta en el estudio de la normalidad son los gráficos probabilísticos, que consisten en dibujar los estadísticos ordenados de los valores obtenidos frente a los cuantiles de la distribución.

Si los puntos aparecen alineados podemos suponer que la distribución de los errores es normal.

Ahora vamos a realizar un gráfico Q-Q Normal de los residuos.



Los puntos deben estar alineados sobre la recta, como se puede comprobar.

Esto quiere decir que, el cuantil que estima y el cuantil que ajusta la distribución normal, se parece mucho. Y por tanto los errores siguen una distribución normal.

También se puede hacer un contraste no paramétrico para estudiar la normalidad.

La significación nos indica que no se rechaza la hipótesis nula, y que los errores siguen una distribución normal.

Realizaremos la Prueba de Kolmogorov-Smirnov para una muestra.

$$\begin{cases} H_0 -> N(\mu, \sigma) \\ H_1: X\ no\ sigue\ una\ Normal \end{cases}$$

Prueba de Kolmogorov-Smirnov para una muestra

		Ruido residual de interes-Modelo_1
N		119
Parámetros normales[a,b]	Media	-.0002
	Desviación típica	.08955
Diferencias más extremas	Absoluta	.065
	Positiva	.065
	Negativa	-.050
Z de Kolmogorov-Smirnov		.704
Sig. asintót. (bilateral)		.705

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Como el nivel de significación (0.705) mayor que 0.05, se acepta la hipótesis nula y por tanto, los residuos siguen una normal.

ESTUDIO DE LA HOMOCEDASTICIDAD

Para comprobar que los residuos cumplen la homocedasticidad es decir, el error comedito por el modelo tiene siempre la misma varianza, aplicaremos el test de Levene.

$$\begin{cases} H_0: Las\ varianzas\ de\ los\ grupos\ no\ son\ diferentes \\ H_1: Las\ varianzas\ de\ los\ grupos\ son\ diferentes \end{cases}$$

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									95% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Ruido residual de interes-Modelo_1	Se han asumido varianzas iguales	.682	.420	1.001	18	.330	.02597	.02595	-.02856	.08050
	No se han asumido varianzas iguales			1.001	16.771	.331	.02597	.02595	-.02885	.08078

Como el valor de significación (0.420) es mayor que 0.05, se acepta la hipótesis nula, y por tanto se cumple la hipótesis de homocedasticidad en los residuos.

Ahora vamos a realizar un sobreajuste.

SOBREAJUSTE

Esta técnica consiste en ajustar un modelo más complejo con más parámetros y estudiar la significación de los mismos. Incluiremos un parámetro, en la parte autorregresiva y luego en la de medias móviles. Así probaremos los modelos ARIMA (1,1,1) y ARIMA (0,1,2).

Probamos primero con el modelo <u>ARIMA (1,1,1)</u>

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
interes-Modelo_1	interes	Sin transformación	Constante		.018	.006	3.308	.001
			AR	Retardo 1	.126	.308	.407	.685
			Diferencia		1			
			MA	Retardo 1	.415	.282	1.475	.143

Vemos arriba claramente que, al ser dos de los parámetros mayores que 0.05 no son significativamente distintos de 0, por lo que deberíamos de eliminar algún parámetro.

Veámoslo ahora con el modelo <u>ARIMA (0,1,2)</u>

Parámetros del modelo ARIMA

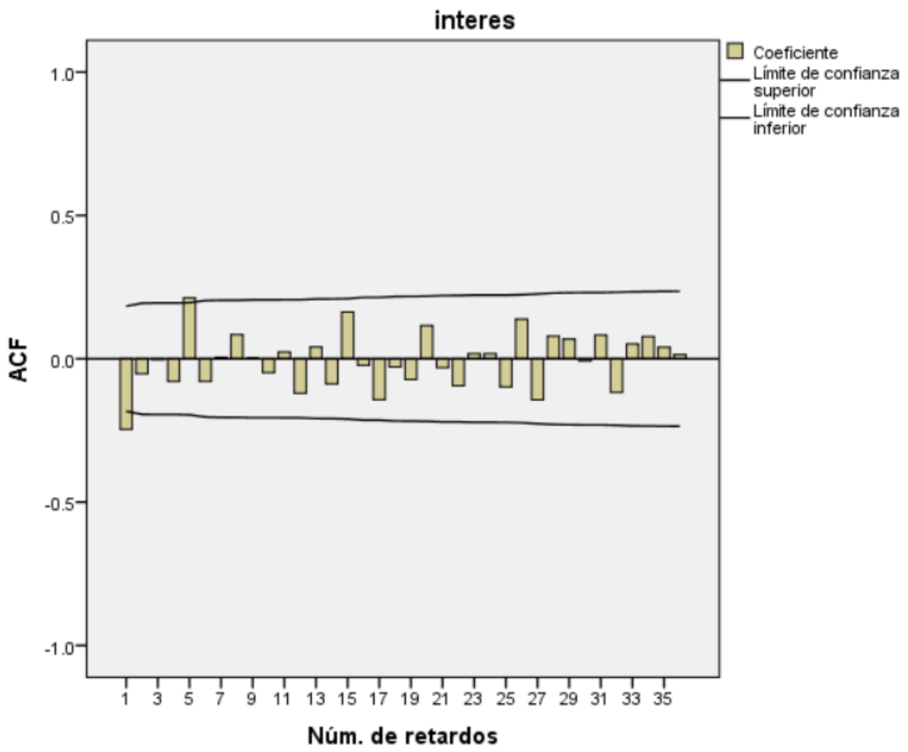
					Estimación	ET	t	Sig.
interes-Modelo_1	interes	Sin transformación	Constante		.018	.006	3.315	.001
			Diferencia		1			
			MA	Retardo 1	.288	.094	3.055	.003
				Retardo 2	.046	.097	.473	.637

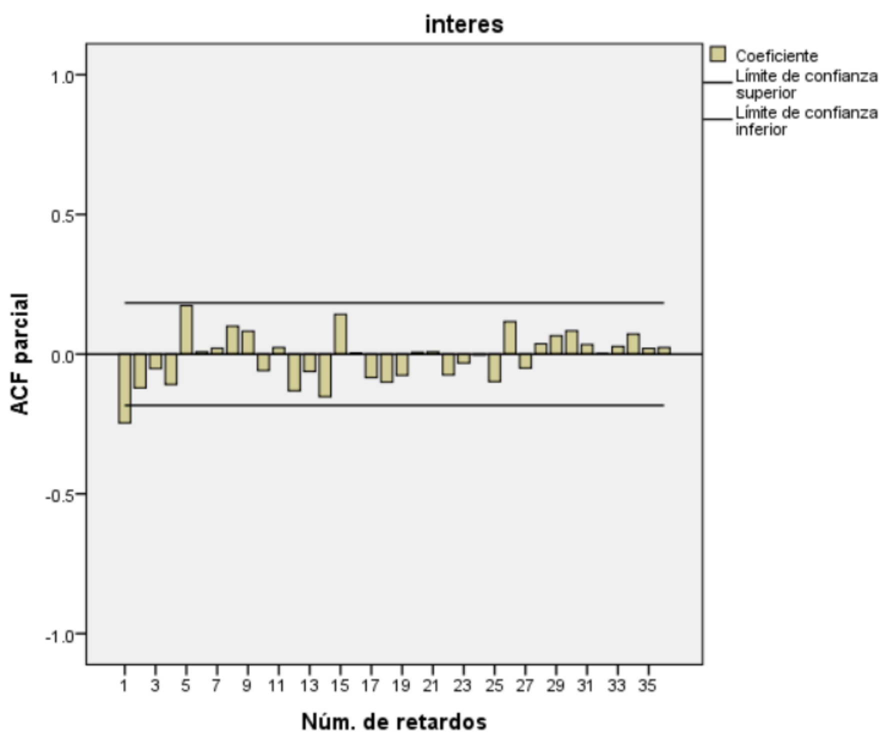
Aquí también hay un parámetro mayor que 0.05, por lo que ningún modelo es válido. Esto hace que volvamos a optar por el modelo anteriormente estimado.

CORRELACIÓN

Por último, debemos estudiar si la correlación entre los estimadores es pequeña, ya que si no se cumple esto las estimaciones serán inestables, muy sensibles a la muestra y no muy buenas.

Esto lo hemos visto antes en la gráfica de ACF y ACF parcial que hemos visto anteriormente.



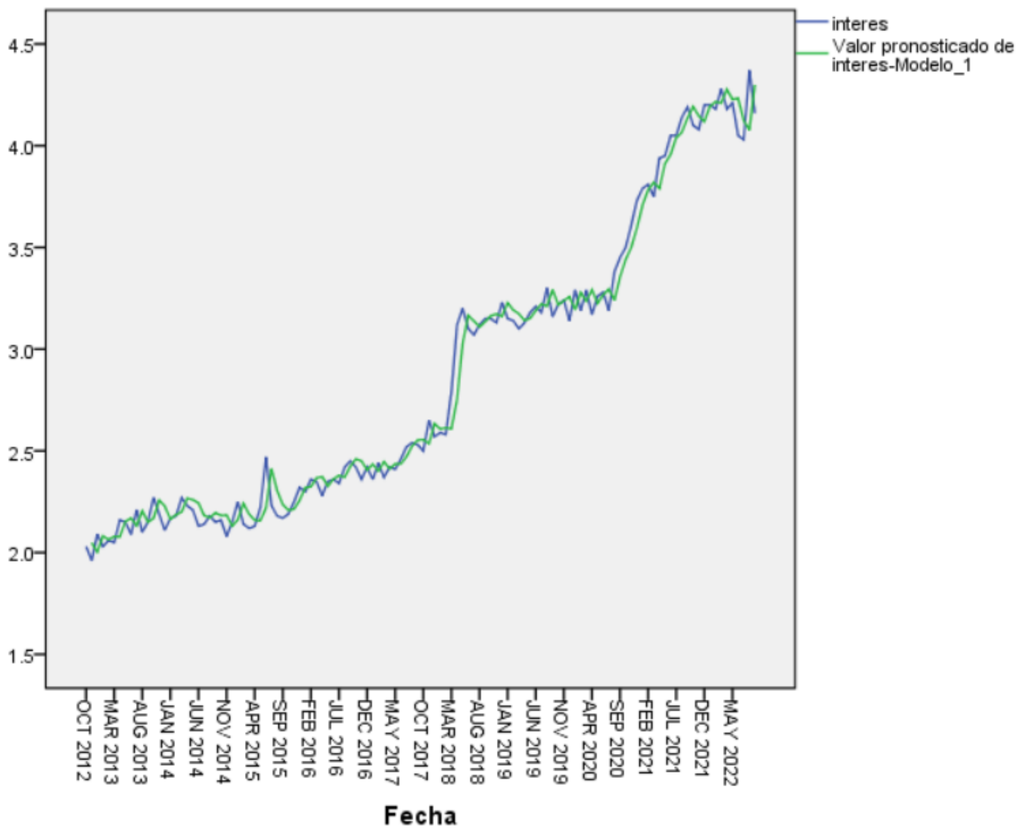


Se cumple, por lo que el modelo es válido.

4. PREDICCIÓN

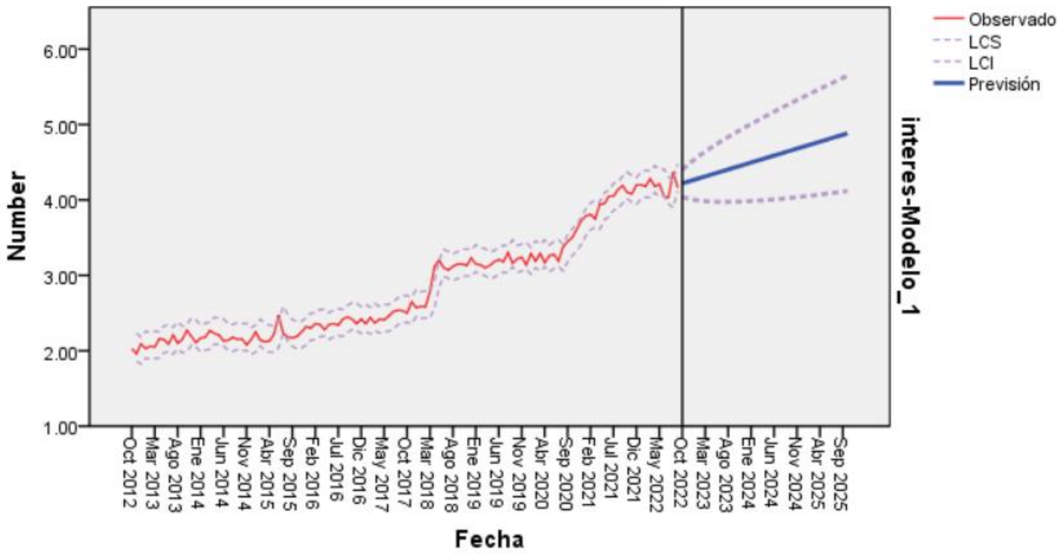
Veamos si los datos pronosticados se ajustan a los valores reales.

Realicemos el gráfico de secuencias:



Vemos que se ajustan bastante bien.

Por último, vamos a realizar una predicción de la serie para los tres años siguientes usando el modelo ajustado.



Vemos que los intervalos de confianza en la predicción son más largas y más amplias. Las bandas de confianza no se estabilizan con en el caso ARMA.

Al crear el modelo, en la pestaña "Guardar" hemos marcado la casilla de "Valores pronosticados". Así ahora, al ver nuestros datos, tenemos una variable llamada "Pronosticado_interes_Modelo" que son los datos pronosticados. Así podemos ver cuáles serán los datos pronosticados hasta octubre de 2025, (en la imagen solo aparecen hasta abril de 2024) como podemos ver a continuación:

	interes	YEAR_	MONTH_	DATE_	Pronosticado_interes_Modelo
112	4.20	2022	1	JAN 2022	4.19
113	4.18	2022	2	FEB 2022	4.22
114	4.28	2022	3	MAR 2022	4.21
115	4.18	2022	4	APR 2022	4.28
116	4.21	2022	5	MAY 2022	4.23
117	4.05	2022	6	JUN 2022	4.23
118	4.03	2022	7	JUL 2022	4.12
119	4.37	2022	8	AUG 2022	4.08
120	4.16	2022	9	SEP 2022	4.30
121	.	2022	10	OCT 2022	4.22
122	.	2022	11	NOV 2022	4.24
123	.	2022	12	DEC 2022	4.26
124	.	2023	1	JAN 2023	4.28
125	.	2023	2	FEB 2023	4.29
126	.	2023	3	MAR 2023	4.31
127	.	2023	4	APR 2023	4.33
128	.	2023	5	MAY 2023	4.35
129	.	2023	6	JUN 2023	4.37
130	.	2023	7	JUL 2023	4.39
131	.	2023	8	AUG 2023	4.40
132	.	2023	9	SEP 2023	4.42
133	.	2023	10	OCT 2023	4.44
134	.	2023	11	NOV 2023	4.46
135	.	2023	12	DEC 2023	4.48
136	.	2024	1	JAN 2024	4.50
137	.	2024	2	FEB 2024	4.52
138	.	2024	3	MAR 2024	4.53
139	.	2024	4	APR 2024	4.55

MODELO

Finalmente, después de probar el modelo y comprobar que verifica los requisitos, podemos decir que hemos encontrado un modelo ARIMA(0,1,1).

Este lo escribimos como:

$$X_t = X_{t-1} + \varepsilon_t - \theta\epsilon_{t-1}$$

En nuestro caso,

$$(1-\theta)X_t = (1-0.305)\varepsilon_t \rightarrow X_t - X_{t-1} = \varepsilon_t - 0.305\theta\, \varepsilon_{t-1}$$