

Trabajo Fin de Grado

Evolución y tendencias de la tasa de paro
femenina en España: Un estudio econométrico.

Autora

Andrea Jaca Martín

Directora

María Isabel Ayuda Bosque

Facultad de Economía y Empresa. Universidad de Zaragoza

Año 2024

Autora del trabajo: Andrea Jaca Martín

Directora del trabajo: María Isabel Ayuda Bosque

Título del trabajo: Evolución y tendencias de la tasa de paro femenina en España: Un estudio econométrico.

Title of the work: Evolution and Trends of the Female Unemployment Rate in Spain: An Econometric Study

Titulación: Grado en Economía

Resumen

La finalidad de este trabajo es realizar un análisis econométrico univariante de una serie temporal. La variable objeto de estudio es la tasa de desempleo de las mujeres a nivel nacional, con datos trimestrales, desde el primer trimestre de 2002 hasta el cuarto trimestre de 2023, es decir, 88 observaciones, cuyos datos han sido obtenidos del Instituto Nacional de Estadística (INE).

El objetivo de este trabajo es la aplicación de la metodología de Box-Jenkins para examinar el comportamiento de dicha serie temporal, identificar posibles patrones estacionales y tendencias a largo plazo, y poder identificación posibles modelos autorregresivos integrados de medias móviles (ARIMA) que han podido generar dicha serie con objeto de predecir de la mejor manera posible el comportamiento del paro femenino en España corto plazo.

Abstract

The purpose of this work is to perform a univariate econometric analysis of a time series. The variable under study is the national unemployment rate of women, with quarterly data from the first quarter of 2002 to the fourth quarter of 2023, i.e., 88 observations, obtained from the National Institute of Statistics (INE).

The objective of this work is to apply the Box-Jenkins methodology to examine the behavior of this time series, identify possible seasonal patterns and long-term trends, and identify potential autoregressive integrated moving average (ARIMA) models that may have generated this series, in order to best predict the short-term behavior of female unemployment in Spain.

Índice

1.-Introducción	2
2.-Búsqueda de datos en fuentes estadísticas	3
3.-Explicación de la metodología de Box – Jenkins.....	4
4.-Aplicación de la metodología de Box – Jenkins a la serie objeto de estudio.....	5
Identificación de los posibles procesos estocásticos lineales discretos	5
Estimación de los procesos estocásticos	17
Chequeo	19
Predicción	35
6.- Conclusiones	37
7.-Bibliografía	38
Anexo I	40

1.-Introducción

La tasa de desempleo en España es uno de los indicadores macroeconómicos que mayor impacto genera en la economía y uno de los indicadores que más preocupa a la sociedad, por estos motivos, se han realizado numerosos estudios y artículos sobre esta variable. Algunos de estos trabajos son Dueñas Fernández, Iglesias Fernández y Llorente Heras (2016). donde a través de modelos ARIMA se pronostica el número de mujeres ocupadas y en paro en los últimos tres trimestres de 2014 en España y a través de modelos logit se analiza la probabilidad de que encuentren empleo o Millán Vázquez de la Torre, Santos Pita y Pérez Naranjo (2015) en el que se estudia la distribución laboral femenina en el año 2012 en el mercado de trabajo español según la metodología de descomposición Oaxaca-Blinder sobre las probabilidades que tienen hombres y mujeres de pertenecer a una ocupación masculinizada o feminizada.

Actualmente las mujeres tienen un papel fundamental en el mercado laboral, pero a pesar de esto sigue siendo uno de los sectores más vulnerables en situaciones de crisis. La crisis económica y financiera de 2008 afectó especialmente a España lo que provocó un aumento de la tasa de paro, llegando a alcanzar en 2013 un 26%. Este aumento de la tasa de paro tuvo un impacto desigual ya que los más vulnerables a la pérdida de empleo fueron los jóvenes y en concreto las mujeres debido a la mayor precariedad laboral en sectores como la hostelería, el comercio minorista y los servicios, donde las mujeres estaban sobrerrepresentadas. A medida que la economía española se recuperaba de la crisis, se implementaron diversas políticas y programas para abordar las disparidades de género en el empleo. Se promulgaron leyes que promovían la igualdad de género y se incorporaron políticas de conciliación laboral y familiar, como el permiso parental igualitario y medidas de apoyo al cuidado infantil.

Otro gran impacto en la economía española y en concreto en el mercado laboral se debe a la pandemia del Covid-19. Se produjo una pérdida masiva de empleos debido a las medidas de confinamiento y al cierre temporal o permanente de muchas empresas, los sectores como la hostelería y el turismo fueron los más afectados en este sentido. Pero sin duda el sector más afectado debido a la pandemia fue el sector sanitario, que está compuesto mayormente por mujeres, este sector se vio desbordado debido a la gran cantidad de personas afectadas que hubo. Además, con el confinamiento muchas mujeres

se vieron obligadas a asumir una mayor carga de trabajo en el hogar debido al cierre de escuelas y guarderías.

A pesar de los avances que ha habido en los últimos años, siguen existiendo desafíos significativos en cuanto a la igualdad de género en el empleo en España. Las mujeres siguen enfrentándose a obstáculos para acceder a empleos de calidad y para avanzar en sus carreras, además la brecha salarial de género sigue siendo una preocupación importante.

El trabajo, además de esta primera introducción se estructura en cuatro partes. En el apartado segundo se indica la fuente de los datos y se define la variable objeto de estudio. En un apartado tercero se explica la metodología utilizada para posteriormente en el apartado cuarto aplicarla a la variable objeto de estudio y finalmente acabar con unas conclusiones.

2.-Búsqueda de datos en fuentes estadísticas

La medición de la tasa de paro en España se realiza mediante la Encuesta de Población Activa que realiza trimestralmente el INE, y el paro registrado, es decir, el número de personas inscritas en el SEPE (Servicio Público de Empleo Estatal).

Los datos de la muestra que vamos a analizar se pueden encontrar en la base de datos del INE. El INE es un organismo público español, el cual pertenece al Ministerio de Economía, Comercio y Empresa. Este organismo tiene como finalidad la realización de estudios estadísticos sobre los datos estadísticos más notorios de nuestro país.

Por lo tanto, nuestra muestra se ha obtenido con los datos de la Encuesta de Población Activa. Dicha encuesta divide a la población en dos grupos:

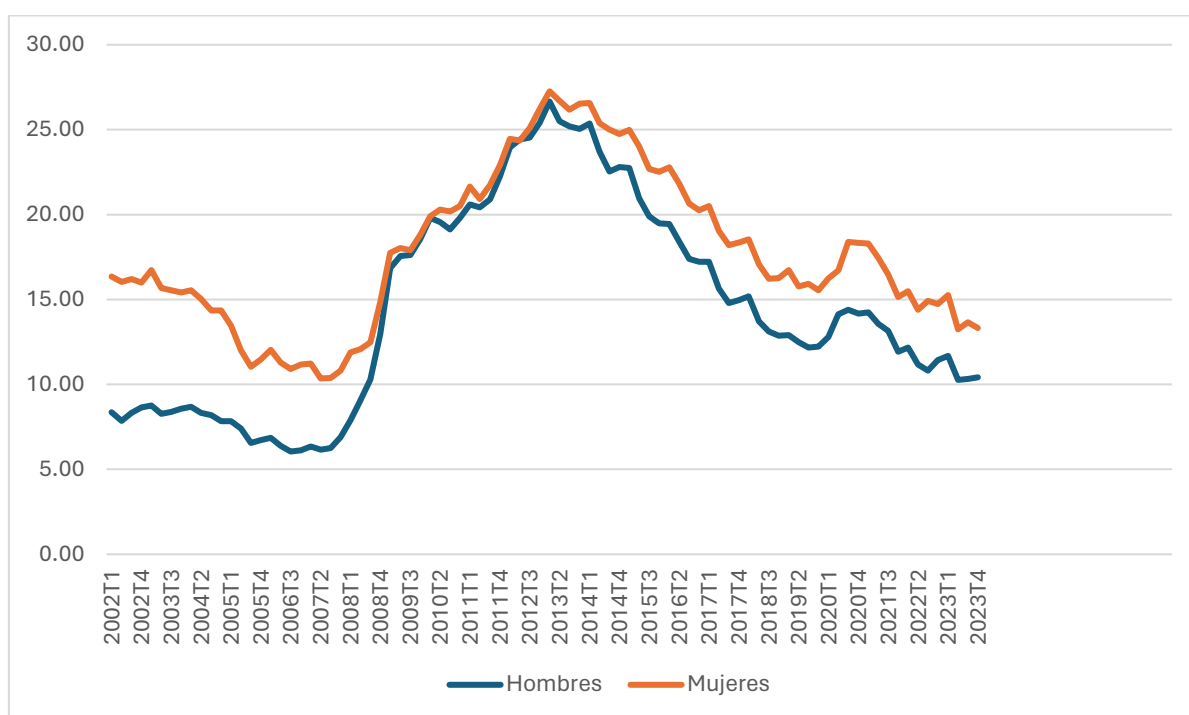
- Población Económicamente Activa: dentro de este grupo están las personas en edad de trabajar que están empleadas o desempleadas, pero buscan empleo. Por lo tanto, se divide en otros dos subgrupos:
 - Ocupados: personas con un empleo remunerado.
 - Desempleados: personas que no están ocupadas, pero buscan empleo activamente.

- Población No Económicamente Activa: dentro de este grupo están las personas de 16 años o más que están estudiando, realizan tareas domésticas, jubiladas, imposibilitadas para trabajar o que simplemente no están buscando trabajo.

Y la tasa de desempleo se calcula dividiendo el número de desempleados dividido entre la población económicamente activa, expresado en porcentaje.

En el Figura 1 se muestra la tasa de paro femenino y masculino para el período de estudio y se observa como en general la tasa de desempleo femenina tiende a ser más alta que la masculina excepto en el período de la crisis financiera de 2008 donde ambas tasas se disparan, llegando la tasa de paro masculina a una convergencia con la femenina.

Figura 1: Tasa de paro trimestral femenino y masculino en España (2002:1-2023:4)



3.-Explicación de la metodología de Box – Jenkins

La metodología Box-Jenkins, también conocida como el enfoque de Modelos Autorregresivos Integrados de Medias Móviles (ARIMA por sus siglas en inglés), es una técnica ampliamente utilizada en el análisis de series temporales para modelar y predecir patrones y tendencias. Desarrollada por Box y Jenkins (1973) en la década de 1970, esta metodología se ha convertido en un estándar en la modelización de datos temporales

debido a su flexibilidad y capacidad para capturar la complejidad de las series temporales y su eficiencia en predicciones a corto plazo. Dicha metodología está basada en 4 etapas:

1.- Etapa de IDENTIFICACIÓN: consiste en utilizar los datos recogidos de la serie temporal e identificar el posible modelo ARIMA que sigue la serie. Para ello, el primer requisito que debe tenerse en cuenta es si la serie es estacionaria o no.

Si no lo fuera, habría que decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria en media y en varianza.

Una vez que esta serie transformada es estacionaria, se tratará de averiguar los posibles ordenes tanto de la parte regular del modelo (autorregresiva, p , y medias móviles, q) como de la parte estacional (autorregresiva, P y medias móviles, Q). Por último, tendríamos que determinar si el modelo tiene término constante o no.

2- Etapa de ESTIMACIÓN: una vez identificados los posibles modelos que han podido generar la serie temporal, en esta etapa, trataremos de cuantificar los parámetros de los mismos a través de la estimación por máxima verosimilitud.

3.- Etapa de CHEQUEO: en esta etapa se comprueba la correcta especificación del modelo. Para ello comprobaremos el cumplimiento de las estimaciones de los parámetros de las condiciones de estacionariedad e invertibilidad y el análisis de los residuos, esto es, si puede afirmarse que son semejantes a un ruido blanco.

4.- Etapa de PREDICCIÓN: una vez superada la etapa de chequeo, el modelo ya puede ser utilizado para la obtención de predicciones puntuales y/o por intervalo de la serie objeto de estudio.

4.-Aplicación de la metodología de Box – Jenkins a la serie objeto de estudio

Identificación de los posibles procesos estocásticos lineales discretos

Como hemos explicado antes, uno de los requisitos que debe cumplir la serie objeto de estudio es que sea estacionaria, en este contexto la estacionariedad se define como un proceso que se mantiene constante a lo largo del tiempo y no muestra ninguna tendencia

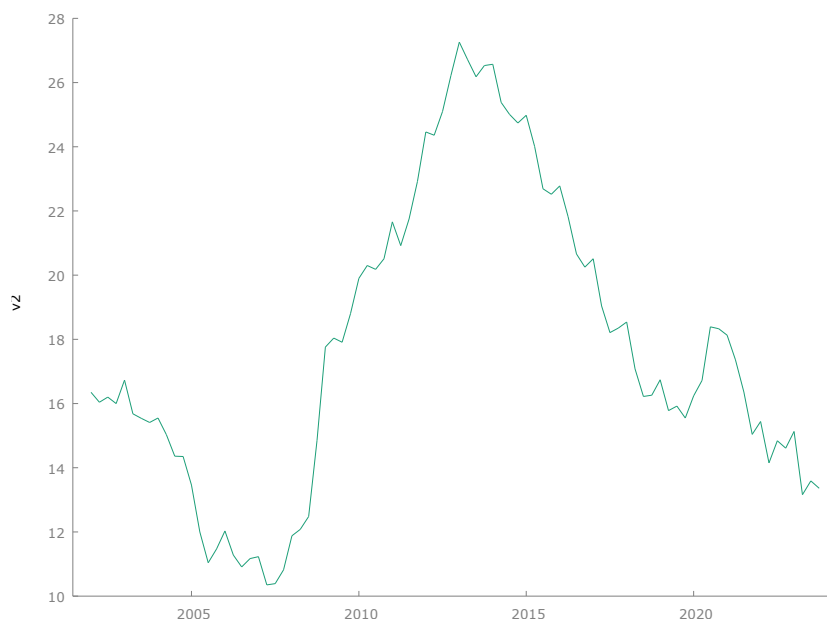
o patrón discernible en términos de media, varianza o covarianza. Por lo tanto, se puede distinguir dos tipos de procesos en función de la estacionariedad:

- Proceso estacionario en sentido estricto: la función de distribución conjunta del proceso se mantiene invariante ante desplazamientos en el tiempo.
- Proceso estacionario en sentido débil: se verifica simultáneamente las siguientes tres condiciones sobre los momentos de primer y segundo orden:
 - $E[y_t] = \mu \quad \forall t$
 - $V[y_t] = \sigma^2 \quad \forall t$
 - $\text{Cov}[y_t, y_{t \pm s}] = \gamma_s \quad \forall t$

En nuestro caso vamos a verificar que se cumple que la serie temporal sea estacionaria en media para ello tiene que mantenerse el supuesto de que existe una única media para toda la serie temporal, es decir cuando esta fluctúe en torno a una única media. Y que sea estacionaria en varianza, es decir que la dispersión de la serie permanezca invariable a lo largo del tiempo y que las covarianzas sólo dependen de la diferencia temporal, s pero no de t.

Una vez explicados estos conceptos, vamos a analizar si nuestra serie temporal objeto de estudio es estacionaria, para ello vamos a utilizar gráficos, correlogramas de la serie original y de distintas transformaciones de esta y el contraste del orden de integración de Dickey – Fuller.

Figura 2: Gráfico de la tasa de paro femenina en España (1T de 2002 – 4T de 2022)



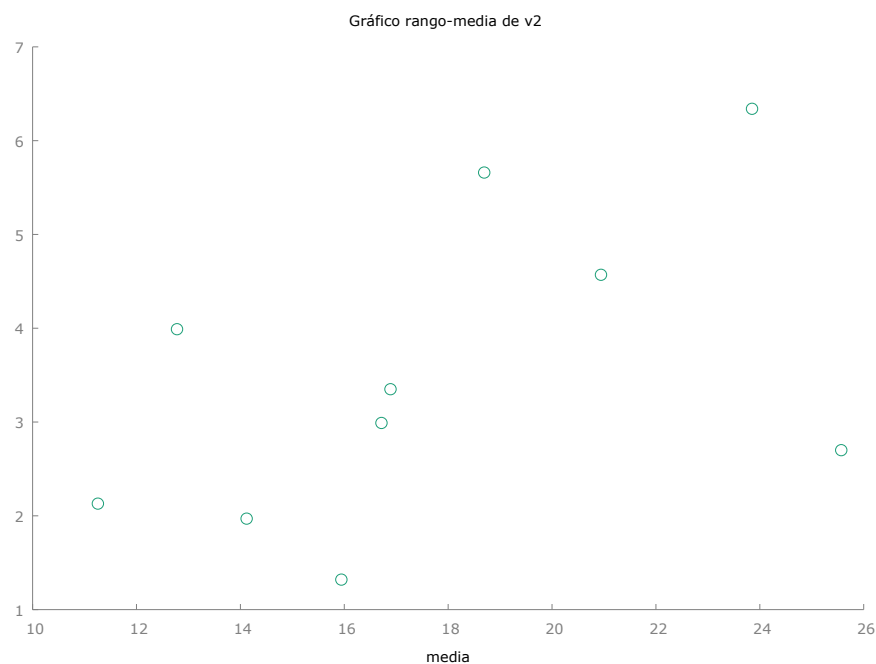
Observando la Figura 2 diríamos que la serie original no es estacionaria en media y presenta una tendencia estocástica ya que presenta periodos de crecimiento y decrecimiento a lo largo del tiempo, por lo que habría que diferenciar la serie el número de veces que fuera necesario hasta que desaparezca dicha tendencia.

En la figura se ve como antes de que estallara la crisis de 2008, la tasa de desempleo estaba disminuyendo como consecuencia del crecimiento económico que se estaba experimentando en aquella época, además del empleo que se estaba creando en sectores como la construcción y el turismo. También hay que destacar que, tras la crisis financiera de 2008, la tasa de paro se dispara llegando a alcanzar un 26%. Por lo que se puede observar lo que se conoce en econometría como cambio estructural, que no es más que cuando existe un cambio inesperado en una serie temporal macroeconómica.

Por último, se observa como a partir de 2014 se empieza a recuperar la economía, al igual que empieza a disminuir gradualmente la tasa de desempleo hasta llegar a 2020, año en el que comienza la pandemia del covid-19 causando un aumento de la tasa de paro como consecuencia del confinamiento y las medidas restrictivas que afectaron a mucho de los sectores clave en el mercado laboral español como la hostelería o el turismo.

En la Figura 3 presentamos el diagrama rango-media con el cuál se pretende analizar la estacionariedad en varianza.

Figura 3: Gráfico rango – media de la serie original



Este gráfico ajusta una recta y se contrasta si la pendiente es significativamente distinta de cero. Si el coeficiente de la recta estimada es significativamente distinto de cero, entonces la serie no es estacionaria en varianza.

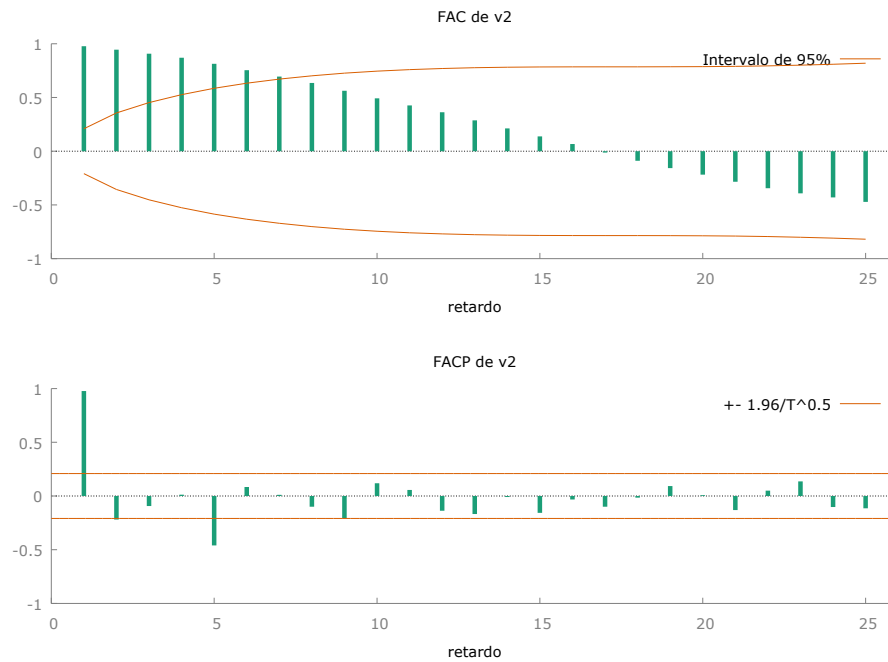
Figura 4: Estadísticos y contraste de rango – media de la serie original

Estadísticos de rango-media para v2		
Utilizando 10 submuestras de tamaño 9		
	Rango	media
2002:1 - 2004:1	0.810000	15.9086
2004:2 - 2006:2	3.08000	12.7071
2006:3 - 2008:3	1.69000	11.2114
2008:4 - 2010:4	2.54000	18.9843
2011:1 - 2013:1	4.56000	23.7814
2013:2 - 2015:2	1.83000	25.6271
2015:3 - 2017:3	3.65000	21.0700
2017:4 - 2019:4	2.57000	16.6214
2020:1 - 2022:1	2.89000	16.9400
2022:2 - 2023:4	1.48000	14.1100

Pendiente de 'rango' con respecto a 'media' = 0.10026
 El valor p para H0: [Pendiente = 0 es 0.222201]

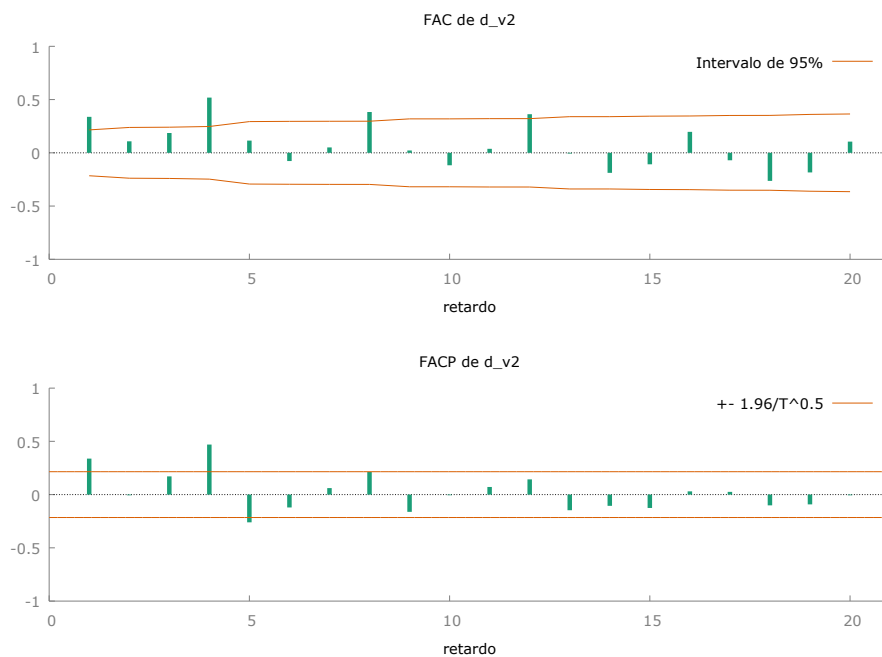
Como el p-valor es $0.222 > 0.05$ podemos concluir que la serie es estacionaria en varianza y por lo tanto no habrá que transformarla con logaritmos. Sin embargo, sí que hay que diferenciarla, el correlograma de la serie original y de sus diferencias nos puede también ayudar a determinar el número de diferencias que tenemos que aplicar a la serie para convertirla en estacionaria.

Figura 5: Correlograma de la serie original



En la Figura 5 se presenta el correlograma y el correlograma parcial de la serie original, como la serie no es estacionaria la FACM disminuye muy lentamente y la FACPM tiene el primer coeficiente muy grande, cercano a 1 y los demás coeficientes son prácticamente nulos.

Figura 6: Correlograma de la serie diferenciada una vez



Por último, como la tasa de paro es una variable que puede presentar un comportamiento estacional, debido a que en España hay épocas como el verano en la que disminuye la tasa de desempleo, vamos a analizar si la serie presenta este comportamiento estacional. Para ello observamos los correlogramas de la serie diferenciada, presentados en la Figura 6, en el que podemos observar un patrón repetitivo en el cuarto trimestre, es decir, un comportamiento estacional. Para poder seguir con el análisis del orden de integración de la serie temporal con el contraste de Dickey-Fuller, que está desarrollado para series sin componente estacional, el primer paso que vamos a realizar es desestacionalizar la serie para ello se ha utilizado la función X13ARIMA de Gretl.

Figura 7: Gráfico de la tasa de paro femenina en España desestacionalizada (1T de 2002 – 4T de 2022)



Si comparamos la Figura 7 que es el gráfico de la serie desestacionalizada con la figura 2, no observamos grandes cambios a primera vista, vemos que la serie desestacionalizada sigue sin ser estacionaria en media y que al igual que antes presenta periodos de crecimiento y decrecimiento a lo largo del tiempo, por lo que habría que diferenciar la serie el número de veces que fuera necesario hasta que desaparezca dicha tendencia.

En la Figura 8 presentamos el diagrama rango-media con el cuál analizamos la estacionariedad en varianza.

Figura 8: Gráfico rango – media de la serie desestacionalizada

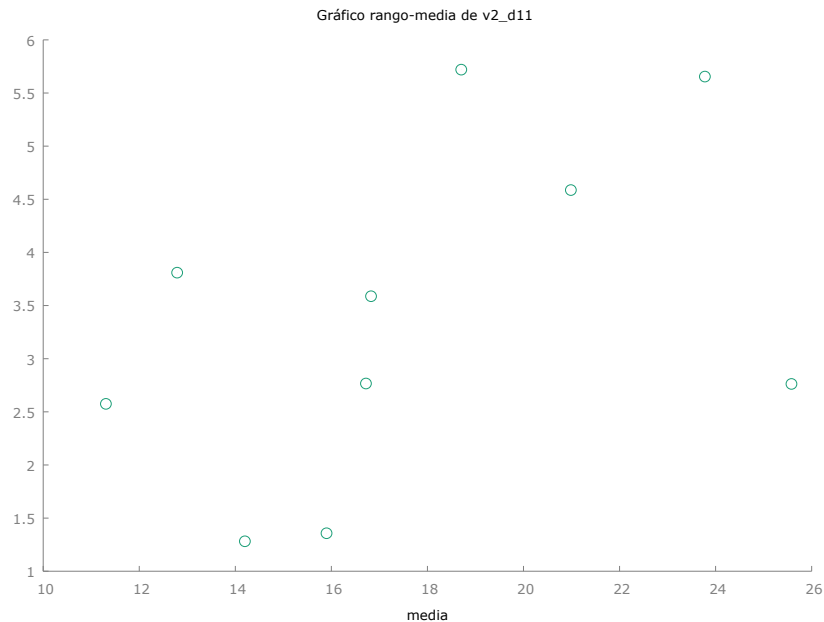


Figura 9: Estadísticos y contraste de rango – media de la serie desestacionalizada

Estadísticos de rango-media para v2_d11
Utilizando 10 submuestras de tamaño 9

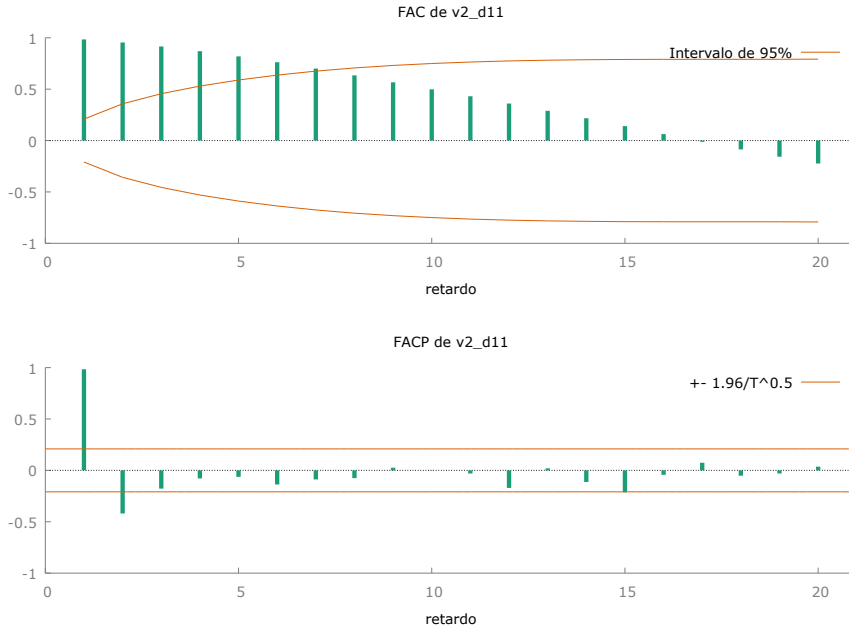
	Rango	media
2002:1 - 2004:1	0.992655	15.9250
2004:2 - 2006:2	3.27741	12.6458
2006:3 - 2008:3	1.45297	11.1929
2008:4 - 2010:4	3.50401	18.9617
2011:1 - 2013:1	5.23800	23.7645
2013:2 - 2015:2	2.21100	25.6114
2015:3 - 2017:3	3.43794	21.0259
2017:4 - 2019:4	1.95940	16.6221
2020:1 - 2022:1	3.20093	16.8676
2022:2 - 2023:4	1.22183	14.2316

Pendiente de 'rango' con respecto a 'media' = 0.145824
El valor p para H0: [Pendiente = 0 es 0.125472]

Como el p-valor es $0.125 > 0.05$ podemos concluir que la serie desestacionalizada es estacionaria en varianza.

En la figura 10 se presenta el correlograma de la serie desestacionalizada, el cual sigue siendo el propio de una serie no estacionaria, al disminuir la FACM muy lentamente y al tener la FACPM un corte brusco en los primeros coeficientes.

Figura 10: Correlograma de la serie desestacionalizada



Por último, vamos a realizar el contraste de Dickey-Fuller para la serie desestacionalizada. Este contraste nos va a servir para determinar el orden de integración en la parte regular de la serie, que determina el número de veces que es necesario diferenciar una serie para eliminar la tendencia estocástica en la parte regular.

En primer lugar, elegimos el modelo de referencia que mejor se ajuste a nuestra serie y se obtiene la ecuación de contraste asociada.

$$M_B: y_t = \delta + \phi_1 y_{t-1} + u_t$$

$$\Delta y_t = \delta + (\phi_1 - 1)y_{t-1} + u_t \rightarrow \Delta y_t = \delta + \phi_1^* y_{t-1} + u_t$$

Se establece la hipótesis nula y alternativa y el estadístico de contraste

H0: $y_t \sim I(1)$ la serie desestacionalizada presenta al menos una raíz unitaria

Ha: $y_t \sim I(0)$ la serie desestacionalizada no presenta raíz unitaria

$$t_{DF}: \frac{\hat{\phi}_1 - 1}{\hat{\sigma}_{\hat{\phi}_1}}$$

Figura 11: Contraste Dickey – Fuller para la serie desestacionalizada

```
Contraste aumentado de Dickey-Fuller para v2_d11
contrastar hacia abajo desde 9 retardos, con el criterio estadístico t
tamaño muestral 78
la hipótesis nula de raíz unitaria es: [a = 1]

con constante y tendencia
incluyendo 9 retardos de (1-L)v2_d11
modelo: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valor estimado de (a - 1): -0.0287093
estadístico de contraste: tau_ct(1) = -1.87714
valor p asintótico 0.6663
Coef. de autocorrelación de primer orden de e: -0.013
diferencias retardadas: F(9, 66) = 8.238 [0.0000]
```

Como el p- valor es $0.6663 > 0.05$ se acepta la hipótesis nula, al menos la serie es $I(1)$, por lo tanto, se confirma que la serie no es estacionaria en varianza, o que tiene una tendencia estocástica.

Volvemos a realizar el contraste de Dickey- Fuller, pero usando las primeras diferencias de la variable para comprobar si la serie podría ser $I(2)$ o si por el contrario se confirma que es $I(1)$.

Figura 12: Contraste Dickey – Fuller para la serie desestacionalizada diferenciada

```
Contraste aumentado de Dickey-Fuller para d_v2_d11
contrastar hacia abajo desde 5 retardos, con el criterio estadístico t
tamaño muestral 86
la hipótesis nula de raíz unitaria es: [a = 1]

contraste sin constante
incluyendo 0 retardos de (1-L)d_v2_d11
modelo: (1-L)y = (a-1)*y(-1) + e
valor estimado de (a - 1): -0.348138
estadístico de contraste: tau_nc(1) = -4.23834
valor p asintótico 2.382e-05
Coef. de autocorrelación de primer orden de e: -0.076
```

Como el p- valor es $0.0062 < 0.05$ no se acepta la hipótesis nula de que al menos la serie es $I(2)$, por lo tanto se confirma que la serie es $I(1)$ y solo hay que realizar una diferencia regular.

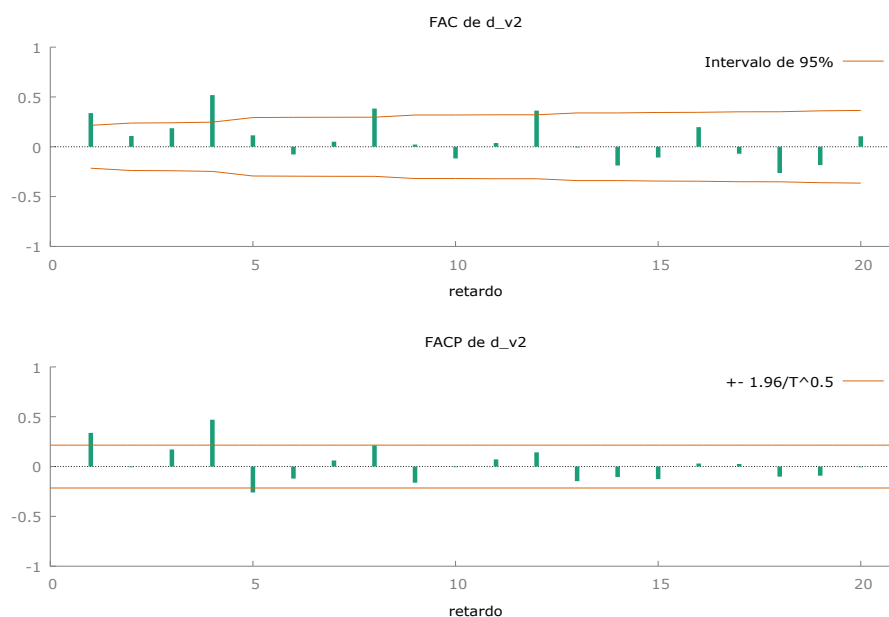
Para analizar si es necesario diferenciar estacionalmente la serie utilizamos al gráfico de la serie sin desestacionalizar, con una diferencia regular, Figura 13.

Figura 13: Gráfico de la serie original diferenciada



Aparentemente la representación gráfica de la serie diferenciada, que puede verse en la figura 13, evoluciona como una serie estacionaria en media, ya que, sí que parece que fluctúa en torno a una media, pero presenta una tendencia estocástica ya que sigue habiendo periodos de crecimiento y decrecimiento.

Figura 14: Correlograma de la serie original diferenciada



En el correlograma representado en la figura 14, se ve que hay un patrón repetitivo por el componente estacional de la tasa de desempleo y que la parte estacional no se anula tan rápidamente como una serie estacionaria, por lo tanto, volvemos a transformar la serie, esta vez con una diferencia estacional. La serie transformada mediante una diferencia regular ($d=1$) y una diferencia estacional ($D=1$) quedaría así:

$$w_t = (1 - L^4)(1 - L)y_t$$

Figura 15: Gráfico de la serie original con una diferencia regular y una estacional

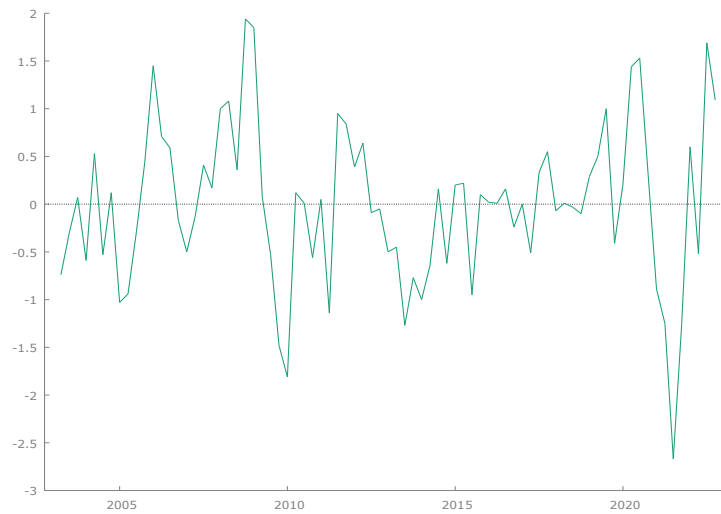
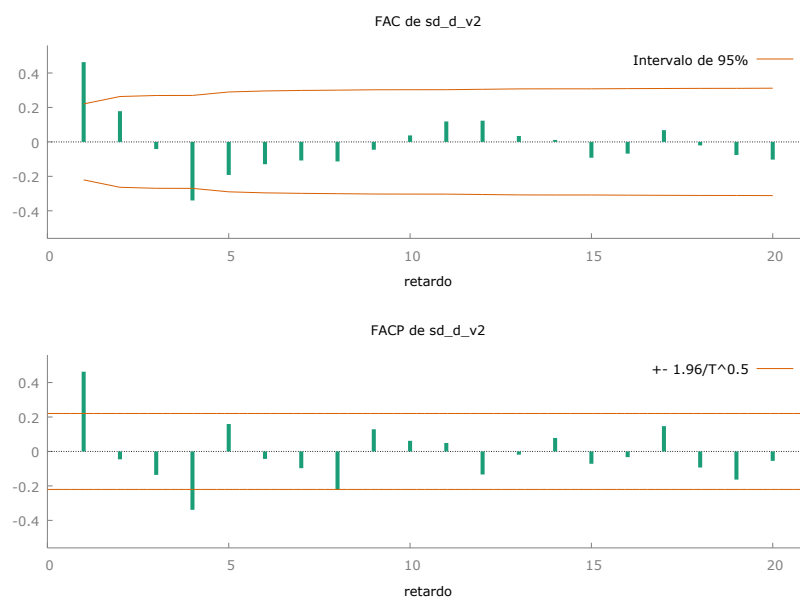


Figura 16: Gráfico de la serie original con una diferencia regular y una estacional



Para la identificación del Modelo ARIMA (p, d, q) x ARIMA (P, D, Q), se ha de realizar un estudio de los correlogramas de la figura 16.

Observando la parte regular llegamos a la conclusión de que $p=1$ y $q=0$ ya que el comportamiento es típico de un esquema AR (1), es decir, un corte brusco en la FACP con un primer valor muy significativo y decrecimiento rápido en la FAC. Por lo tanto, la serie seguiría un proceso ARI (1,1). Sin embargo, si observamos la parte estacional no llegamos a una conclusión tan clara, por lo tanto, podemos identificar un $Q=1$ o $Q=2$ o $P=1$, en cualquier caso, $D=1$. Por tanto, identificamos tres modelos multiplicativos estacionales alternativos:

$$\text{ARIMA}(p, d, q) \times \text{ARIMA}(P, D, Q)_4$$

- Modelo 1: ARIMA (1,1,0) x ARIMA (0,1,1)₄
- Modelo 2: ARIMA (1,1,0) x ARIMA (1,1,0)₄
- Modelo 3: ARIMA (1,1,0) x ARIMA (0,1,2)₄

Antes de realizar la estimación de los tres modelos identificados, vamos a discutir si hay que incluir término independiente, para ello observamos la Figura 15, en principio, como la serie deambula alrededor de cero, los modelos no tendrían término constante, aunque para confirmar esto realizamos el contraste de si la media poblacional o esperanza es cero:

$$H_0: E[w_t] = 0 \quad \delta = 0$$

$$H_a: E[w_t] \neq 0 \quad \delta \neq 0$$

Figura 17: Contraste de si la media poblacional es cero

```
Hipótesis nula: [Media poblacional = 0]
Tamaño muestral: n = 79
Media muestral = -0.0102532, desv. típica = 0.834567
Estadístico de contraste: t(78) = (-0.0102532 - 0)/0.0938962 = -0.109197
Valor p a dos colas = 0.9133
(a una cola = 0.4567)
```

Como $t_{\bar{w}} = \frac{\bar{w}}{\sqrt{V(\bar{w})}} \sim N(0; 1)$ es igual a $-0.109 < N_{0.025} = 1.96$ se acepta la hipótesis nula y por lo tanto los modelos no deberían incorporar término independiente.

Estimación de los procesos estocásticos

Primer modelo

$$\text{ARIMA}(1,1,0) \times \text{ARIMA}(0,1,1)_4: (1 - L^4)(1 - L)(1 - \phi_1 L)y_t = (1 - \Theta_1 L^4)u_t$$

En el siguiente cuadro mostramos la estimación máximo-verosímil del Modelo 1:

Cuadro 1: Estimación del Modelo 1

Modelo 1: ARIMA, usando las observaciones 2003:2-2023:4 (T = 83)				
Estimado usando AS 197 (MV exacta)				
Variable dependiente: (1-L) (1-Ls) v2				
Desviaciones típicas basadas en el Hessiano				
	coeficiente	Desv. típica	z	valor p

phi_1	0.521612	0.0952593	5.476	4.36e-08 ***
Theta_1	-0.670216	0.125680	-5.333	9.67e-08 ***
Media de la vble. dep.	-0.019639	D.T. de la vble. dep.	0.817849	
Media de innovaciones	-0.017181	D.T. innovaciones	0.628264	
R-cuadrado	0.983946	R-cuadrado corregido	0.983748	
Log-verosimilitud	-80.49461	Criterio de Akaike	166.9892	
Criterio de Schwarz	174.2457	Crit. de Hannan-Quinn	169.9045	
	Real	Imaginaria	Módulo	Frecuencia

AR				
Raíz 1	1.9171	0.0000	1.9171	0.0000
MA (estacional)				
Raíz 1	1.4921	0.0000	1.4921	0.0000

Segundo modelo

$$\text{ARIMA}(1,1,0) \times \text{ARIMA}(1,1,0)_4: (1 - \Phi L^4)(1 - \phi_1 L)(1 - L^4)(1 - L)y_t = u_t$$

En el siguiente cuadro mostramos la estimación máximo-verosímil del Modelo 2:

Cuadro 2: Estimación del Modelo 2

Modelo 2: ARIMA, usando las observaciones 2003:2-2023:4 (T = 83)
 Estimado usando AS 197 (MV exacta)
 Variable dependiente: (1-L)(1-Ls) v2
 Desviaciones típicas basadas en el Hessiano

	coeficiente	Desv. típica	z	valor p	
phi_1	0.480117	0.0962983	4.986	6.17e-07	***
Phi_1	-0.358943	0.101582	-3.534	0.0004	***
Media de la vble. dep.	-0.019639	D.T. de la vble. dep.		0.817849	
Media de innovaciones	-0.009788	D.T. innovaciones		0.669673	
R-cuadrado	0.981839	R-cuadrado corregido		0.981615	
Log-verosimilitud	-84.87970	Criterio de Akaike		175.7594	
Criterio de Schwarz	183.0159	Crit. de Hannan-Quinn		178.6747	

	Real	Imaginaria	Módulo	Frecuencia
AR				
Raíz 1	2.0828	0.0000	2.0828	0.0000
AR (estacional)				
Raíz 1	-2.7860	0.0000	2.7860	0.5000

Tercer modelo

ARIMA (1,1,0) x ARIMA (0,1,2)₄: $(1 - L^4)(1 - L)(1 - \phi_1 L)y_t = (1 - \theta_1 L^4 - \theta_2 L^8)u_t$

En el siguiente cuadro mostramos la estimación máximo-verosímil del Modelo 3:

Cuadro 3: Estimación del Modelo 3

Modelo 3: ARIMA, usando las observaciones 2003:2-2023:4 (T = 83)
 Estimado usando AS 197 (MV exacta)
 Variable dependiente: (1-L)(1-Ls) v2
 Desviaciones típicas basadas en el Hessiano

	coeficiente	Desv. típica	z	valor p	
phi_1	0.505605	0.0957247	5.282	1.28e-07	***
Theta_1	-0.594657	0.117942	-5.042	4.61e-07	***
Theta_2	-0.189706	0.128664	-1.474	0.1404	
Media de la vble. dep.	-0.019639	D.T. de la vble. dep.		0.817849	
Media de innovaciones	-0.023080	D.T. innovaciones		0.616568	
R-cuadrado	0.984190	R-cuadrado corregido		0.983794	
Log-verosimilitud	-79.43447	Criterio de Akaike		166.8689	
Criterio de Schwarz	176.5443	Crit. de Hannan-Quinn		170.7560	

	Real	Imaginaria	Módulo	Frecuencia
AR				
Raíz 1	1.9778	0.0000	1.9778	0.0000
MA (estacional)				
Raíz 1	1.2126	0.0000	1.2126	0.0000
Raíz 2	-4.3472	0.0000	4.3472	0.5000

Chequeo

Una vez estimados los parámetros de los tres procesos estocásticos alternativos y siguiendo las etapas de la metodología Box-Jenkins, se trata de responder a la cuestión de si dichos modelos son adecuados.

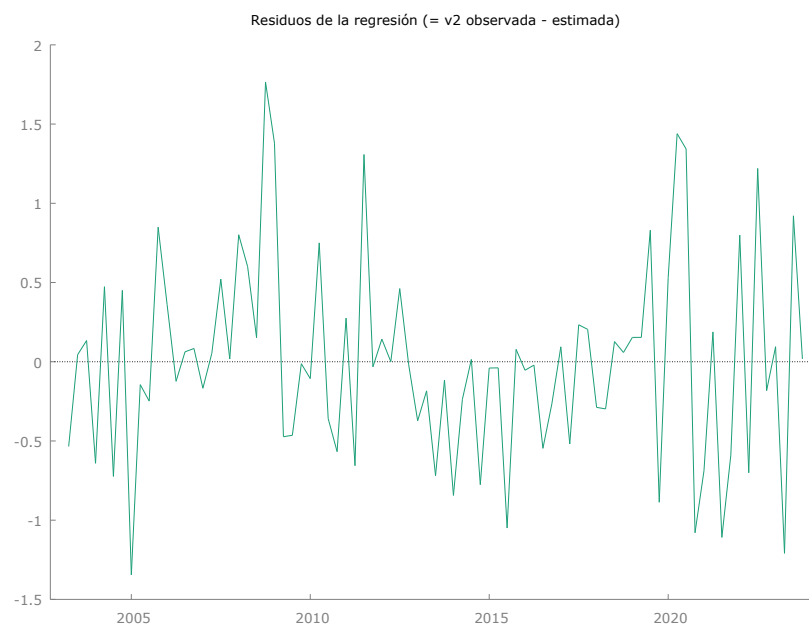
Para poder afirmar que los modelos resultan adecuados deben de cumplir: que las estimaciones de sus parámetros sean significativas, las condiciones de estacionariedad e invertibilidad y que los residuos se comporten como un ruido blanco, es decir que cumplan las condiciones de media nula, varianza constante u homocedasticidad, no autocorrelación y distribución normal.

MODELO 1

Empezamos estudiando si los residuos del Modelo 1 se comportan como un ruido blanco.

- Media nula

Figura 18: Evolución de la serie de residuos del modelo 1



Para que la media de los residuos fuera nula, estos deberían fluctuar en torno a cero, sin embargo, en la Figura 18 vemos como estos llegan a alcanzar tanto pico por arriba

llegando hasta 1.5 tanto por abajo llegando hasta -1.5 Por lo tanto, vamos a realizar un contraste de significatividad sobre el valor esperado de la serie de residuos:

$$H_0: E[\tilde{u}_t] = 0$$

$$H_A: E[\tilde{u}_t] \neq 0$$

Siendo el estadístico de contraste:

$$t = \frac{\bar{\tilde{u}}}{\sqrt{V(\tilde{u})}} \sim N(0,1)$$

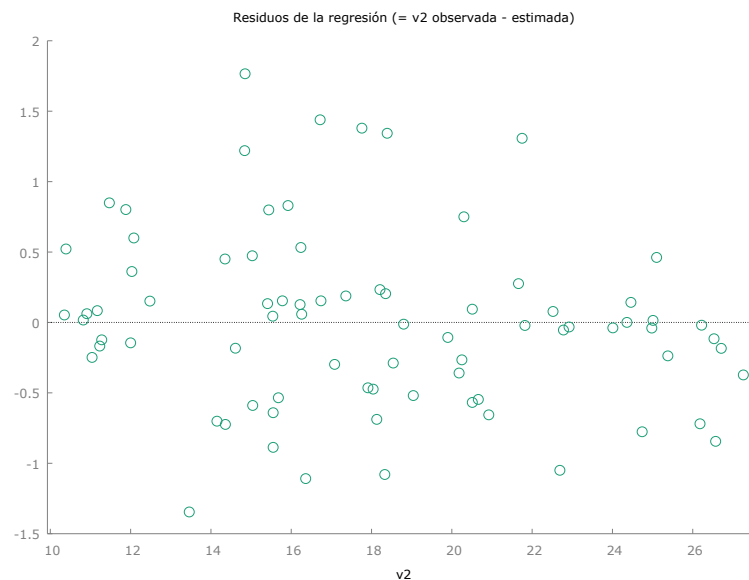
Figura 19: Contraste de significatividad de la media de los residuos del modelo 1

```
Hipótesis nula: [Media poblacional = 0]
Tamaño muestral: n = 79
Media muestral = -0.000856197, desv. típica = 0.624424
Estadístico de contraste: t(78) = (-0.000856197 - 0)/0.0702532 = -0.0121873
Valor p a dos colas = 0.9903
(a una cola = 0.4952)
```

En conclusión, podemos aceptar que la media no difiere significativamente de cero.

- Varianza constante

Figura 20: Gráfica dispersión de los residuos Modelo 1



Para que la varianza de los residuos fuera constante, la dispersión de la Figura 20 debería ser constante a lo largo del tiempo, es decir tendría que haber homocedasticidad. Como la dispersión se mantiene constante, pero hay algunos datos atípicos realizamos el contraste de ARCH para contrastar si no hay heterocedasticidad condicionada a un autorregresivo hasta el orden 4:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_4 u_{t-4}^2$$

$H_0: \sigma_t^2 = \alpha_0 \forall t \rightarrow$ Homocedasticidad

$H_A: \sigma_t^2 \neq \alpha_0 \forall t \rightarrow$ Heterocedasticidad

Con el estadístico de contraste:

$$LM = TR_{R.A.}^2 \sim \chi^2(p)$$

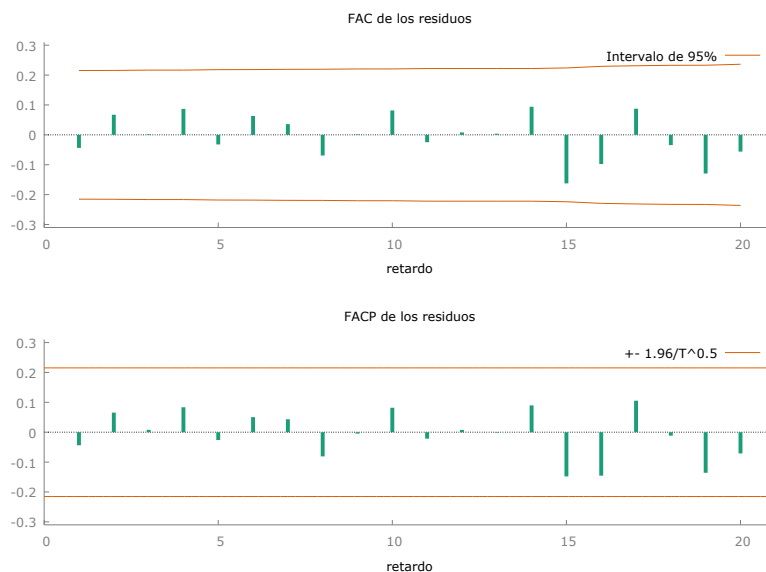
Figura 21: Contraste de ARCH Modelo 1

```
Hipótesis nula: [No hay efecto ARCH]
Estadístico de contraste: LM = 5.73323
con valor p = P(Chi-cuadrado(4) > 5.73323) = 0.219976
```

Como el p-valor $0.219 > 0.05$ se acepta la hipótesis nula de homocedasticidad.

- No autocorrelación

Figura 22: Correlograma de los residuos del Modelo 1



Con el correlograma muestral de los residuos ya podemos confirmar que los coeficientes de autocorrelación poblacional de los residuos no son significativamente distintos de cero, es decir, que se cumple la no autocorrelación de los residuos.

Esto se puede confirmar observando la Figura 22 ya que ningún valor de $r_j(\hat{u})$ supera las bandas de significatividad, para un nivel de significación del 5%.

Otra manera de comprobar la ausencia de autocorrelación es realizando el contraste de significatividad conjunta de Ljung-Box (1978).

Donde las hipótesis planteadas son:

$H_0: \rho_1(\tilde{u}_t) = \rho_2(\tilde{u}_t) = \dots = \rho_M(\tilde{u}_t) = 0 \rightarrow$ No existe autocorrelación

$H_A: \text{Algún } \rho_j(\tilde{u}_t) \neq 0 \rightarrow$ Existe autocorrelación

El estadístico de contraste es:

$$Q^* = T(T+2) \sum_{j=1}^M \frac{r_j^2(\tilde{u}_t)}{(T-j)} \sim \chi^2(M-k)$$

Figura 23: Contraste de Ljung – Box Modelo 1

```
Contraste de autocorrelación hasta el orden 4

Ljung-Box Q' = 1.19022,
con valor p = P(Chi-cuadrado(2) > 1.19022) = 0.5515
```

Atendiendo a los resultados del contraste de Ljung-Box de no autocorrelación hasta el orden 4 en la Figura 23, dado que el p-valor es $0.552 > 0.05$ no se rechaza la hipótesis nula de no autocorrelación, para un nivel de significación del 5%.

Por último, para terminar de confirmar esta condición, observamos la función de autocorrelación de los residuos.

Figura 24: Función de autocorrelación de los residuos del Modelo 1

Función de autocorrelación de los residuos
 ***, ** y * indica significatividad a los niveles del 1%, 5% y 10%
 utilizando desviaciones típicas de Bartlett para la FAC

RETARDO	FAC	FACP	Estad-Q. [valor p]	
1	-0.0438	-0.0438		
2	0.0662	0.0644		
3	0.0023	0.0078	0.5477	[0.459]
4	0.0848	0.0815	1.1902	[0.552]
5	-0.0323	-0.0263	1.2844	[0.733]
6	0.0625	0.0503	1.6427	[0.801]
7	0.0363	0.0438	1.7651	[0.881]
8	-0.0699	-0.0811	2.2244	[0.898]
9	0.0023	-0.0045	2.2249	[0.946]
10	0.0822	0.0825	2.8773	[0.942]
11	-0.0237	-0.0205	2.9324	[0.967]
12	0.0081	0.0075	2.9389	[0.983]
13	0.0044	-0.0014	2.9409	[0.991]
14	0.0949	0.0908	3.8617	[0.986]
15	-0.1616	-0.1471	6.5710	[0.923]
16	-0.0971	-0.1447	7.5634	[0.911]
17	0.0888	0.1064	8.4071	[0.906]
18	-0.0333	-0.0101	8.5273	[0.932]
19	-0.1282	-0.1354	10.3386	[0.889]
20	-0.0549	-0.0707	10.6758	[0.908]

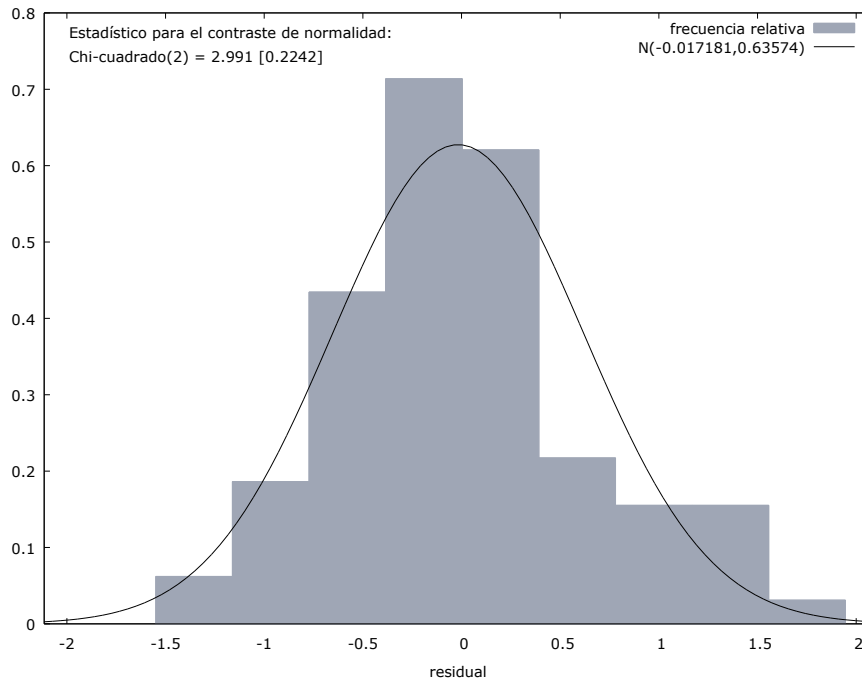
Dado que para cualquier retardo el contraste de Ljung-Box presenta unos p- valores muy grandes, esto nos lleva a no rechazar la hipótesis de no autocorrelación, para cualquier nivel de significación de los normalmente utilizados.

- Distribución Normal

La última condición que tiene que cumplir el modelo para pasar la etapa de chequeo en lo referente a la distribución de las perturbaciones es que estas se distribuyan de acuerdo a una Normal.

$$u_t \sim N(0, \sigma^2)$$

Figura 25: Gráfico contraste de normalidad de los residuos del Modelo 1



Observando la Figura 24 vemos como los residuos se puede considerar que se aproximan a una distribución normal.

Otra manera de comprobar si los residuos se distribuyen como una normal será calculando los coeficientes de asimetría y curtosis y realizar el contraste de Jarque-Bera (1980).

$$g_1 = \frac{m_3}{m_2^{3/2}} = 0.44879 ; \quad g_2 = \frac{m_4}{m_2^2} - 3 = 0.28642$$

$H_0: g_1 = g_2 = 0 \rightarrow$ Normalidad

H_A : No normalidad

Y siendo el estadístico de contraste y la distribución de este la siguiente:

$$LM_N = T\left(\frac{g_1^2}{6} + \frac{g_2^2}{24}\right) \sim X^2(2)$$

Figura 26: Contraste de normalidad de los residuos del Modelo 1

```
Contraste de Normalidad de uhat1:  
  
Contraste de Doornik-Hansen = 2.99083, con valor p 0.224156  
  
W de Shapiro-Wilk = 0.971581, con valor p 0.0625802  
  
Contraste de Lilliefors = 0.124673, con valor p ~ = 0  
  
Contraste de Jarque-Bera = 3.01594, con valor p 0.221359
```

Como el p-valor $0.2212 > 0.05$ se acepta la hipótesis nula de normalidad.

Por último, para analizar si los coeficientes estimados son significativos tenemos que efectuar contrastes de significatividad de dichos coeficientes del modelo.

$$H_0: \phi_1 = 0 \qquad H_0: \theta_1 = 0$$

$$H_A: \phi_1 \neq 0 \qquad H_A: \theta_1 \neq 0$$

Y como para ambos coeficientes el p-valor del t-ratio es muy pequeño se rechazan las hipótesis nulas individuales de que los parámetros estimados no son significativos individualmente.

Además, como este primer modelo tiene una parte AR y una MA, se deberá cumplir tanto la condición de estacionariedad como la de invertibilidad. Si el modelo no fuera estacionario habría que diferenciar una vez más y si no fuera invertible sería síntoma de sobrediferenciación.

Para que el modelo sea estacionario se tiene que cumplir que $|\hat{\phi}_1| < 1$ y para que sea invertible se tiene que cumplir que $|\hat{\theta}_1| < 1$, por lo tanto, ambas condiciones se cumplen ya que $0.522 < 1$ y $0.670 < 1$.

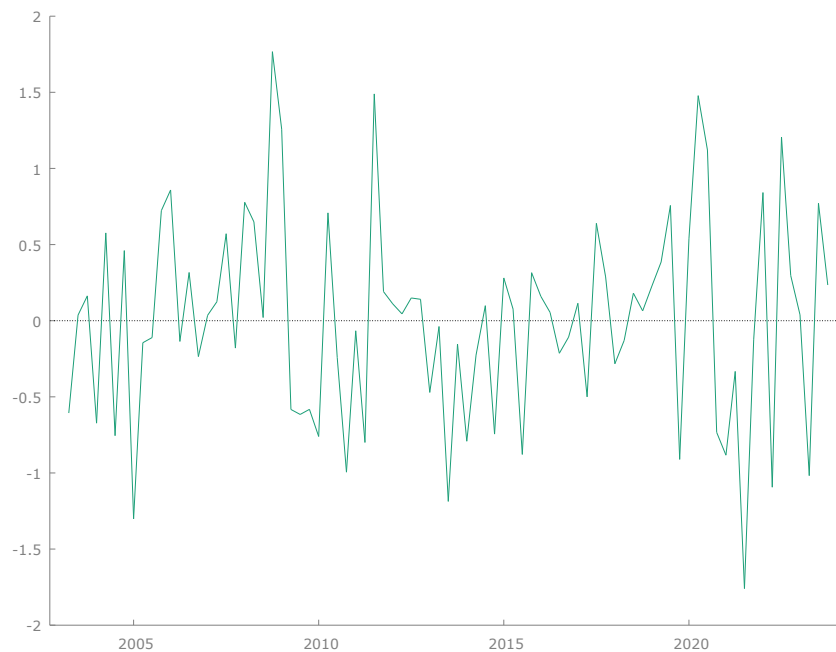
Otra manera de comprobar estas condiciones es ver que las raíces características de los polinomios, tanto el correspondiente a la parte autorregresiva como el de medias móviles, caen fuera del círculo unitario ya que $1.917 > 1$ y $1.492 > 1$.

Modelo 2

Empezamos estudiando si los residuos del Modelo 2 se comportan como un ruido blanco.

- Media nula

Figura 27: Evolución de la serie de residuos del modelo 2



Como en la figura 26, los residuos sí que fluctúan en torno a cero, pero llegan a alcanzar en algunos puntos pico hasta 2 y -2 realizamos el contraste de significatividad sobre el valor esperado de la serie de residuos.

$$H_0: E[\tilde{u}_t] = 0$$

$$H_A: E[\tilde{u}_t] \neq 0$$

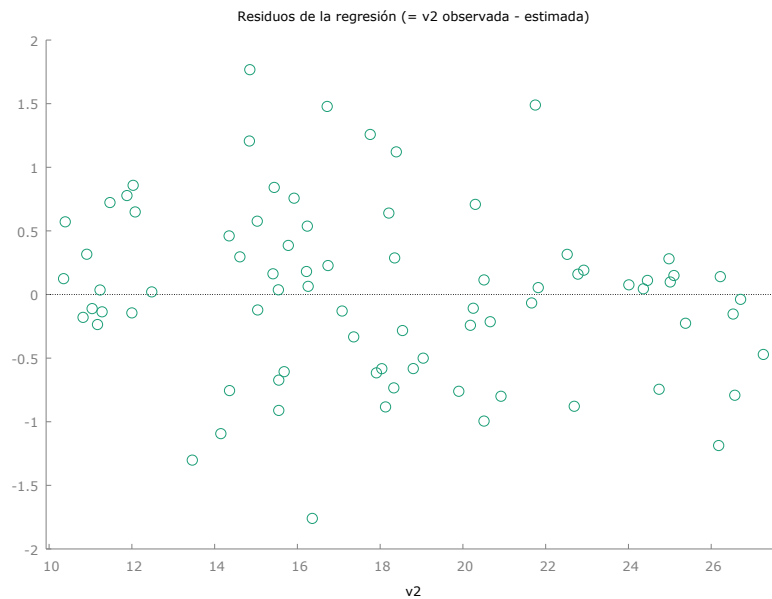
Figura 28: Contraste de significatividad de la media de los residuos del modelo 2

```
Hipótesis nula: [Media poblacional = 0]
Tamaño muestral: n = 79
Media muestral = -0.00069567, desv. típica = 0.674895
Estadístico de contraste: t(78) = (-0.00069567 - 0)/0.0759316 = -0.00916179
Valor p a dos colas = 0.9927
(a una cola = 0.4964)
```

Tras este contraste aceptamos la hipótesis nula de media nula.

- Varianza constante

Figura 29: Gráfico de dispersión de los residuos del Modelo 2



Observando la figura 29, no podemos concluir que la dispersión de los residuos se mantenga constante a lo largo del tiempo ya que existen datos atípicos propios de momentos como la crisis financiera de 2008. Por lo tanto, para comprobar la presencia o no de homocedasticidad realizamos el contraste de ARCH.

$H_0: \sigma_t^2 = \alpha_0 \forall t \rightarrow$ Homocedasticidad

$H_A: \sigma_t^2 \neq \alpha_0 \forall t \rightarrow$ Heterocedasticidad

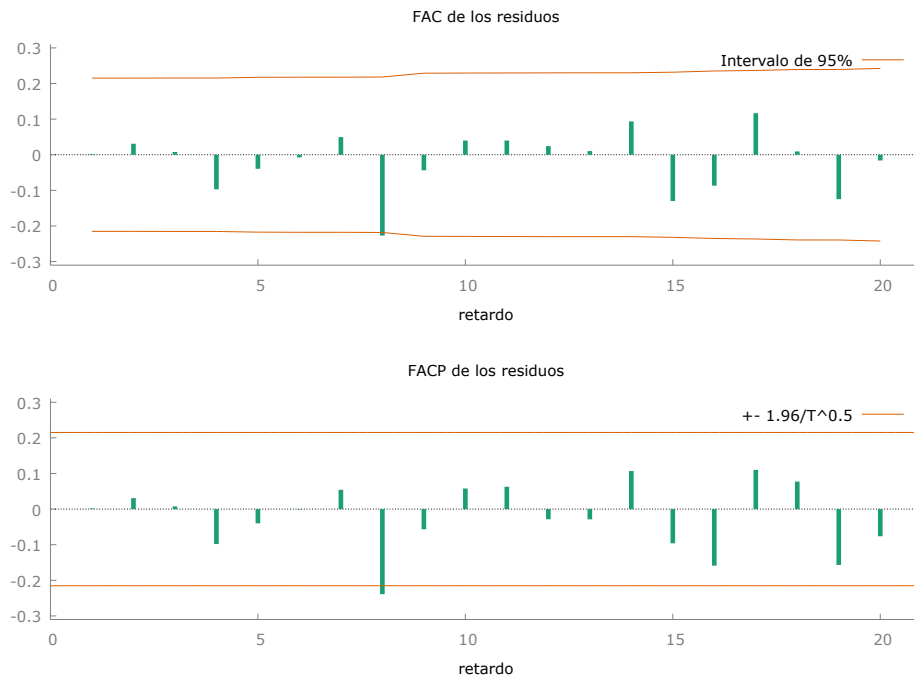
Figura 30: Contraste ARCH Modelo 2

```
Hipótesis nula: [No hay efecto ARCH]
Estadístico de contraste: LM = 3.16111
con valor p = P(Chi-cuadrado(4) > 3.16111) = 0.531234
```

Como el p-valor $0.531 > 0.05$ se acepta la hipótesis nula de homocedasticidad.

- No autocorrelación

Figura 31: Correlograma de los residuos del Modelo 2



Observando la figura 31 vemos que hay algún valor de $r_j(\hat{u})$ que supera las bandas de significatividad. Esto no quiere decir que no se vaya a cumplir la no autocorrelación.

Para comprobar esto realizamos el contraste de Ljung – Box, para $M = 4$:

$H_0: \rho_1(\tilde{u}_t) = \rho_2(\tilde{u}_t) = \dots = \rho_M(\tilde{u}_t) = 0 \rightarrow$ No existe autocorrelación

$H_A: \text{Algún } \rho_j(\tilde{u}_t) \neq 0 \rightarrow$ Existe autocorrelación

Figura 32: Contraste de Ljung – Box Modelo 2

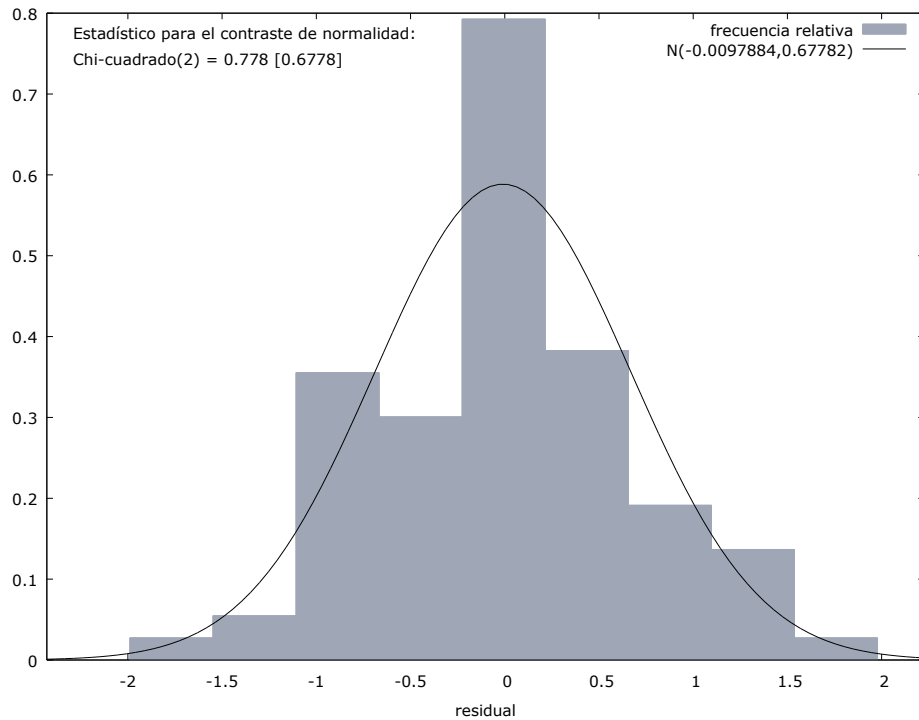
```
Contraste de autocorrelación hasta el orden 4

Ljung-Box Q' = 0.92629,
con valor p = P(Chi-cuadrado(2) > 0.92629) = 0.6293
```

Como el p-valor es $0.629 > 0.05$ se acepta la hipótesis nula de no autocorrelación.

- Distribución Normal

Figura 33: Gráfico contraste de Normalidad de los residuos del Modelo 2



Observando la figura 33 vemos como los residuos sí que se aproximan a una distribución normal. Para confirmar esto calculamos los coeficientes de asimetría y curtosis y realizamos el contraste de Jarque-Bera (1980)

$$g_1 = \frac{m_3}{m_2^{3/2}} = 0.119; \quad g_2 = \frac{m_4}{m_2^2} - 3 = 0.103$$

$H_0: g_1 = g_2 = 0 \rightarrow$ Normalidad

H_A : No normalidad

Y siendo el estadístico de contraste y la distribución de este:

$$LM_N = T\left(\frac{g_1^2}{6} + \frac{g_2^2}{24}\right) \sim X^2(2)$$

Figura 34: Contraste de Normalidad de los residuos del Modelo 2

```
Contraste de Normalidad de uhat2:

Contraste de Doornik-Hansen = 0.777764, con valor p 0.677814

W de Shapiro-Wilk = 0.988191, con valor p 0.653476

Contraste de Lilliefors = 0.0781179, con valor p ~= 0.23

Contraste de Jarque-Bera = 0.234007, con valor p 0.889582
```

Como el p-valor $0.889 > 0.05$ se acepta la hipótesis nula de normalidad.

Por último, para analizar si los coeficientes estimados son significativos tenemos que efectuar contrastes de hipótesis respecto a dichos coeficientes del modelo.

$$H_0: \phi_1 = 0 \qquad H_0: \Phi_1 = 0$$

$$H_A: \phi_1 \neq 0 \qquad H_A: \Phi_1 \neq 0$$

Y como para ambos coeficientes el p-valor es muy pequeño se rechazan las hipótesis nulas de que los parámetros estimados no son significativos.

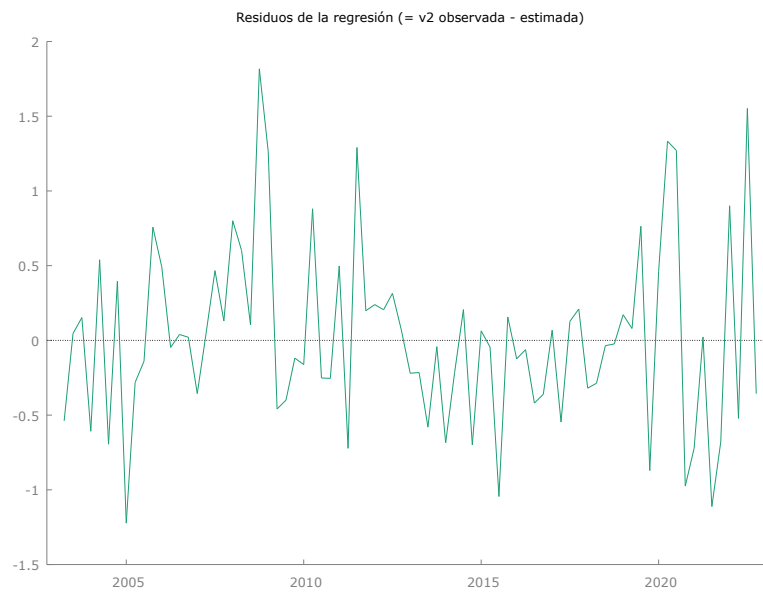
Este segundo modelo solo tiene parte AR por lo tanto será invertible, y para que se cumpla la condición de estacionariedad se tiene que cumplir que $|\hat{\phi}_1| < 1$ y $|\hat{\Phi}_1| < 1$. Dicha condición se cumple ya que $0.480 < 1$ y $0.358 < 1$. Otra manera de comprobar estas condiciones es ver que las raíces características de los polinomios, las correspondiente a la parte autorregresiva caen fuera del círculo unitario ya que $2.083 > 1$ y $2.786 > 1$.

Modelo 3

Empezamos estudiando si los residuos del Modelo 3 se comportan como un ruido blanco.

- Media nula

Figura 35: Evolución de la serie de residuos del modelo 3



En la Figura 35 los residuos fluctúan en torno a cero, pero con ciertos picos en algunos períodos por eso realizamos el contraste de si el valor esperado de la serie de residuos es nulo.

$$H_0: E[\tilde{u}_t] = 0$$

$$H_A: E[\tilde{u}_t] \neq 0$$

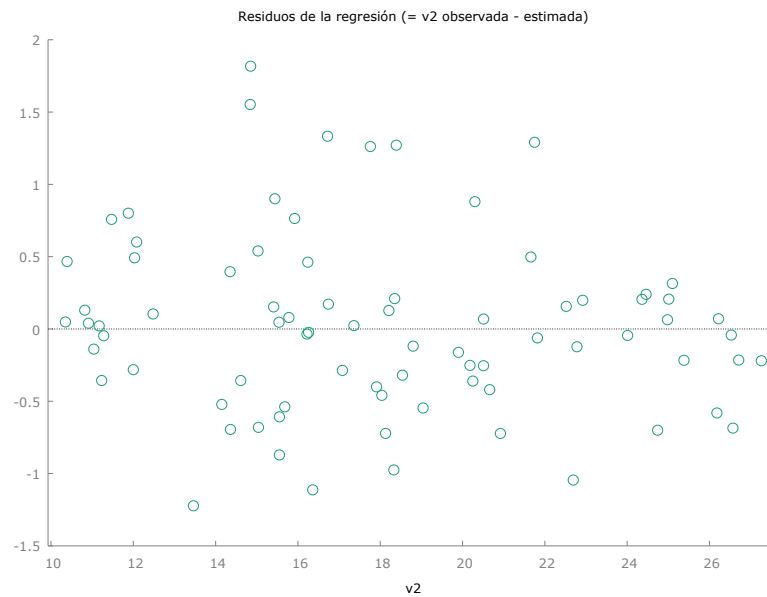
Figura 36: Contraste de significatividad de la media de los residuos del modelo 3

```
Hipótesis nula: [Media poblacional = 0]
Tamaño muestral: n = 79
Media muestral = 0.0169976, desv. típica = 0.615282
Estadístico de contraste: t(78) = (0.0169976 - 0)/0.0692247 = 0.245542
Valor p a dos colas = 0.8067
(a una cola = 0.4033)
```

Por lo tanto, se acepta la hipótesis nula de media nula de los residuos.

- Varianza constante

Figura 37: Gráfico de dispersión de los residuos del Modelo 3



Como los residuos fluctúan alrededor de cero, pero no podemos llegar a una conclusión sobre que la dispersión de los residuos se mantenga constante a lo largo del tiempo, realizamos el contraste de ARCH:

$H_0: \sigma_t^2 = \alpha_0 \forall t \rightarrow$ Homocedasticidad

$H_A: \sigma_t^2 \neq \alpha_0 \forall t \rightarrow$ Heterocedasticidad

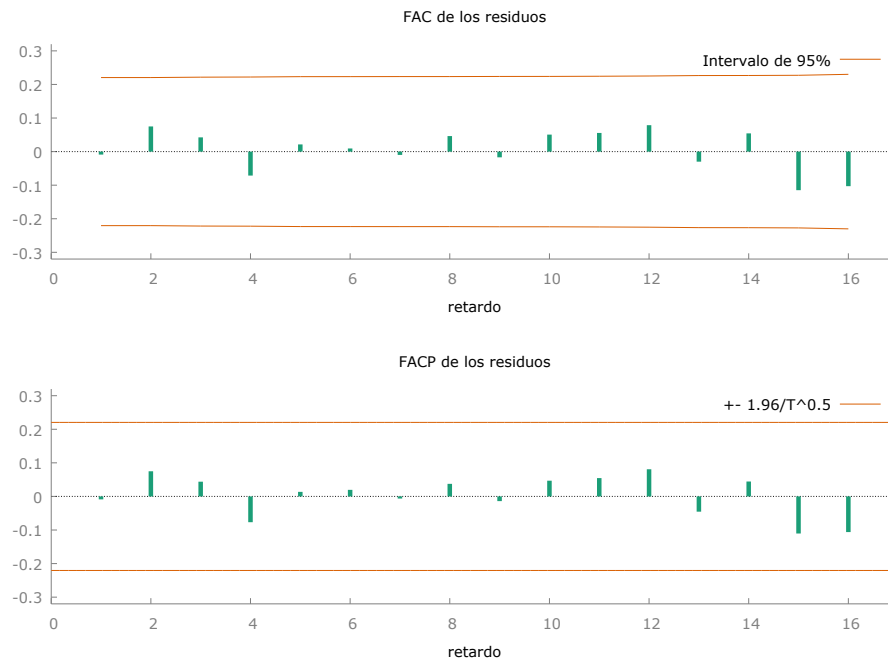
Figura 38: Contraste ARCH Modelo 3

```
Contraste de ARCH de orden 4 -
Hipótesis nula: [No hay efecto ARCH]
Estadístico de contraste: LM = 3.79972
con valor p = P(Chi-cuadrado(4) > 3.79972) = 0.433789
```

Como el p- valor $0.454 > 0.05$ se acepta la hipótesis nula de homocedasticidad.

- No autocorrelación

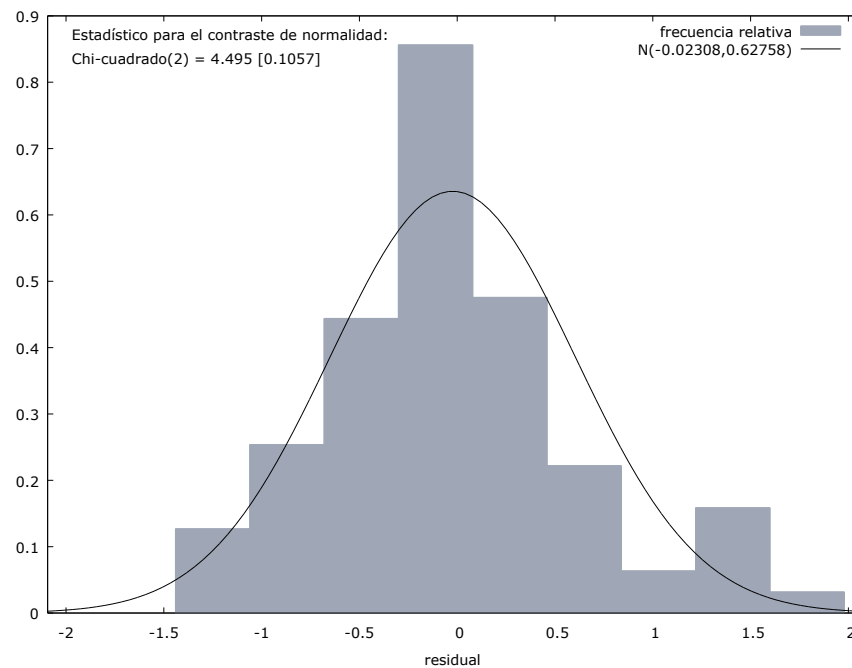
Figura 39: Correlograma de los residuos del Modelo 3



Con la Figura 39 ya podemos concluir que se cumple la condición de no autocorrelación ya que todos los valores de $r_j(\hat{u})$ caen dentro de las bandas de no significatividad.

- Distribución normal

Figura 40: Gráfico contraste de Normalidad de los residuos del Modelo 3



Observando la Figura 40 ya podríamos afirmar que los residuos sí que siguen una distribución normal, pero para confirmar esto realizamos el contraste de Jarque Bera:

$H_0: g_1 = g_2 = 0 \rightarrow$ Normalidad

H_A : No normalidad

Figura 41: Contraste de normalidad de los residuos del modelo 3

```
Contraste de Normalidad de uhat3:

Contraste de Doornik-Hansen = 4.49489, con valor p 0.105669
W de Shapiro-Wilk = 0.969318, con valor p 0.0444022
Contraste de Lilliefors = 0.125084, con valor p ~ 0
Contraste de Jarque-Bera = 4.81654, con valor p 0.0899708
```

Como el p-valor es $0.089 > 0.05$ se acepta la hipótesis nula de normalidad de los residuos al 5% pero al 10% se rechazaría.

Para analizar si los coeficientes estimados son significativos tenemos que efectuar contrastes de hipótesis respecto a dichos coeficientes del modelo.

$$\begin{array}{lll} H_0: \phi_1 = 0 & H_0: \theta_1 = 0 & H_0: \theta_2 = 0 \\ H_A: \phi_1 \neq 0 & H_A: \theta_1 \neq 0 & H_A: \theta_2 \neq 0 \end{array}$$

Tanto para ϕ_1 como θ_1 el p-valor es muy pequeño y se rechazan las hipótesis nulas de que los parámetros estimados no son significativos. Sin embargo, para θ_2 el p-valor es 0.140, es decir, se acepta la hipótesis nula de que el parámetro estimado no es significativo, así que no pasaría el chequeo.

Además, como el modelo tiene una parte AR y una MA, se deberá cumplir tanto la condición de estacionariedad como la de invertibilidad. Es decir, para que el modelo sea estacionario se tiene que cumplir que $|\hat{\phi}_1| < 1$ y para que sea invertible se tiene que cumplir que; $|\hat{\theta}_2| < 1$; $\hat{\theta}_1 + \hat{\theta}_2 < 1$; $\hat{\theta}_2 - \hat{\theta}_1 < 1$ condiciones que se cumplen ya que $0.500 < 1$, $0.206 < 1$, $(-0.602) + (-0.206) < 1$, $(-0.206) - (-0.602) < 1$

Una vez realizada la etapa de chequeo para los tres modelos estimados y al ver que los dos primeros pasan dicha etapa y que el modelo 3 no la pasa porque uno de sus parámetros no es significativo, tendremos que realizar un análisis de los criterios Schwarz, Akaike y Hannan-Quinn para tomar la decisión de que modelo resulta más adecuado para realizar las predicciones de la tasa de paro femenino para el año siguiente.

	Criterio Schwarz	Criterio Akaike	Criterio Hannan-Quinn
Modelo 1	174.24	166.989	169.905
Modelo 2	183.016	175.759	178.675

Para los tres criterios el modelo que resulta más adecuado es el modelo 1 ya que es el que menor valor de dichos criterios obtiene.

Predicción

Una vez identificados, estimados y validados los procesos estocásticos compatibles con la estructura de los datos se tratará de obtener predicciones para la serie temporal objeto de estudio, con el modelo seleccionado que es el Modelo 1:

$$\text{ARIMA}(1,1,0) \times \text{ARIMA}(0,1,1)_4$$

$$(1 - L^4)(1 - L)(1 - \phi_1 L)y_t = (1 - \theta_1 L^4)u_t$$

Expresamos el modelo en términos de las variables originales:

$$(1 - L^4)(1 - L)(1 - 0,521L)y_t = (1 - (-0,670)L^4)u_t$$

$$(1 - L^4 - 1,52L + 1,52L^5 + 0,52L^2 - 0,52L^6)y_t = (1 + 0,67L^4)u_t$$

$$y_t = 1,52y_{t-1} - 0,52y_{t-2} + y_{t-4} - 1,52y_{t-5} + 0,52y_{t-6} + u_t - 0,67u_{t-4}$$

Y a través de Gretl obtenemos las predicciones a un intervalo del 95% de confianza para el año siguiente en este caso 2024

Figura 42: Predicción puntual para el año 2024 Modelo 1

Para intervalos de confianza 95%, $z(0.025) = 1.96$

v2	predicción	Desv. Típica	Intervalo de confianza 95%	
2024:1	13.8282	0.628264	12.5968	15.0596
2024:2	12.6522	1.14394	10.4102	14.8943
2024:3	12.8921	1.60578	9.74487	16.0394
2024:4	12.5590	2.01429	8.61110	16.5070

A continuación, para evaluar la capacidad predictiva de los dos modelos que han pasado la etapa de chequeo vamos a utilizar el error absoluto medio porcentual, el error cuadrático medio y el error absoluto medio. Pero para ello antes, como no tenemos los datos reales de 2024, vamos a reducir la muestra eliminando las 4 últimas observaciones, es decir las correspondientes al año 2023, quedándose de esta manera la muestra reducida

desde el primer trimestre de 2002 hasta el cuarto trimestre de 2022 y hacemos la predicción para el año 2023

Para esta nueva muestra volvemos a estimar el modelo 1 modelo 2 y a calcular las predicciones para el año 2023. Nos aparecerán los principales estadísticos de evaluación de la predicción:

Error Cuadrático Medio (ECM)

$$ECM(H) = \frac{1}{H} \sum_{l=1}^H e_{T-H}^2(l) = \frac{1}{H} \sum_{l=1}^H \left(y_{T-H+l} - \hat{y}_{T-H}(l) \right)^2$$

Error Absoluto Medio (EAM)

$$EAM(H) = \frac{1}{H} \sum_{l=1}^H |e_{T-H}(l)| = \frac{1}{H} \sum_{l=1}^H |y_{T-H+l} - \hat{y}_{T-H}(l)|$$

Error Absoluto Porcentual Medio (EAPM)

$$EAPM(H) = \frac{1}{H} \sum_{l=1}^H \frac{|e_{T-H}(l)|}{y_{T-H+l}} \times 100 = \frac{1}{H} \sum_{l=1}^H \frac{|y_{T-H+l} - \hat{y}_{T-H}(l)|}{y_{T-H+l}} \times 100$$

Figura 43: Análisis de predicciones Modelo 1 (1T 2002 - 4T 2022)

Para intervalos de confianza 95%, $z(0.025) = 1.96$

	v2	predicción	Desv. Típica	Intervalo de confianza 95%	
2023:1	15.13	15.03	0.616	13.82 -	16.24
2023:2	13.16	14.24	1.143	12.00 -	16.48
2023:3	13.59	14.30	1.625	11.11 -	17.48
2023:4	13.36	13.97	2.060	9.94 -	18.01

Estadísticos de evaluación de la predicción tilizando 4 observaciones

Error medio	-0.57359
Raíz del Error cuadrático medio	0.71463
Error absoluto medio	0.62431
Porcentaje de error medio	-4.3238
Porcentaje de error absoluto medio	4.6591
U2 de Theil	0.7381
Proporción de sesgo, UM	0.64422
Proporción de regresión, UR	0.23132
Proporción de perturbación, UD	0.12446

En general un menor valor de dichos estadísticos indicará unas predicciones más precisas. Por lo tanto, también resulta importante calcular estos estadísticos a la hora de elegir el modelo más adecuado.

	ECM	EAM	EAPM
Modelo 1	0.511	0.624	4.659
Modelo 2	0.309	0.434	3.235

Hemos calculado los principales estadísticos de evaluación de la predicción para el modelo 2 en el Anexo I y según estos estadísticos el modelo que resulta más adecuado es el modelo 2, por lo que realizamos la predicción puntual para el año 2024 con este modelo también.

Figura 44: Predicción puntual para el año 2024 Modelo 2

Para intervalos de confianza 95%, $z(0.025) = 1.96$

	v2	predicción	Desv. Típica	Intervalo de confianza 95%	
2024:1		14.0248	0.669673	12.7122 -	15.3373
2024:2		12.3890	1.19621	10.0445 -	14.7336
2024:3		12.9557	1.65627	9.70943 -	16.2019
2024:4		12.7465	2.05690	8.71500 -	16.7779

6.- Conclusiones

La finalidad de este trabajo era realizar un análisis de la serie temporal, tasa de desempleo femenina en España desde 2002 hasta 2023 con el objeto de estudiar su comportamiento y predecir con la mayor exactitud el posible comportamiento futuro a corto plazo de dicha variable. Para ello ha sido necesario transformar la serie ya que en un principio no era estacionaria y presentaba un comportamiento estacional, propio de la tasa de desempleo.

Una vez transformada la serie he identificado y estimado tres modelos multiplicativos estacionales alternativos, de los cuales el tercero no ha pasado la etapa de chequeo al ser uno de los parámetros no significativo. Como el primer y el segundo modelo si que han pasado la etapa de chequeo, me he basado en los criterios de información para elegir cuál de ellos era el más adecuado, cuya conclusión ha sido el primer modelo. Sin embargo,

según los estadísticos de evaluación de la capacidad de predicción el modelo más adecuado es el modelo 2.

Para decantarnos por un modelo u otro y resolver dicha contradicción podemos fijarnos en el dato real de la tasa de desempleo femenina para el primer trimestre de 2024 en España que ha sido del 13.73% y por tanto el modelo que más se acerca a dicho dato es el primer modelo.

Las predicciones realizadas para el año 2024 nos indican que la tasa de desempleo femenina en España va a seguir el mismo comportamiento que tiene desde 2013, es decir decrecer ligeramente respecto al año anterior.

Cabe destacar que para mejorar la capacidad predictiva habría que eliminar los datos atípicos que hemos podido observar en algunos gráficos y que se corresponden con la etapa de la pandemia del covid-19.

7.-Bibliografía

Aznar, A. y Trívez Bielsa, F.J. (1993). *Métodos de predicción en Economía II*. Editorial Ariel, Barcelona.

Box, G. E. P., y Jenkins, G. M. (1973). Some comments on a paper by Chatfield and Prothero and on a review by Kendall. *Journal of the Royal Statistical Society. Series A (General)*, 136(3), 337-352.

Dueñas Fernández, D., Iglesias Fernández, C. y Llorente Heras, R. (2016). ¿Por qué las mujeres no se distribuyen de forma homogénea en el mercado de trabajo Español? El “efecto rechazo” y el “efecto atracción”. *El Trimestre Económico*, vol. 83, n. 330, p.339-369. Disponible en <https://www.redalyc.org/articulo.oa?id=31345267004>

González Casimiro, M. P. (2009). *Análisis de series temporales: Modelos ARIMA*. Editorial: Sarriko-On, País Vasco.

Jarque, C. M. y Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6 (3): 255-259. [doi:10.1016/0165-1765\(80\)90024-5](https://doi.org/10.1016/0165-1765(80)90024-5)

Ljung, G. M. y G. E. P. Box (1978). On a Measure of a Lack of Fit in Time Series Models». *Biometrika* 65 (2): 297-303. [doi:10.1093/biomet/65.2.297](https://doi.org/10.1093/biomet/65.2.297)

Millán Vázquez de la Torre, M.G., Santos Pita, M.P. y Pérez Naranjo, L. (2015). Análisis del mercado laboral femenino en España: evolución y factores socioeconómicos determinantes del empleo. Papeles de población, Toluca, v. 21, n. 84, p. 197-225. Disponible en: <https://www.scielo.org.mx/pdf/pp/v21n84/v21n84a8.pdf>

Anexo I

Modelo 1 para (2002:1 – 2022:4)

Evaluaciones de la función: 29
Evaluaciones del gradiente: 10

Modelo 9: ARIMA, usando las observaciones 2003:2-2022:4 (T = 79)
Estimado usando AS 197 (MV exacta)
Variable dependiente: (1-L) (1-Ls) v2
Desviaciones típicas basadas en el Hessiano

	coeficiente	Desv. típica	z	valor p
phi_1	0.560796	0.0953095	5.884	4.01e-09 ***
Theta_1	-0.735772	0.125412	-5.867	4.44e-09 ***
Media de la vble. dep.	-0.010253	D.T. de la vble. dep.	0.834567	
Media de innovaciones	-0.015971	D.T. innovaciones	0.616462	
R-cuadrado	0.984650	R-cuadrado corregido	0.984451	
Log-verosimilitud	-75.55147	Criterio de Akaike	157.1029	
Criterio de Schwarz	164.2113	Crit. de Hannan-Quinn	159.9508	
	Real	Imaginaria	Módulo	Frecuencia
AR				
Raíz 1	1.7832	0.0000	1.7832	0.0000
MA (estacional)				
Raíz 1	1.3591	0.0000	1.3591	0.0000

Modelo 2 para (2002:1 – 2022:4)

Evaluaciones de la función: 17
Evaluaciones del gradiente: 5

Modelo 10: ARIMA, usando las observaciones 2003:2-2022:4 (T = 79)
Estimado usando AS 197 (MV exacta)
Variable dependiente: (1-L) (1-Ls) v2
Desviaciones típicas basadas en el Hessiano

	coeficiente	Desv. típica	z	valor p
phi_1	0.506141	0.0977007	5.181	2.21e-07 ***
Phi_1	-0.425911	0.109022	-3.907	9.36e-05 ***
Media de la vble. dep.	-0.010253	D.T. de la vble. dep.	0.834567	
Media de innovaciones	-0.011430	D.T. innovaciones	0.667788	
R-cuadrado	0.982189	R-cuadrado corregido	0.981958	
Log-verosimilitud	-80.71794	Criterio de Akaike	167.4359	
Criterio de Schwarz	174.5442	Crit. de Hannan-Quinn	170.2837	
	Real	Imaginaria	Módulo	Frecuencia
AR				
Raíz 1	1.9757	0.0000	1.9757	0.0000
AR (estacional)				
Raíz 1	-2.3479	0.0000	2.3479	0.5000

Figura 45: Análisis de predicciones Modelo 2 (1T 2002 - 4T 2022)

Para intervalos de confianza 95%, $z(0.025) = 1.96$					
	v2	predicción	Desv. Típica	Intervalo de confianza 95%	
2023:1	15.13	15.03	0.668	13.73 -	16.34
2023:2	13.16	14.11	1.207	11.74 -	16.47
2023:3	13.59	14.15	1.686	10.85 -	17.45
2023:4	13.36	13.49	2.107	9.36 -	17.62

Estadísticos de evaluación de la predicción utilizando 4 observaciones	
Error medio	-0.38595
Raíz del Error cuadrático medio	0.55633
Error absoluto medio	0.43368
Porcentaje de error medio	-2.919
Porcentaje de error absoluto medio	3.2345
U2 de Theil	0.56416
Proporción de sesgo, UM	0.48127
Proporción de regresión, UR	0.053253
Proporción de perturbación, UD	0.46547