



Universidad Politécnica  
de Madrid



**Escuela Técnica Superior de  
Ingenieros Informáticos**

Doble grado en Ingeniería Informática y ADE

Trabajo Fin de Grado

**Análisis de Series Temporales: estudio de  
casos prácticos en Economía**

Autor: José María Martínez de la Casa Rico

Tutora: María Isabel Rodríguez Galiano

Madrid, junio 2024

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

*Trabajo Fin de Grado*

*Doble grado en Ingeniería Informática y ADE*

*Título: Análisis de Series Temporales: Estudio de Casos Prácticos en Economía*  
Junio 2024

*Autor:* José María Martínez de la Casa Rico

*Tutor:*

María Isabel Rodríguez Galiano  
Departamento de Inteligencia Artificial (DIA)  
ETSI Informáticos  
Universidad Politécnica de Madrid

# Resumen

En este trabajo se llevará a cabo un análisis exhaustivo de una serie temporal utilizando diversos modelos con el objetivo de ajustar los datos y realizar predicciones futuras. Se emplearán varias técnicas para comparar su desempeño y determinar qué modelo se adapta mejor a nuestra serie, facilitando así pronósticos más precisos. Comenzaremos con el estudio teórico de enfoques clásicos como el de Holt-Winters, y avanzaremos hacia métodos más recientes como ARIMA, los modelos ARCH-GARCH, que permiten modelizar la varianza condicional y analizar la dispersión de los valores de la serie y, finalmente, el modelo Prophet (desarrollado por Facebook).

La serie temporal que analizaremos es el Índice de Precios al Consumo (IPC) en España desde enero de 2002 hasta mayo de 2024. Estudiaremos el comportamiento de los valores a lo largo del tiempo y los ajustaremos a los métodos mencionados anteriormente. Para facilitar este análisis utilizaremos una interfaz desarrollada en el trabajo sinérgico de fin de grado de Ingeniería Informática, que permite realizar estudios de series temporales de manera sencilla e intuitiva. Esta interfaz ha sido creada utilizando las bibliotecas disponibles para el análisis de series temporales en Python.

De esta manera, esperamos identificar el modelo que mejor se ajuste a la serie del IPC y realizar predicciones precisas sobre su evolución futura.

# Abstract

In this paper, a comprehensive analysis of a time series will be carried out using various models with the aim of fitting the data and making future forecasts. Several techniques will be employed to compare their performance and determine which model best fits our series, thus providing more accurate forecasts. We will start with the theoretical study of classical approaches such as Holt-Winters, and move towards more recent methods such as ARIMA, ARCH-GARCH models, which allow modelling the conditional variance and analysing the dispersion of the values of the series and, finally, Prophet (developed by Facebook).

The time series we will analyse is the Consumer Price Index (CPI) in Spain from January 2002 to May 2024. We will study the behaviour of the values over time and adjust them to the methods mentioned above. To facilitate this analysis, we will use an interface developed in the synergic work of the Computer Engineering degree final project, which allows us to carry out time series studies in a simple and intuitive way. This interface has been created using the libraries available for time series analysis in Python.

In this way, we hope to identify the model that best fits the CPI series and make accurate predictions about its future evolution.

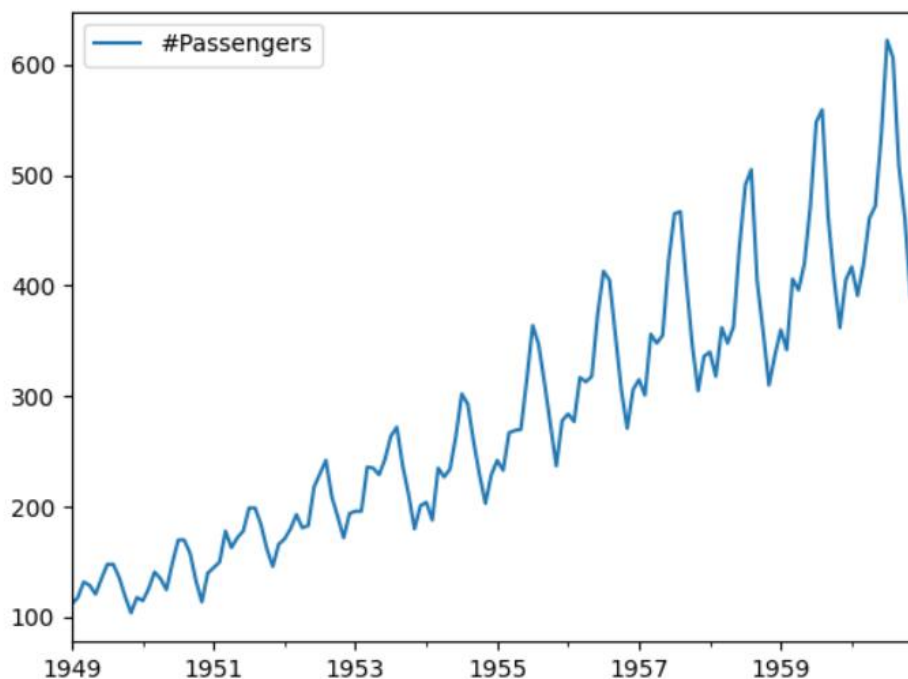
# Tabla de contenidos

<b>1</b>	<b>Introducción</b>	<b>1</b>
1.1	Objetivos	2
<b>2</b>	<b>Glosario</b>	<b>3</b>
<b>3</b>	<b>Modelos</b>	<b>4</b>
3.1	Métodos de descomposición	4
3.2	Método de Suavizado Exponencial	5
3.2.1	Suavizado exponencial simple	6
3.2.2	Suavizado exponencial de Holt	6
3.2.3	Modelo de Holt-Winters	7
3.3	Modelo ARIMA	8
3.3.1	Modelo AR	8
3.3.2	Modelo MA	12
3.3.3	Modelo ARMA	14
3.3.4	Proceso Integrado	14
3.3.5	ARIMA	15
3.4	Modelo SARIMA	16
3.5	Ajuste del modelo ARIMA	17
3.5.1	Identificación del modelo	18
3.5.2	Estimación de los parámetros	20
3.5.3	Validación del modelo	21
3.5.4	Predicción	22
3.6	Modelo ARCH	22
3.7	Modelo GARCH	23
3.8	Modelo Prophet	24
<b>4</b>	<b>Estudio de la serie temporal</b>	<b>25</b>
4.1	Explicación de la serie temporal	25
4.2	Estudio mediante el modelo Holt-Winters	26
4.3	Estudio mediante el modelo ARIMA	29
4.3.1	Identificación	29
4.3.2	Estimación de los parámetros	31
4.3.3	Validación del modelo	33
4.3.4	Predicción futura	37
4.4	Estudio mediante el modelo ARCH-GARCH	37
4.4.1	Análisis de heterocedasticidad	37
4.4.2	Ajuste del modelo	40
4.4.3	Predicciones	41
4.5	Estudio mediante el modelo Prophet	44

<b>5</b>	<b>Resultados y conclusiones .....</b>	<b>47</b>
<b>6</b>	<b>Bibliografía .....</b>	<b>48</b>

# 1 Introducción

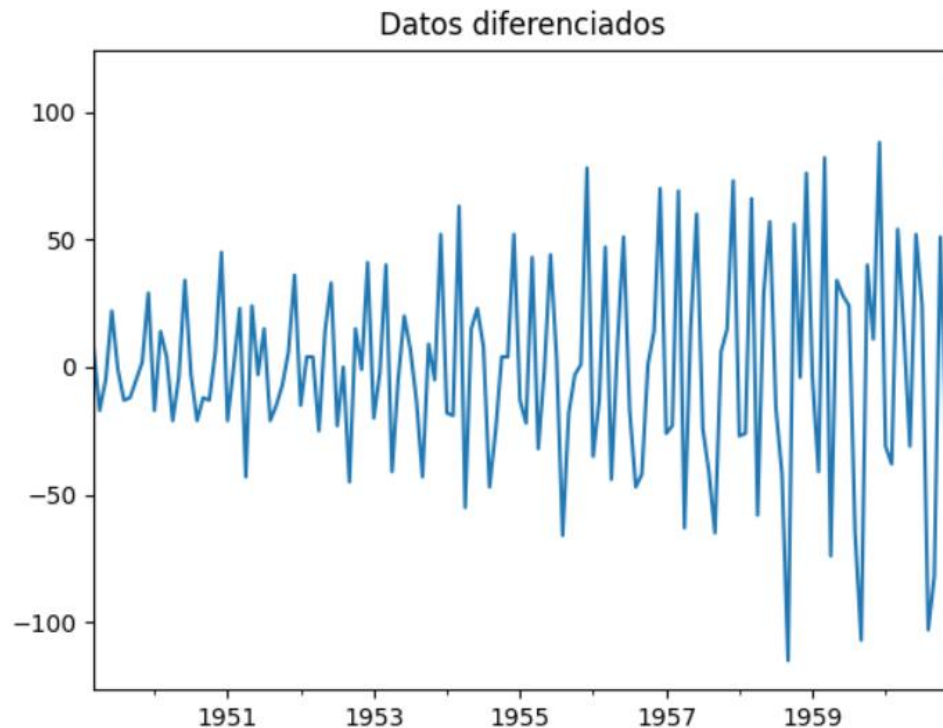
El análisis de las series temporales es fundamental en el campo de la estadística y la ciencia de datos. Su finalidad es la de modelar el comportamiento de estas series a lo largo del tiempo. Las series temporales son un conjunto de observaciones de una variable  $Y_t$  grabadas en un momento específico  $t$  [1]. Como ejemplo de serie temporal podemos observar la representada en la Figura 1.1, que muestra el número de pasajeros de aviones a lo largo de los años en Estados Unidos. Observando la evolución de la serie temporal se pueden analizar las distintas propiedades que posee una serie temporal.



*Figura 1.1: Viajeros en vuelos americanos de 1949 a 1961.*

En esta serie se observan dos propiedades muy claras: estacionalidad y tendencia. La estacionalidad hace referencia a patrones que se repiten en intervalos regulares de tiempo (horas, días, semanas, meses, trimestres, etc.). En este caso se observa una clara estacionalidad debido a que habitualmente los viajes en avión aumentan en verano y disminuyen en invierno. Por otra parte, la tendencia se puede entender como la dirección en la cual los datos cambian a lo largo del tiempo [2]. En esta serie se observa que la dirección de los datos ha ido en aumento, por lo que la tendencia es alcista. Otra propiedad de las

series temporales es la estacionariedad, la cual no es observable en la Figura 1.1, pero sí en la serie mostrada en la Figura 1.2.



*Figura 1.2: Datos de la Figura 1.1 diferenciados*

Los datos de la Figura 1.2 se corresponden con la diferenciación de la serie mostrada en la Figura 1.1. El proceso de diferenciación será analizado más adelante y se utiliza para conseguir que una serie sea estacionaria, es decir, con la media y la varianza constante a lo largo del tiempo.

El análisis de las series temporales se llevará a cabo a través de los diferentes modelos que se han estudiado a lo largo del tiempo: Desde los métodos clásicos, hasta el modelo Prophet ofrecido por Facebook [3], pasando por los modelos más populares como el modelo ARIMA o los modelos ARCH y GARCH. Finalmente, realizaremos predicciones sobre estas series temporales, estimando la evolución de los datos de la serie a lo largo del tiempo [4].

## **1.1 Objetivos**

Los objetivos de este trabajo son:

- **Comprensión y análisis de las series temporales:** Se busca entender la naturaleza de las series temporales, incluyendo la identificación de sus componentes clave como tendencia, estacionalidad y ciclos. Esto se logra mediante técnicas de visualización, como gráficos de líneas y

correlogramas, y pruebas estadísticas para la estacionariedad, como la prueba de Dickey-Fuller aumentada.

- Análisis de modelos clásicos: Aquí se exploran desde métodos de descomposición y modelo de Holt Winters hasta modelos como el de medias móviles (MA), autorregresivo (AR), y el modelo autorregresivo de promedio móvil (ARMA).
- Análisis de modelos ARIMA, ARCH o GARCH: El modelo ARIMA (Autorregresivo Integrado de Promedio Móvil) se utiliza para datos no estacionarios al incorporar diferenciación. Los modelos ARCH (Autorregresivo de Heterocedasticidad Condicional) y GARCH (Generalizado ARCH) se utilizan para modelizar la varianza condicional en los datos.
- Análisis de modelos Prophet: El modelo Prophet, desarrollado por Facebook, es adecuado para datos con fuertes componentes de tendencia y estacionalidad.
- Ajuste de los modelos y predicciones: En esta etapa, los modelos seleccionados se ajustan a los datos preparados. Esto incluye dividir los datos en conjuntos de entrenamiento y prueba, ajustar los parámetros del modelo usando el conjunto de entrenamiento, y evaluar el desempeño del modelo en el conjunto de prueba. Se realizan predicciones y se comparan los modelos utilizando métricas como el error cuadrático medio (RSME) y las bandas de confianza, para evaluar la precisión y la incertidumbre de las predicciones. Este ajuste se realizará en una serie temporal concreta, en este caso el Índice de Precios al Consumo (IPC).

## 2 Glosario

El siguiente capítulo explicará conceptos a conocer para que resulte más fácil la comprensión del trabajo:

- Serie temporal: Una serie temporal es una secuencia de datos observados, medidos o registrados en momentos sucesivos a lo largo del tiempo. Estos datos se recopilan generalmente en intervalos regulares, como horas, días, semanas, meses, trimestres o años. La serie temporal se organiza cronológicamente reflejando cómo evolucionan los datos a lo largo del tiempo.

- **Tendencia:** La tendencia en una serie temporal se refiere a la dirección general en la que los datos están cambiando a lo largo del tiempo. Es una característica importante que indica si los valores de la serie están aumentando, disminuyendo o permaneciendo relativamente estables en el tiempo.
- **Estacionalidad:** La estacionalidad se refiere a patrones recurrentes o cíclicos en los datos que se repiten a intervalos regulares. Estos pueden ser diarios, semanales, mensuales, estacionales, etc.
- **Ciclo:** Fluctuación que se repite a lo largo del tiempo, pero que no tiene una periodicidad fija.
- **Componente de error:** La componente de error es la parte no explicada por la tendencia, la estacionalidad o los ciclos en una serie temporal. Representa la aleatoriedad o la variabilidad no sistemática en los datos, lo que también se conoce como ruido blanco.
- **Proceso estocástico:** Secuencia de variables aleatorias que evolucionan a lo largo del tiempo.
- **Heterocedástico:** Varianza de los errores no constante a lo largo del tiempo.
- **Ruido blanco:** Secuencia de valores aleatorios independientes con media cero y varianza constante a lo largo del tiempo con una distribución normal.
- **RMSE:** El error cuadrático medio es una métrica comúnmente utilizada para evaluar la precisión de un modelo de predicción. Proporciona una medida de la diferencia entre los valores predichos por un modelo y los valores observados (reales).

## **3 Modelos**

### **3.1 Métodos de descomposición**

Al analizar las series temporales podemos encontrar diferentes patrones, estos patrones se pueden ver reflejados en las componentes que pueden conformar una serie temporal [5], estas componentes son:

- Tendencia
- Estacionalidad

- Cíclica
- Residual

Estos métodos intentan construir un modelo en función de las anteriores componentes. Las componentes de tendencia y ciclo se suelen unir en tendencia-cíclica. Se encontrarán diferentes modelos de descomposición dependiendo de la manera en la cual relacionemos las componentes de las series temporales.

- Modelo aditivo:  $Y_t = S_t + T_t + E_t$
- Modelo multiplicativo:  $Y_t = S_t \times T_t \times E_t$
- Modelo mixto:  $Y_t = S_t T_t + E_t$

Donde  $S_t$  es la componente estacional,  $E_t$  es la componente residual y por último  $T_t$  es la componente de tendencia-cíclica en el periodo  $t$ . Se supone independencia de las variaciones residuales respecto a las demás componentes.

El modelo aditivo es el más adecuado si la amplitud de las fluctuaciones estacionales se mantiene constante durante toda la serie temporal. Por el contrario, cuando esta amplitud no es constante y parece estar directamente relacionado con el nivel de la serie temporal, encontramos más adecuado utilizar el modelo multiplicativo. En el caso de no poder categorizar los modelos como aditivos o multiplicativos se utiliza el modelo mixto al ser una combinación de ambos modelos.

Dentro de estos métodos de descomposición, nos encontramos el suavizado de medias móviles que nos ayuda a estimar la tendencia y los ciclos, la fórmula es la siguiente:

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j}$$

En este caso  $m = 2k + 1$ , por lo que la estimación de la tendencia y del ciclo en  $t$ , se calcula mediante la media de los valores de la serie en  $k$  periodos de  $t$  (tiempo hacia ambos lados). Lo que se consigue es eliminar la aleatoriedad de los datos obteniendo una componente de tendencia-cíclica suave.

### 3.2 Método de Suavizado Exponencial

Estos métodos son posiblemente los más utilizados para suavizar las series temporales y conseguir predecir el futuro a un corto o medio plazo. Son métodos

sencillos, precisos y eficientes y permiten adaptarse a los cambios que ocurren en el proceso que se está realizando, de ahí su gran popularidad. En el método de suavizado exponencial la previsión se construye a partir de una media ponderada exponencialmente de los valores pasados [6]. El método de suavizado está elaborado para poder ilustrar las características de una serie temporal, como son la tendencia o la estacionalidad, de manera más sencilla y visual, permitiendo identificar patrones o tendencias significativas [7].

### 3.2.1 Suavizado exponencial simple

Este método se ajusta a las series temporales que no presentan ni tendencia ni estacionalidad. El concepto detrás del suavizado exponencial simple se basa en darle más peso a las observaciones más recientes que a las más antiguas. Es un método sencillo cuya fórmula es la siguiente:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots,$$

Siendo  $\alpha$  el parámetro de suavizado para  $0 \leq \alpha \leq 1$ .  $T + 1$  es una media ponderada de todas las observaciones que ha habido en nuestra serie temporal, desde  $y_1$  hasta  $y_T$ . Dependiendo de cuál sea el parámetro  $\alpha$  los pesos de cada observación pasada serán mayores o menores. Se puede observar a partir de la anterior ecuación que si  $\alpha$  tiene un valor pequeño, se les da un mayor peso a las observaciones más lejanas en el tiempo. Al contrario, si  $\alpha$  tiene un valor mayor y cercano a 1, se les da un mayor peso a las observaciones más recientes de la variable, siendo el caso extremo el método conocido como naïve en el cual si sustituimos  $\alpha$  por 1 se puede ver como el valor de la predicción es igual al último valor observado de la serie.

### 3.2.2 Suavizado exponencial de Holt

El suavizado exponencial de Holt nos permite predecir series temporales con una tendencia. En este caso, el modelo está formado por tres ecuaciones: una para la predicción y dos ecuaciones de suavizado. Las dos ecuaciones de suavizado se dividen a su vez en una para el nivel y otra para la tendencia, refiriéndonos al nivel como estimación del valor medio sobre el cual la serie fluctúa en un periodo de tiempo determinado [8].

- Ecuación de predicción:  $\hat{y}_{t+h|t} = l_t + hb_t$
- Ecuación de nivel:  $l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1})$
- Ecuación de tendencia:  $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*) b_{t-1}$

$\epsilon_t$  es una estimación del nivel de la serie en el momento  $t$ ,  $b_t$  hace referencia a la tendencia en el momento  $t$ ,  $\alpha$  es el parámetro de suavizado para el nivel con  $0 \leq \alpha \leq 1$ . Por último,  $\beta^*$  es el parámetro de suavizado para la tendencia donde  $0 \leq \beta^* \leq 1$ .

### 3.2.3 Modelo de Holt-Winters

El modelo de Holt-Winters es una extensión del modelo de Holt que permite además de la tendencia, modelizar la estacionalidad, por lo que aparte de las tres ecuaciones que teníamos en el método de Holt, se añade una más que refleja la componente estacional. En este caso los parámetros de suavizado serán  $\alpha$ ,  $\beta$ ,  $\gamma$ . Debemos de establecer la periodicidad en la variable  $m$ , por ejemplo, si la estacionalidad es cuatrimestral  $m = 4$  y en el caso de estacionalidad mensual,  $m = 12$ .

En este caso, al igual que en los métodos de descomposición, encontramos dos variaciones del modelo dependiendo de la componente estacional: el modelo aditivo y el multiplicativo. El modelo aditivo es más adecuado cuando las variaciones de la componente estacional son aproximadamente constantes a lo largo de la serie temporal, mientras que el modelo multiplicativo se elige en los casos en los cuales la componente estacional varía en proporción al nivel de la serie.

#### Modelo aditivo

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t-m+h_m^+}$$

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*) b_{t-1}$$

$$s_t = \gamma(y_t - l - b_{t-1}) + (1 - \gamma) s_{t-m}$$

Donde  $h_m^+ = [(h - 1) \text{ mod } m] + 1$ <sup>1</sup> que asegura que las estimaciones de los índices estacionales que se emplean en las previsiones provienen exclusivamente de los datos del año más reciente dentro de nuestra muestra. La ecuación de nivel muestra una media ponderada entre la observación estacionalizada ( $y_t - s_{t-m}$ ) y la previsión no estacional ( $l_{t-1} + b_{t-1}$ ).

#### Modelo multiplicativo

---

<sup>1</sup> Donde  $[x]$  indica la parte entera de  $x$

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t-m+h_m^+}$$

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*) b_{t-1}$$

$$s_t = \gamma \frac{y_t}{(l_{t-1} - b_{t-1})} + (1 - \gamma) s_{t-m}$$

### 3.3 Modelo ARIMA

#### 3.3.1 Modelo AR

El Modelo Autorregresivo es un tipo de modelo estadístico, que utiliza las observaciones pasadas de una variable de una serie temporal para explicar su valor actual [4]. La fórmula del modelo es la siguiente:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + e_t$$

Nos referimos a estos modelos como AR ( $p$ ) siendo  $p$  el número de periodos hacia atrás que son necesarios para pronosticar el valor actual, donde  $y_t$  es la observación en el tiempo  $t$ ,  $c$  es una constante,  $\varphi_i$ ,  $i = 1, \dots, p$ , son los parámetros del modelo,  $y_{t-1}$ ,  $y_{t-2}$ ,  $\dots$ ,  $y_{t-p}$  son las observaciones previas de la serie temporal y  $e_t$  es un término de error que debe ser ruido blanco.

Un modelo autorregresivo será estable si los parámetros se encuentran en ciertos rangos y varía según  $p$ . Por ejemplo, en el caso de  $p = 1$ ,  $-1 < \varphi_1 < 1$ , en el caso de  $p = 2$ ,  $-1 < \varphi_2 < 1$ ,  $\varphi_1 + \varphi_2 < 1$ ,  $\varphi_2 - \varphi_1 < 1$ . En el caso de que no se cumplieran estas restricciones la serie no sería estacionaria, debido a que los valores de los parámetros al ser mayores que |1| provocarían que los valores de las variables crecieran en todo momento.

Se puede simplificar la ecuación utilizando el operador de retardo  $B$ , representado por la fórmula:

$$B y_t = y_{t-1}$$

Esta expresión implica retrasar el valor de  $y_t$  un tiempo. Al aplicar esta fórmula, nuestro modelo autorregresivo se reescribe de la siguiente manera:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p) y_t = e_t$$

Para simplificar aún más, llamamos a  $(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)$  como  $\varphi(B)$ , lo cual nos permite expresar el modelo de manera más reducida:

$$\varphi(B) y_t = e_t$$

La ecuación característica del modelo AR deriva del polinomio característico del modelo. Esta ecuación se utiliza para determinar las raíces del polinomio, que son fundamentales para analizar la estacionariedad del modelo.

El polinomio es:

$$\Phi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \dots - \varphi_p z^p$$

La ecuación característica se obtiene al igualar este polinomio a 0.

$$\Phi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \dots - \varphi_p z^p = 0$$

En términos prácticos esto significa que se buscan las raíces del polinomio  $\Phi(z)$  que hacen que la ecuación sea cero. La estacionariedad de nuestro modelo autorregresivo está vinculada a esta ecuación característica. Se dice que el modelo es estacionario si las raíces del polinomio  $\Phi(z)$  caen fuera del círculo unitario, es decir, verifican  $|z| > 1$ . Esta condición asegura que las propiedades estadísticas del modelo permanecen constantes a lo largo del tiempo, contribuyendo así a la utilidad y estabilidad del modelo autorregresivo.

Si se tiene un modelo AR de orden 1 y estable,  $y_t = \varphi y_{t-1} + e_t$  e iteramos hacia atrás  $k$  veces:

$$y_t = \varphi y_{t-1} + e_t = \varphi(\varphi y_{t-2} + e_{t-1}) + e_t = \varphi^2 y_{t-2} + \varphi e_{t-1} + e_t$$

⋮

$$y_t = \varphi^k y_{t-k} + \sum_{j=0}^{k-1} \varphi^j e_{t-j}$$

De esta manera si continuamos la iteración hacia atrás, asegurando que  $|\varphi| < 1$  e  $y_t$  sea estacionaria, el modelo  $AR(1)$  se puede representar como  $MA(\infty)$ :

$$y_t = \sum_{j=0}^{\infty} \varphi^j e_{t-j}$$

El proceso es estacionario con media:

$$E(y_t) = \sum_{j=0}^{\infty} \varphi^j E(e_{t-j}) = 0$$

Podemos calcular las funciones de autocovarianza y autocorrelación del  $AR(1)$  mediante el procedimiento Yule-Walker, el cual consiste en multiplicar a toda la ecuación por  $y_{t-k}$  y tomar valor esperado. De esta forma se consiguen las autocovarianzas:

$$y_t = \varphi_1 y_{t-1} + e_t$$

$$y_t y_{t-k} = \varphi_1 y_{t-1} y_{t-k} + e_t y_{t-k}$$

$$E(y_t y_{t-k}) = \varphi_1 E(y_{t-1} y_{t-k}) + E(e_t y_{t-k})$$

Para simplificar llamamos  $R(k)$  a  $E(y_t y_{t-k})$  que es la autocovarianza entre  $y_t y_{t-k}$ . En otras palabras, estamos calculando el valor esperado del producto de las observaciones  $y_t y_{t-k}$ , lo cual da una medida de la covariación entre estas dos observaciones separadas por un retraso de  $k$  periodos. El término de error será 0 al ser ruido blanco por lo que no contribuirá a la autocovarianza:

$$R(k) = \varphi_1 R(k-1) + 0$$

Tenemos una ecuación de diferencias de primer orden y se sustituye recursivamente

$$R(k-1) = \varphi_1 R(k-2)$$

$$R(k) = \varphi_1 \varphi_1 R(k-2)$$

$$R(k-2) = \varphi_1 R(k-3)$$

$$R(k) = \varphi_1 \varphi_1 \varphi_1 R(k-3)$$

⋮

$$R(k) = \varphi_1^k R(k-k) = \varphi_1^k R(0)$$

De esta forma se consigue la función de autocovarianza de un AR (1), al ser  $|\varphi_1| < 1$  la autocovarianza irá decayendo lentamente hasta 0. En base a esto podemos calcular la función de autocorrelación.

$$r(k) = \frac{R(k)}{R(0)} = \frac{\varphi_1^k R(0)}{R(0)} = \varphi_1^k$$

Por lo mismo que se acaba de explicar anteriormente, la función de autocorrelación decaerá lentamente hacia 0. Determinar el orden de un AR es difícil mediante la función de autocorrelación, por lo que nos servimos de la función de autocorrelación parcial (fap), la cual es una herramienta útil para determinar el orden de un proceso autorregresivo (AR). En un proceso  $AR(p)$ , cada coeficiente de autocorrelación parcial en la FAP corresponde a la correlación directa entre dos observaciones, controlando el efecto de todas las observaciones intermedias.

Para un proceso  $AR(p)$ , los primeros  $p$  coeficientes de autocorrelación parcial son distintos de cero, lo que indica que hay correlación directa entre la observación actual y las  $p$  observaciones pasadas. Después del  $p$ -ésimo coeficiente, los coeficientes de autocorrelación parcial son nulos, lo que significa que no hay correlación directa entre la observación actual y las observaciones pasadas más allá de  $p$  periodos.

Por lo tanto, al observar la fap de un proceso AR, podemos identificar el orden del proceso AR observando cuántos coeficientes de autocorrelación parcial son distintos de cero. Por ejemplo, si los primeros dos coeficientes de autocorrelación parcial son distintos de cero y los demás son nulos, se puede concluir que el proceso es un  $AR(2)$ .

En resumen, la fap nos permite identificar el orden de un proceso AR al observar cuántos coeficientes de autocorrelación parcial son distintos de cero. Esto nos ayuda a entender la estructura de dependencia temporal en los datos y a seleccionar el orden apropiado del modelo autorregresivo. A continuación, observamos ejemplos de diferentes modelos autorregresivos para comprobar lo anteriormente explicado (ver Figura 3.1 y 3.2) [9].

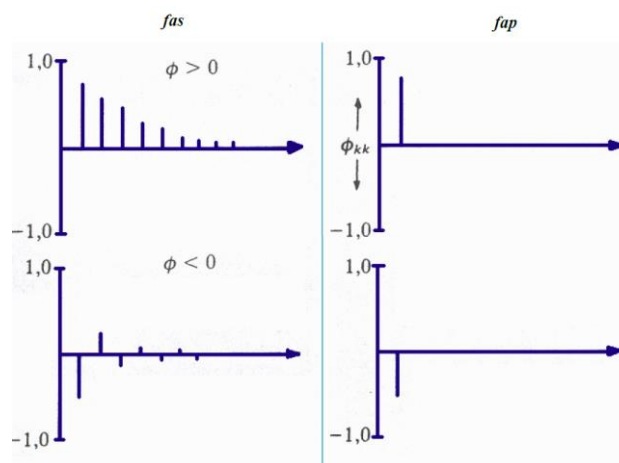


Figura 3.1: Funciones de autocorrelación de un  $AR(1)$

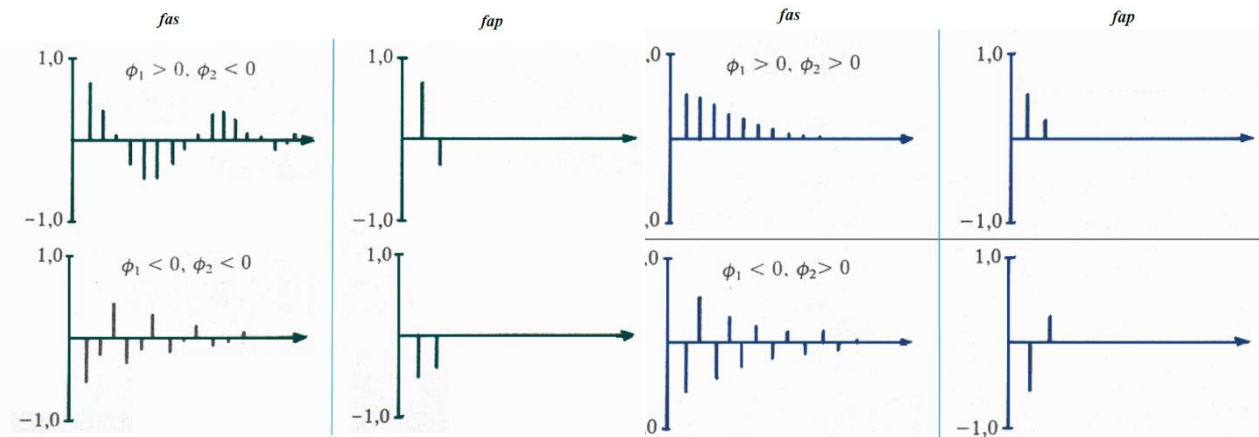


Figura 3.2: Funciones de autocorrelación de un AR(2)

### 3.3.2 Modelo MA

El modelo de Medias Móviles, a diferencia del modelo autorregresivo, utiliza los errores residuales de las anteriores predicciones para hallar el valor de la variable actual. Su fórmula es la siguiente:

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

Nos referimos a estos modelos como MA(q), donde  $y_t$  es la observación en el tiempo  $t$ ,  $c$  es una constante,  $\theta_i$ ,  $i = 1, \dots, q$ , son los parámetros del modelo,  $e_{t-1}, e_{t-2}, \dots, e_{t-q}$  son las observaciones previas de la serie temporal y  $e_t$  es un término de error que debe ser ruido blanco. Si nos damos cuenta, el modelo siempre será estacionario al estar formado por  $q$  procesos estacionarios  $e_{t-1}, e_{t-2}, \dots, e_{t-q}$ , cosa que no ocurre en los modelos autorregresivos. Utilizando el operador de retardo, la ecuación del modelo queda de la siguiente forma:

$$y_t = c + e_t(1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$$

Los modelos MA también tienen una ecuación característica, aunque su propósito y forma son diferentes de las del modelo autorregresivo (AR). La ecuación característica de un modelo MA es útil para analizar la invertibilidad del modelo, que es un concepto análogo a la estabilidad en los modelos AR. La ecuación característica deriva del siguiente polinomio:

$$\Theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_p z^p$$

Para analizar la invertibilidad del modelo MA, se iguala este polinomio a cero y resolvemos para  $z$ :

$$\Theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_p z^p = 0$$

Un modelo MA es invertible si todas las raíces del polinomio característico  $\Theta(z)$  están fuera del círculo unitario en el plano complejo, es decir, si todas las raíces tienen un módulo mayor que uno,  $|z| > 1$ . La invertibilidad de un modelo MA es la propiedad que permite representar el modelo MA en términos de un modelo autorregresivo (AR) infinito, lo cual indica que hay una sola forma de expresar el modelo MA como un modelo autorregresivo (AR) infinito. Esto evita ambigüedades y asegura que el modelo ajustado describe de manera consistente la dependencia temporal en los datos sin que existan múltiples conjuntos de parámetros diferentes que expliquen la misma serie temporal.

Al igual que hicimos con el modelo autorregresivo, se calculará la autocovarianza y la autocorrelación mediante el método de Yule-Walker [9] para un modelo MA(1).

$$\begin{aligned} E(y_t y_{t-k}) &= E(e_t y_{t-k}) + \theta_1 E(e_{t-1} y_{t-k}) \\ R(k) &= E(e_t(e_{t-k} + \theta_1 e_{t-1-k})) + \theta_1 E(e_{t-1}(e_{t-k} + \theta_1 e_{t-1-k})) \\ &= E(e_t e_{t-k}) + \theta_1 E(e_t e_{t-1-k}) + \theta_1 E(e_{t-1} e_{t-k}) + \theta_1 E(e_{t-1} e_{t-1}) \end{aligned}$$

Sustituiremos  $k$  por diferentes valores, comenzamos por  $k = 0$

$$R(0) = \sigma_e^2 + 0 + 0 + \theta_1^2 \sigma_e^2 = \sigma_e^2 (1 + \theta_1^2)$$

Para  $k = 1$

$$R(1) = 0 + 0 + \theta_1 \sigma_e^2 + 0$$

Para  $k = 2$  o  $k > 1$

$$R(k) = 0$$

La función de autocorrelación sería

$$\begin{aligned} r(k) &= \frac{R(k)}{R(0)} = \frac{R(k)}{\sigma_e^2 (1 + \theta_1^2)} \\ r(1) &= \frac{\theta_1}{(1 + \theta_1^2)} = \frac{R(k)}{\sigma_e^2 (1 + \theta_1^2)} \\ r(2) &= 0 \\ &\vdots \\ r(k) &= 0 \end{aligned}$$

De esta manera vemos como la función de autocorrelación decae a 0 abruptamente después del primer retardo. Para observar lo anteriormente

explicado, se verá a continuación ejemplos de modelos de medias móviles y sus correspondientes funciones de autocorrelación (ver Figura 3.3 y Figura 3.4) [9].

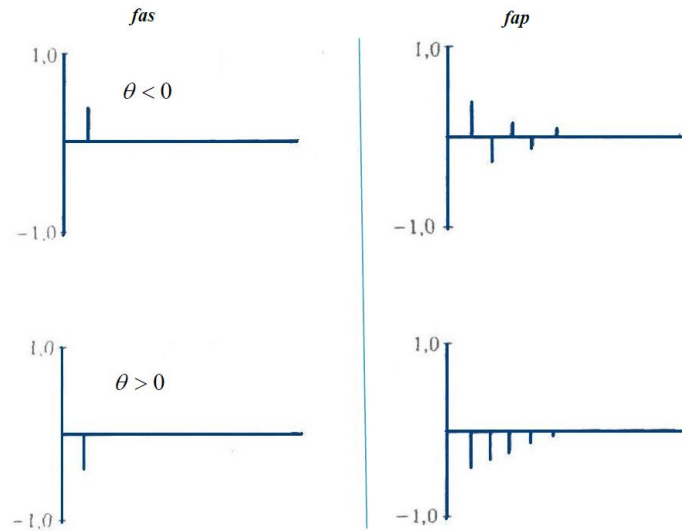


Figura 3.3: Funciones de autocorrelación de un MA(1)

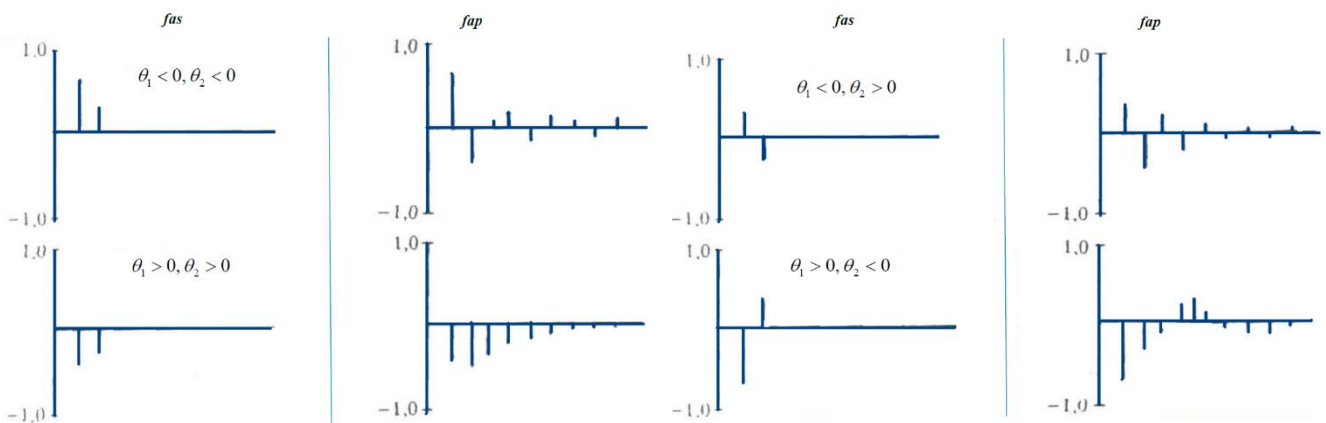


Figura 3.4: Funciones de autocorrelación de un MA(2)

### 3.3.3 Modelo ARMA

El modelo autorregresivo de medias móviles, es la combinación de los dos modelos vistos anteriormente. La combinación del modelo es la siguiente:

$$\varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + e_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

El proceso será estacionario en el caso de que lo sea la componente autorregresiva del modelo y será invertible si lo es la parte de medias móviles.

### 3.3.4 Proceso Integrado

Si la serie no es estacionaria, se diferencia la serie hasta conseguir que sea estacionaria. La diferenciación implica calcular las diferencias entre los valores

sucesivos de la serie. Por ejemplo, si se tiene una serie de tiempo  $y_t$ , la diferencia de primer orden sería  $\Delta y_t = y_t - y_{t-1}$ . Esta operación se repite hasta que la serie resultante se vuelve estacionaria.

Si la serie diferenciada  $d$  veces es estacionaria y se ajusta a un modelo ARMA( $p, q$ ), diremos que la serie original sigue un modelo ARIMA( $p, d, q$ ).

### 3.3.5 ARIMA

El modelo resulta de coordinar la diferenciación con el modelo autorregresivo y el modelo de medias móviles, siendo su fórmula la siguiente:

$$y'_t = c + \varphi_1 y'_{t-1} + \varphi_2 y'_{t-2} + \dots + \varphi_p y'_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

En este caso  $y'_t$  es la serie diferenciada, pudiendo haber sido diferenciada más de una vez, esto es lo que llamamos el modelo ARIMA( $p, d, q$ ), siendo  $p$  el orden de la componente autorregresiva,  $d$  el grado de diferenciación necesario y  $q$  el orden de la componente de media móvil. La ecuación se puede expresar de una forma más sencilla utilizando el operador de retardo:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)(1 - B)^d y_t = c + e_t(1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$$

$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)$  se corresponde con la parte AR( $p$ ),  $(1 - B)^d y_t$  son las diferencias y  $c + e_t(1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$  es la parte MA( $q$ ).

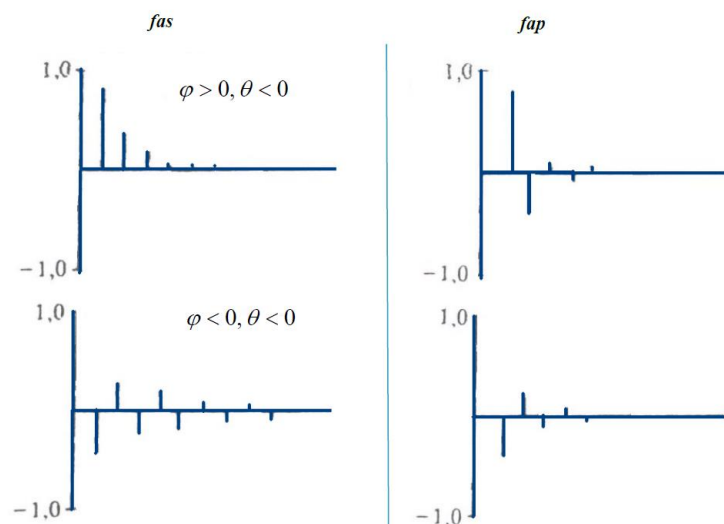


Figura 3.5: Funciones de autocorrelación de un ARMA(1,1)

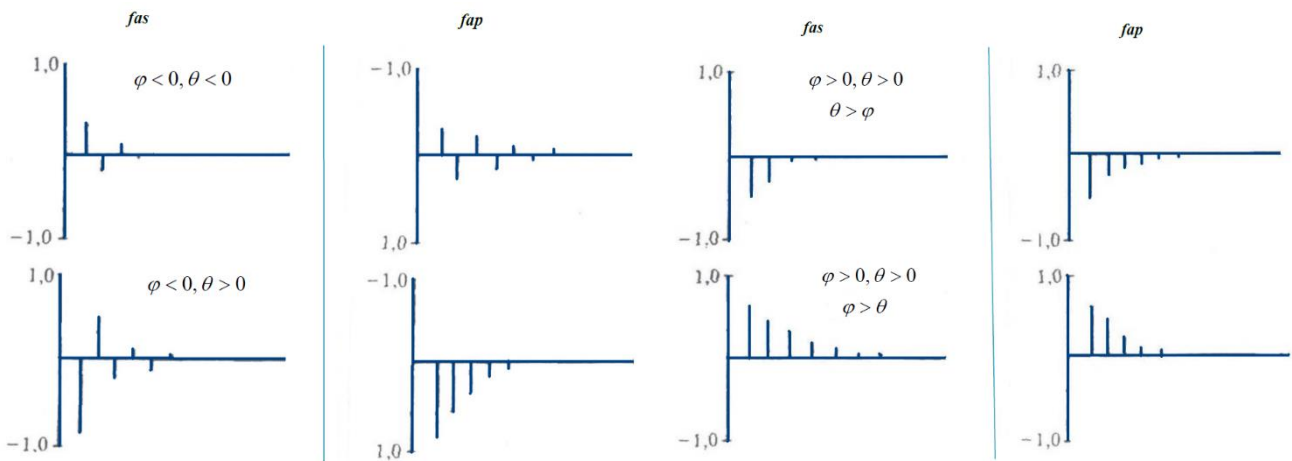


Figura 3.6: Funciones de autocorrelación de un ARMA(1,1)

	FAS	FAP
$AR(p)$	Decrecimiento rápido exponencial atenuado u ondas sinusoidales	$p$ primeros coeficientes distintos de 0 y el resto 0 (coef. 0 para retardo $> p$ )
$MA(q)$	$q$ primeros coeficientes distintos de 0 y el resto 0 (coef. 0 para retardo $> q$ )	Decrecimiento rápido exponencial atenuado u ondas sinusoidales
$ARMA(p,q)$	Decreciendo hacia 0	Decreciendo hacia 0

Figura 3.7: Resumen

### 3.4 Modelo SARIMA

Los modelos SARIMA o ARIMA estacionales, son modelos ARIMA que presentan estacionalidad, por lo cual introducimos a nuestro modelo ARIMA componentes estacionales:

$$ARIMA(p, d, q)(P, D, Q)_s$$

En este caso  $s$  representa la periodicidad de la serie de tiempo, por ejemplo, si se está trabajando con datos mensuales sería 12 porque el patrón estacional se repite cada 12 meses, o si estuviéramos trabajando por trimestres sería 4.

Un ejemplo de cómo quedaría un modelo  $ARIMA(1,1,1)(1,1,1)_{12}$  es:

$$(1 - \varphi_1 B)(\omega_1 B^{12})(1 - B)(1 - B^{12})y_t = (1 + \theta_1 B)(1 + \Theta_1^{12})e_t$$

La estacionalidad se puede observar en los gráficos de las funciones fas y fap. En el ejemplo que se acaba de realizar arriba, la componente estacional  $P = 1$  la observaríamos en la fap, concretamente en el retardo número 12 del gráfico donde la correlación estaría fuera del intervalo de confianza. En el caso del  $Q = 1$  sería en la fas donde en el retardo número 12, la correlación sobrepasaría el intervalo de confianza, de esta manera se muestra que hay autocorrelación entre el valor actual y el mismo hace 12 meses ( $s = 12$ ).

### **3.5 Ajuste del modelo ARIMA**

El ajuste de un Modelo Autorregresivo Integrado de Medias Móviles se realiza mediante el método de Box y Jenkins [11], el cual se estructura en cuatro etapas que se ayudarán de herramientas que se irán viendo a continuación.

1. Durante esta etapa identificamos el modelo ARIMA.
  - Lo primero es identificar si la serie es estacionaria, en el caso de que no lo sea realizamos como hemos visto anteriormente el proceso de diferenciación, determinando de esta manera el valor de  $d$ . En el caso de que la serie sea estacional, además debemos identificar  $D$ .
  - Una vez diferenciada, se determina  $p$  y  $q$  y en el caso de que la serie sea estacional  $P$  y  $Q$ .
2. En la segunda etapa, una vez que tenemos seleccionado un modelo para nuestra serie temporal estacionaria, pasamos a la estimación de los parámetros de los modelos AR y MA los cuales se estiman por los métodos de mínimos cuadrados y de máxima verosimilitud, obteniendo sus errores estándar y los residuos de nuestro modelo.
3. Llegamos a la fase de validación del modelo en la cual se debe de comprobar la calidad de las estimaciones realizadas en la etapa anterior. Se analizan los residuos comprobando que estos son ruido blanco, en el caso de que no lo fueran sería necesario repetir las etapas anteriores del proceso de ajuste para encontrar un modelo que sí que tengan residuos que sean ruido blanco.
4. Finalmente, estamos en la última etapa en la cual una vez que se ha conseguido un modelo que cumpla con los requisitos se pueden llevar a cabo las predicciones sobre este.

### 3.5.1 Identificación del modelo

El proceso de identificación se basa, como hemos visto, en identificar los parámetros del modelo  $ARIMA(p, d, q)(P, D, Q)_s$ . En primer lugar, se comienza estudiando el parámetro  $d$ . Si la serie es inestable como en la figura Figura 3.8 que presenta tendencia creciente, la serie no es estacionaria

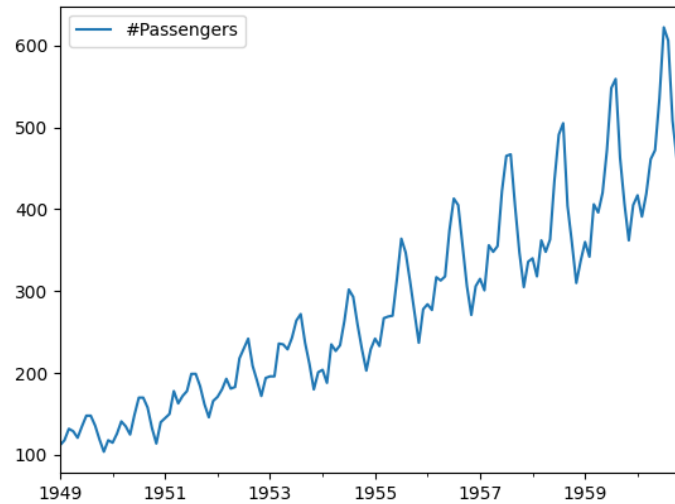


Figura 3.8: Viajeros en vuelos americanos de 1949 a 1961.

Que la serie no sea estacionaria también se puede observar mediante los siguientes métodos:

- Gráfico de la función de autocorrelación simple (fas). Si las correlaciones decaen a cero (ver Figura 3.9), la serie no es estacionaria y necesita diferenciación para poder serlo.

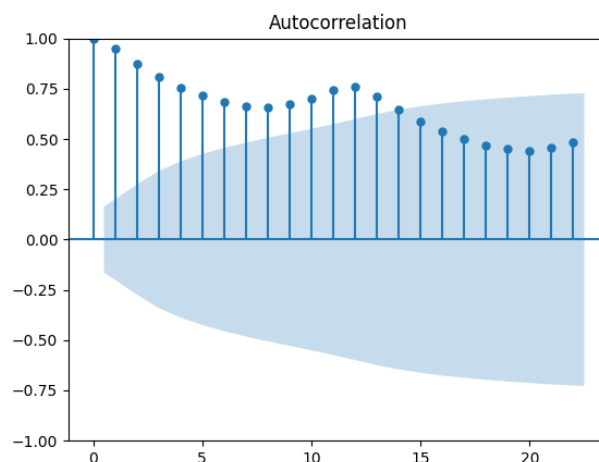


Figura 3.9: Gráfico de función de de autocorrelación de Figura 3.8

- Test de Dickey-Fuller o test de raíces unitarias: El test de Dickey-Fuller, nos permite observar la estabilidad de la serie a lo largo del tiempo.

Contrasta la hipótesis nula de que una serie temporal tiene una raíz unitaria, lo cual indica no estacionariedad. La presencia de una raíz unitaria sugiere que la serie temporal exhibe cierto grado de dependencia en el tiempo y no es estacionaria. Si el p-valor asociado al test es menor que un nivel de significación predefinido, que generalmente suele ser 0,05 entonces se rechaza la hipótesis nula. Esto significa que hay evidencia suficiente para afirmar que la serie es estacionaria. En cambio, si el p-valor es mayor que el nivel de significación, no se puede rechazar la hipótesis nula, lo que sugiere que la serie es no estacionaria, por lo que se procede a llevar a cabo la diferenciación de la serie [12].

El proceso de diferenciación se realiza mediante el método de transformación de Box-Cox. Las transformaciones propuestas por el método vienen definidas en la siguiente función.

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases}$$

Donde:

- $y$  son los datos originales.
- $y^{(\lambda)}$  son los datos transformados.
- $\lambda$  es el parámetro de transformación.

La Figura 3.10 muestra cómo queda la serie de la Figura 3.8 una vez diferenciada.

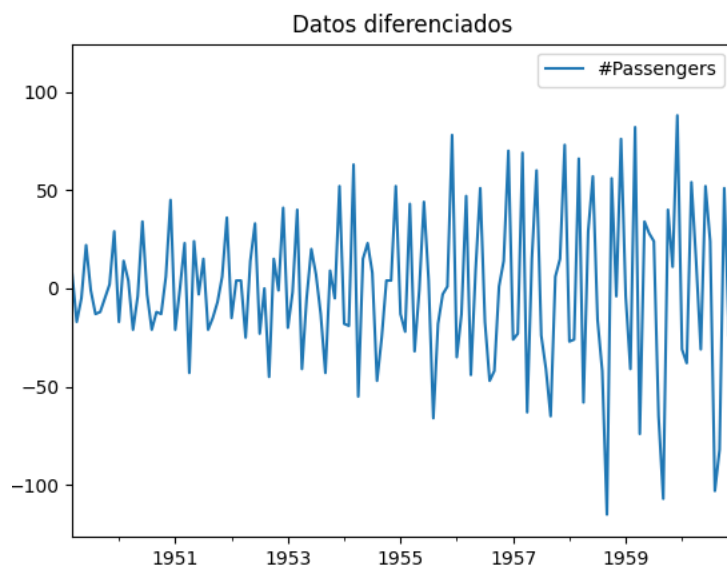


Figura 3.10: Serie de la Figura 3.8 diferenciada

También es importante analizar el gráfico de los valores de la serie temporal (ver Figura 3.8) para observar la estacionalidad, debido a que esta no se puede observar en los gráficos estacionarios. Se puede identificar fácilmente la estacionalidad al ver los ciclos repetitivos de cada año, dividiendo los datos en cuatrimestres, meses, semanas... También se pueden observar las funciones de autocorrelación y observar si hay correlación significativa en los múltiplos de  $s$ . Otro método es llevar a cabo la descomposición de la serie temporal en sus componentes y observar la componente de estacionalidad. Finalmente, si hay estacionalidad, se debe determinar también el orden de diferenciación estacional  $D$ . Esto se puede realizar con los mismos procesos explicados para la parte no estacional, pero aplicados a los ciclos estacionales.

Una vez ya hemos llevado a cabo la diferenciación, se procede a identificar  $p$  y  $q$  y  $P$  de la parte regular y  $Q$  en el caso de que la serie sea estacional. Para hallar estos parámetros lo realizaremos como se ha visto en la explicación del modelo AR y del modelo MA gracias a las funciones de autocorrelación simple y parcial.

### 3.5.2 Estimación de los parámetros

Al comenzar con la estimación de los parámetros se debe comprobar si es necesario incluir un término independiente, para ello se realiza un contraste paramétrico para ver si la media de la serie una vez diferenciada es significativamente distinto de 0. Si es así incluimos la constante. Después se comprobará si hay alguna correlación entre los estimadores debido a que en el

caso de que haya una correlación alta (superior a 0,8) nuestro modelo sería inestable y deberíamos comenzar el proceso de nuevo y buscar otro modelo.

También es conveniente realizar un sobreajuste del modelo que se ha identificado. Para llevar a cabo este proceso se suelen utilizar el Criterio de Información de Akaike (AIC) o el Criterio de Información Bayesiano (BIC). Si nuestro modelo es  $ARIMA(p, d, q)$  el sobreajuste lo realizaríamos con los modelos  $ARIMA(p + 1, d, q)$  y  $ARIMA(p, d, q + 1)$ .

### 3.5.3 Validación del modelo

Durante el diagnóstico del modelo debemos de comprobar si los residuos de nuestro modelo son ruido blanco. Hay diferentes métodos para llevar a cabo la validación del modelo seleccionado.

Para verificar si los residuos son ruido blanco podemos utilizar el método de contraste de Ljung-Box. Las hipótesis de este contraste son las siguientes:

- $H_0$  = Las correlaciones son 0 entre los residuos de la serie.
- $H_1$  = Hay correlación entre los residuos de la serie.

La medida de discrepancia es:

$$Q = N(N + 2) \sum_{k=1}^h \frac{\hat{r}_k^2}{N - k}$$

Siendo  $N$  el tamaño de la muestra,  $\hat{r}_k$  la autocorrelación en el retraso  $k$  y  $h$  el número de retardos. El estadístico  $Q$  sigue una distribución chi-cuadrado con  $h$  grados de libertad bajo la hipótesis nula de no autocorrelación. Entonces fijado un nivel de significación, se compara el valor de  $Q$  calculado con el valor crítico correspondiente de la distribución chi-cuadrado para determinar si se rechaza o no la hipótesis nula, si el valor de  $Q$  calculado es significativamente mayor que el valor crítico de la distribución chi-cuadrado, se rechaza la hipótesis nula, lo que indica la presencia de autocorrelación en los retardos seleccionados de la serie temporal.

También se puede dibujar tanto el gráfico de autocorrelación simple como parcial de los residuos, para comprobar si hay algún retardo fuera de los límites de confianza en cuyo caso los residuos estarían correlacionados. Por último, comprobaremos si nuestros residuos siguen una distribución normal.

### 3.5.4 Predicción

Antes de llevar a cabo una predicción a futuro, se comprueba si el modelo sirve para ello, esto se realiza haciendo una predicción sobre las últimas observaciones de la serie y comparando si los resultados son parecidos a los datos observados en la misma. Si los resultados tienen sentido, se puede comenzar con las predicciones de los valores no observados de la serie temporal.

### 3.6 Modelo ARCH

Los modelos ARCH (modelos de heterocedasticidad condicional autorregresiva), desarrollados por Robert Engle en 1980, son una clase de modelos econométricos utilizados para modelizar la volatilidad sobre todo en series temporales financieras y económicas. La principal diferencia de estos modelos con los modelos ARIMA radica principalmente, en que estos últimos modelos tienen la necesidad de que las series presenten una varianza constante en el tiempo. Esto no ocurre en el modelo ARCH que fue creado para analizar las series temporales donde la varianza condicional no es constante, si no que varía según los datos pasados, por lo que al aparecer nuevos datos nos encontraremos con una nueva varianza.

#### Varianza condicional

La varianza condicional de nuestro término de error se expresa como  $\sigma_t^2$

$$\sigma_t^2 = \text{var}(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-k}) = E[(\varepsilon_t - E(\varepsilon_t))^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-k}]$$

Pero como sabemos, la media del término de error es 0,  $E(\varepsilon_t) = 0$ , por lo que:

$$\sigma_t^2 = \text{var}(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-k}) = E[\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-k}]$$

Lo que nos indica esta ecuación es que la varianza condicional de una variable aleatoria es igual a la varianza condicional del residuo al cuadrado. De esta forma, el modelo *ARCH* ( $p$ ) [13] se especifica de la siguiente manera:

$$\varepsilon_t = z_t \sqrt{\sigma_t^2}$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$$

Siendo  $\varepsilon_t$  el error en el tiempo  $t$ . Este error sigue una distribución normal con media cero y varianza condicional  $\sigma_t^2$ ,  $z_t$  es un término de ruido blanco, generalmente se asume que  $z_t \sim N(0,1)$ , es decir, sigue una distribución normal

estándar con media cero y varianza uno,  $\sigma_t^2$  es la desviación estándar condicional del error  $\varepsilon_t$  en el tiempo  $t$ . En la segunda ecuación  $\sigma_t^2$  es la varianza condicional de la serie temporal en  $t$ ,  $\varepsilon_{t-i}$  son los residuos de la serie en tiempos anteriores,  $\alpha_i$  son los parámetros del modelo ARCH que indican el impacto de los residuos pasados en la volatilidad actual, y por último  $p$  es el orden del modelo, es decir, el número de retardos incluidos en nuestro modelo ARCH.

Debe cumplirse:

$$\alpha_0 > 0, \alpha_p \geq 0$$

$$\sum_{i=1}^p \alpha_i \leq 1$$

### 3.7 Modelo GARCH

El modelo GARCH (modelo de heterocedasticidad condicional autorregresiva generalizada) fue introducido por Tim Bollerslev en 1986 y es una extensión de los modelos ARCH. Se encarga de introducir en el modelo, además de los errores, varianzas condicionales retardadas. El modelo se especifica de la siguiente manera:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

Donde  $\sigma_t^2$  es la varianza condicional de la serie temporal en  $t$ ,  $\varepsilon_{t-i}$  son los residuos de la serie en tiempos anteriores,  $\alpha_i$  son los parámetros del modelo GARCH que indican el impacto de los residuos pasados en la volatilidad actual,  $\beta_j$  son los coeficientes GARCH que indican el impacto de las varianzas condicionadas pasadas en la volatilidad actual y por último tanto  $p$  como  $q$  son el orden del modelo, es decir el número de retardos incluidos en nuestro modelo GARCH.

Debe cumplirse:

$$\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0$$

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j \leq 1$$

### 3.8 Modelo Prophet

Prophet es una librería que ha sido desarrollada por Facebook, específicamente, fue creado por Sean J. Taylor y Ben Letham [14], utiliza el modelo de descomposición con cuatro componentes como se muestra en la siguiente ecuación:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t)$$

$g(t)$  es la tendencia que modeliza los cambios no periódicos en el valor de la variable de la serie temporal,  $s(t)$  es la componente estacional,  $h(t)$  representa los efectos de los días festivos los cuales ocurren en calendarios irregulares durante uno o más días. Finalmente,  $\varepsilon(t)$  es el término de error [14].

La componente de tendencia se fundamenta en las siguientes ecuaciones, si el crecimiento no es lineal, se le conoce como crecimiento logístico, cuya ecuación es:

$$g(t) = \frac{C}{1 + \exp(-k(t - m))}$$

Siendo  $C$  la capacidad de carga máxima que se puede soportar, es decir, el límite superior hacia el cual tiende el crecimiento,  $k$  la tasa de crecimiento o que tan rápida se alcanza la capacidad de carga y  $m$  el desplazamiento o el tiempo en el que se alcanza la mitad de la capacidad de carga.

En el caso de que el crecimiento sea lineal:

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma)$$

Donde  $k$  es la tasa de crecimiento,  $m$  es el desplazamiento, y  $a(t)$  es un vector indicativo que permite cambios en la tasa de crecimiento en puntos de cambio específicos  $\delta$  y  $\gamma$ .

Para capturar la estacionalidad, Prophet usa las series de Fourier, las cuales pueden capturar patrones estacionales de cualquier periodo (anuales, semanales...) siendo la ecuación la siguiente:

$$s(t) = \sum_{n=1}^N [a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)]$$

$P$  es el periodo regular que se espera que la serie tenga (por ejemplo,  $P = 365.25$  para datos anuales o  $P = 7$  para datos semanales).  $N$  es el número de términos de Fourier.  $a_n$  y  $b_n$  son los coeficientes de Fourier que deben ser estimados.

En cuanto a la componente de vacaciones, se identifican las fechas de las vacaciones. Una vez identificadas las fechas, Prophet genera una matriz de regresores utilizando funciones indicadoras. Para cada vacación  $i$ , se define un conjunto de fechas  $D_i$  y se utiliza una función indicadora  $1(t \in D_i)$  que toma el valor de 1 si el día  $t$  está dentro del conjunto de fechas de las vacaciones  $i$ , y 0 en caso contrario. La matriz de regresores  $Z(t)$  se construye con estas funciones indicadoras para todas las vacaciones consideradas, lo que permite modelizar el impacto individual de cada evento en la serie temporal.

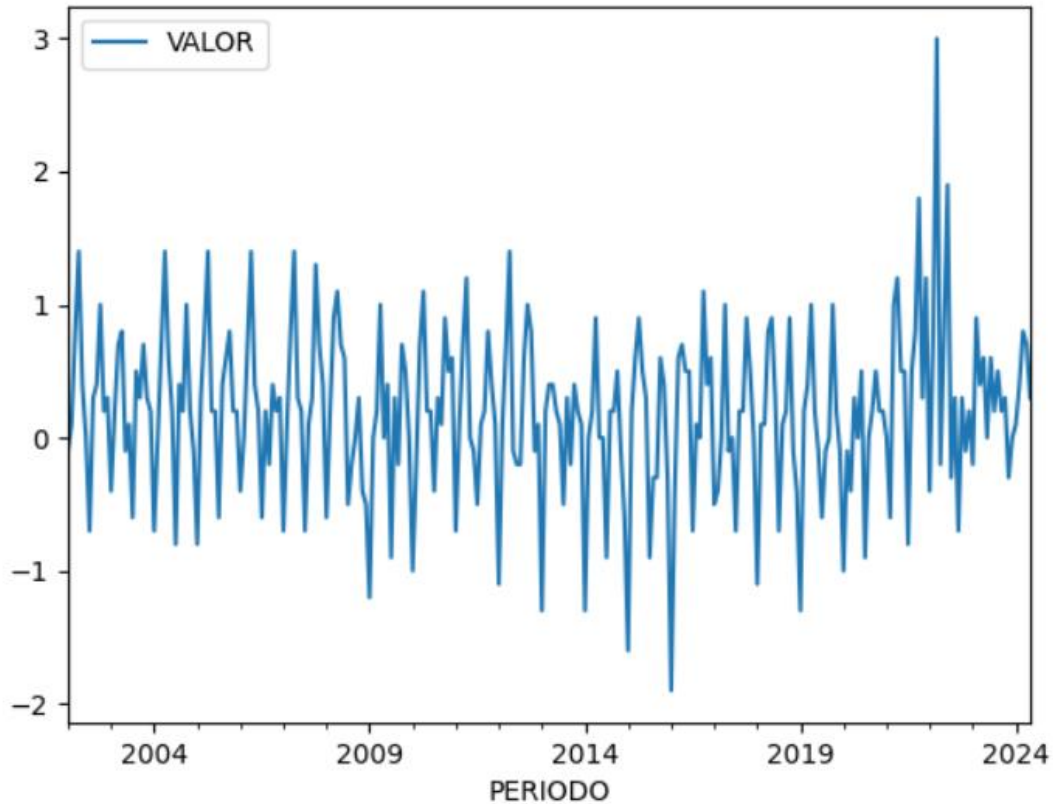
La componente de vacaciones se modela de manera aditiva mediante la ecuación  $h(t) = Z(t)k$ , donde  $k$  son los parámetros que representan el cambio en la predicción debido a las vacaciones. Estos parámetros se estiman como parte del proceso de ajuste del modelo, y permiten cuantificar cómo cada evento afecta los valores de la serie temporal. Este enfoque permite a Prophet capturar tanto el efecto directo de los días festivos como el impacto de los días cercanos, mediante la inclusión de ventanas adicionales alrededor de cada vacación.

## **4 Estudio de la serie temporal**

Se procederá al estudio de una serie temporal mediante los diferentes modelos que se han ido estudiando a lo largo de este trabajo. La herramienta que nos ayudará a realizar el estudio será la interfaz desarrollada para el trabajo de fin de grado de Ingeniería Informática, la cual está pensada para poder seleccionar cualquier serie temporal que se encuentre en un archivo .csv y poder estudiarla a través de uno de los modelos vistos con anterioridad.

### **4.1 Explicación de la serie temporal**

La serie seleccionada para el estudio es el Índice de Precios al Consumo (IPC) en España, abarcando desde enero de 2002 (momento de la entrada del euro en España) hasta mayo de 2024 (ver Figura 4.1). El IPC es un indicador clave que refleja la evolución de los precios de los bienes y servicios consumidos por la población residente en viviendas en España. Los datos se presentan con una periodicidad mensual y muestran la variación intermensual, es decir, el cambio en el valor de la cesta de la compra de un mes a otro, y no su valor absoluto [15].



*Figura 4.1: Valores del IPC (enero 2002-mayo 2024)*

## **4.2 Estudio mediante el modelo Holt-Winters**

En primer lugar, después de seleccionar la periodicidad, que en este caso será mensual, se realiza la descomposición de las componentes de la serie temporal para poder analizar la tendencia y la estacionalidad y así poder identificar si el modelo es aditivo o multiplicativo (ver Figura 4.2).

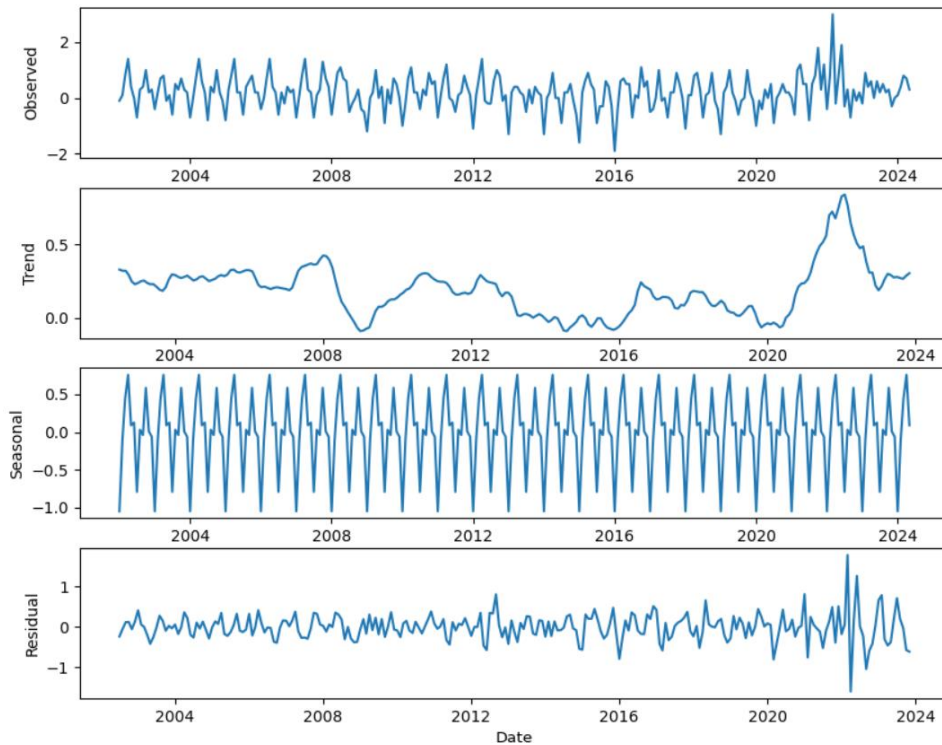


Figura 4.2: Descomposición de las componentes

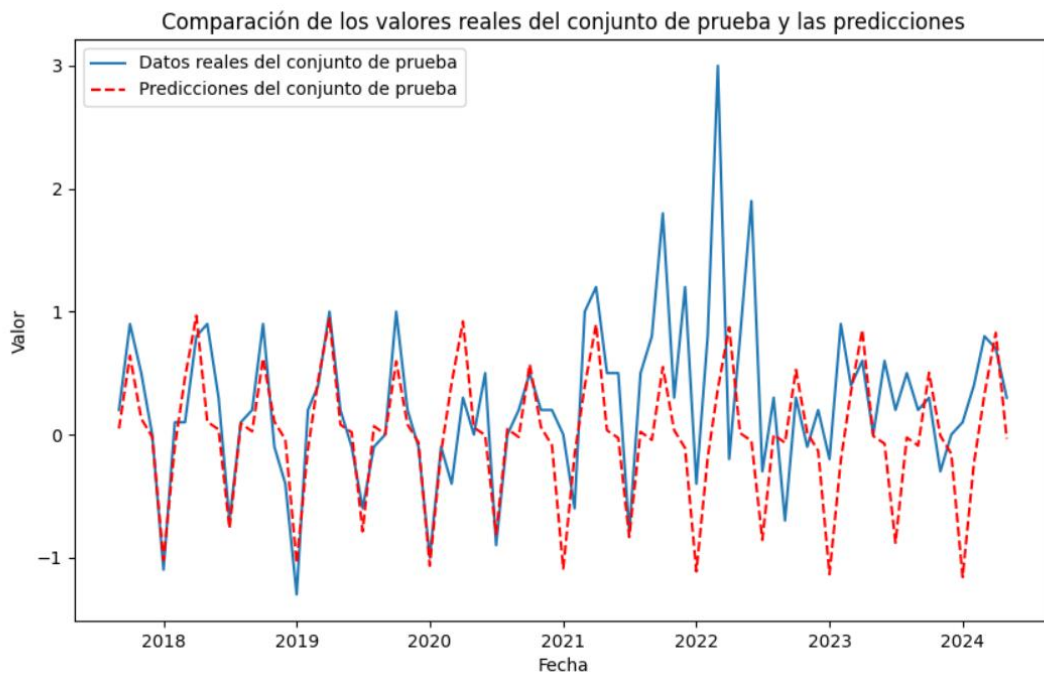
Como sabemos, si las observaciones estacionales son constantes, como podemos observar en la *Figura 4.2*, el modelo a elegir en la parte estacional es el aditivo [16]. A continuación, se estudia la tendencia, en este caso observamos que la tendencia no es constante pero aun así no presenta un crecimiento exponencial, es por eso por lo que se elige el modelo aditivo [17]. De esta manera llevamos a cabo la predicción a futuro de 24 periodos hacia adelante (ver *Figura 4.3*)

Parámetro tendencia (add/mul):	<input type="text" value="add"/>
Parámetro estacional (add/mul):	<input type="text" value="add"/>
Número de periodos a predecir:	<input type="text" value="24"/>
<input type="button" value="Predecir"/>	

Figura 4.3: Introducción valores

En nuestra interfaz se dividen los datos entre datos de entrenamiento y de prueba, siendo el 70% de entrenamiento y el 30% final de test. Llevada a cabo esta diferenciación de los datos, se ajusta el modelo con los datos de entrenamiento y se lleva a cabo la predicción de este modelo con una longitud igual a la de los datos de prueba, de esta manera se realiza la comparación entre

los datos que se han predicho y los datos reales de la serie temporal (ver Figura 4.4) y se halla el error cuadrático medio para ver el ajuste del modelo.



*Figura 4.4: Valores reales de test vs predicciones de test*

Finalmente, se realiza la predicción a futuro que será representada gráficamente con el número de periodos que hayan sido especificados, en este caso 24 (ver Figura 4.5). Podemos observar el error cuadrático medio para poder compararlo con el ajuste de otros modelos, y también el valor de las predicciones en cada periodo futuro (ver Figura 4.6).

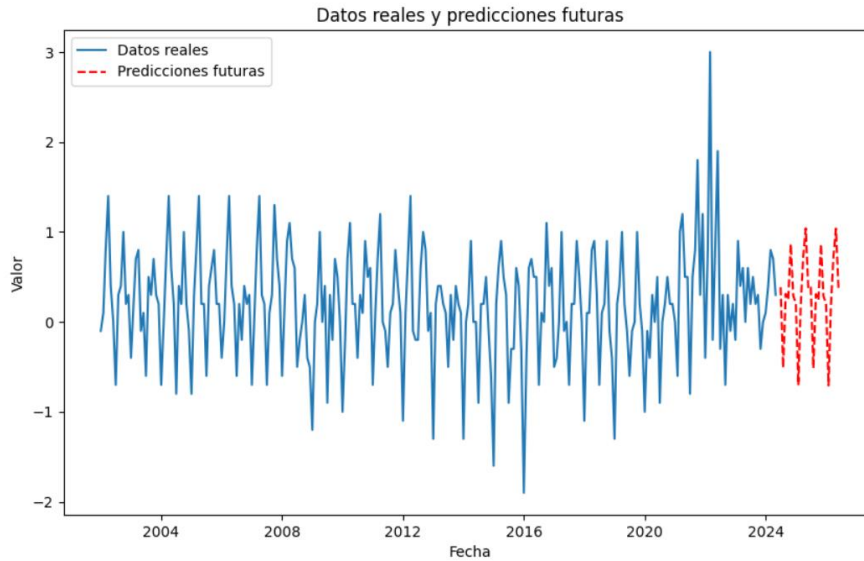


Figura 4.5: Predicciones a futuro

```

RMSE: 0.6242275925964106
Predicciones futuras:
Fecha: 2024-06-30, Predicción: 0.3851327260716604
Fecha: 2024-07-31, Predicción: -0.5103419889538285
Fecha: 2024-08-31, Predicción: 0.312418386115663
Fecha: 2024-09-30, Predicción: 0.24419493713791063
Fecha: 2024-10-31, Predicción: 0.8669534661456375
Fecha: 2024-11-30, Predicción: 0.28968333007257735
Fecha: 2024-12-31, Predicción: 0.20784273250900823
Fecha: 2025-01-31, Predicción: -0.7041076678891218
Fecha: 2025-02-28, Predicción: 0.18284867113902875
Fecha: 2025-03-31, Predicción: 0.730675079629908
Fecha: 2025-04-30, Predicción: 1.0393712092336203
Fecha: 2025-05-31, Predicción: 0.3654584462634871
Fecha: 2025-06-30, Predicción: 0.38220895536155713
Fecha: 2025-07-31, Predicción: -0.5132657596639318
Fecha: 2025-08-31, Predicción: 0.3094946154055598
Fecha: 2025-09-30, Predicción: 0.24127116642780738
Fecha: 2025-10-31, Predicción: 0.8640296954355343
Fecha: 2025-11-30, Predicción: 0.2867595593624741

```

Figura 4.6: Valor del Error Cuadrático Medio y de las predicciones

### 4.3 Estudio mediante el modelo ARIMA

Comenzamos con el ajuste del modelo ARIMA, los pasos a seguir van a ser los explicados anteriormente mediante el proceso de Box-Jenkins, por lo tanto, se inicia con el primer paso que es el de Identificación.

#### 4.3.1 Identificación

En primer lugar, se estudia el parámetro  $d$  del modelo, comenzaremos observando si la serie es estable o inestable. Fijándonos en el gráfico de la serie (ver Figura 4.1) se puede intuir a simple vista que la serie es estable, aun así, se llevará a cabo el test de Dickey-Fuller para observar la estabilidad de la serie

a lo largo del tiempo. Se parte de la hipótesis nula de que nuestra serie tiene una raíz unitaria, y por lo tanto no es estacionaria como hemos visto en la explicación teórica. Ayudándonos de nuestra interfaz, la cual realiza automáticamente el test, vemos como la serie es estacionaria (ver Figura 4.7) debido a que el valor p es menor que 0,05 por lo que se rechaza la hipótesis nula de que existe una raíz unitaria y de que por lo tanto la serie no es estacionaria. De esta forma no hay necesidad de realizar ninguna diferenciación de los datos de la serie, por lo tanto, el valor del parámetro  $d$  es 0.

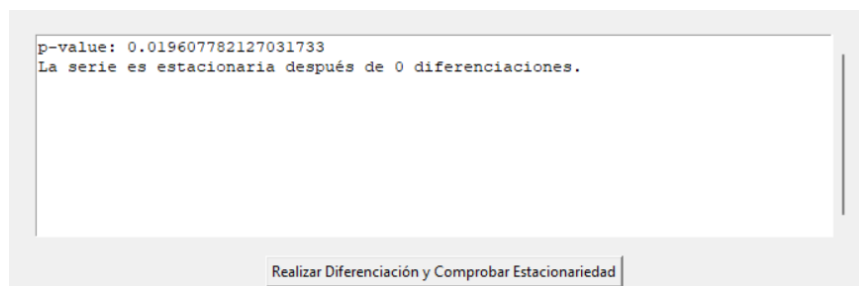


Figura 4.7: Resultado test de Dickey-Fuller

Antes de dar valor a los parámetros del modelo debemos especificar el valor de  $s$ . En el caso de esta serie es 12 al ser valores mensuales (ver Figura 4.8).

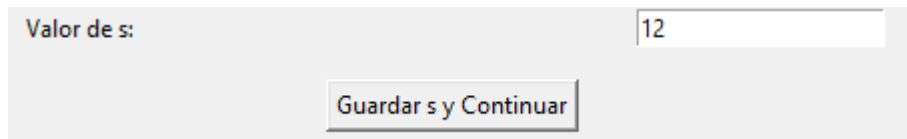


Figura 4.8: Introducción de periodicidad

A continuación, estudiamos la estacionalidad tomando una periodicidad mensual, para ello se analiza la descomposición de las componentes de la serie, en concreto la estacional (ver Figura 4.9).

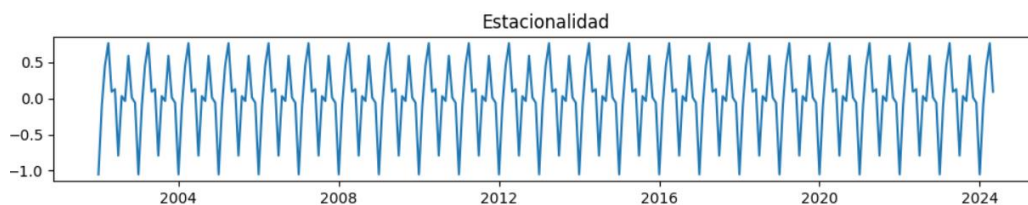


Figura 4.9: Componente estacional de la serie temporal

Se observa que, sí que hay estacionalidad, la cual también se puede deducir analizando la fas. Analizamos los múltiplos de 12 para ver si la serie es estacional (ver Figura 4.10).

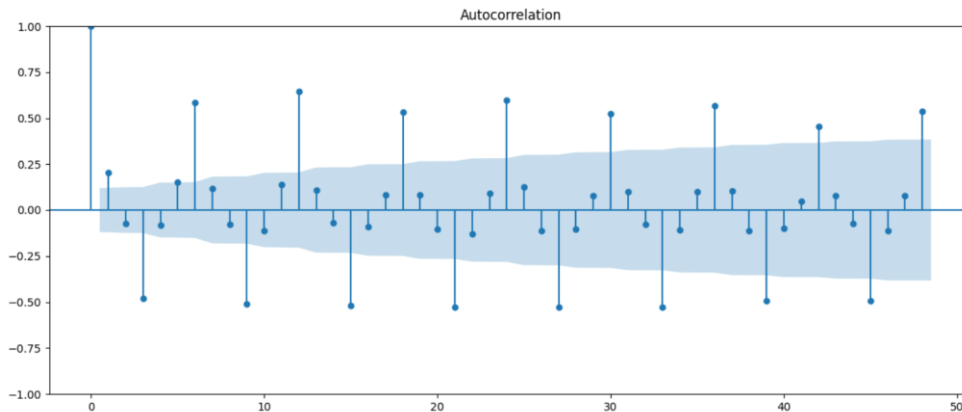


Figura 4.10: Gráfico de función de autocorrelación

Continuamos estudiando los valores de los parámetros  $p$  y  $q$ . Comenzamos con  $p$ , el cual se estudia analizando el gráfico de autocorrelación parcial (ver Figura 4.11), el gráfico nos sugiere elegir entre el valor 1 o 3. Siguiendo con  $q$  observando el gráfico de autocorrelación (ver Figura 4.10) se observa como los valores van decreciendo lentamente a 0. En el caso de que se mostraran más retardos en el gráfico se vería claramente como estos valores siguen decayendo hasta llegar a 0.

En cuanto a la parte estacional, al no realizar ninguna diferenciación los valores de  $P$  y de  $Q$  serán los mismos que hemos elegido para  $p$  y  $q$ .

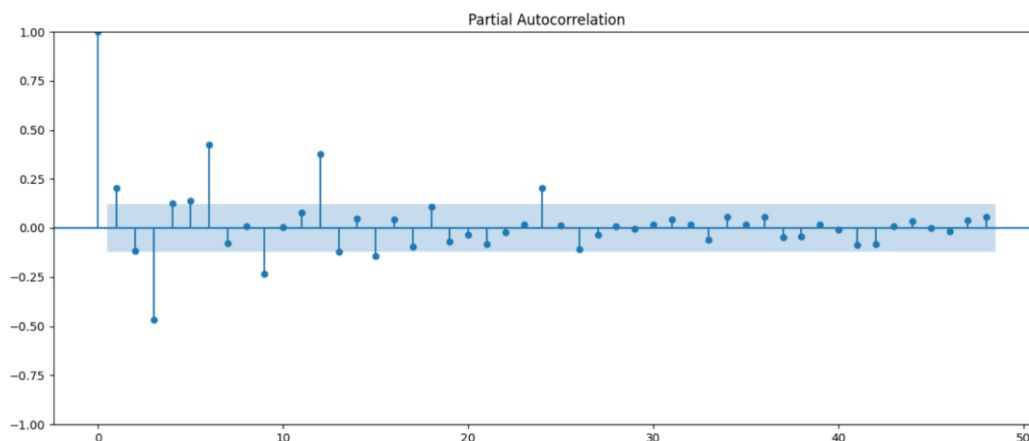


Figura 4.11: Gráfico de función de autocorrelación parcial

### 4.3.2 Estimación de los parámetros

El modelo elegido ha sido el  $SARIMA(1,0,0)x(1,0,0)_{12}$ , al ayudarnos de la interfaz creada en el trabajo de fin de grado de Ingeniería Informática. Esta nos permite

realizar diferentes combinaciones de los parámetros de los modelos, por lo tanto, podremos conocer las combinaciones de los modelos que mejor se ajustan.

Nuestra interfaz elige el modelo que menor error cuadrático medio presenta al comparar los valores del test con los valores predichos a partir del ajuste del modelo con los datos de entrenamiento, aun así, se pueden observar los valores AIC o BIC (ver Figura 4.12).

En caso de que se desee ajustar un modelo específico basándose en los valores de AIC o BIC, la interfaz permite realizar predicciones con dicho modelo sin riesgo de sobreajuste. Esto se debe a que la interfaz realiza el ajuste del modelo utilizando únicamente los valores introducidos inicialmente, sin realizar una comparación exhaustiva con otros modelos disponibles.

En resumen, nuestra interfaz no solo facilita la selección del modelo con el mejor rendimiento predictivo según el error cuadrático medio, sino que también ofrece la flexibilidad de elegir y ajustar modelos específicos basándose en criterios adicionales como AIC y BIC. Esto asegura que podamos encontrar un equilibrio adecuado entre precisión y simplicidad del modelo, evitando el sobreajuste.

```

p=1, d=0, q=0, P=1, D=0, Q=0, m=12, RMSE=0.619178578280074,AIC=142.96842183067034, BIC=152.67774771916018
p=1, d=0, q=0, P=1, D=0, Q=1, m=12, RMSE=0.5748141373161301,AIC=84.48339728186937, BIC=97.42916513318917
p=1, d=0, q=0, P=2, D=0, Q=0, m=12, RMSE=0.5833059022335413,AIC=110.21288732746851, BIC=123.15865517878831
p=1, d=0, q=0, P=2, D=0, Q=1, m=12, RMSE=0.5766444640514415,AIC=86.34925971453741, BIC=102.53146952868715
p=1, d=0, q=1, P=1, D=0, Q=0, m=12, RMSE=0.6189899908818333,AIC=144.01468335336787, BIC=156.96045120468767
p=1, d=0, q=1, P=1, D=0, Q=1, m=12, RMSE=0.5753491304560603,AIC=85.11001632861527, BIC=101.29222614276502
p=1, d=0, q=1, P=2, D=0, Q=0, m=12, RMSE=0.5835383513087028,AIC=110.70207777199454, BIC=126.88428758614428
p=1, d=0, q=1, P=2, D=0, Q=1, m=12, RMSE=0.5775909069034247,AIC=86.90474883015631, BIC=106.323400607136
p=2, d=0, q=0, P=1, D=0, Q=0, m=12, RMSE=0.6190780075676523,AIC=143.5532994578165, BIC=156.4990673091363
p=2, d=0, q=0, P=1, D=0, Q=1, m=12, RMSE=0.575366288566974,AIC=85.0444105577863, BIC=101.22662037193605
p=2, d=0, q=0, P=2, D=0, Q=0, m=12, RMSE=0.5831728746609528,AIC=110.4880965791893, BIC=126.67030639333905
p=2, d=0, q=0, P=2, D=0, Q=1, m=12, RMSE=0.5775807560752809,AIC=86.83833519003642, BIC=106.2569869670161
p=2, d=0, q=1, P=1, D=0, Q=0, m=12, RMSE=0.6191771245991351,AIC=145.34276633917315, BIC=161.5249761533229
p=2, d=0, q=1, P=1, D=0, Q=1, m=12, RMSE=0.5753721116650778,AIC=87.03898270433064, BIC=106.45763448131032
p=2, d=0, q=1, P=2, D=0, Q=0, m=12, RMSE=0.5831185415412641,AIC=112.48440260389577, BIC=131.90305438087546
p=2, d=0, q=1, P=2, D=0, Q=1, m=12, RMSE=0.5775944175464882,AIC=88.8322998636516, BIC=111.48739360346124

```

Figura 4.12: RMSE, AIC y BIC de todos los modelos

En este caso, el modelo elegido  $SARIMA(1,0,0) \times (1,0,1)_{12}$  coincide en ser aquel con menor error y con menor valor de AIC y BIC, y el resumen de este modelo (ver Figura 4.13) nos indica los estimadores de los parámetros del modelo:  $AR_1 = 0.4127$ ,  $SAR_1 = 0.9945$ ,  $MA_1 = -0.7963$ . También podemos observar la columna  $P > |z|$ , en esta en el caso de que el valor sea menor que 0.05 significa que existe diferencia suficiente para rechazar la hipótesis nula de que el coeficiente es igual a cero. Esto implica que la variable asociada al coeficiente tiene un impacto significativo en la variable dependiente en el modelo. Se observa que todos los

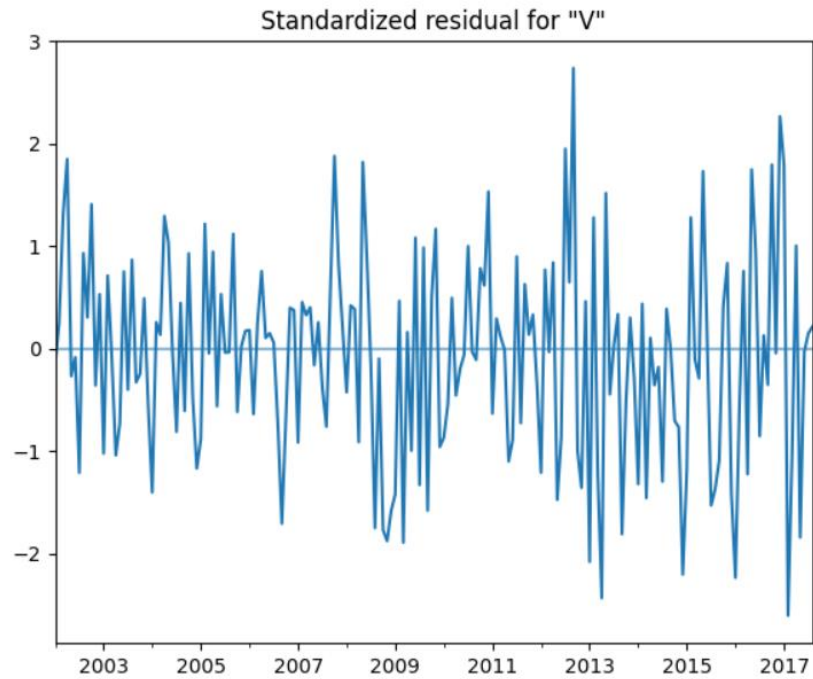
valores son 0.000, por lo tanto, todas las variables tienen un impacto significativo.

Dep. Variable:		VALOR	No. Observations:			
Model:	SARIMAX(1, 0, 0)x(1, 0, [1], 12)		Log Likelihood	-38.242		
Date:	Tue, 11 Jun 2024	AIC	84.483			
Time:	00:09:50	BIC	97.429			
Sample:	01-01-2002	HQIC	89.729			
	- 08-01-2017					
Covariance Type:		opg				
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.4127	0.059	7.018	0.000	0.297	0.528
ar.S.L12	0.9945	0.004	234.613	0.000	0.986	1.003
ma.S.L12	-0.7963	0.048	-16.630	0.000	-0.890	-0.702
sigma2	0.0755	0.009	8.774	0.000	0.059	0.092
Ljung-Box (L1) (Q):		0.14	Jarque-Bera (JB):	0.88		
Prob(Q):		0.71	Prob(JB):	0.65		
Heteroskedasticity (H):		2.61	Skew:	-0.06		
Prob(H) (two-sided):		0.00	Kurtosis:	2.69		

Figura 4.13: Resumen del modelo

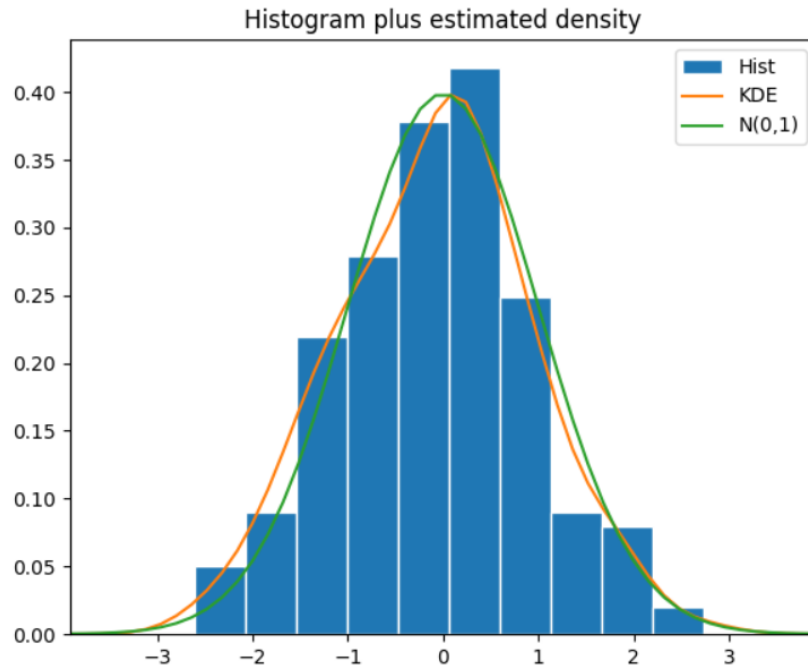
### 4.3.3 Validación del modelo

Para llevar a cabo la validación del modelo, se deben analizar los residuos de este, para ello, vamos a analizar 4 gráficos. En primer lugar, nos fijamos en el gráfico de los residuos (ver Figura 4.14), en el cual se muestran los residuos a lo largo del tiempo, buscando patrones en el gráfico. Si los residuos se distribuyen aleatoriamente alrededor de cero, sin patrones claros, esto sugiere que el modelo se ajusta bien a los datos. Nuestro modelo parece ajustarse bien debido a que la media de los residuos es muy cercana a 0.



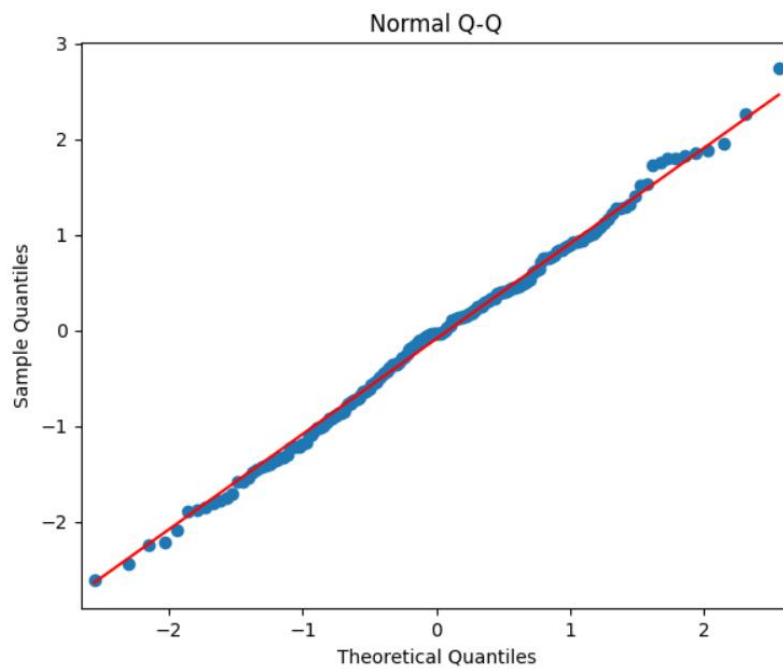
*Figura 4.14: Gráfico de los residuos*

En el siguiente gráfico se muestra la distribución de los residuos comparada con una distribución normal (ver Figura 4.15). La línea KDE (Kernel Density Estimate) y la curva de una distribución normal  $N(0,1)$  permiten comparar la distribución de los residuos con la normal. Si los residuos siguen una distribución normal, la KDE y la curva normal deberían coincidir aproximadamente. Desviaciones significativas pueden indicar problemas de ajuste del modelo. En nuestro modelo, los residuos parecen ajustarse muy bien a una distribución normal.



*Figura 4.15: Distribución de los residuos*

Continuamos con el gráfico Q-Q (ver Figura 4.16), en el que también se compara la distribución de los residuos con una distribución normal. Si los puntos caen aproximadamente a lo largo de la línea roja (la línea de referencia), indica que los residuos siguen una distribución normal. También se observa que nuestro modelo se ajusta bien a una distribución normal.



*Figura 4.16: Gráfico Q-Q*

Por último, estudiamos las autocorrelaciones de los residuos a diferentes retardos (ver Figura 4.17). La presencia de autocorrelaciones significativas indica que los residuos no son completamente aleatorios, en este caso es un gráfico de autocorrelación normal y se ve que nuestro modelo no presenta ninguna autocorrelación por encima de la banda de confianza.

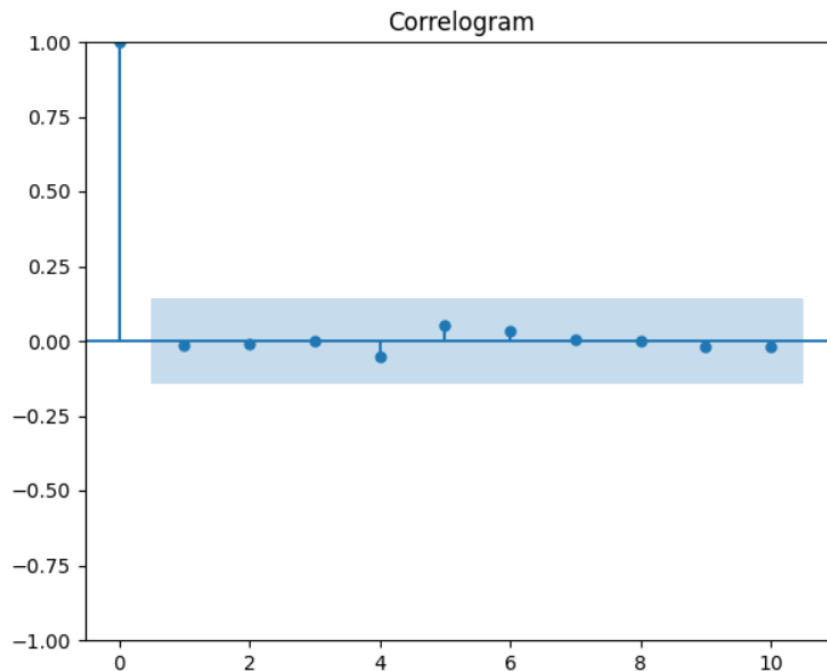


Figura 4.17: Gráfico de función de autocorrelación de los residuos

Otro de los análisis que podemos realizar es el test de Ljung-Box, el cual ha sido explicado anteriormente. Vemos (ver Figura 4.18) que, en nuestro modelo al realizar el test, el valor  $p$  es 0.62, como hemos visto con anterioridad, al ser  $p > 0.05$  no se rechaza la hipótesis nula lo cual sugiere que no hay autocorrelación significativa en los residuos, lo cual refuerza la validez del modelo.

```
Modelo 1 ljungbox:      lb_stat  lb_pvalue
10  8.049423   0.624009
```

Figura 4.18: Resultado test de Ljung-Box

Después de realizar todos estos análisis de los residuos, se puede concluir que el modelo seleccionado se ajusta adecuadamente a la serie temporal.

### 4.3.4 Predicción futura

El último paso del proceso es llevar a cabo las predicciones futuras en función del modelo ajustado. Los periodos futuros por predecir van a ser 24 meses y una vez hecha la predicción los resultados son los mostrados en la Figura 4.19.

	Prediction		
		2025-06-01	0.153420
2024-06-01	0.154270	2025-07-01	-0.628170
2024-07-01	-0.631647	2025-08-01	0.143480
2024-08-01	0.144274	2025-09-01	0.078508
2024-09-01	0.078942	2025-10-01	0.691829
2024-10-01	0.695658	2025-11-01	0.231189
2024-11-01	0.232469	2025-12-01	0.062422
2024-12-01	0.062767	2026-01-01	-1.042179
2025-01-01	-1.047948	2026-02-01	-0.078749
2025-02-01	-0.079185	2026-03-01	0.413653
2025-03-01	0.415942	2026-04-01	0.885913
2025-04-01	0.890817	2026-05-01	0.179060
2025-05-01	0.180051	2026-06-01	0.152576

Figura 4.19 Valores de las predicciones a futuro

Finalmente observamos la predicción en el gráfico de la Figura 4.20.

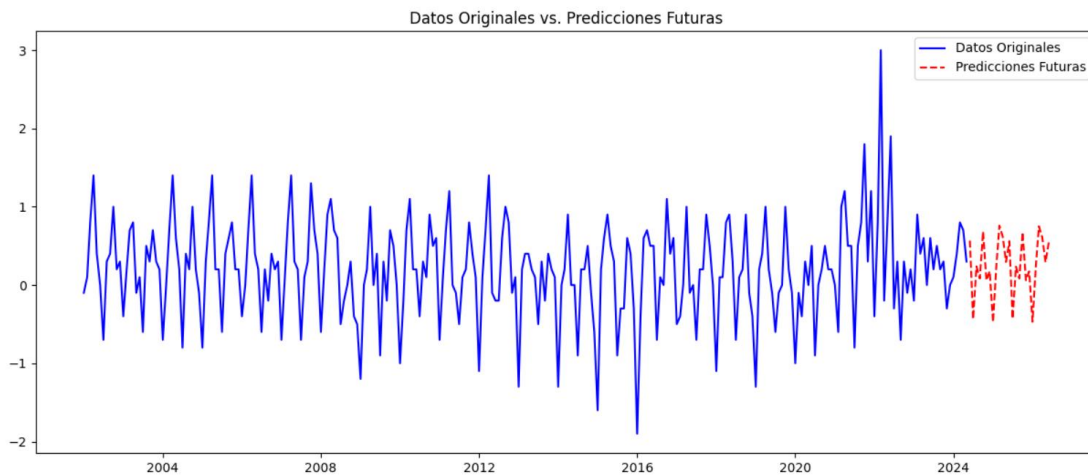


Figura 4.20: Gráfico con predicciones futuras

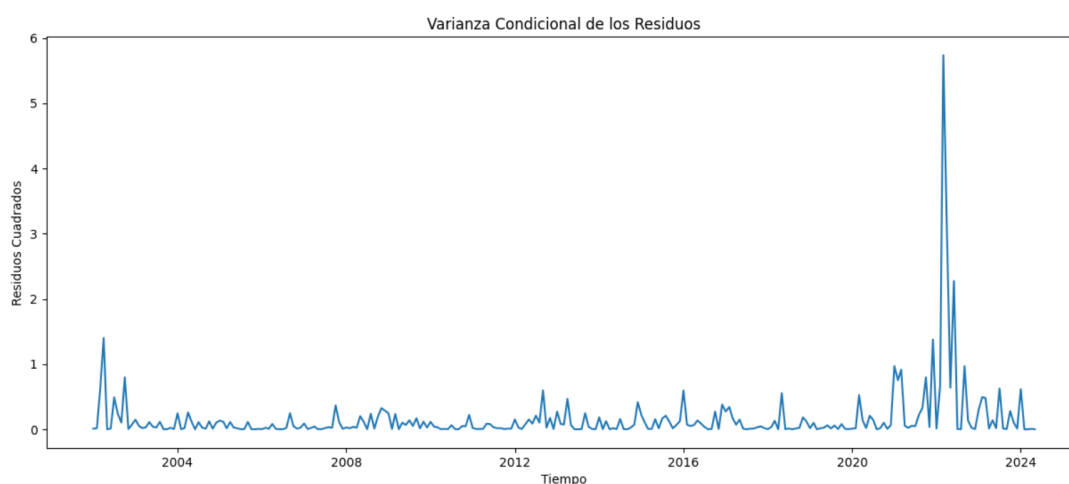
## 4.4 Estudio mediante el modelo ARCH-GARCH

### 4.4.1 Análisis de heterocedasticidad

Una vez que se ha ajustado nuestro modelo SARIMA, si deseamos estudiar la serie temporal a través de un modelo ARCH o GARCH, es necesario analizar si los residuos del ajuste de nuestro modelo presentan heterocedasticidad, es decir, que la varianza de estos residuos no es constante a lo largo del tiempo. En la

interfaz desarrollada son varios los métodos utilizados para estudiar si los residuos presentan o no constancia en la varianza. Por lo tanto, vamos a estudiar si el modelo  $SARIMA(1,0,0)\times(1,0,1)_{12}$  presenta o no heterocedasticidad en sus residuos.

En primer lugar, comenzaremos con el estudio de la varianza condicional de los residuos del modelo, representada a través de los residuos al cuadrado. Al elevar al cuadrado los residuos observamos la variabilidad o la dispersión de los errores a lo largo del tiempo. Si la varianza es constante a lo largo del tiempo se espera que los cuadrados de los residuos sean aproximadamente uniformes. Sin embargo, si la varianza cambia, se observarán picos y valles en los residuos al cuadrado, indicando períodos de alta y baja volatilidad. En nuestro gráfico (ver Figura 4.21), se observan picos que sugieren que los residuos tienen una alta variabilidad, sobre todo lo vemos entre los años 2020 y 2023 debido a la pandemia. Por lo tanto, a partir del análisis de este gráfico se deduce que los residuos presentan heterocedasticidad.



*Figura 4.21: Varianza condicional de los residuos*

Podemos continuar el estudio analizando los gráficos de autocorrelación (fas) y autocorrelación parcial (fap) de los residuos al cuadrado. Estos gráficos nos ayudarán a identificar la presencia de heterocedasticidad, mostrando si hay autocorrelación significativa en los residuos. Si observamos valores significativos en los primeros retardos, esto indicará dependencia temporal en la varianza, sugiriendo la necesidad de un modelo ARCH o GARCH. En particular, la fap nos permitirá determinar el orden  $p$  del modelo ARCH, mostrando cuántos retardos deben incluirse en el modelo.

En nuestros gráficos (ver Figura 4.22 y Figura 4.23) observamos que sí que hay autocorrelación en los residuos al cuadrado y concretamente de nuestro gráfico de autocorrelación parcial se puede deducir el orden  $p$  del modelo que en este caso sería 1, debido a que en el siguiente retardo el valor cae abruptamente a 0.

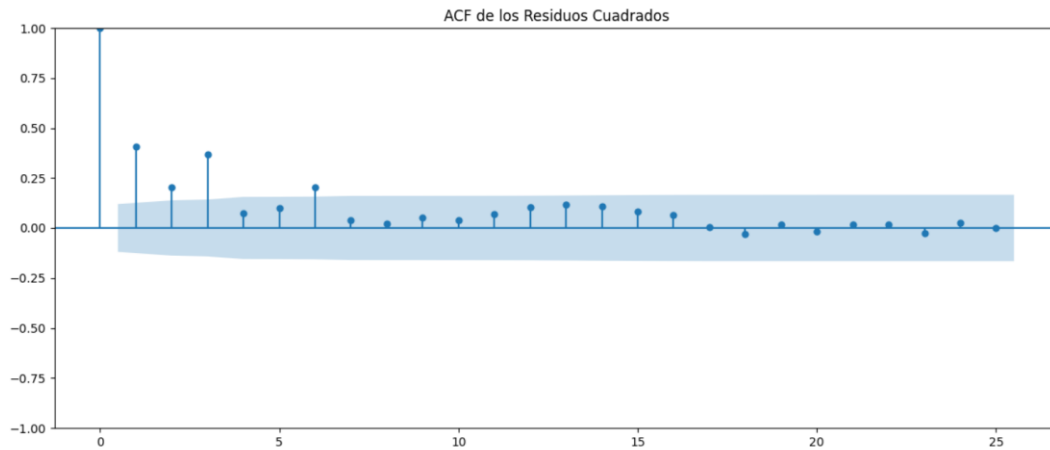


Figura 4.22: Gráfico de función de autocorrelación de los residuos al cuadrado

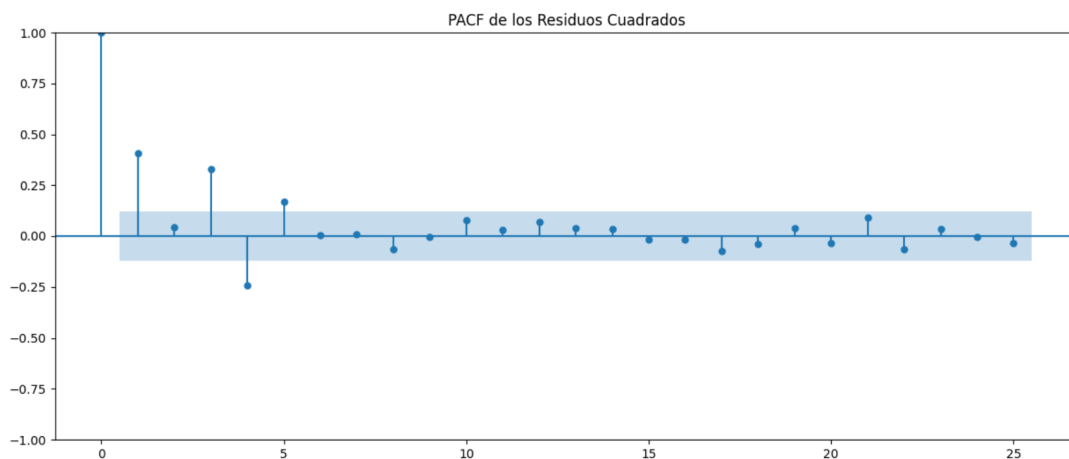


Figura 4.23: Gráfico de función de autocorrelación parcial de los residuos al cuadrado

Por último y para asegurarnos firmemente la existencia de heterocedasticidad, utilizamos la prueba de Engle para heterocedasticidad o también conocida como prueba ARCH. Nuestra interfaz nos devuelve directamente el valor  $p$  asociado con el estadístico de la prueba, en nuestro caso el valor es menor que 0,05 por lo tanto se rechaza la hipótesis nula sugiriendo la presencia de heterocedasticidad (ver Figura 4.24)

```

Prueba de Engle para Heterocedasticidad ARCH:
LM: 69.04742674126152
Valor p (LM): 6.826752040780489e-15
F: 30.61722848110063
Valor p (F): 5.320661497398741e-17
Conclusión: Hay evidencia de heterocedasticidad en los residuos del modelo.

```

Figura 4.24: Resultado de la prueba de Engle

#### 4.4.2 Ajuste del modelo

Vamos a comenzar ajustando el modelo  $GARCH(1,1)$ , como se ha visto anteriormente el gráfico de autocorrelación parcial nos sugiere utilizar el valor 1 para el parámetro  $p$ , por lo tanto, probaremos también con el valor 1 para  $q$  y analizaremos si estos valores son significativos para el ajuste del modelo. De esta manera después de ajustar el modelo  $GARCH(1,1)$  analizamos el resumen del modelo para observar si se ha ajustado bien (ver Figura 4.25). Observando la columna  $P > |t|$  se ve que tanto el valor 1 para el parámetro  $q$  como el valor 1 del parámetro  $p$  son significativos, por lo tanto, sería un modelo válido.

```

=====
Zero Mean - GARCH Model Results
=====
Dep. Variable:          None      R-squared:              0.000
Mean Model:            Zero Mean  Adj. R-squared:         0.004
Vol Model:            GARCH      Log-Likelihood:        -96.6862
Distribution:          Normal     AIC:                   199.372
Method:               Maximum Likelihood  BIC:                   210.157
Date:                 Tue, Jun 18 2024  No. Observations:      269
Time:                 18:21:01      Df Residuals:          269
                                           Df Model:              0
                                           Volatility Model
=====
              coef      std err          t      P>|t|      95.0% Conf. Int.
-----
omega         0.0118   5.954e-03     1.987   4.690e-02  [1.624e-04, 2.350e-02]
alpha[1]      0.1807   7.616e-02     2.372   1.769e-02  [3.139e-02, 0.330]
beta[1]       0.7331   8.376e-02     8.751   2.109e-18  [ 0.569, 0.897]
=====
Covariance estimator: robust

```

Figura 4.25: Resumen del modelo  $GARCH(1,1)$

Para reafirmar que nuestro modelo  $GARCH(1,1)$  es adecuado, ajustaremos un modelo  $GARCH(3,3)$  debido a que observando el gráfico de autocorrelación parcial (ver Figura 4.23) es después del retardo 3 donde los retardos abruptamente pasan a valer 0, y observaremos si los valores de los parámetros son significativos. Se puede ver (ver Figura 4.26) que los parámetros son significativos, aun así, los valores de AIC y de BIC son mayores que en el modelo  $GARCH(1,1)$ , por lo que nos quedamos con este último modelo. Por lo tanto, nuestro modelo sería el siguiente:

$$\sigma_t^2 = 0.0118 + 0.1807\varepsilon_{t-1}^2 + 0.7331\sigma_{t-1}^2$$

Zero Mean - GARCH Model Results

---

Dep. Variable:	None	R-squared:	0.000
Mean Model:	Zero Mean	Adj. R-squared:	0.004
Vol Model:	GARCH	Log-Likelihood:	-95.2451
Distribution:	Normal	AIC:	204.490
Method:	Maximum Likelihood	BIC:	229.653
		No. Observations:	269
Date:	Tue, Jun 18 2024	Df Residuals:	269
Time:	18:32:11	Df Model:	0

Volatility Model

---

	coef	std err	t	P> t	95.0% Conf. Int.
omega	0.0272	1.221e-02	2.226	2.599e-02	[3.253e-03, 5.111e-02]
alpha[1]	0.1688	7.273e-02	2.321	2.026e-02	[2.629e-02, 0.311]
alpha[2]	0.0389	8.362e-02	0.466	0.642	[-0.125, 0.203]
alpha[3]	0.1760	9.840e-02	1.789	7.363e-02	[-1.683e-02, 0.369]
beta[1]	0.0000	7.940e-02	0.000	1.000	[-0.156, 0.156]
beta[2]	0.0000	0.149	0.000	1.000	[-0.292, 0.292]
beta[3]	0.4104	0.134	3.058	2.226e-03	[ 0.147, 0.673]

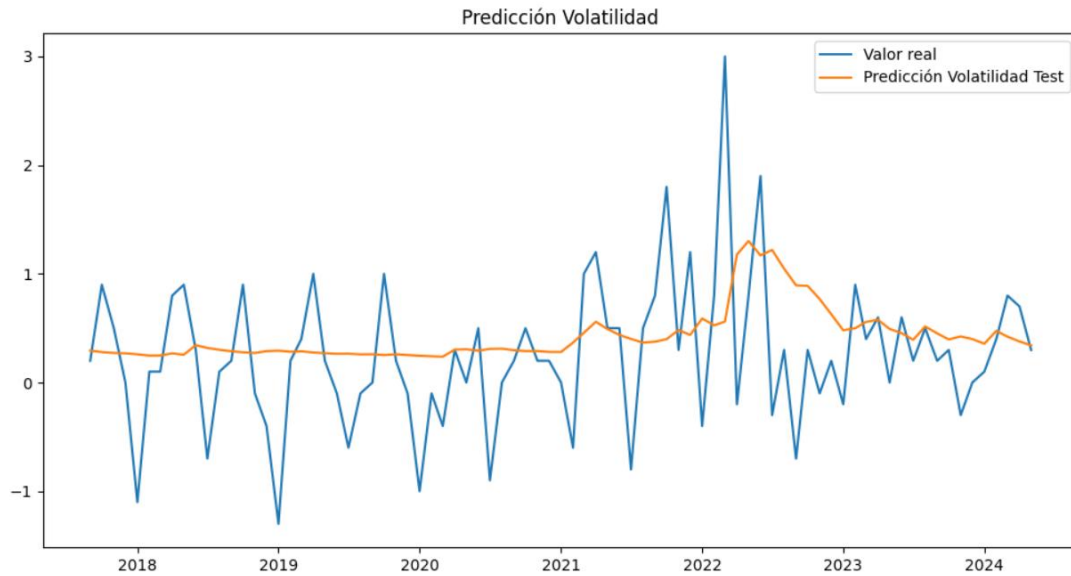
---

Covariance estimator: robust

*Figura 4.26: Resumen del modelo GARCH (3,3)*

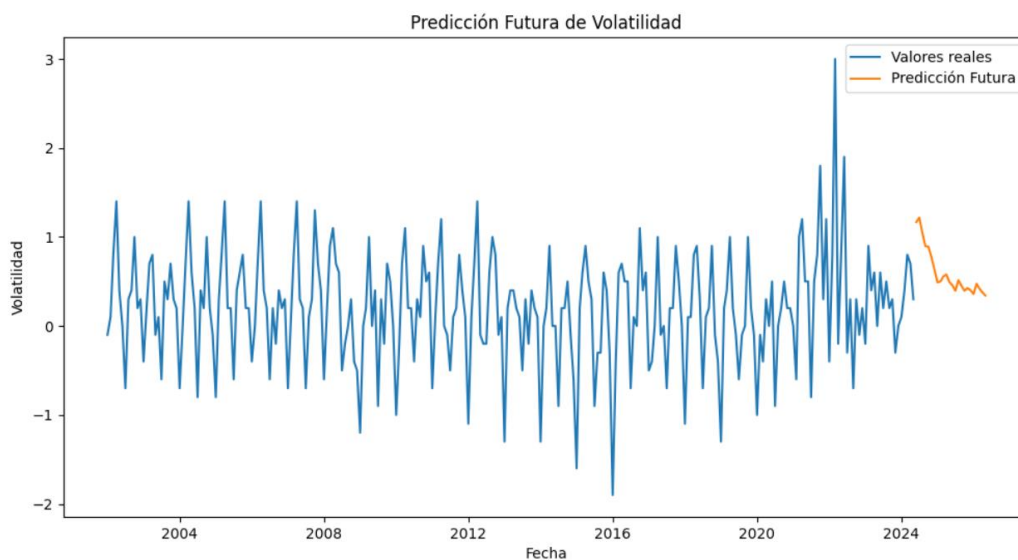
#### 4.4.3 Predicciones

Una vez que el modelo está ajustado se llevará a cabo la predicción de la volatilidad (ver Figura 4.27). Para ello, hemos dividido los datos de nuestros residuos en dos, en datos de entrenamiento y datos de test. La volatilidad se mantuvo estable hasta la pandemia, donde aumentó significativamente. Utilizamos el método rolling para capturar la volatilidad, prediciendo el primer valor de prueba con los datos de entrenamiento y luego añadiendo este primer residuo al conjunto de entrenamiento. Este proceso se repite, actualizando continuamente el conjunto de entrenamiento con cada nuevo residuo de prueba, hasta que se realizan predicciones para todos los valores de prueba, capturando así la tendencia general de la volatilidad a lo largo del tiempo.



*Figura 4.27: Predicción de volatilidad en los valores de test*

También se pueden realizar predicciones a futuro. Para igualar el número de predicciones con el resto de los modelos, hemos realizado una predicción con 24 pasos hacia adelante. En la gráfica (ver Figura 4.28) observamos una volatilidad inicialmente alta, atribuida a los efectos de la pandemia, seguida de una tendencia decreciente en la volatilidad. Este comportamiento sugiere que, aunque la volatilidad fue muy alta durante los años de la pandemia, se espera que disminuya en el futuro cercano. También se pueden conocer los valores exactos de la predicción (ver Figura 4.29).



*Figura 4.28: Predicción de volatilidad a futuro*

Predicciones Futuras:	
2024-05-31	1.167399
2024-06-30	1.217092
2024-07-31	1.046583
2024-08-31	0.895626
2024-09-30	0.892180
2024-10-31	0.773453
2024-11-30	0.635877
2024-12-31	0.488806
2025-01-31	0.500769
2025-02-28	0.556083
2025-03-31	0.578723
2025-04-30	0.494306
2025-05-31	0.457250
2025-06-30	0.395029
2025-07-31	0.514332
2025-08-31	0.451016
2025-09-30	0.395401
2025-10-31	0.425205
2025-11-30	0.401870
2025-12-31	0.359355
2026-01-31	0.474117
2026-02-28	0.421198
2026-03-31	0.377101
2026-04-30	0.341675

Figura 4.29: Valores de predicción de volatilidad

Por último, unimos las predicciones realizadas con el modelo  $SARIMA(1,0,0)x(1,0,1)_{12}$  y las predicciones de volatilidad de nuestro modelo  $GARCH(1,1)$  el cual se utiliza para generar bandas de confianza para las predicciones del modelo  $SARIMA(1,0,0)x(1,0,1)_{12}$ . Las bandas de confianza sirven para proporcionar un rango dentro del cual esperamos que caigan las predicciones futuras con un cierto nivel de confianza (en este caso, 95%). Se crean calculando la desviación estándar de las predicciones de volatilidad y multiplicándola por 1.96, ya que 1.96 es el valor crítico para un intervalo de confianza del 95% en una distribución normal estándar debido a que esta es la que presentan nuestros residuos (ver Figura 4.15). Esto asegura que aproximadamente el 95% de los datos futuros caerán dentro de este rango, proporcionando una medida de incertidumbre en nuestras predicciones. Cuanto mayor es la volatilidad, más grande es el intervalo de confianza, ya que una mayor volatilidad indica una mayor variabilidad en los datos, lo que aumenta la posible dispersión de las predicciones. Esto lo podemos observar en las primeras predicciones y al haber mayor volatilidad (ver Figura 4.28) las bandas de confianza son mayores (ver Figura 4.30). Si un valor observado escapa de las bandas de confianza, puede indicar que el modelo no ha capturado todos los factores influyentes, que ha ocurrido un evento anómalo o que la volatilidad ha sido subestimada, lo que puede requerir una revisión y ajuste del modelo.

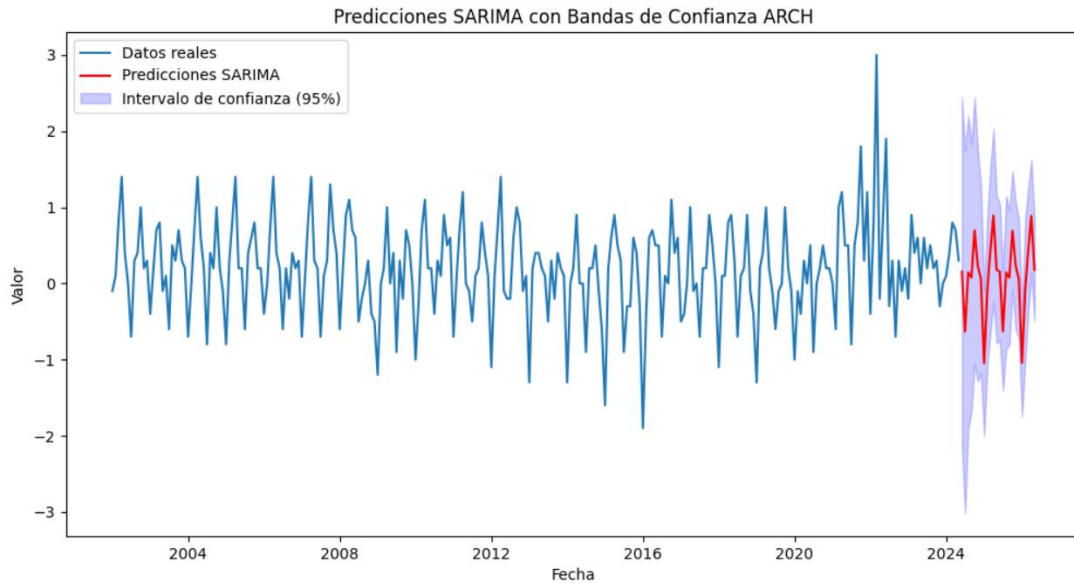


Figura 4.30: Predicciones SARIMA(1,0,0)x(1,0,1)<sub>12</sub> con bandas de confianza

## 4.5 Estudio mediante el modelo Prophet

Para realizar predicciones utilizando el modelo Prophet, es necesario seguir una serie de pasos sencillos. Primero, se debe seleccionar la periodicidad de los datos, que en este caso será mensual. Además, es importante definir el número de periodos que se desean predecir (ver Figura 4.31).

La serie temporal se dividirá en datos de entrenamiento y datos de test. Los datos de entrenamiento se utilizarán para ajustar el modelo, mientras que los datos de test se emplearán para evaluar la precisión de las predicciones realizadas por el modelo.

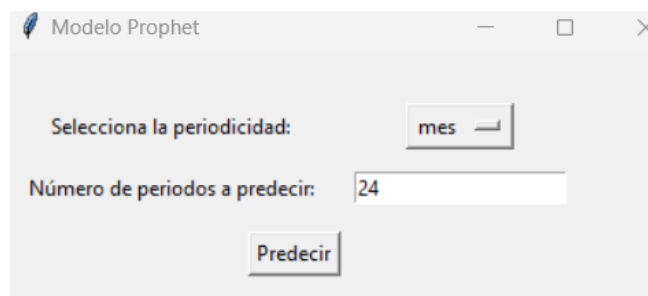


Figura 4.31: Introducción de valores

Una vez ajustados los datos de entrenamiento al modelo, el propio algoritmo de Prophet hace las predicciones para hacer la comparación con los valores reales de los datos de test, que como podemos comprobar no se ajusta bien durante los años de Covid al ser un evento imprevisto (ver Figura 4.32). Esta comparación nos permite hallar el RSME y así poder comparar el valor de este

con los demás modelos ajustados. También el modelo Prophet nos da la opción de estudiar algunas de las componentes de la serie, en este caso observamos la tendencia y la estacionalidad (ver Figura 4.33).

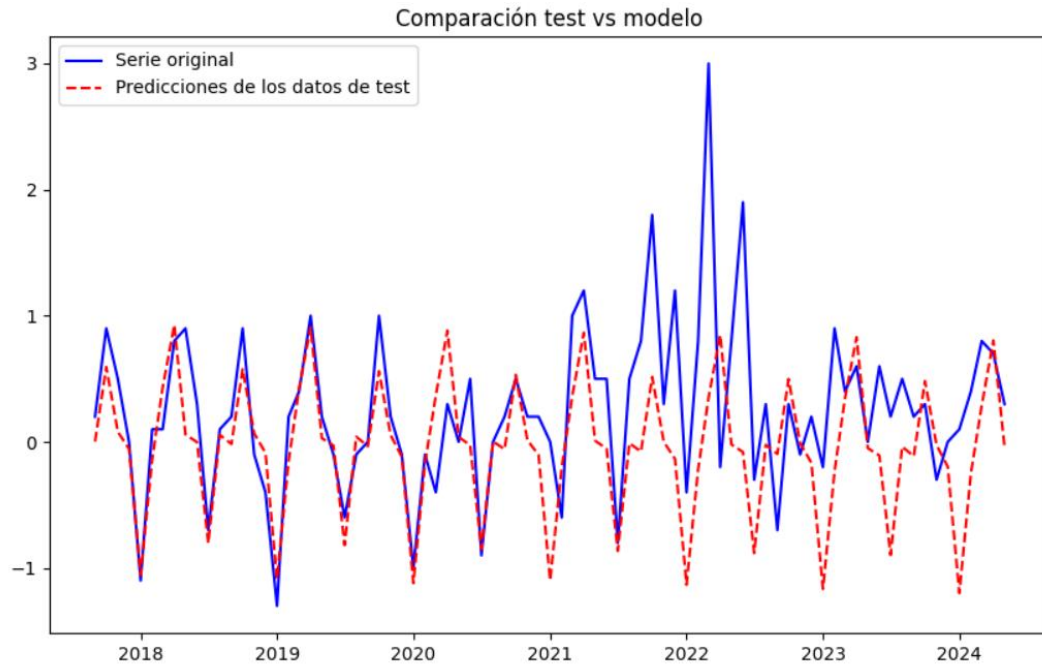


Figura 4.32: Valores reales de test vs predicciones de test

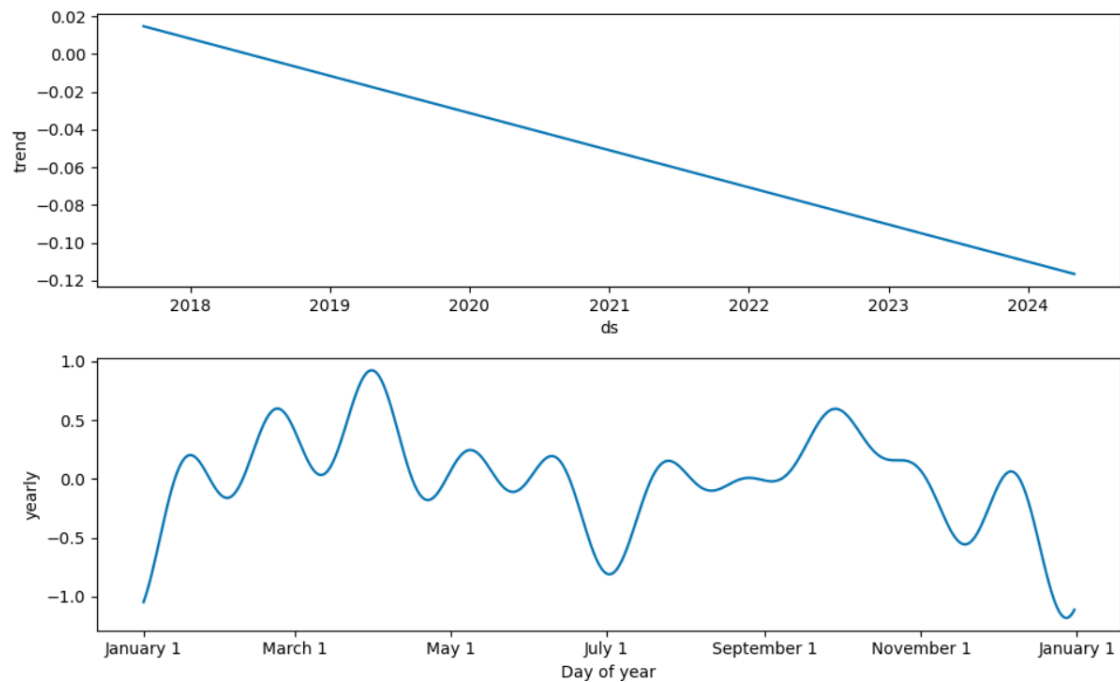


Figura 4.33: Tendencia y estacionalidad

Por último, hacemos las predicciones a futuro, en este caso al igual que en el resto 24 periodos hacia adelante (ver Figura 4.35). Podemos observar el valor concreto de las predicciones para hacer una comparación con el resto de modelos, además también tenemos el valor del RSME (ver Figura 4.35).

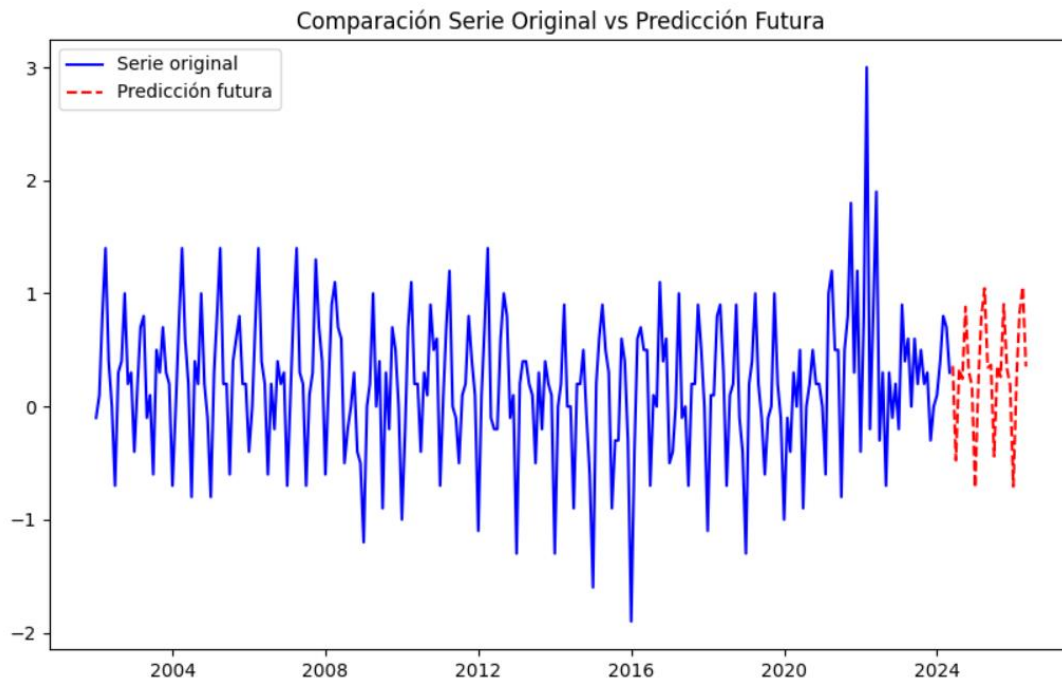


Figura 4.34: Predicciones a futuro

```

RMSE: 0.6360821570309662
MAE: 0.4424529407547169
Predicciones futuras:
Fecha: 2024-05-31, Predicción: 0.358229922743862
Fecha: 2024-06-30, Predicción: -0.4760267821537513
Fecha: 2024-07-31, Predicción: 0.3330074926237138
Fecha: 2024-08-31, Predicción: 0.24447477254596928
Fecha: 2024-09-30, Predicción: 0.8817355908019613
Fecha: 2024-10-31, Predicción: 0.3104587540540356
Fecha: 2024-11-30, Predicción: 0.18436373451346702
Fecha: 2024-12-31, Predicción: -0.7177006827067534
Fecha: 2025-01-31, Predicción: 0.22253511903120107
Fecha: 2025-02-28, Predicción: 0.7953045283499653

Fecha: 2025-03-31, Predicción: 1.0448775641407988
Fecha: 2025-04-30, Predicción: 0.34443332359120776
Fecha: 2025-05-31, Predicción: 0.3684348509829297
Fecha: 2025-06-30, Predicción: -0.44073441015869763
Fecha: 2025-07-31, Predicción: 0.3597111933281991
Fecha: 2025-08-31, Predicción: 0.2626684927464487
Fecha: 2025-09-30, Predicción: 0.904612713496395
Fecha: 2025-10-31, Predicción: 0.3348299994582746
Fecha: 2025-11-30, Predicción: 0.18987654372047666
Fecha: 2025-12-31, Predicción: -0.7072792118126252
Fecha: 2026-01-31, Predicción: 0.2541008286750148
Fecha: 2026-02-28, Predicción: 0.8271608340613466
Fecha: 2026-03-31, Predicción: 1.0589377749178244
Fecha: 2026-04-30, Predicción: 0.3528473810378073

```

Figura 4.35: Valor de las predicciones y RSME

## 5 Resultados y conclusiones

A lo largo de este trabajo hemos explorado varios métodos para realizar predicciones futuras de series temporales, comenzando con el análisis de sus componentes y luego estudiando las particularidades de cada modelo. En los casos donde el ajuste del modelo SARIMA presentó heterocedasticidad, utilizamos modelos ARCH para predecir la varianza condicional de la serie temporal, lo que nos permitió anticipar el comportamiento futuro de la serie.

Nuestra serie temporal utilizada es el Índice de Precios al Consumidor (IPC). Tras ajustar todos los modelos y realizar las predicciones comparamos su desempeño utilizando el error cuadrático medio (RMSE).

El modelo Holt-Winters presentó un error de 0.6242, el modelo  $SARIMA(1,0,0)x(1,0,1)_{12}$  un error de 0.5748, y el modelo Prophet un error de 0.6360.

De estos resultados, podemos concluir que el modelo  $SARIMA(1,0,0)x(1,0,1)_{12}$  es el que mejor se ajusta a nuestros datos. Una vez los modelos han sido ajustados hemos podido llevar a cabo las predicciones a futuro, en concreto 24 meses hacia delante.

Además, al ajustar un modelo  $GARCH(1,1)$  a los residuos del modelo SARIMA, logramos modelar la volatilidad y obtener bandas de confianza al 95%. Esto nos proporciona un rango dentro del cual esperamos que se encuentren las predicciones del modelo  $SARIMA(1,0,0)x(1,0,1)_{12}$ , cuantificando de manera efectiva la incertidumbre asociada a estas predicciones.

Se ha observado que un suceso inesperado, como el Covid-19, puede alterar significativamente el curso natural de una serie temporal. En este trabajo, no se ha abordado de manera específica el impacto de dicho suceso. Por lo tanto, esta omisión representa un área de mejora importante. Es crucial tratar eventos individuales de manera específica para ajustar los modelos de forma más precisa y mejorar la calidad de las predicciones.

En resumen, el uso combinado de SARIMA y GARCH no solo mejora la precisión de nuestras predicciones, sino que también nos permite conocer mejor la incertidumbre inherente a estas. Este enfoque integral, que incluye el análisis de componentes y la gestión de la volatilidad, proporciona una robusta herramienta para la predicción de series temporales.


## 6 Bibliografía

- [1] P. J. & D. R. A. Brockwell, Introduction to Time Series and Forecasting, Nueva York: Springer, 2016.
- [2] C. Chatfield, The Analysis of Time Series: An Introduction (7th ed.), Chapman and Hall/CRC, 2019.
- [3] S. J. Taylor y B. Letham, «Prophet,» 2017. [En línea]. Available: <https://facebook.github.io/prophet/>.
- [4] R. J. Hyndman y G. Athanasopoulos, Forecasting principles and practice, OTexts, 2014.
- [5] R. J Hyndman y G. Athanasopoulos, Forecasting Principles and Practice, O Texts, 2014.
- [6] E. Ostertagová y O. Ostertag, «THE SIMPLE EXPONENTIAL SMOOTHING MODEL,» Košice, 2011.
- [7] Statgraphics, «Statgraphics,» 25 mayo 2007. [En línea]. Available: <https://www.statgraphics.com/>.
- [8] openforecast, «openforecast,» 8 mayo 2024. [En línea]. Available: <https://openforecast.org/adam/tsComponents.html>.
- [9] G. Eshel, «The Yule Walker Equations for the AR Coefficients,» [En línea]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://www-stat.wharton.upenn.edu/~stele/Courses/956/Resource/YWSourceFiles/YW-Eshel.pdf.
- [10] M. I. R. Galiano, Tema 6: Series Temporales.
- [11] B. Gómez, «RPubs,» 20 enero 2020. [En línea]. Available: <https://rpubs.com/brianz0r/574656>.
- [12] Minitab, «Minitab,» 2024. [En línea]. Available: <https://support.minitab.com/es-mx/minitab/help-and-how->

to/statistical-modeling/time-series/how-to/augmented-dickey-fuller-test/before-you-start/example/.

- [13] R. d. Arce, «INTRODUCCIÓN A LOS MODELOS AUTORREGRESIVOS CON HETEROCEDASTICIDAD CONDICIONAL (ARCH),» diciembre 1998. [En línea]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.uam.es/uam/media/doc/1606862179240/doctra9806.pdf.
- [14] Meta, «Meta,» febrero 2023. [En línea]. Available: <https://facebook.github.io/prophet/>.
- [15] INE, «Instituto Nacional de Estadística,» mayo 2024. [En línea]. Available: [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176802&menu=ultiDatos&idp=1254735976607](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176802&menu=ultiDatos&idp=1254735976607).
- [16] PennState, «PennState: Eberly College of Science,» [En línea]. Available: <https://online.stat.psu.edu/stat510/lesson/5/5.1>.
- [17] L. A. Maydana, «Entendiendo las Tendencias Aditivas y Multiplicativas en Series Temporales,» 30 agosto 2023. [En línea]. Available: <https://medium.com/@leomay282/entendiendo-las-tendencias-aditivas-y-multiplicativas-en-series-temporales-e8dca7f1e863>.

Este documento esta firmado por

	<b>Firmante</b>	CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES
	<b>Fecha/Hora</b>	Thu Jul 11 23:26:18 CEST 2024
	<b>Emisor del Certificado</b>	EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES
	<b>Numero de Serie</b>	561
	<b>Metodo</b>	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)