



Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

Grado en Economía
2016-2020

Trabajo Fin de Grado

“Análisis comparativo de las tasas de
desempleo de España y otros países de
la Unión Europea”

Luis Cadenas Alonso

RESUMEN

En este trabajo voy a realizar un estudio sobre la existencia (o no) de convergencia en niveles de desempleo entre España y algunos países de la Unión Europea (UE). He escogido como variable para realizar este análisis la tasa de desempleo de los países que voy a estudiar, ya que me parece un tema relevante, de actualidad y muy característico de la economía española. Es llamativo ver las diferencias de tasas de desempleo entre los diferentes países de la UE, por lo que me gustaría llegar a ver si estas diferencias tienen carácter estructural y si seguirán existiendo en un futuro próximo.

Con este trabajo pretendo ver si existe o no convergencia en esta variable, analizar la situación individual y conjunta de los países seleccionados a comienzos de siglo, durante la crisis y en 2018. Trataré, también, de realizar una predicción de los valores futuros de este indicador para cada uno de los países sometidos a estudio, observando si en el futuro habrá convergencia o no.

ABSTRACT

This work focuses on the study of convergence in the unemployment rates of Spain and other European Union (EU) countries. I chose unemployment rate as the variable because it is a current topic nowadays and it is a very characteristic indicator of the Spanish economy. The difference among the EU countries' unemployment rates is remarkable, so I want to analyse if these differences had always been and if they will continue to exist in the next years.

With this study, I intend to detect if there is convergence or not. I will analyse the situation of this indicator of each county separately and jointly at the beginning of this century, the evolution in the crisis period and in 2018. Then, I will try to predict the values of this indicator for each country in the future. Finally, I will conclude with the existence of convergence (or the lack of it) analysing the forecast values.

ÍNDICE

1. INTRODUCCIÓN.....	3
2. INFORMACIÓN ESTADÍSTICA	4
3. METODOLOGÍA.....	6
4. APLICACIÓN EMPÍRICA.....	7
4.1 ANÁLISIS UNIVARIANTE	7
4.1.1 TASA DE DESEMPLEO DE ESPAÑA	7
4.1.2 TASA DE DESEMPLEO DE FRANCIA	20
4.1.3 TASA DE DESEMPLEO DE ALEMANIA.....	26
4.1.4 TASA DE DESEMPLEO DE ITALIA.....	31
4.2 ANÁLISIS MULTIVARIANTE.....	35
5. CONCLUSIONES.....	43
6. BIBLIOGRAFÍA	47

1. INTRODUCCIÓN

En 1997 se puso en marcha la Estrategia Europea de Empleo (EEE), en la cual se establecieron objetivos comunes en esta materia. Su objetivo principal es la mejora del mercado laboral europeo mediante la creación de puestos de trabajos de calidad y la mejora de los ya existentes.

Esta Estrategia es uno de los puntos que se revisan anualmente en el Semestre Europeo. Cada año, se establecen unas pautas a seguir, junto con las prioridades y objetivos de este periodo propuestos por la Comisión Europea. También se realiza el Informe Conjunto sobre el Empleo (*Joint Employment Report, JER*), que pone sobre la mesa la situación del empleo en Europa, la implementación de las pautas antes mencionadas y los indicadores sociales y de empleo más importantes. Además, en el Semestre Europeo se incluyen los Programas Nacionales de Reforma (PNR), los cuales recogen las medidas que implementará cada país miembro para la consecución de los objetivos comunes. Finalmente, también se realizan informes nacionales que analizan la situación de cada país miembro, en los que, además, se incluyen recomendaciones personalizadas.

Todo este seguimiento de la situación del mercado de trabajo europeo implicaría la implementación de políticas que, en principio, supondrían un acercamiento de los mercados laborales de cada país hacia un objetivo común. De esta manera, los niveles de desempleo y las condiciones laborales de cada Estado miembro deberían estar convergiendo. Por ejemplo, el mercado laboral de un país como Alemania estaría mejorando, pero el de un país como España, Italia o Portugal debería estar mejorando a un ritmo superior para alcanzar esa convergencia. En este trabajo analizaré esta cuestión en profundidad, estudiando si realmente la convergencia existe.

El mercado laboral y el empleo son temas de suma relevancia en la actualidad. A nivel comunitario constituyen uno de los puntos clave de la EEE y, además, a nivel nacional, en el caso de España, se trata de un tema de suma importancia debido a las tasas de desempleo tan altas que hemos tenido desde la crisis de 2008.

En estos momentos estamos viviendo una situación excepcional que afectará notablemente a los mercados de trabajo mundiales. Hay países más afectados que otros, por lo que la convergencia antes mencionada podría no llegar a darse por el momento. Si bien no hay datos suficientes como para incorporar esta situación al modelo, parece claro que la actual crisis va a suponer un profundo cambio en la economía global, con importantes repercusiones en el ámbito laboral.

Este trabajo se estructura en siete apartados. Comienza con una introducción en la que expongo el marco institucional y social relacionado con el estudio a realizar.

En segundo lugar, el apartado de información estadística contiene la descripción de las variables del estudio, incluyendo la fuente de la que provienen las series y también del programa informático que utilizaré para su tratamiento.

El tercero de los apartados es de metodología e incluye una breve descripción de los métodos utilizados, sobre los que se profundizará en los apartados siguientes.

El cuarto apartado es de aplicación empírica y contiene la descripción y modelización individual de cada una de las series temporales, así como la modelización conjunta de todas ellas.

En el quinto apartado realizaré un análisis de los resultados obtenidos en los anteriores apartados, interpretando los principales resultados del estudio empírico realizado.

A continuación, el apartado sexto resume las conclusiones del trabajo y, finalmente, el séptimo apartado recopila las referencias bibliográficas.

2. INFORMACIÓN ESTADÍSTICA

Las variables objeto de análisis en este trabajo son las tasas de desempleo de los países seleccionados. Más concretamente, se trata de series temporales mensuales cuyo recorrido es suficientemente amplio como para realizar un análisis fiable de su evolución y llevar a cabo las correspondientes predicciones.

Las variables investigadas miden la tasa de desempleo calculada como la proporción de la población activa que se encuentra desempleada y aparecen en tantos por ciento. Se consideran las series: “u_rate_Germany”, referida a la tasa de desempleo mensual de Alemania; “u_rate_Spain”, referida a la tasa de desempleo mensual de España; “u_rate_Italy”, referida a la tasa de desempleo mensual de Italia; “u_rate_France” referida a la tasa de desempleo mensual de Francia.

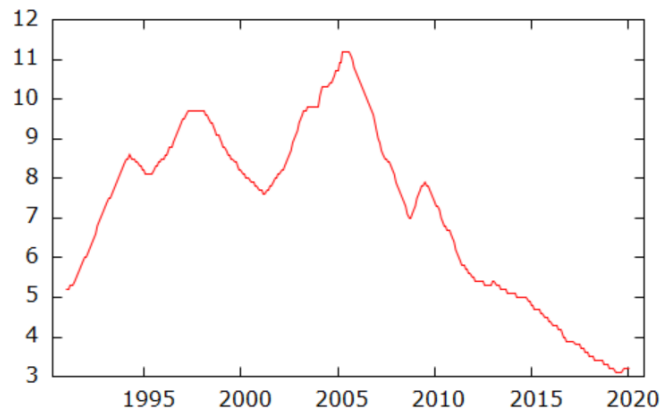
La información de todas estas variables proviene de la base de datos de la Oficina Europea de Estadística (Eurostat) cuya frecuencia es mensual. Esta fuente estadística proporciona una doble ventaja: por una parte, la homogeneidad de los datos que, al proceder de una misma fuente, garantizan una mayor consistencia al estudio y, por otra parte, la frecuencia de los datos es mensual, siendo más elevada que las series publicadas por los respectivos Institutos Nacionales de Estadística, incluido el de España (INE), que publica trimestralmente la Encuesta de Población Activa (EPA).

Cabe añadir, además, que las series analizadas tienen un recorrido temporal bastante amplio, de tal manera que permiten observar adecuadamente el comportamiento a largo plazo de cada serie temporal estudiada.

El programa econométrico utilizado para el análisis es **Gretl** (*Gnu Regression Econometrics and Time-series Library*), software libre que proporciona opciones adecuadas para el tratamiento univariante y multivariante de las series temporales analizadas en este trabajo.

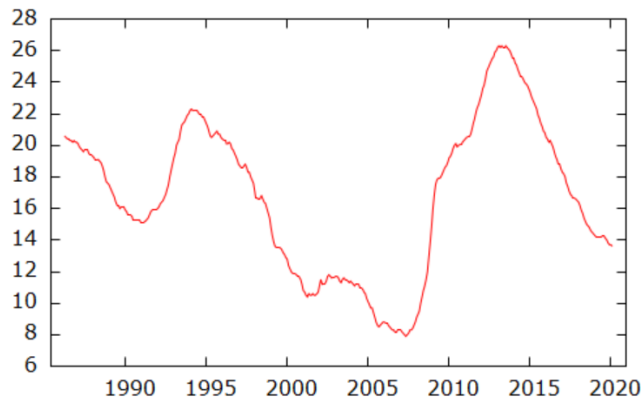
La primera de las variables mencionadas, la relativa a Alemania, tiene datos de tasas de desempleo recogidos desde enero de 1991 hasta febrero de 2020 (Figura 1.1). Esta serie temporal muestra una tendencia ascendente desde el primer dato que tenemos hasta finales de 2005 (salvando un valle entre principios de 1998 y mediados de 2003). Después de ese punto, se observa una tendencia descendente hasta el momento actual, salvando un pequeño repunte durante la crisis de 2008.

Figura 1.1: Evolución de la tasa de desempleo de Alemania, frecuencia mensual (1991-2020).



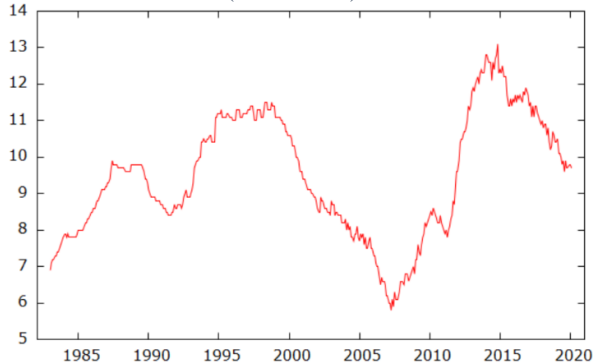
En cuanto a la serie relativa a la tasa de desempleo de España, el recorrido es algo mayor. Estos van desde abril de 1986 hasta febrero de 2020 (Figura 1.2). Los datos de esta variable muestran subidas y bajadas, dando la sensación de que puede ser una serie estacionaria en media, e incluso podría serlo en varianza. Esto es algo que analizaré posteriormente.

Figura 1.2: Evolución de la tasa de desempleo de España, frecuencia mensual (1986-2020).



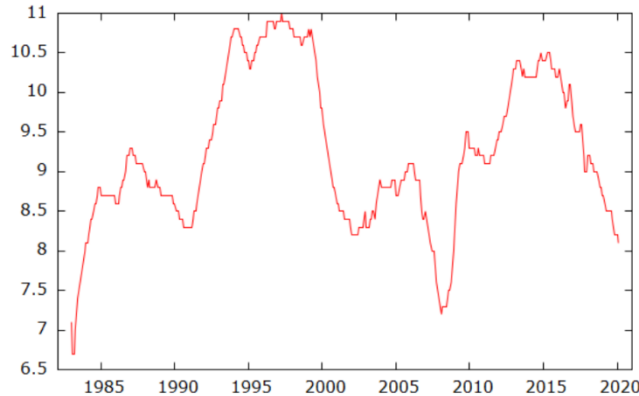
Respecto a la variable referida a la tasa de desempleo de Italia, el tamaño de la muestra es aún mayor, siendo el primer dato de enero de 1983. Su recorrido llega hasta febrero de 2020 (Figura 1.3). La evolución de la variable sugiere un comportamiento similar al de la española, solo que con unas tasas de desempleo más bajas en la mayor parte del recorrido.

Figura 1.3: Evolución de la tasa de desempleo de Italia, frecuencia mensual (1983-2020).



Finalmente, la variable referida a la tasa de desempleo francesa tiene el mismo recorrido que la italiana (Figura 1.4). La evolución de esta variable sugiere, también, una posible estacionariedad en media, y la estacionariedad en varianza podría darse, aunque el análisis gráfico no resulta concluyente.

Figura 1.4: Evolución de la tasa de desempleo de Francia, frecuencia mensual (1983-2020).



3. METODOLOGÍA

El análisis de las series investigadas se llevará a cabo tanto desde la óptica univariante como multivariante, utilizando las funciones proporcionadas por el paquete econométrico Gretl.

Para la consecución de este trabajo, utilizaré la *Metodología Box-Jenkins* [G.E.P. Box y G.M. Jenkins (1970)], que consiste en un análisis estocástico de series temporales mediante modelos univariantes, y la *Metodología VAR (Vectores Autorregresivos)* [C. Sims (1980)], consistente en un análisis de series temporales mediante modelos multivariantes.

Previamente a la implementación de ambos análisis, presentaré brevemente los principales conceptos relativos a las series temporales y sus componentes.

Una serie temporal se define como un conjunto de observaciones de una variable aleatoria que se recogen secuencialmente en el tiempo. Los datos de una serie temporal tienen una frecuencia fija, con lo cual dichos datos pueden ser diarios, semanales, mensuales... pero nunca pueden combinarse esas frecuencias.

Una serie temporal tiene unos componentes que explican su comportamiento en el tiempo. Estos son:

- *Tendencia*: es el comportamiento a largo plazo de una serie temporal. Es la evolución de la serie temporal en plazos de 10, 20 ó 30 años. Se suele representar con la letra " T_t ".
- *Ciclo*: son las oscilaciones en torno a la tendencia que se producen en plazos más cortos de tiempo que los del anterior componente, pudiendo ser estos de 3, 5, 8 o incluso más años. Por ejemplo, la tendencia del Producto Interior Bruto (PIB) de una economía puede ser ascendente o descendente en el largo plazo, pero, a su vez, tener oscilaciones respecto a la tendencia en periodos menores de tiempo (medio plazo). El componente cíclico se suele representar

con la letra “ C_t ” y, en algunas ocasiones, aparece asociado con la tendencia, dando lugar al componente tendencia-ciclo.

- *Componente estacional*: este componente, también denominado “variación estacional” y representado por la letra “ e_t ”, recoge las variaciones que se producen alrededor de la tendencia en el corto plazo, definido como un periodo inferior a un año. Son variaciones identificables y medibles, y se suelen repetir de forma regular anualmente. En el caso del desempleo, estas variaciones estacionales pueden aparecer asociadas a periodos de vacaciones, en los que suele aumentarse la contratación en el sector servicios debido al turismo (fenómeno que se magnifica en nuestro país).
- *Componente irregular*: comúnmente denominado “ruido”, valores atípicos u “*outliers*”. Son las alteraciones en la serie, que suelen ser pequeñas y no tienen ninguna pauta de comportamiento fija y cuyas causas son, a priori, desconocidas.

La obtención y análisis de estos componentes la detallaré en cada uno de los apartados de este trabajo, incluyendo una explicación de cada uno de los métodos según vaya utilizándolos.

4. APLICACIÓN EMPÍRICA

4.1 ANÁLISIS UNIVARIANTE

4.1.1 Tasa de desempleo de España

En primer lugar, para el correcto análisis de esta serie, me dispongo a la detección de “*outliers*” o valores atípicos. Decimos que un valor es un outlier cuando éste se desvía demasiado de la media. Para su detección, el criterio utilizado es establecer un número “ k ” de desviaciones típicas que consideramos admisible que se desvíe de la media. Normalmente se suelen considerar valores para “ k ” iguales a 3 ó 4. Algebraicamente:

$$|Y - \bar{Y}| \geq k * S_Y$$

Los valores atípicos pueden ser de diversos tipos, siendo los más habituales los siguientes:

- *Tipo impulso*: este tipo de valores atípicos se corresponden con una única observación atípica en su entorno. Es decir, podría ser un incremento o decremento brusco de los valores de la serie en una sola observación, volviendo tras este valor a los valores normales de la serie. Sería una variación atípica en un momento determinado del tiempo. Como mencioné antes, el impacto de los valores atípicos suele medirse mediante la utilización de variables dicotómicas. En este caso, para medir su impacto, se utilizaría una variable dicotómica que tomaría valor unitario si “ t ” coincide con la observación “ i ” y, en el resto de los casos, tomaría valor nulo:

$$I\{i\}_t = \begin{cases} 1 & \text{si } t \text{ coincide con la observación } i \\ 0 & \text{en cualquier otro caso} \end{cases}$$

- *Tipo escalón*: en este caso, el valor atípico no corresponde a un momento determinado del tiempo, sino que a una variación a partir de cierto momento del tiempo. Para calcular su impacto en la serie, se utilizaría una variable dicotómica que tomaría valor unitario a partir del momento en el que se produce el cambio, es decir:

$$E\{i\}_t = \begin{cases} 1 & \text{si } t \text{ es mayor o igual a la observación } i \\ 0 & \text{en cualquier otro caso} \end{cases}$$

A partir de estos dos tipos de variables, se observa que se cumple:

$$(1 - L) * E\{i\}_t = I\{i\}_t$$

Para la incorporación de valores atípicos a una serie temporal, al existir la posibilidad de que haya varios valores, su tratamiento consiste en incorporarlos secuencialmente de mayor a menor discrepancia.

Una vez explicada la manera de obtener estos valores atípicos, me dispongo a su análisis sobre las series objeto de estudio. Para ello, utilizo el programa Gretl y más concretamente, el paquete de funciones *EAG_Outliers* (Pérez, R. y López A.J.). Los valores atípicos obtenidos en los que me basaré son los propuestos por **TRAMO** (*Time series Regression with ARIMA noise, Missing values and Outliers*), **SEATS** (*Signal Extraction in ARIMA Time Series*) y **X13-ARIMA** (US Census Bureau y Banco de España).

Adoptando como referencia un margen $k = 2,5$, se obtienen los resultados de la Tabla 4.1.1.1.

Tabla 4.1.1.1: Outliers detectados en la tasa de desempleo de España.

Tipo	Periodo	Parámetro estimado	Error estándar	Valor t
OA	2002:02	0,2945	0,06461	4,56

Como se puede ver en esta salida, en esta serie existe un valor atípico de tipo OA. Este tipo de valores atípicos se corresponde con los descritos de tipo impulso.

Una vez identificados los valores atípicos correspondientes a esta serie, TRAMO los incorpora al modelo de mayor a menor discrepancia y confecciona una serie corregida de outliers, la cual se denomina “u_rate_S_xl”. Esta variable será la que utilice para el análisis de los siguientes contrastes que realizaré para esta serie temporal.

Una vez incorporada esta nueva variable a la base de datos, procedo a la detección de tendencia en esta serie. Para ello, se han de realizar el Contraste de Rachas y el Contraste de Rangos.

El primero de ellos, el Contraste de Rachas, es una aplicación del test de aleatoriedad, cuya hipótesis nula es que los datos de la serie son una muestra aleatoria simple (m.a.s.). Este contraste consiste en la generación de una variable dicotómica, que denominaré “ Z_t ”, que toma valores nulos si las observaciones están por debajo de la mediana (Me) y unitarios si estos están por encima de esta:

$$Z_t = \begin{cases} 0 & \text{si } Y_t \leq Me \\ 1 & \text{si } Y_t > Me \end{cases}$$

De esta manera, obtendremos una variable que tendrá rachas de valores iguales. Si esta variable contiene un número bajo de rachas (es decir, tiene muchos valores unitarios o nulos seguidos), tendremos que existe crecimiento o decrecimiento de la serie a lo largo del tiempo, pudiendo evaluar así la existencia de tendencia determinista. En caso de contar con muchas rachas, la variable nos acercaría a la hipótesis nula de este contraste.

Como medida de discrepancia se utiliza la siguiente:

$$d_R = \frac{R - 2np(1 - p)}{2\sqrt{np(1 - p)}} \approx N(0,1)$$

$$P(|N(0,1)| > |d_{R*}|) = p$$

De esta manera, si obtenemos un valor “p” lo suficientemente pequeño, rechazaríamos la hipótesis nula de la serie, llegando a la conclusión de que existe tendencia.

Utilizando la función de Gretl *EAG_Test_Tendencia* (Pérez, R. y López A.J.), se obtiene la salida reflejada en la Tabla 4.1.1.2. Esta función del programa se utilizará para la realización de ambos contrastes de tendencia.

Tabla 4.1.1.2: Salida del Contraste de Rachas de la tasa de desempleo de España.

Número de rachas	Discrepancia	Valor "p" a dos colas
6	-19,7028	2,0396e-086

Observando el valor “p” a dos colas, podemos concluir que, para esta serie, se rechazaría la hipótesis nula del contraste, lo que quiere decir que la serie, en principio, tendría tendencia.

El segundo de los contrastes previamente mencionados es el Contraste de Rangos. Este contraste se basa en la hipótesis nula de que la serie no tiene tendencia, siendo la hipótesis alternativa la existencia de esta. Este contraste tiene como medida de discrepancia la siguiente:

$$\hat{\theta} = \sqrt{T - 1}r \rightarrow N(0,1)$$

$$P(|N(0,1)| > |\hat{\theta}_*|) = p$$

“r” se corresponde con el Coeficiente de Correlación de Spearman, que tiene la siguiente forma:

$$r = 1 - \frac{6 \sum_{t=1}^T d_t^2}{T(T^2 - 1)}$$

Dentro de este, “ d_i ” se corresponde con:

$$d_i = \text{rango}(Y_i) - t$$

En caso de que el valor “p” sea lo suficientemente pequeño, se podría llegar a la conclusión de rechazo de la hipótesis nula, teniendo que la serie a estudiar tendría tendencia.

Aplicando este contraste a la serie en cuestión, obtengo la salida recogida en la Tabla 4.1.1.3.

Tabla 4.1.1.3: Salida del Contraste de Rangos de la tasa de desempleo de España.

Coef. de rangos	Discrepancia	Nivel crítico
-0,0395859	-0,797633	0,425083

Fijándonos en el nivel crítico, vemos que su valor no es lo suficientemente pequeño como para rechazar la hipótesis nula, llegando a la conclusión de que, según este contraste, la serie no tendría tendencia.

Los resultados obtenidos en ambos contrastes son contradictorios, por lo que no podríamos concluir con certeza si esta serie tiene tendencia o no. Para evitar restringir

los modelos que realizaré posteriormente, considero la existencia de tendencia de esta serie temporal.

Una vez realizados estos contrastes, me dispongo al análisis y cuantificación del factor estacional. Para ello, en primer lugar, realizaré el Contraste de Kruskal-Wallis (Kruskal, W. y Allen Wallis, W.). A continuación, realizaré el Contraste de Rango-Media para obtener la hipótesis de composición de la serie. De esta manera, con estos contrastes veré si hay necesidad de desestacionalizar la serie o no.

El test de Kruskal-Wallis contrasta la hipótesis nula de que la serie no tiene componente estacional determinista. El contraste tiene la siguiente forma:

$$d_{KW} = \frac{12}{T(T+1)} \left(\sum_{i=1}^{Freq} \frac{R_i^2}{T_i} \right) - 3(T+1) \rightarrow \chi_{Freq-1}^2$$

$$P(\chi_{Freq-1}^2 > |d_{KW}^*|) = p$$

“*Freq*” se refiere a la frecuencia (meses, trimestres, etc.), “*T_i*” es el número de observaciones del *i*-ésimo periodo en el que se define la frecuencia, “*R_i*” se refiere a la suma de rangos del *i*-ésimo periodo en el que se define la frecuencia y “*T*” es el total de observaciones.

Si el valor “*p*” es lo suficientemente pequeño, se rechazaría la hipótesis nula de no existencia de componente estacional, por lo que se concluye que sí existiría dicho componente.

Aplicando este contraste a la serie en cuestión y utilizando la función *EAG_Test_Estacionalidad* (Pérez, R. y López A.J.), se obtiene la salida recogida en la Tabla 4.1.1.4.

Tabla 4.1.1.4: Salida del Contraste de Kruskal-Wallis de la tasa de desempleo de España.

Nº observaciones	Grados de libertad	Valor del estadístico	Nivel crítico
407	11	0,0675859	1,00

Como se ve, el nivel crítico de este contraste es igual a 1, por lo que se podría decir que la serie no tendría componente estacional. Si bien este resultado puede sorprender teniendo en cuenta que el nivel de desempleo en España suele estar ligado al mes en el que estamos (descendiendo en verano debido al turismo), la representación gráfica recogida en las Figuras 1.1 a 1.4 no muestra evidencia clara de estacionalidad.

Siguiendo con la *Metodología Box-Jenkins*, me dispongo a la introducción de la modelización **ARIMA** (*Autoregressive Integrated Moving Average*).

Según el *Teorema de Wold* (1938), cualquier proceso estocástico y estacionario (al menos débilmente) se puede representar como la suma de dos procesos incorrelados: X_t , que se corresponde con un proceso linealmente determinista; U_t , que se corresponde con un proceso aleatorio MA (Moving Average) infinito.

$$Y_t = X_t + U_t$$

Este teorema arroja una idea de cuál es la composición de una serie temporal (en el caso de que se cumplan las dos condiciones mencionadas en este). Si consideramos los datos

de las series estudiadas como una muestra, es evidente que el proceso general podría haber generado estos datos u otros similares.

El siguiente paso, por tanto, sería tratar de obtener el modelo generador de datos. Un Proceso Generador de Datos (PGD) ha de estar en equilibrio estadístico, lo que quiere decir que se han de dar unas condiciones estables para poder modelizar. La principal condición que ha de darse es la estacionariedad del proceso.

Se dice que un proceso es débilmente estacionario cuando se cumplen las siguientes condiciones:

$$\begin{aligned}\mu &= E(Y_t) = E(Y_{t-1}) \\ \sigma_0 &= Var(Y_t) = Var(Y_{t-s}) \\ \sigma_s &= Cov(Y_t, Y_{t-s}) = Cov(Y_{t+r}, Y_{t+r-s}) \\ \rho_s &= Corr(Y_t, Y_{t-s}) = Corr(Y_{t+r}, Y_{t+r-s})\end{aligned}$$

Si se dan estas condiciones (resultando en un proceso débilmente estacionario) y, adicionalmente, se dan condiciones para la existencia de normalidad en la distribución, podemos decir que el proceso es estacionario.

Para analizar la estacionariedad de una serie temporal hay que fijarse en las Funciones de Autocorrelación y Autocorrelación Parcial.

La Función de Autocorrelación (FAC) se expresa de la siguiente manera:

$$Corr(Y_t, Y_{t-s}) = \rho_s$$

Siempre se da que:

$$\begin{aligned}\rho_s &= \rho_{-s} \\ |\rho_s| &\leq \rho_0 = 1\end{aligned}$$

En cuanto a la Función de Autocorrelación Parcial (FACP), esta se expresa mediante “ ϕ_{ss} ”, coeficiente que indica la correlación neta existente entre “ Y_t ” e “ Y_{t-s} ”, omitiendo el efecto de las variables intermedias.

$$Y_t = \phi_{s1}Y_{t-1} + \phi_{s2}Y_{t-2} + \dots + \phi_{ss}Y_{t-s} + u_t$$

Siempre se cumple que:

$$\phi_{11} = \rho_1$$

Estas dos funciones son determinantes a la hora de analizar la estacionariedad y la identificación de modelos ARIMA para una serie temporal.

Antes de proceder con la modelización ARIMA, es necesario realizar una introducción conceptual para su completo entendimiento.

Un modelo ARIMA (p, d, q) tiene diferentes componentes:

- AR(p): es la parte autorregresiva del modelo. Se puede encontrar expresado de la siguiente forma:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + u_t$$

Su polinomio característico se puede expresar de la siguiente manera:

$$\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

O bien:

$$\Phi_p(L)Y_t = c + u_t$$

Este modelo es siempre invertible y, para que sea estacionario, las “p” raíces de su polinomio característico han de ser mayores que la unidad. Más adelante introduciré el concepto de proceso invertible.

Este modelo se puede ver reflejado en las Funciones de Autocorrelación y de Autocorrelación Parcial. En la primera de estas funciones, se observaría una convergencia a cero lenta o amortiguada.

$$\rho_s = \phi_1^s$$

En cambio, en la FACP se verían valores no nulos en los “p” primeros retardos, teniendo valores nulos el resto de los retardos.

$$\phi_{ss} = 0 \text{ si } \forall s \geq p$$

- I(d): se corresponde con la parte de las diferenciaciones necesarias para hacer que la serie original sea estacionaria. El proceso de diferenciación consiste en generar una serie temporal nueva que tenga la siguiente forma:

$$Z_t = (1 - L)^d Y_t$$

El proceso de identificación del número de diferencias necesarias lo introduciré más adelante.

- MA(q): se corresponde con la parte de media móvil. Se puede encontrar expresado se la siguiente forma:

$$Y_t = \mu + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}$$

Su polinomio característico puede expresarse de la siguiente forma:

$$\Theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

Este modelo será siempre estacionario, pero, para que sea invertible, las “q” raíces del polinomio característico antes identificado han de ser mayores que la unidad.

En cuanto a su identificación en las Funciones de Autocorrelación y Autocorrelación Parcial, la FAC tendrá valores no nulos en los “q” primeros retardos.

$$\rho_r = 0 \text{ si } r \geq q$$

La FACP, en cambio, no se anulará nunca, descendiendo esta hacia cero indefinidamente.

Dentro de estos componentes queda un concepto por definir. Este es el de proceso invertible. Se dice que si un proceso está expresado según la forma descrita en el Teorema de Wold (descrita anteriormente), éste está escrito en forma directa. En cambio, un proceso expresado en función de su pasado se considera escrito en forma inversa. Si un proceso escrito en forma directa puede expresarse en forma inversa, se dice que éste es invertible.

Es necesario que se cumpla este punto para poder predecir los valores futuros o ausentes de una serie temporal. Esta característica se comprueba atendiendo a los requisitos de invertibilidad de cada componente anteriormente descrito.

Una vez analizados los componentes de un modelo ARIMA (p, d, q), me dispongo al análisis teórico conjunto de dichos componentes, con el fin de entender el concepto completamente.

Combinando todos los polinomios característicos de los componentes anteriormente mencionados, se obtendría la siguiente igualdad:

$$\Phi_p(L)(1-L)^d Y_t = \Theta_q(L)u_t$$

La cual se puede expresar también de la siguiente forma:

$$\frac{\Phi_p(L)}{\Theta_q(L)}(1-L)^d Y_t = u_t$$

Dentro de estas expresiones, podemos diferenciar cada polinomio característico, correspondiéndose $\Phi_p(L)$ a la parte AR (p), $\Theta_q(L)$ a la parte MA (q) y $(1-L)^d Y_t$ a la serie en diferencias.

Una serie temporal que no es estacionaria es homogénea o integrada de orden “d” siempre que mediante “d” diferencias esta se convierta en estacionaria.

En el caso de que el proceso sea no estacionario en varianza, hay que transformar la serie original en otra mediante las *transformaciones Box-Cox* (Box, C. E. P. y Cox, D., 1964). Este procedimiento se lleva a cabo de la siguiente forma:

$$Y_t^\lambda = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log Y_t & \text{si } \lambda = 0 \end{cases}$$

En casos como los de las series temporales estudiadas en este trabajo, la estacionariedad puede que se dé respecto al mismo periodo de los años anteriores. Este tipo de estacionariedad ha de ser comprobado aparte. Para esta circunstancia, pueden darse dos tipos de modelos:

- Modelos estacionales puros: en este caso, las correlaciones se producen solo entre observaciones separadas por un número “s” de retardos o por múltiplos de este número.
- Modelos estacionales multiplicativos: en este caso, además de las correlaciones regulares del modelo, existen correlaciones de tipo múltiple que se acentúan en retardos de orden “s” o múltiplos de este. En este caso, las correlaciones normales interactúan con las estacionales de manera multiplicativa.

Para su implementación a los modelos de series temporales, se utilizan los modelos ARIMA (p, d, q) x SARIMA (P, D, Q). La parte referida a SARIMA se refiere a la parte estacional de los componentes mencionados de los modelos ARIMA (p, d, q). El polinomio característico de este tipo de modelos se expresa de la siguiente forma:

$$\Phi_p(L)(1-L)^d \Phi_P(L^s)(1-L^s)^D Y_t = \Theta_q(L)\Theta_Q(L^s)u_t$$

O, alternativamente:

$$\frac{\Phi_p(L)\Phi_P(L^s)(1-L)^d(1-L^s)^D}{\Theta_q(L)\Theta_Q(L^s)} Y_t = u_t$$

Las expresiones remarcadas en negrita se corresponden con la parte SARIMA (P, D, Q).

La *Metodología Box-Jenkins* [G.E.P. Box y G.M. Jenkins (1970)] sigue un esquema que comienza con la obtención de los datos muestrales. A continuación, se identificaría el

modelo, lo que lleva a la consiguiente estimación y validación de este. Tras esto, se procede a la elaboración de predicciones, que, analizando su calidad, nos llevará a la elección de un modelo definitivo.

La primera de las fases ya la he realizado, por lo que me dispongo a comenzar con la fase de identificación.

Para ello, el primer paso es el análisis de estacionariedad de la serie. Este análisis consta de dos partes: la estacionariedad en varianza y la estacionariedad en media.

Una serie temporal es estacionaria en varianza cuando su dispersión se mantiene estable a lo largo del tiempo. Para comprobar si una serie temporal es estacionaria en varianza se utiliza el Contraste de Rango-Media. La hipótesis nula de este contraste es la estacionariedad en varianza de la serie temporal. En este contraste consiste en la comparación del rango (eje de ordenadas) de un número de submuestras y un tamaño definidos con sus respectivas medias (eje de abscisas), pudiendo observar así si hay alguna tendencia o patrón de comportamiento que nos pueda indicar la no existencia de estacionariedad en varianza.

Al realizar este contraste para la tasa de desempleo de España, se obtienen los resultados reflejados en la Tabla 4.1.1.6.

Tabla 4.1.1.6: salida del Contraste Rango-Media para la tasa de desempleo de España.

Submuestras	Tamaño muestral	Pendiente	Valor p
21	20	0,0693398	0,460381

Estos resultados y el valor “p” asociado al contraste no resultan significativos para rechazar el supuesto de estacionariedad en varianza, por lo que no sería necesario realizar una transformación Box-Cox de esta serie temporal.

Continuando con el análisis de estacionariedad, queda por analizar la estacionariedad en media. Una serie temporal es estacionaria en media cuando los valores de las esperanzas de esta serie son invariantes respecto al tiempo. Este análisis lo efectuaré mediante la realización de los contrastes ADF (Contraste Aumentado de Dickey-Fuller), ADF-GLS (modificación más eficiente del anterior contraste) y KPSS.

Los dos primeros contrastes (ADF y ADF-GLS) tienen como hipótesis nula la existencia de una raíz unitaria. En caso de aceptar esta hipótesis, estaríamos ante una serie no estacionaria en media, por lo que, si el contraste KPSS no dice lo contrario, habría que diferenciar la serie temporal para tratar de conseguir esta estacionariedad.

La hipótesis nula del contraste KPSS es la estacionariedad de la serie y por tanto difiere de la planteada por los dos contrastes anteriores, con lo cual se espera que las conclusiones sean las opuestas (cuando el resultado de los test ADF o ADF-GLS conducen al rechazo se espera que la conclusión del test KPSS sea no rechazar y viceversa).

La realización de estos contrastes en Gretl (siguiendo la metodología habitual de estimar modelos auxiliares más a menos amplios), arroja los resultados recogidos en la Tabla 4.1.1.7.

Tabla 4.1.1.7: Salidas de los contrastes de estacionariedad en media para la tasa de desempleo de España.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	NO	SÍ	-	-2,27634	0,1798
ADF-GLS	-	-	-	-1,70174	0,08417
KPSS	-	-	-	0,789214	<0,01

Al realizar el primer contraste, se obtiene que la tendencia no sería significativa, por lo que se resolvería el problema relacionado con la existencia de tendencia en esta serie temporal. De todos modos, se obtiene que la constante es significativa, por lo que tomaré como referencia el análisis con constante para el resto de los contrastes. El valor “p” de este contraste no es lo suficientemente pequeño como para rechazar la hipótesis nula de existencia de raíz unitaria, por lo que este contraste indicaría la necesidad de realizar una diferencia para que la serie fuera estacionaria en media.

El segundo de los contrastes (realizado con constante y sin tendencia), nos lleva a rechazar la hipótesis nula al 10% pero no al 5%. Por ello, para que el posterior modelo tenga más consistencia, no rechazo la hipótesis nula por falta de rigor. De esta manera, se obtiene la misma conclusión que en el Contraste ADF.

El último contraste (realizado con constante y sin tendencia) llevaría a la misma conclusión que los dos anteriores contrastes. Con la hipótesis nula de estacionariedad de la serie temporal, al ser el valor “p” muy pequeño, se rechazaría la hipótesis nula, por lo que se concluye que la serie no es estacionaria y habría que realizar una diferencia para garantizar la estacionariedad en media.

Tras analizar las conclusiones derivadas de los anteriores contrastes, me dispongo a la realización de estos para la serie diferenciada, con el fin de ver si esta es estacionaria en media. Los resultados obtenidos están recogidos en la Tabla 4.1.1.8.

Tabla 4.1.1.8: Contrastes de estacionariedad en media de la tasa de desempleo d España diferenciada.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	NO	NO	SÍ	-3,42375	0,000607
ADF-GLS	-	-	-	-2,4319	0,01454
KPSS	-	-	-	0,308877	>0,10

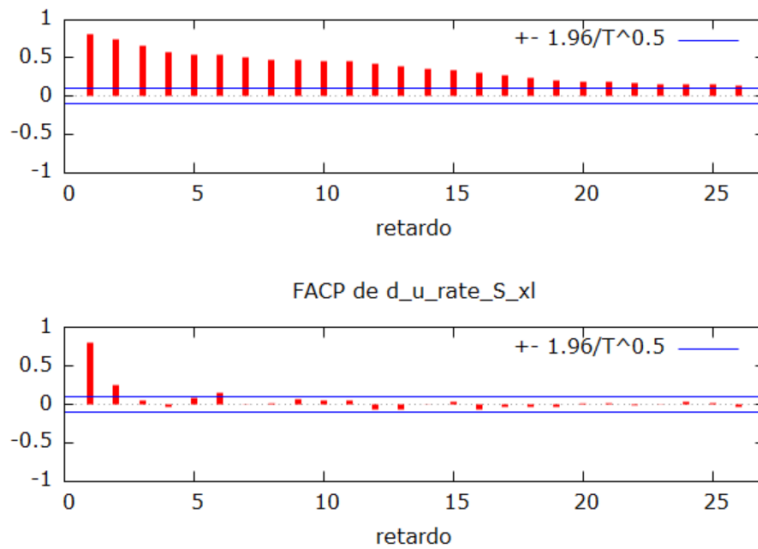
Las salidas de los contrastes muestran que la serie diferenciada podría ser estacionaria en media. En primer lugar, en el Contraste ADF no fueron significativas ni la tendencia ni la constante, por lo que utilizaré el análisis sin constante para el resto de los contrastes.

Con los valores “p” de los dos primeros contrastes, al ser lo suficientemente pequeños, se rechazaría la hipótesis nula de estos, lo que indica que no es necesario realizar más diferencias en la variable para que ésta sea estacionaria en media. El valor “p” del Contraste KPSS confirma la estacionariedad en media de la serie diferenciada, ya que el resultado de p no es lo suficientemente pequeño como para rechazar su hipótesis nula.

Con todos estos contrastes llegaríamos a la conclusión de que la serie estudiada en este apartado es integrada de orden 1, ya que su primera diferencia es estacionaria.

A continuación, me dispongo a la identificación de las partes AR(p) y MA(q) con el fin de elaborar un modelo ARIMA para la serie temporal estudiada en este apartado. Para ello, como dije antes, hay que observar las Funciones de Autocorrelación y de Autocorrelación Parcial de la serie diferenciada. Estas funciones están reflejadas en la Figura 4.1.1.1.

Figura 4.1.1.1: FAC y FACP de la tasa de desempleo de España diferenciada.



La identificación del componente AR se basa en la FACP de la serie diferenciada, que en este caso sugiere inicialmente $p = 2$, al observarse dos correlaciones parciales significativas, si bien la primera de ellas $p = 1$ en mucho mayor medida.

En cuanto a la parte MA(q), la FAC muestra muchos elementos significativos que podrían sugerir un proceso de memoria larga o diferencias fraccionales.

La principal característica de un proceso de memoria larga es la lenta convergencia a cero de su FAC para retardos lejanos. Como mencioné antes, sería necesario realizar una diferencia fraccional. Estas diferencias pueden ser:

$$\begin{cases} -0,5 < d < 0 & \text{memoria larga antipersistente} \\ d = 0 & \text{proceso de memoria corta} \\ 0 < d < 0,5 & \text{memoria larga persistente} \end{cases}$$

Por lo tanto, el valor de esta diferencia estaría entre $-0,5$ y $0,5$.

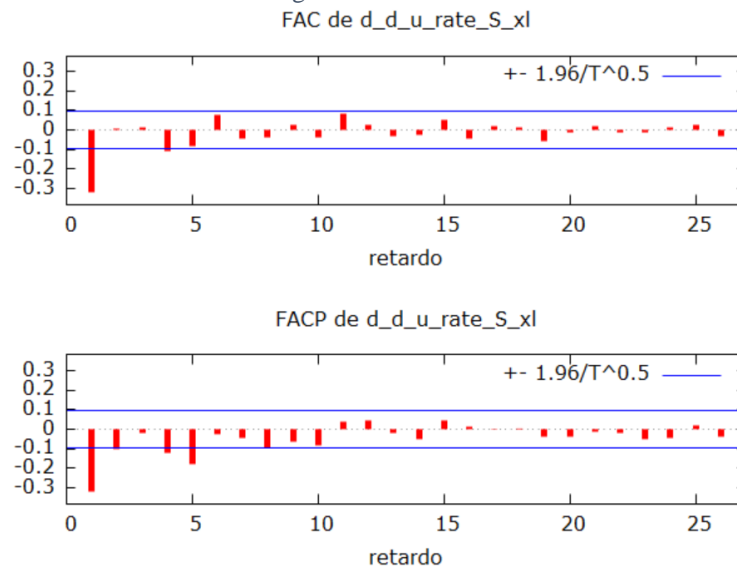
Para la identificación de su valor utilizaré el Coeficiente H de Hurst, el cual toma valores entre cero y la unidad cuando el número de observaciones tiende a infinito. El valor de las diferencias será el valor de este coeficiente menos $0,5$.

$$d = H - 0,5$$

Aplicando este contraste a la serie diferenciada, el valor obtenido para este coeficiente es de $0,910107$. De esta manera, siguiendo la igualdad anterior, concluyo que el valor de la diferencia fraccional necesaria para convertir la serie en totalmente estacionaria sería

de 0,410107. De todos modos, considero más adecuada la realización de una diferencia entera adicional para corregir el problema de no estacionariedad de la serie.

Figura 4.1.1.2: FAC y FACP de la tasa de desempleo de España en segundas diferencias.



En estas dos funciones se puede observar que se ha solucionado el problema anterior de la FAC. Podría decirse, de esta manera, que la tasa de desempleo de España es integrada de orden 2.

En cuanto a la parte AR(p), en la FACP se ve que el primer retardo es muy significativo, mientras que el segundo apenas lo es. Por ello, creo conveniente considerar la posibilidad de que “p” sea igual a 2 ó a 1. Es decir, tengo en cuenta ambas posibilidades.

Respecto a la parte MA(q), en la FAC se puede ver de manera clara que solamente el primer retardo es notablemente significativo, por lo que considero que “q” puede ser igual a 1.

De esta manera, tras haber realizado ambas diferencias, estaríamos ante un modelo ARIMA (2,2,1) o ARIMA (1,2,1).

Para comprobar qué modelo es más adecuado para las predicciones, hay que comparar tanto el R^2 corregido, los criterios de información de Akaike, Schwarz, Hannan-Quinn y el Logaritmo de Verosimilitud.

El R^2 corregido se refiere a la capacidad explicativa del modelo. Es decir, es el porcentaje de las variaciones de la variable que vendrían explicadas por el modelo. Este estadístico no es utilizado habitualmente en este tipo de modelos de series temporales, pero es interesante incluirlo de todos modos.

Los criterios de información recogen la pérdida de información del modelo. Interesa que su valor sea lo más pequeño posible. El Logaritmo de Verosimilitud es otra medida de bondad del modelo, cuyo valor interesa que sea lo mayor posible.

Todos estos resultados referidos a los dos modelos estimados quedan recogidos en la Tabla 4.1.1.9. En esta tabla resumen puede verse claramente que el modelo ARIMA (2,2,1) mejora todas las medidas consideradas. Tiene mayor capacidad explicativa, con

resultados inferiores de los criterios de información y superiores para el logaritmo de verosimilitud. Por ello, considero que este modelo sería el más adecuado.

Tabla 4.1.1.9: Comparativa modelos ARIMA de la tasa de desempleo de España.

	ARIMA (2,2,1)	ARIMA (1,2,1)
R-cuadrado corregido	0,999306	0,99929
Akaike	-496,2635	-490,579
Schwarz	-480,2677	-478,5821
Hannan-Quinn	-489,9308	-485,8295
Log-verosimilitud	252,1317	248,2895

Tras la selección del modelo más adecuado, antes de realizar predicciones me dispongo a chequear el modelo. Para ello, me dispongo al análisis de los residuos del modelo. En teoría, si el modelo es adecuado, sus residuos convergerían asintóticamente hacia un ruido blanco. Entre todas las características que habrían de tener, destacan la no autocorrelación, la ausencia de patrones reconocibles en su correlograma y que siguen una distribución normal.

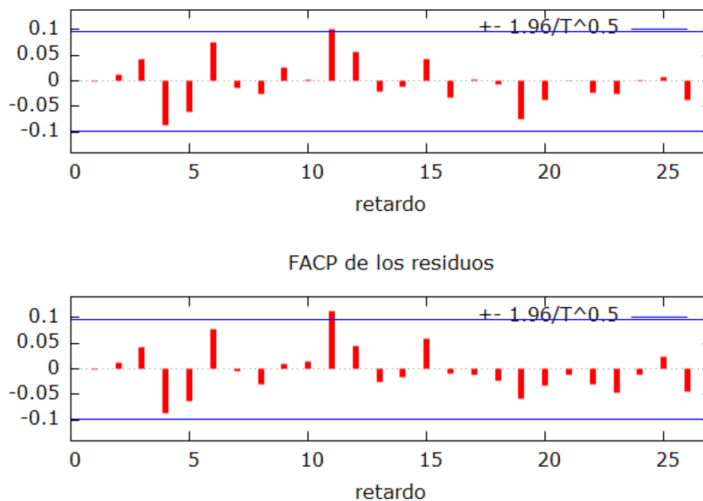
Para comprobar la primera de estas características, el programa realiza el contraste de Ljung-Box (1978). Este contraste parte de la hipótesis nula de no autocorrelación, teniendo su estadístico la siguiente forma:

$$Q^* = T(T + 2) \left(\sum_{i=1}^j \frac{r_i^2(\hat{u})}{T - i} \right) \rightarrow \chi_j^2$$

El valor “p” de este contraste para el modelo analizado es igual a 0,1174. Al no ser este lo suficientemente pequeño como para rechazar la hipótesis nula del contraste, concluyo que los residuos no tendrían autocorrelación.

En cuanto al correlograma de los residuos, en la Figura 4.1.1.3 puede verse que, aunque haya retardos significativos, no puede detectarse la existencia de un patrón o modelo ARMA (p, q) dentro de este.

Figura 4.1.1.3: correlograma de los residuos del modelo ARIMA (2,2,1) para la tasa de desempleo de España.



El siguiente y último de los puntos esenciales que han de cumplirse es la distribución normal de los residuos del modelo. Para ello, realizo el contraste de normalidad, cuya hipótesis nula es la normalidad. El valor “p” obtenido para este contraste es igual a 0,169784, no siendo éste lo suficientemente pequeño como para rechazar la hipótesis nula del contraste. Por ello, es posible asumir que los residuos del modelo seguirían una distribución normal.

Tras comprobar que el modelo es adecuado, me dispongo a la obtención de predicciones. Para su obtención, he reducido el recorrido muestral hasta diciembre de 2019 con el fin de poder compararlas con valores reales de la serie temporal. He realizado predicciones hasta diciembre de 2020, las cuales quedan recogidas en la Tabla 4.1.1.10.

Tabla 4.1.1.10: Predicciones de la tasa de desempleo de España.

Periodo	Valor
ene-20	13,6
feb-20	13,5
mar-20	13,4
abr-20	13,2
may-20	13,1
jun-20	13
jul-20	12,9
ago-20	12,8
sep-20	12,7
oct-20	12,6
nov-20	12,5
dic-20	12,4

Las predicciones de los valores de los dos primeros meses del año se aproximan bastante a los datos reales. Solo divergen de estos en un 0,1%, siendo los valores previstos menores de los reales.

Para evaluar las predicciones, me dispongo al análisis de los estadísticos de evaluación de predicciones. En concreto, haré hincapié en el porcentaje de error absoluto medio, en la U de Theil y en la proporción de sesgo. Los valores de estos estadísticos están resumidos en la Tabla 4.1.1.11.

Tabla 4.1.1.11: Estadísticos de evaluación de predicciones de la tasa de desempleo de España.

Estadístico	Valor
Porcentaje de error absoluto medio	0,9239
U de Theil	1,3229
Proporción de sesgo	0,99798

El valor del primero de los indicadores es bastante bueno, indicando que el porcentaje de error absoluto medio es únicamente del 0,9239%.

El índice U de Theil se obtiene con el cociente de las raíces del error cuadrático medio del modelo propuesto y de otro modelo ingenuo que asigna como predicción el valor actual ($\hat{Y}_{t+1} = Y_t$).

$$U = \sqrt{\frac{\frac{1}{T} \sum_{t=1}^{T-1} \left(\frac{Y_{t+1} - \hat{Y}_{t+1}}{Y_t} \right)^2}{\frac{1}{T} \sum_{t=1}^{T-1} \left(\frac{Y_{t+1} - Y_t}{Y_t} \right)^2}}$$

Es cierto que este estadístico no puede ser negativo, pero no tiene ninguna cota superior. De todos modos, se considera una “cota superior psicológica” igual a la unidad. Esto es debido a que un valor unitario para este estadístico supondría que la raíz del error cuadrático medio del modelo propuesto sería igual a la raíz del error cuadrático medio del modelo ingenuo antes mencionado, lo que querría decir que las predicciones del modelo propuesto no mejorarían las ingenuas ($\hat{Y}_{t+1} = Y_t$).

Por otro lado, un valor nulo para este contraste supondría el caso contrario. Para que este estadístico se anule sería necesario que el valor futuro real (Y_{t+1}) sea igual al valor futuro predicho (\hat{Y}_{t+1}). De esta manera, tendríamos que la predicción del modelo propuesto se ha ajustado perfectamente a la realidad, por lo que se dice que, si este estadístico toma valor nulo, las predicciones son perfectas.

El valor obtenido para estas predicciones no dista mucho de esta cota, y además la proporción de sesgo es elevada, por lo cual las parece detectarse algún problema sistemático en las predicciones.

4.1.2 Tasa de desempleo de Francia

En primer lugar, siguiendo los mismos pasos que con la tasa de desempleo de España, me dispongo a la detección de valores atípicos de la tasa de desempleo de Francia.

Tras realizar el mismo procedimiento para la detección de outliers que en el caso español, obtengo los resultados reflejados en la Tabla 4.1.2.1.

Tabla 4.1.2.1: Outliers detectados en la tasa de desempleo de Francia.

Tipo	Periodo	Parámetro estimado	Error estándar	Valor t
AO	1983:01	0,5458	0,06754	8,12
AO	1983:03	-0,1501	0,04140	-3,63
LS	2003:01	-0,2409	0,06243	-3,86
AO	2003:08	-0,1490	0,04129	-3,61
LS	2005:01	-0,2218	0,06243	-3,55
LS	2010:01	-0,2283	0,06243	-3,66
AO	2016:07	-0,1510	0,04130	-3,66
LS	2016:10	0,2618	0,06244	4,19
LS	2018:01	0,2587	0,06243	4,14

Para esta serie temporal se obtienen nueve valores atípicos de dos tipos. Los hay de tipo AO (tipo impulso) y de tipo LS (tipo escalón). Estos valores quedan corregidos en la nueva serie generada por el programa, denominada “u_rate_F_xl”, que se corresponde con una serie linealizada de outliers mediante TRAMO.

Una vez incorporada al conjunto de datos, me dispongo a la detección de tendencia y estacionalidad de la serie. En primer lugar, realizo los contrastes de tendencia, continuando después con los de estacionalidad.

El Contraste de Rachas, para esta serie, arroja la salida de la Tabla 4.1.2.2.

Tabla 4.1.2.2: Salida del Contraste de Rachas de la tasa de desempleo de Francia.

Número de rachas	Discrepancia	Valor "p" a dos colas
11	-20,1944	1,09678e-090

Observando el valor “p” a dos colas, puedo concluir que, al ser este lo suficientemente pequeño, la serie tendría tendencia.

A su vez, realizando el Contraste de Rangos, se obtiene la salida recogida en la Tabla 4.1.2.3.

Tabla 4.1.2.3: Salida del Contraste de Rangos de la tasa de desempleo de Francia.

Coef. de rangos	Discrepancia	Nivel crítico
0,214933	4,53401	5,78735e-006

En este caso, el nivel crítico del Contraste de Rangos, al ser lo suficientemente pequeño, indica que se rechazaría la hipótesis nula del contraste, teniendo que la serie tendría tendencia. Como se puede observar, en esta serie no se muestra ningún tipo de contradicción en estos contrastes, llegando a la conclusión de que, efectivamente, puede decirse que esta serie tendría tendencia.

Siguiendo con el procedimiento llevado a cabo en la serie española, me dispongo a la realización de los contrastes de estacionalidad. El Contraste de Kruskal-Wallis aplicado a esta serie temporal arroja la salida recopilada en la Tabla 4.1.2.4.

Tabla 4.1.2.4: Salida del Contraste de Kruskal-Wallis de la tasa de desempleo de Francia.

Nº observaciones	Grados de libertad	Valor del estadístico	Nivel crítico
446	11	0,259825	1,00

Observando el nivel crítico de este contraste, se puede llegar a la conclusión de que la serie no presenta estacionalidad determinista.

A continuación, para analizar la estacionariedad de esta serie temporal realizo en primer lugar el contraste de rango media con el fin de analizar si esta serie es estacionaria en varianza o no. Los resultados de este contraste quedan recogidos en la Tabla 4.1.2.6.

Tabla 4.1.2.6: Salida del contraste de rango-media para la tasa de desempleo de Francia.

Submuestras	Tamaño muestral	Pendiente	Valor p
21	21	-0,221645	0,0665532

La salida de este contraste llevaría al no rechazo de la hipótesis nula de este al 5%, pero sí que se rechazaría al 10%. Como nivel crítico de rigor he establecido el 5%, por lo que considero la aceptación de la hipótesis nula de este contraste, o lo que es lo mismo, la serie es estacionaria en varianza.

Para el análisis de la estacionariedad en media, hay que realizar los contrastes de raíz unitaria (ADF y ADF-GLS) y el contraste de estacionariedad en media KPSS, tal y como hice anteriormente. Las salidas de estos contrastes están resumidas en la Tabla 4.1.2.7.

Tabla 4.1.2.7: Salidas de los contrastes ADF, ADF-GLS y KPSS para la tasa de desempleo de Francia.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	NO	SÍ	-	-2,68693	0,07625
ADF-GLS	-	-	-	-1,05608	0,2633
KPSS	-	-	-	0,739904	0,01

Las salidas de los contrastes recogidos en la Tabla 4.1.2.7 conducen a la conclusión de existencia de una raíz unitaria y de no estacionariedad en media. De esta manera, sería necesaria la diferenciación de la serie estudiada en este apartado. Al diferenciar y repetir los contrastes, se obtienen las salidas resumidas en la Tabla 4.1.2.8.

Tabla 4.1.2.8: Salidas de los contrastes ADF, ADF-GLS y KPSS para la tasa de desempleo de Francia diferenciada.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	NO	NO	SÍ	-4,74901	2,441E-6
ADF-GLS	-	-	-	-1,53456	0,1174
KPSS	-	-	-	0,538033	0,039

Como se puede observar, los resultados de los contrastes inducen a una contradicción, ya que en el contraste ADF, al ser el valor “p” lo suficientemente pequeño, se rechazaría la hipótesis nula, mientras que en el ADF-GLS se da el caso contrario. Este segundo contraste es una extensión del anterior, por lo que considero más acertada la salida de este. Además, el contraste KPSS llevaría a rechazar la hipótesis nula de estacionariedad en media, lo que aclara el problema. Sería necesario realizar una diferencia adicional para conseguir la estacionariedad en media.

Diferenciando otra vez la serie y realizando los mismos contrastes, se obtienen las salidas resumidas en la Tabla 4.1.2.9.

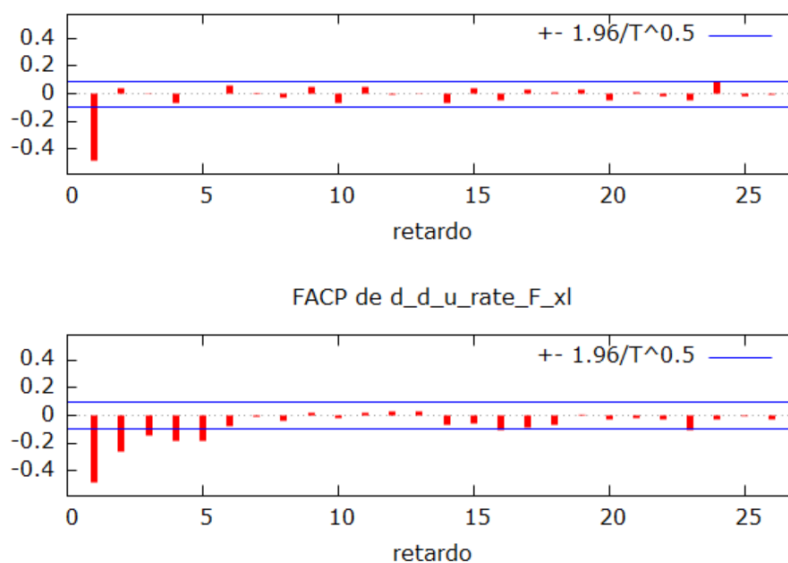
Tabla 4.1.2.9: Contrastes ADF, ADF-GLS y KPSS para la tasa de desempleo de Francia en segundas diferencias.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	NO	NO	SÍ	-13,4796	1,293E-28
ADF-GLS	-	-	-	-12,9088	3,465E-27
KPSS	-	-	-	0,0166249	>0,10

Los resultados obtenidos conducen a la conclusión de que la serie en segundas diferencias es estacionaria en media. Esto quiere decir que la tasa de desempleo de Francia es integrada de orden 2.

A continuación, me dispongo al análisis de las Funciones de Autocorrelación y Autocorrelación Parcial, con el fin de identificar los componentes AR(p) y MA(q) de esta serie. Estas dos funciones están recogidas en la Figura 4.1.2.1.

Figura 4.1.2.1: FAC y FACP de la tasa de desempleo de Francia en segundas diferencias.
FAC de d_d_u_rate_F_xl



Como expliqué en el anterior apartado, la identificación AR(p) y MA(q) se consigue mediante la observación de las FACP y FAC respectivamente.

Observando la FACP, se puede ver que hay dos retardos bastante significativos y otros bastante cerca de no ser significativos. Esto podría interpretarse como que $p = 2$.

En cambio, si se mira hacia la FAC, se puede ver claramente que solo hay un retardo significativo, por lo que se trataría de un MA(1).

Por tanto, en resumen, el modelo identificado sería un ARIMA (2,2,1). Además, no parece que en las FAC y FACP haya algún indicio que pudiera indicar la existencia de raíces estacionales.

A continuación, me dispongo a la modelización ARIMA de esta serie temporal. En primer lugar, estimo el modelo propuesto anteriormente, que se corresponde con un ARIMA (2,2,1). Al estimar este modelo se observa que ni las raíces AR ni la constante son significativas. Elimino en primer lugar la constante, ya que es la que cuenta con un mayor valor “p” (0,6166).

Al eliminarla, siguen sin salir significativas las raíces AR. Por ello, en primer lugar, las elimino de una en una, viendo que ninguna acaba siendo significativa. De esta manera, resultaría un modelo ARIMA (0,2,1). Los resultados más importantes de este modelo comparados con los de un ARIMA (2,2,1) quedan resumidos en la Tabla 4.1.2.10.

Tabla 4.1.2.10: Comparativa de modelos ARIMA para la tasa de desempleo de Francia.

	ARIMA (0,2,1)	ARIMA (2,2,1)
R-cuadrado corregido	0,994272	0,994265
Akaike	-1064,559	-1061,639
Schwarz	-1.056,377	-1045,274
Hannan-Quinn	-1061,332	-1055,184
Log-verosimilitud	534,2796	534,8195

Como se puede ver en la tabla, los valores de los estadísticos recogidos son similares. Si bien es cierto que el modelo ARIMA (0,2,1) maximiza el R-cuadrado corregido y minimiza los criterios de información, el modelo ARIMA (2,2,1) maximiza el logaritmo de verosimilitud. Al ser los resultados tan ajustados, analizaré ambos modelos.

Antes de pasar a las predicciones es necesario llevar a cabo contrastes sobre los modelos. En primer lugar, analizaré la normalidad de los residuos, en segundo lugar, los correlogramas de los residuos y, finalmente, analizaré la autocorrelación de los modelos. Todo este análisis lo realizaré siguiendo los mismos pasos que con la serie española.

Comenzando con la normalidad de los residuos, se obtiene que el contraste para el modelo ARIMA (0,2,1) toma un valor “p” igual a 0,0395536. Este valor, por tanto, sería lo suficientemente pequeño como para rechazar la hipótesis nula de normalidad de los residuos. En cambio, el mismo contraste para el modelo ARIMA (2,2,1) arroja un valor “p” igual a 0,0508826, que llevaría al rechazo de la hipótesis nula al 10% pero no al 5%.

A continuación, me dispongo a analizar los correlogramas de ambos modelos. Estos están recogidos en las Figuras 4.1.2.2 y 4.1.2.3. En ambas figuras no puede detectarse ningún tipo de patrón que pudiera llevar a la obtención de un modelo ARMA (p,q).

Figura 4.1.2.2: correlograma de los residuos del modelo ARIMA (0,2,1) para la tasa de desempleo de Francia.

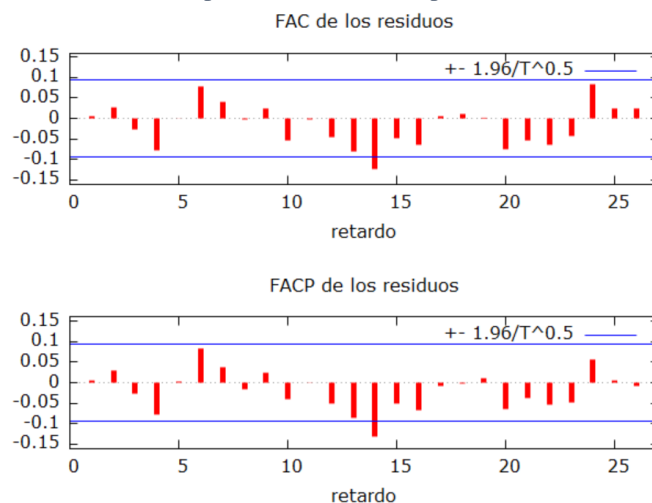
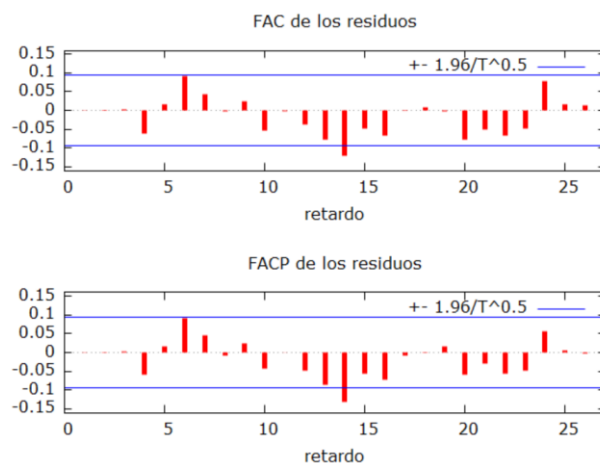


Figura 4.1.2.3: correlograma de los residuos del modelo ARIMA (2,2,1) para la tasa de desempleo de Francia.



Finalmente, el Contraste Ljung-Box de autocorrelación lleva a la misma conclusión en ambos modelos. Para el modelo ARIMA (0,2,1), se obtiene un valor “p” igual a 0,5754, lo que llevaría al no rechazo de la hipótesis nula del contraste, es decir, no habría autocorrelación en este. Para el modelo ARIMA (2,2,1), se llegaría a la misma conclusión con un valor “p” igual a 0,4784.

Una vez concluido el análisis de ambos modelos, me dispongo a realizar predicciones con cada modelos. Los valores de estas predicciones quedan recogidos en la Tabla 4.1.2.11. La evaluación de estas está resumida en la Tabla 4.1.2.12.

Tabla 4.1.2.11: Predicciones de los modelos ARIMA para la tasa de desempleo de Francia.

Periodo	ARIMA (0,2,1)	ARIMA (2,2,1)
ene-20	8,134337	8,135259
feb-20	8,084149	8,090038
mar-20	8,033961	8,041557
abr-20	7,983773	7,99316
may-20	7,933584	7,944536
jun-20	7,883396	7,895904
jul-20	7,833208	7,847255
ago-20	7,78302	7,798604
sep-20	7,732831	7,749953
oct-20	7,682643	7,701301
nov-20	7,632455	7,652648
dic-20	7,582267	7,603996

En cuanto a los valores de las predicciones para esta serie temporal, puede verse que en el modelo ARIMA (0,2,1) estos son ligeramente menores que los del otro modelo estudiado.

Tabla 4.1.2.12: Estadísticos para la evaluación de las predicciones de los modelos ARIMA propuestos para la tasa de desempleo de Francia.

Estadístico	ARIMA (0,2,1)	ARIMA (2,2,1)
Porcentaje de error absoluto medio	0,30893	0,33506
U de Theil	0,0037663	0,055123
Proporción de sesgo	0,5075	0,38949

En cuanto a la evaluación de las predicciones, puede verse que el modelo ARIMA (0,2,1) mejora los estadísticos U de Theil y porcentaje de error absoluto medio, mientras que el modelo ARIMA (2,2,1) mejora la proporción de sesgo. De todos modos, los valores de estos estadísticos indican unas muy buenas predicciones, viendo que la U de Theil se aproxima mucho a cero, el porcentaje de error absoluto medio no pasa de 0,35 y la proporción de sesgo es bastante más pequeña que en el caso español en ambos modelos. Considero que las predicciones del modelo ARIMA (0,2,1) son algo mejores que las del ARIMA (2,2,1).

4.1.3 Tasa de desempleo de Alemania.

En primer lugar y como en el resto de las series temporales, me dispongo al análisis de los valores atípicos de ésta. Los resultados de la obtención de outliers están reflejados en la Tabla 4.1.3.1.

Tabla 4.1.3.1: Outliers detectados en la tasa de desempleo de Alemania.

Tipo	Periodo	Parámetro estimado	Error estándar	Valor t
LS	2004:02	0,0328	0,00652	5,03
LS	2008:08	-0,0256	0,00652	-3,93
LS	2013:01	0,0234	0,00652	3,59
LS	2018:04	-0,0276	0,00670	-4,12
LS	2018:10	-0,0323	0,00695	-4,65
LS	2019:01	-0,0348	0,00683	-5,10
LS	2019:05	-0,0286	0,00691	-4,15
LS	2019:10	0,0310	0,00710	4,37

En este caso, la serie tendría ocho valores atípicos, que se corresponden con las observaciones señaladas en la tabla. Para su introducción a la base de datos, TRAMO confecciona la serie linealizada de outliers “u_rate_G_xl”.

A continuación, me dispongo a la realización de los contrastes de tendencia y estacionalidad. En primer lugar, realizaré los de tendencia (rachas y rangos) y, posteriormente, los de estacionalidad (Kruskal-Walis).

Comenzando con el Contraste de Rachas, al aplicarlo a la serie estudiada en este apartado, se obtiene la siguiente salida:

Tabla 4.1.3.2: Salida del Contraste de Rachas de la tasa de desempleo de Alemania.

Número de rachas	Discrepancia	Valor "p" a dos colas
7	-18,0926	3,64307e-073

La salida de este contraste me lleva a rechazar la hipótesis nula de este, ya que el valor “p” a dos colas que se obtiene es lo suficientemente pequeño. Es decir, podría concluir que esta serie, según este contraste, tendría tendencia.

A continuación, el Contraste de Rangos nos lleva a la misma conclusión que el anterior. La salida que se obtiene al realizarlo es la siguiente:

Tabla 4.1.3.3: Salida del Contraste de Rangos de la tasa de desempleo de Alemania.

Coef. de rangos	Discrepancia	Nivel crítico
-1,06938	-19,9777	8,61046e-089

Al tener un nivel crítico lo suficientemente pequeño, podría rechazarse la hipótesis nula de este contraste, llegando a la conclusión de que, según este, la serie tendría tendencia.

Siguiendo con el procedimiento mencionado con anterioridad, me dispongo a analizar la presencia de estacionalidad. El Contraste de Kruskal-Wallis arroja la siguiente salida:

Tabla 4.1.3.4: Salida del Contraste de Kruskal-Wallis de la tasa de desempleo de Alemania.

Nº observaciones	Grados de libertad	Valor del estadístico	Nivel crítico
350	11	0,142736	1,00

El elevado nivel crítico del contraste lleva a no rechazar la hipótesis nula de no estacionalidad en la serie.

La salida arrojada por el contraste de rango-media queda recogida en la Tabla 4.1.3.6. En esta tabla puede verse que, al contrario de como sucede con las anteriores series temporales, el valor “p” del contraste es lo suficientemente pequeño como para rechazar la hipótesis nula de estacionariedad en varianza. De esta manera, sería necesario realizar una transformación logarítmica Box-Cox.

Tabla 4.1.3.6: Salida del Contraste de rango-media de la tasa de desempleo de Alemania.

Submuestras	Tamaño muestral	Pendiente	Valor p
29	12	0,062522	0,0280022

Para la realización de esta transformación, he generado una nueva variable denominada “l_u_rate_G_xl”, la cual se corresponde con la transformación logarítmica de la serie temporal estudiada en este apartado. De esta manera, puede verse que los resultados del contraste de rango-media para esta serie transformada (reflejados en la Tabla 4.1.3.7) indican estacionariedad en varianza.

Tabla 4.1.3.7: Salida del contraste de rango-media del logaritmo de la tasa de desempleo de Alemania.

Submuestras	Tamaño muestral	Pendiente	Valor p
29	12	0,00778059	0,751175

Una vez conseguida la estacionariedad en varianza, me dispongo a realizar los contrastes de raíces unitarias y estacionariedad en media. Las salidas de estos contrastes están resumidas en la Tabla 4.1.3.8.

Tabla 4.1.3.8: Salida de los contrastes ADF, ADF-GLS y KPSS del logaritmo de la tasa de desempleo de Alemania.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	SÍ	-	-	-2,68081	0,2446
ADF-GLS	-	-	-	-0,722888	>0,10
KPSS	-	-	-	1,26325	<0,01

Para esta serie temporal, a diferencia de las anteriores, la tendencia es significativa. Por ello, todos los contrastes los he realizado incluyendo la tendencia. Como bien se puede ver en la Tabla 4.1.3.8, los valores “p” de todos los contrastes llevan a la misma conclusión que en las anteriores series temporales. Los contrastes ADF y ADF-GLS llevan al no rechazo de sus hipótesis nulas, lo que indicaría la necesidad de tomar diferencias en la serie. El valor “p” del Contraste KPSS lleva al rechazo de su hipótesis nula, lo que confirma que la serie temporal no es estacionaria en media.

Tras tomar una diferencia y realizar los mismos contrastes se obtienen las salidas reflejadas en la Tabla 4.1.3.9. Los valores “p” de los contrastes ADF y ADF-GLS llevan a la conclusión de que no habría que diferenciar esta nueva serie ya que sería estacionaria en media. En cambio, el contraste KPSS indicaría lo contrario.

Tabla 4.1.3.9: Salida de los contrastes ADF, ADF-GLS y KPSS de la diferencia del logaritmo de la tasa de desempleo de Alemania.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	SÍ	-	-	-4,41735	0,001997
ADF-GLS	-	-	-	-3,78423	<0,01
KPSS	-	-	-	0,167627	0,039

Al darse esta contradicción entre contrastes, considero conveniente la realización de una segunda diferencia con el fin de que el posterior modelo tenga suficiente rigor. Las salidas de los contrastes de la variable en segundas diferencias están resumidas en la Tabla 4.1.3.10.

Tabla 4.1.3.10: Salida de los contrastes ADF, ADF-GLS y KPSS de la segunda diferencia del logaritmo de la tasa de desempleo de Alemania.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	NO	NO	SÍ	-6,95959	2,231e-11
ADF-GLS	-	-	-	-11,3446	4,333e-23
KPSS	-	-	-	0,0153909	>0,10

Los resultados obtenidos tras esta segunda diferenciación indican claramente que la serie en segundas diferencias es estacionaria en media. De esta manera, podría decirse que la serie de logaritmos de la tasa de desempleo de Alemania es integrada de orden 2, ya que, para que esta serie sea estacionaria en media, sería necesaria la realización de dos diferencias.

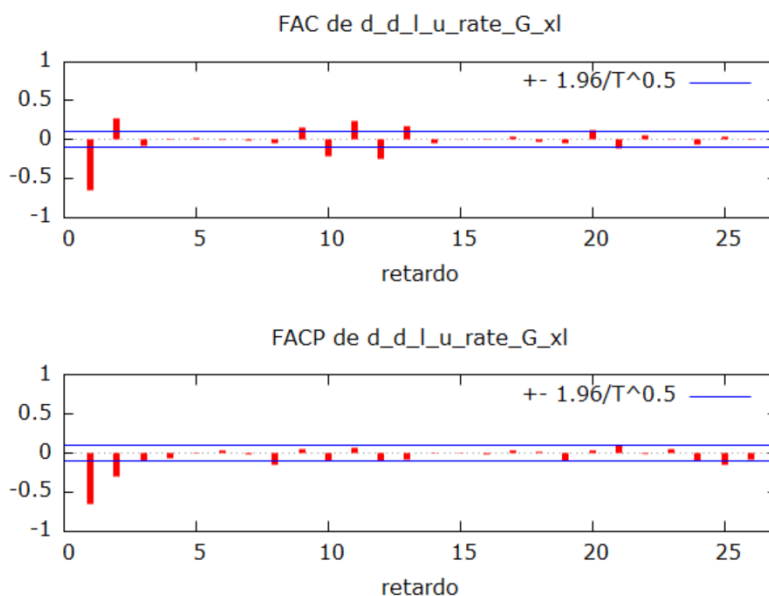
Continuando con el procedimiento aplicado en las anteriores series temporales, me dispongo a la identificación de los componentes AR(p) y MA(q) para la realización de un modelo ARIMA (p, d, q). Para ello, me fijo en las FAC y FACP de la serie temporal transformada (Figura 4.1.3.1).

En la figura mencionada puede verse con claridad que en ambas funciones los dos retardos iniciales son significativos, lo que podría significar que estamos ante un modelo ARIMA (2,2,2).

Además, puede observarse un patrón en la FAC, cada más o menos, 10 retardos, el cual hace que haya retardos significativos entre el número 9 y el 13. Al no repetirse este suceso en el tiempo, podría deducirse que se debe a una raíz estacional MA (Q), siendo $Q = 1$.

Por ello, el modelo que selecciono para realizar predicciones en esta serie temporal es un ARIMA (2,2,2) x SARIMA (0,0,1).

Figura 4.1.3.1: FAC y FACP de la tasa de desempleo de Alemania Box-Cox en segunda diferencia.



Al realizar el modelo seleccionado, todas las raíces son significativas salvo la constante. Una vez eliminada, todas siguen siendo significativas, por lo que no habría necesidad de eliminar ninguna raíz, pudiendo significar esto que el modelo es bastante correcto. Los valores de los estadísticos principales están recogidos en la Tabla 4.1.3.11.

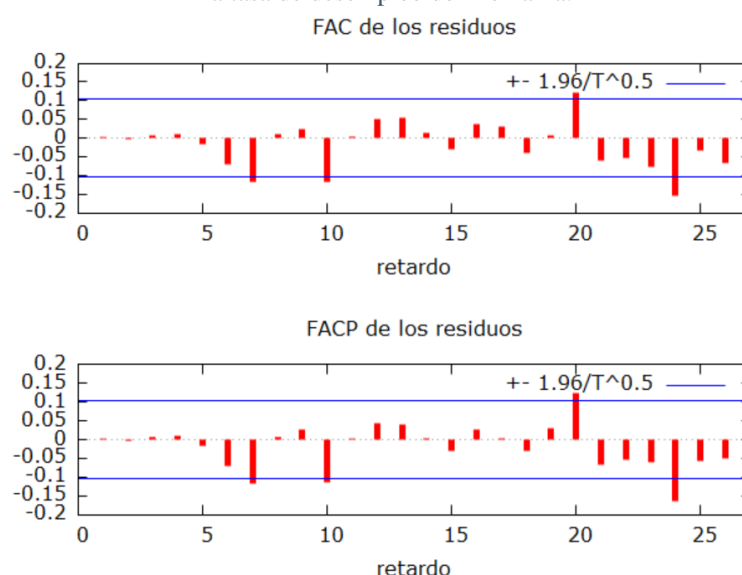
Tabla 4.1.3.11: Estadísticos principales del modelo ARIMA (2,2,2) x SARIMA (0,0,1) para la serie de logaritmos de la tasa de desempleo de Alemania.

Estadístico	Valor
R-cuadrado corregido	0,999451
Akaike	-2355,01
Schwarz	-2.331,9320
Hannan-Quinn	-2345,82
Log-verosimilitud	1183,505

Para comprobar la bondad del modelo, me dispongo a efectuar el análisis realizado en las anteriores series temporales. En primer lugar, me dispongo a contrastar la normalidad de los residuos de este modelo. El valor “p” obtenido (0,08320) llevaría al rechazo de la hipótesis nula de normalidad al 10% pero no al 5%. Al ser, dentro de lo que cabe, elevado, considero que los residuos siguen una distribución normal.

En cuanto al correlograma de los residuos, como bien se ve en la Figura 4.1.3.2, es cierto que hay retardos significativos, pero no muestran aparentemente un patrón de comportamiento que llevase a la identificación de un modelo ARMA (p,q).

Figura 4.1.3.2: correlograma de los residuos del modelo propuesto para la tasa de desempleo de Alemania.



Finalmente, en cuanto a la autocorrelación, al realizar el Contraste Ljung-Box se obtiene un valor “p” igual a 0,08015. Este valor es lo suficientemente pequeño como para rechazar la no autocorrelación al 10% pero no al 5%. Como el criterio de rechazo que he utilizado en este trabajo ha sido el 5%, decido no rechazar la hipótesis nula de este contraste, obteniendo que no habría autocorrelación en el modelo.

Una vez comprobados estos puntos, me dispongo a la realización de predicciones. Al haber transformado el modelo en logaritmos, es necesario deshacer este cambio calculando los antilogaritmos de las predicciones. Los resultados quedan resumidos en la Tabla 4.1.3.12.

Tabla 4.1.3.12: Predicciones para la tasa de desempleo de Alemania.

Periodo	Predicción Tasa desempleo Alemania
ene-20	3,281973768
feb-20	3,286972735
mar-20	3,287308024
abr-20	3,289359944
may-20	3,288557438
jun-20	3,287876777
jul-20	3,285076699
ago-20	3,284232543
sep-20	3,280746539
oct-20	3,278654091
nov-20	3,274610727
dic-20	3,270549456

En los valores predichos para esta serie, se puede ver un muy ligero repunte en los primeros datos acompañado de una tendencia descendente a partir de abril de 2020 (sin contar éste).

En cuanto a la evaluación de estas predicciones, los valores de los estadísticos principales quedan recogidos en la Tabla 4.1.3.13. Puede verse que el porcentaje de error absoluto medio es muy pequeño, lo que es algo bastante positivo para el modelo. Sin embargo, el índice U de Theil es muy elevado, que refleja limitaciones en estas predicciones. En cuanto a la proporción de sesgo, se sitúa en valores intermedios. De todos modos, hay que tener en cuenta que estos estadísticos han sido obtenidos para la serie en logaritmos, por lo que son una aproximación que no va referida directamente a la tasa de desempleo de Alemania.

Tabla 4.1.3.13: Estadísticos de evaluación de predicciones para la tasa de desempleo de Alemania en logaritmos.

Estadístico	Valor
Porcentaje de error absoluto medio	0,091603
U de Theil	14,395
Proporción de sesgo	0,63345

4.1.4 Tasa de desempleo de Italia.

Finalmente, la cuarta y última serie temporal que voy a analizar en este trabajo es la tasa de desempleo de Italia.

En primer lugar, y para seguir con la misma estructura de análisis que con las anteriores series temporales, llevo a cabo la detección de los valores atípicos de la serie en cuestión. Aplicando el mismo procedimiento que para el resto, se obtienen los resultados de la Tabla 4.1.4.1.

Tabla 4.1.4.1: Outliers detectados en la tasa de desempleo de Italia.

Tipo	Periodo	Parámetro estimado	Error estándar	Valor t
LS	1994:10	0,7102	0,13274	5,35
AO	2014:06	-0,4092	0,09883	-4,14
LS	2014:12	-0,6639	0,13287	-5,00
LS	2015:07	-0,5800	0,13454	-4,31

Esta serie temporal, como se puede ver, tiene cuatro valores atípicos, de los cuales tres son de tipo escalón y uno de tipo impulso. La incorporación de estos outliers a la base de datos se efectúa mediante la incorporación de la variable "u_rate_I_xl". Esta variable se corresponde con la linealización de la serie original mediante TRAMO.

Siguiendo con el procedimiento utilizado en el resto de las series temporales, me dispongo a la realización de los contrastes de tendencia y, posteriormente, los de estacionalidad.

El primero de los contrastes de tendencia que voy a aplicar es el Contraste de Rachas. Aplicándolo de la misma forma que en el resto de las series temporales, se obtiene la siguiente salida:

Tabla 4.1.4.2: Salida del Contraste de Rachas de la tasa de desempleo de Italia.

Número de rachas	Discrepancia	Valor "p" a dos colas
6	-20,6684	6,66432e-095

Observando el valor “p” a dos colas de la salida del contraste, se puede llegar a rechazar la hipótesis nula del contraste, lo que querría decir que la serie estudiada en este apartado tendría tendencia determinista.

El Contraste de Rangos me lleva a la misma conclusión, siendo su salida la siguiente:

Tabla 4.1.4.3: Salida del Contraste de Rangos de la tasa de desempleo de Italia.

Coef. de rangos	Discrepancia	Nivel crítico
0,207407	4,37525	1,212953-005

A continuación, el Contraste de estacionalidad de Kruskal-Wallis arroja la siguiente salida:

Tabla 4.1.4.4: Salida del Contraste de Kruskal-Wallis de la tasa de desempleo de Italia.

Nº observaciones	Grados de libertad	Valor del estadístico	Nivel crítico
446	11	0,269755	1,00

Este resultado es el mismo que para el resto de las series temporales, obteniendo la misma conclusión de que, según este contraste, la serie no tendría estacionalidad.

Continuando con el procedimiento seguido en el resto de los apartados, realizo el contraste de rango-media con el fin de identificar si la serie temporal es estacionaria en varianza o no. La salida de este contraste queda resumida en la Tabla 4.1.4.6. El valor “p” de este contraste llevaría a la conclusión de no rechazo de su hipótesis nula, lo que indicaría que la serie sería estacionaria en varianza. De esta manera, no sería necesario realizar la transformación Box-Cox en esta serie.

Tabla 4.1.4.6: Salida del contraste rango-media de la tasa de desempleo de Italia.

Submuestras	Tamaño muestral	Pendiente	Valor p
21	21	-0,0997429	0,231833

A continuación, para analiza la estacionariedad en media realizo los mismos contrastes que en el resto de las series temporales, obteniendo las salidas resumidas en la Tabla 4.1.4.7. Los valores “p” de los contrastes realizados indican la necesidad de realizar una diferencia en la variable estudiada ya que esta no sería estacionaria en media.

Tabla 4.1.4.7: Salidas de los contrastes ADF, ADF-GLS y KPSS de la tasa de desempleo de Italia.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	NO	SÍ	-	-1,9842	0,294
ADF-GLS	-	-	-	-1,03945	0,2697
KPSS	-	-	-	0,828412	<0,01

Realizando los mismos contrastes para la variable diferenciada se obtienen los resultados resumidos en la Tabla 4.1.4.8. Los valores “p” obtenidos con la realización de estos contrastes llevan a conclusiones contradictorias. Por un lado, los contrastes ADF y KPSS llevan a la misma conclusión, teniendo que no sería necesario realizar más diferencias. Por otro lado, el ADF-GLS indica lo contrario. Al ser éste de mayor consistencia que el Contraste ADF, le doy más valor a su conclusión. De todos modos, sigue habiendo contradicción con el Contraste KPSS, por lo que, para que el posterior modelo tenga más rigor, considero la posibilidad de realizar una diferencia adicional.

Tabla 4.1.4.8: Salidas de los contrastes ADF, ADF-GLS y KPSS de la tasa de desempleo de Italia diferenciada.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	NO	NO	SÍ	-3,9722	7,256e-5
ADF-GLS	-	-	-	-0,996514	0,2866
KPSS	-	-	-	0,258131	>0,10

Las salidas de los contrastes de la serie italiana en segundas diferencias están resumidas en la Tabla 4.1.4.9. Tras la realización de esta segunda diferencia, el análisis de los contrastes lleva a una conclusión única, no existiendo ningún tipo de contradicción. Por tanto, podría decirse que la tasa de desempleo de Italia es integrada de orden 2, ya que sería necesario la realización de segundas diferencias para que esta serie sea estacionaria.

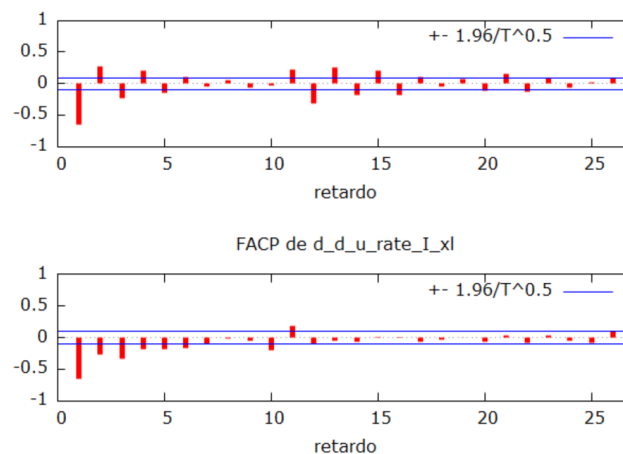
Tabla 4.1.4.9: Salidas de los contrastes ADF, ADF-GLS y KPSS de la tasa de desempleo de Italia en segundas diferencias.

Contraste	¿Tendencia significativa?	¿Constante significativa?	¿Sin constante?	Estadístico	Valor p
ADF	NO	NO	SÍ	-8,46274	2,802e-15
ADF-GLS	-	-	-	-34,5508	5,234e-24
KPSS	-	-	-	0,0148753	>0,10

Siguiendo con el procedimiento aplicado en las demás series temporales, me dispongo a la identificación de un modelo ARIMA (p, d, q) adecuado. Para ello, observo las FAC y FACP recogidas en la Figura 4.1.4.1.

En ambas funciones puede verse que los primeros tres o cuatro retardos son significativos. No hay indicios de que se trate de un proceso de memoria larga, por lo que no habría que realizar ninguna diferencia adicional. Por ello, ante la dificultad de poder deducir qué modelo se ajusta mejor a estas funciones, parto de un modelo ARIMA (2,2,2). Además, puede verse un patrón de retardos significativos en la FAC que se repite entre el retardo 11 y el 16. Por ello, considero que podría haber una parte MA (Q) estacional con $Q = 1$. De esta manera, el modelo de partida sería un ARIMA (2,2,2) x SARIMA (0,0,1).

Figura 4.1.4.1: FAC y FACP de la tasa de desempleo de Italia.



Al proceder con el modelo, puede verse que ni la constante ni la segunda raíz AR son significativas. Por ello, elimino en primer lugar esa segunda raíz y luego la constante siguiendo el criterio de eliminación de mayor a menor discrepancia. De esta manera, se llega a un modelo ARIMA (1,2,2) x SARIMA (0,0,1) sin constante. Además, he probado a incluir más raíces tanto AR como MA para comprobar si éstas salieran significativas, pero no ha sido así.

Los estadísticos principales de este modelo quedan resumidos en la Tabla 4.1.4.10. Como se puede ver en dicha tabla, los estadísticos que se obtienen son bastante positivos: el R-cuadrado corregido es muy elevado, los criterios de información son menores que los de modelos alternativos y el logaritmo de verosimilitud es bastante elevado.

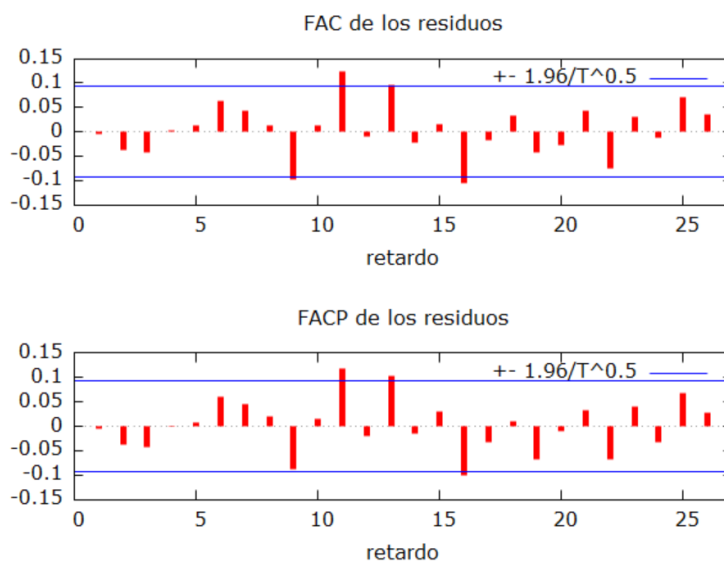
Tabla 4.1.4.10: Estadísticos del modelo ARIMA (1,2,2) x SARIMA (0,0,1) para la tasa de desempleo de Italia.

Estadístico	Valor
R-cuadrado corregido	0,992762
Akaike	-425,439
Schwarz	-404,9825
Hannan-Quinn	-417,3704
Log-verosimilitud	217,7195

Para comprobar la bondad del modelo es necesario realiza los contrastes realizados en los anteriores apartados de este trabajo. En primer lugar, el contraste de normalidad de los residuos arroja un valor “p” igual a 2,2745e-6, el cual es lo suficientemente pequeño como para rechazar la hipótesis nula de normalidad de los residuos.

En cuanto al correlograma de los residuos, tal y como se ve en la Figura 4.1.4.2, no existe ningún patrón de comportamiento que permita la identificación de algún modelo ARMA (p,q), por lo que podría decirse que el modelo inicial estaría bien especificado.

Figura 4.1.4.2: correlograma de los residuos del modelo ARIMA (1,2,2) x SARIMA (0,0,1) para la tasa de desempleo de Italia.



Finalmente, en cuanto a la autocorrelación en el modelo, del contraste Ljung-Box se obtiene un valor “p” igual a 0,04915, el cual es lo suficientemente pequeño como para rechazar la hipótesis nula de no autocorrelación.

Finalmente, me dispongo a la realización de predicciones para este modelo. Los valores de estas predicciones, resumidos en la Tabla 4.1.4.11, muestran pequeñas bajadas y subidas durante este recorrido anual. De todos modos, en líneas generales, la tendencia es decreciente.

Tabla 4.1.4.11: Predicciones para la tasa de desempleo de Italia.

Periodo	Valor
ene-20	9,715801
feb-20	9,644463
mar-20	9,697411
abr-20	9,64478
may-20	9,640482
jun-20	9,627047
jul-20	9,576763
ago-20	9,612095
sep-20	9,483925
oct-20	9,457982
nov-20	9,41637
dic-20	9,353727

En cuanto a los estadísticos de análisis de bondad de las predicciones, sus valores están recogidos en la Tabla 4.1.4.12 y confirman la buena calidad de los resultados ya que el porcentaje de error absoluto medio es bastante bajo y el estadístico U de Theil muestra un valor reducido.

Tabla 4.1.4.12: Estadísticos de análisis de la bondad de las predicciones para la tasa de desempleo de Italia.

Estadístico	Valor
Porcentaje de error absoluto medio	0,2616
U de Theil	0,11132
Proporción de sesgo	0,75944

4.2 ANÁLISIS MULTIVARIANTE

En el anterior apartado realicé un análisis individual de todas las series temporales de tasas de desempleo sin considerar ningún tipo de relación entre éstas. En este apartado, en cambio, me dispongo a abordar un enfoque multivariante, en el que se tienen en cuenta simultáneamente todas las series con su estructura temporal, con el objetivo de detectar posibles estructuras causales.

Los modelos de *Vectores Autorregresivos* (VAR) no cuentan, en principio, con restricciones de exogeneidad. Estos modelos incluyen relaciones dinámicas que están determinadas por los propios valores de las series temporales. Un modelo VAR representa relaciones para un sistema de “m” variables considerando “p” retardos. De esta manera, se obtiene que cada variable estaría explicada por los componentes

deterministas, sus respectivos “p” retardos y otros tantos retardos de las otras variables estudiadas conjuntamente.

Un modelo VAR, por tanto, aumenta su complejidad a medida que se añaden variables o retardos. Para poder ver este hecho de una manera más clara, un modelo VAR con “m” variables y “p” retardos se expresaría de la siguiente forma:

$$\begin{pmatrix} Y_{1,t} \\ \vdots \\ Y_{m,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} + \begin{pmatrix} \phi_{11}^{(1)} & \dots & \phi_{1m}^{(1)} \\ \vdots & \ddots & \vdots \\ \phi_{m1}^{(1)} & \dots & \phi_{mm}^{(1)} \end{pmatrix} * \begin{pmatrix} Y_{1,t-1} \\ \vdots \\ Y_{m,t-1} \end{pmatrix} + \dots + \begin{pmatrix} \phi_{11}^{(p)} & \dots & \phi_{1m}^{(p)} \\ \vdots & \ddots & \vdots \\ \phi_{m1}^{(p)} & \dots & \phi_{mm}^{(p)} \end{pmatrix} * \begin{pmatrix} Y_{1,t-p} \\ \vdots \\ Y_{m,t-p} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ \vdots \\ u_{m,t} \end{pmatrix}$$

Como bien puede verse en esta ecuación, cada nueva variable incorporada al modelo añadiría una fila adicional en las matrices verticales y una fila y una columna en las matrices de coeficientes ($m \times m$). Además, cada retardo adicional, añadiría una matriz de coeficientes adicional con su respectiva matriz de retardos. Un modelo como este se denomina “ $VAR_m(p)$ ”.

Uno de los primeros problemas que surgen a la hora de realizar un modelo como este, es la determinación del número de retardos. Si se cogieran pocos retardos, el modelo podría ser pobre en lo que la especificación se refiere. En cambio, si se consideran demasiados retardos, al aumentar tanto el número de parámetros, se van perdiendo grados de libertad, disminuyendo la calidad de los contrastes.

La determinación correcta del número de retardos se suele llevar a cabo mediante el estudio de los correlogramas de los residuos o mediante el análisis de los criterios de información de Akaike, Schwartz y Hannan-Quinn.

En primer lugar, es recomendable observar la FAC de cada una de las series univariantes y la FAC cruzada con cada dos variables que intervienen en el modelo. Para poder estudiar la FAC cruzada, es necesario que los procesos a estudiar sean conjuntamente estacionarios.

Dos procesos “X” e “Y” son conjuntamente estacionarios si uno de ellos es estacionario y las covarianzas cruzadas dependen del retardo “h” entre las variables y no del periodo “t” observado. Se dice que “X” influye linealmente en “Y” si existe algún valor de un retardo que no sea nulo para el que la covarianza cruzada sea distinta de cero. “Y”, por tanto, influiría en “X” linealmente si para algún retardo menor que cero la autocovarianza cruzada no es nula.

$$\sigma_{XY}(t, t + h) = \sigma_{XY}(h), \forall h = 0, \pm 1, \pm 2, \dots$$

$$\sigma_{XY}(h) = \sigma_{XY}(-h)$$

Una vez comprobado que los procesos son conjuntamente estacionarios, el coeficiente de autocorrelación cruzado para estos dos procesos tiene la siguiente forma:

$$\rho_{XY}(h) = \frac{\sigma_{XY}(h)}{\sqrt{\sigma_X(0)\sigma_Y(0)}}, h = 0, \pm 1, \pm 2, \dots$$

Este coeficiente en valor absoluto es siempre menor o igual que uno. Es un coeficiente no simétrico, dándose que $\rho_{XY}(h) = \rho_{YX}(-h)$ pero $\rho_{XY}(h) \neq \rho_{XY}(-h)$. Para valores positivos de “h”, este coeficiente indica la relación del proceso “X” hacia el proceso “Y”, dándose el caso contrario para valores negativos de este.

La función de autocorrelación cruzada ha de obtenerse de dos en dos variables, por lo que, para las variables estudiadas en este apartado, han de obtenerse tres funciones. Estas son la función de autocorrelación cruzada entre las tasas de desempleo de España y la de Italia (Figura 4.2.1), la función de autocorrelación cruzada entre las tasas de desempleo de España y de Francia (Figura 4.2.2) y la función de autocorrelación cruzada entre las tasas de desempleo de Francia e Italia (Figura 4.2.3). Además, estas funciones han de obtenerse para las variables estacionarias, lo que quiere decir que han de utilizarse las variables diferenciadas.

Figura 4.2.1: Función de autocorrelación cruzada entre las tasas de desempleo de España e Italia.

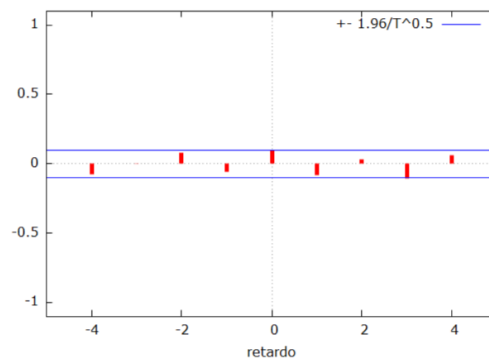


Figura 4.2.2: Función de autocorrelación cruzada entre las tasas de desempleo de España y Francia.

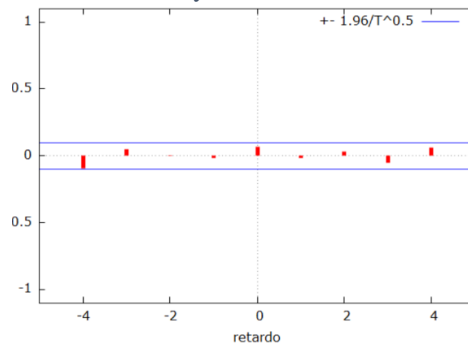
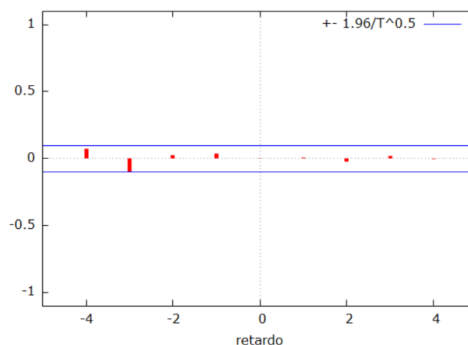


Figura 4.2.3: Función de autocorrelación cruzada entre las tasas de desempleo de Francia e Italia.



Para comprobar la existencia de autocorrelación cruzada entre los residuos de las distintas ecuaciones, se realiza el Contraste de Portmanteau (Hosking y Li-McLeod). Este es una generalización del Contraste de Ljung-Box al caso multivariante. Su hipótesis nula es la no autocorrelación y el estadístico de este contraste tiene la siguiente forma:

$$Q(j) = T^2 \sum_{h=1}^j \frac{1}{T-h} \text{tr} \left(\hat{\Sigma}_y(h) \hat{\Sigma}_y^{-1}(0) \hat{\Sigma}_y(h) \hat{\Sigma}_y^{-1}(0) \right) \sim \chi_{m^2 j}^2$$

donde “m” es el número de variables del vector \mathbf{Y} , es decir, el número de ecuaciones.

Tras haber analizado la FAC de cada una de las series temporales y la FAC cruzada con cada dos variables, habría que estimar varios modelos VAR desde un orden superior hacia abajo, comprobando cuándo se optimizan algunos criterios. Hay dos tipos de criterios: no consistentes y consistentes.

Los no consistentes son el Criterio de Error de Predicción Final (EPF) (Akaike, 1969, 1971) y el Criterio de Información AIC (Akaike 1973, 1974). El primero de los criterios tiene la siguiente forma:

$$EPF(p) = \left[\frac{T + kp + 1}{T - kp - 1} \right]^k \text{Det} \left(\hat{\Sigma}_u(p) \right)$$

La optimización de este criterio se cumple para un número de retardos que minimice el valor de la expresión.

$$EPF(\hat{p}) = \min_{i=1, \dots, j} EPF(i)$$

En cuanto al Criterio de Información AIC, también se optimiza cuando su valor es mínimo. Tiene la siguiente forma:

$$AIC(p) = \log \text{Det} \left(\hat{\Sigma}_u(p) \right) + \frac{2}{T} pk^2$$

Estos dos criterios sobrevaloran el número de retardos. Para valores altos, ambos procedimientos coincidirían en sus conclusiones.

Los criterios consistentes se denominan de esa manera porque cumplen la siguiente condición:

$$\forall \epsilon > 0, \lim_{T \rightarrow \infty} Pt(|\hat{p} - p| > \epsilon) = 0$$

Los criterios que cumplen esta expresión son el Criterio de Información HQ y el Criterio de Información BIC (Schwarz, 1978).

El primero de estos dos criterios tiene la siguiente forma:

$$HQ(p) = \log \text{Det} \left(\hat{\Sigma}_u(p) \right) + 2 \frac{\log \log T}{T} pk^2$$

En cuanto al segundo criterio, este tiene la siguiente forma:

$$BIC(p) = \log \text{Det} \left(\hat{\Sigma}_u(p) \right) + \frac{\log T}{T} pk^2$$

El Criterio de Información BIC estimará siempre los mismos o menos retardos que el Criterio de Información AIC siempre que $T \geq 8$. También estimará siempre los mismos o menos retardos que el Criterio de Información HQ para todo valor de “T”. El Criterio de Información HQ siempre estimará los mismos o menos retardos que el Criterio de Información AIC siempre que $T \geq 16$. Por tanto, se puede afirmar que, para $T \geq 16$, siempre se cumple que:

$$\hat{p}(BIC) \leq \hat{p}(HQ) \leq \hat{p}(AIC)$$

Para la determinación del número de retardos también se suele utilizar el Test de Razón de Verosimilitud. Este test tiene como hipótesis nula que el número de retardos es “ p_0 ”. Su hipótesis alternativa señala que el número de retardos es, en cambio, “ p_1 ”, que es un número de retardos mayor que “ p_0 ”.

Denominando como “ $\hat{\Sigma}_u(p_0)$ ” y “ $\hat{\Sigma}_u(p_1)$ ” a las matrices de varianzas y covarianzas de las hipótesis nula y alternativa respectivamente, la forma del estadístico de contraste sería la siguiente:

$$RV = T \cdot \log \left(\frac{|\hat{\Sigma}_u(p_0)|}{|\hat{\Sigma}_u(p_1)|} \right) \sim \chi_m^2$$

Donde “m” es el número de variables del modelo VAR.

Se recomienda estimar cualquier modelo VAR mediante máxima verosimilitud exacta. No obstante, la mayoría de los programas econométricos utilizan Mínimos Cuadrados Ordinarios (MCO) para la estimación de cada una de las ecuaciones de un modelo VAR.

Para la selección de variables en un modelo VAR, es necesario que estas sean homogéneas. Es decir, han de ser integradas del mismo orden y si están transformadas, han de estarlo todas. Tal y como hemos descrito en los apartados anteriores, las tasas de desempleo de España, Italia y Francia son todas integradas de orden 2 y, además, no están transformadas. En cambio, la tasa de desempleo de Alemania ha de transformarse mediante Box-Cox para que sea estacionaria en varianza, por lo cual esta serie no es homogénea y se excluye del análisis VAR.

Para la selección del número de retardos idóneo, Gretl compara las conclusiones de cuatro estadísticos: el logaritmo de verosimilitud y los Criterios de Información AIC, BIC y HQC. El programa señala el número de retardos para el cual cada criterio de información se minimiza y, a su vez, realiza el test de logaritmo de verosimilitud para cada retardo con el fin de que se pueda interpretar cuál es su conclusión para cada uno de ellos. Los resultados de todos estos criterios quedan resumidos en la Tabla 4.2.1.

Tabla 4.2.1: Selección de número de retardos VAR con los criterios RV, AIC, BIC y HQC.

Retardos	Log. Verosimilitud	p(RV)	AIC	BIC	HQC
1	730,65607	-	-3,772473	-3,64829	-3,723202
2	882,45136	0	-4,522054	-4,304735	-4,43583
3	902,45019	0,00001	-4,579791	-4,269334	-4,456613
4	907,56989	0,33145	-4,559422	-4,155828	-4,399291
5	915,60939	0,06525	-4,554380	-4,057649	-4,357295
6	923,9854	0,5274	-4,551104	-3,961237	-4,317066
7	933,99716	0,01777	-4,556416	-3,873411	-4,285424
8	938,06421	0,52069	-4,530521	-3,754379	-4,222576
9	944,83208	0,13982	-4,518804	-3,649525	-4,173905
10	950,9414	0,20126	-4,503629	-3,541214	-4,121778

Las casillas marcadas indican el número de retardos elegido según cada contraste. Según los criterios AIC, HQC y RV, el número de retardos correcto sería de tres. En cambio, según el criterio BIC, el número de retardos correcto sería de 2. Es conveniente comenzar estimando un modelo VAR con el mayor retardos y luego, según la salida del contraste de razón de verosimilitud, ir disminuyendo su número.

Para un número de retardos igual a 3, se rechazaría la hipótesis nula de que el retardo más largo es 2, llevando a la conclusión de que el número de retardos correcto es igual a 3. Por ello, en resumen, el número de retardos elegido para la realización de modelos VAR es de 3.

Procediendo con la realización del modelo VAR, finalmente, selecciono un modelo VAR (3). En concreto, es un modelo sin constante, ya que ésta no resultaba significativa en los distintos modelos estimados. Para el análisis de este, he seleccionado los indicadores principales en la Tabla 4.2.2.

Como se puede ver en esta tabla, el Contraste de Portmanteau tiene un valor “p” nulo, lo que sugiere sin duda el rechazo de su hipótesis nula de no autocorrelación cruzada entre las distintas ecuaciones que forman el modelo analizado.

Además de este contraste, en la tabla he introducido los contrastes F de restricciones cero. Estos contrastes son los Contrastes de Causalidad de Granger. Se dice que una serie temporal “ Y_2 ” es causa de otra serie temporal “ Y_1 ” si la segunda se puede predecir mejor utilizando los valores pasados de la primera, ceteris paribus el resto de información del modelo. La hipótesis nula de este contraste es que “ Y_2 no es causa de Y_1 en el sentido de Granger”.

Puede verse que la mayor parte de los valores “p” llevan al rechazo de esta hipótesis nula. En concreto, en el modelo de la tasa de desempleo de Italia, se observa que los valores “p” son nulos para la tasa de desempleo retardada de España, lo que llevaría al rechazo de la hipótesis nula de no causalidad (se concluye que la tasa de paro de España causa en el sentido de Granger la de Italia). En cambio, el mismo contraste para los retardos de la tasa de desempleo de Francia arroja un valor “p” que no es lo suficientemente pequeño como para rechazar la hipótesis nula de no causalidad.

Para los retardos en el modelo de la tasa de desempleo de España, los coeficientes de los retardos de la tasa de desempleo de Italia y de la de Francia son significativos para

rechazar al 5%. Así pues, se concluye que las tasas de desempleo de Italia y Francia causan la española en el sentido de Granger.

En cuanto al tercer modelo, el de la tasa de desempleo de Francia, el contraste para los coeficientes de los retardos de España arroja unos valores “p” nulos, lo que llevaría al rechazo de la hipótesis nula de no causalidad, pero en cambio, para los coeficientes de los retardos de la tasa de desempleo de Italia se obtiene un valor “p” demasiado elevado que llevaría al no rechazo de la hipótesis nula. En consecuencia, se concluye que la tasa de desempleo de España causa la de Francia.

Por tanto, el resumen de estos resultados es que existen relaciones de doble causalidad entre las tasas de desempleo de España y Francia y también entre las tasas de desempleo de España e Italia, pero en cambio no se detecta evidencia de causalidad en ningún sentido entre las tasas de desempleo de Francia e Italia.

Finalmente, los contrastes de restricciones relativos a los coeficientes del tercer retardo de todas las variables del modelo proporcionan un valor “p” también nulo, lo que llevaría al rechazo de la hipótesis nula, confirmando la adecuación del retardo seleccionado para el modelo VAR estimado.

Tabla 4.2.2: Salida del modelo VAR (3).

Indicador	Valor	Valor "p"
Contraste de Portmanteau	525,88	0,0000
Contrastes F de restricciones cero modelo Italia	-	-
Retardos de Italia	4739,6	0,0000
Retardos de España	11,032	0,0000
Retardos de Francia	2,0062	0,1126
Todas las variables (retardo 3)	1,4892	0,2170
Contraste F de restricciones cero modelo España	-	-
Retardos de Italia	3,568	0,0143
Retardos de España	75707	0,0000
Retardos de Francia	2,6376	0,0493
Todas las variables (retardo 3)	7,9247	0,0000
Contraste F de restricciones cero modelo Francia	-	-
Retardos de Italia	0,13422	0,9396
Retardos de España	7,7333	0,0000
Retardos de Francia	38139	0,0000
Todas las variables (retardo 3)	6,3914	0,0003

Los resultados en su conjunto son interesantes, por lo que podría decirse que estamos ante un buen modelo con el que continuar con las predicciones del modelo. Estas predicciones se obtienen ecuación por ecuación y pueden ser comparadas con las obtenidas anteriormente mediante modelos ARIMA.

En primer lugar, las predicciones para la tasa de desempleo de Italia quedan recogidas en la Tabla 4.2.3. Estas predicciones apuntan a un descenso continuado de esta variable, continuando con la tendencia decreciente que venía teniendo en anteriores años. Para analizar la bondad de las predicciones, es necesario analizar los estadísticos observados en las predicciones de los modelos ARIMA del anterior apartado.

Estas predicciones presentan un porcentaje de error absoluto medio igual a 0,65917, el cual puede considerarse bajo. La U de Theil toma un valor que se acerca más a cero que a la unidad ($U = 0,35889$), lo que llevaría a pensar que estamos ante predicciones bastante buenas. Dado que este valor de U es reducido en este caso no resulta preocupante la elevada proporción de sesgo (83,76%).

Tabla 4.2.3: Predicciones para la tasa de desempleo de Italia según modelo VAR (3).

Periodo	Valor
ene-20	9,663243
feb-20	9,619707
mar-20	9,547426
abr-20	9,485627
may-20	9,418554
jun-20	9,352973
jul-20	9,28592
ago-20	9,218597
sep-20	9,150483
oct-20	9,081817
nov-20	9,012533
dic-20	8,942705

En cuanto a las predicciones para la tasa de desempleo de España, éstas están recogidas en la Tabla 4.2.4. Las predicciones para esta variable también sugieren un descenso continuado de sus valores, continuando con la tendencia establecida. En cuanto al porcentaje de error absoluto medio, este toma un valor igual a 1,0064, que podría interpretarse como un valor medio/elevado. La U de Theil toma un valor superior a la unidad (1,4866) y presenta como inconveniente la elevada proporción de sesgo, que supera el 99%.

Tabla 4.2.4: Predicciones para la tasa de desempleo de España según modelo VAR (3).

Periodo	Valor
ene-20	13,6
feb-20	13,5
mar-20	13,3
abr-20	13,2
may-20	13
jun-20	12,9
jul-20	12,7
ago-20	12,5
sep-20	12,4
oct-20	12,2
nov-20	12
dic-20	11,9

Finalmente, las predicciones para la tasa de desempleo de Francia (recogidas en la Tabla 4.2.5), indican lo mismo que las de las dos variables anteriores. Estas decrecen

continuando con la tendencia previa a las predicciones. El porcentaje de error absoluto medio para estas predicciones es igual a 0,47246, pudiendo interpretarse este valor como medio/bajo. En cuanto a la U de Theil, ésta toma un valor bastante reducido (0,40759) y apenas existe sesgo (0,0037033).

Tabla 4.2.5: Predicciones para la tasa de desempleo de Francia según modelo VAR (3).

Periodo	Valor
ene-20	8,148451
feb-20	8,125284
mar-20	8,097835
abr-20	8,070887
may-20	8,042357
jun-20	8,012777
jul-20	7,982011
ago-20	7,95023
sep-20	7,917538
oct-20	7,884065
nov-20	7,849923
dic-20	7,815212

Tras este análisis del modelo VAR y las correspondientes predicciones, podría ser interesante llevar a cabo un análisis de cointegración entre las series, que permitiría profundizar aún más en el estudio de las variables seleccionadas detectando si existen relaciones estables de largo plazo entre ellas. De todos modos, este proceso incrementaría bastante la complejidad y la extensión del estudio, por lo que no será abordado en este trabajo.

5. CONCLUSIONES

En este trabajo he podido realizar un análisis de las tasas de desempleo de España, Italia, Francia y Alemania mediante técnicas univariantes y multivariantes. En este apartado resumiré los resultados más importantes y haré una comparación de las predicciones obtenidas para cada serie temporal, haciendo un comentario final sobre convergencia.

Tal y como era previsible las series temporales investigadas eran no estacionarias exigiendo transformaciones para su análisis. En cuanto a la estacionariedad en varianza, todas las series salvo la tasa de desempleo de Alemania eran estacionarias en varianza. Esto se tradujo en la necesidad de realizar una transformación logarítmica de esta serie temporal, haciendo que no fuera posible incluirla en el análisis multivariante.

Todas las series temporales analizadas eran no estacionarias en media. Realizando el análisis individual se ha podido ver que todas eran integradas de orden 2, lo que quiere decir que es necesario diferenciar dos veces cada variable para convertirla en estacionaria en media. Esto ha hecho que haya cierta homogeneidad entre las variables, pudiendo realizar así un análisis multivariante de tres de las series estudiadas en un principio.

Los modelos univariantes realizados han sido los siguientes: para la tasa de desempleo de España, el modelo que he considerado más adecuado ha sido un ARIMA (2,2,1);

para la tasa de desempleo de Francia un ARIMA (0,2,1); para la tasa de desempleo de Alemania (transformada mediante Box-Cox), el modelo que he considerado más conveniente ha sido un ARIMA (2,2,2) x SARIMA (0,0,1) y finalmente, para la tasa de desempleo de Italia, el modelo propuesto ha sido un ARIMA (1,2,2) x SARIMA (0,0,1).

Con estos modelos he obtenido las predicciones recogidas en las tablas 4.1.1.10, 4.1.2.11, 4.1.3.12 y 4.1.4.11. Estas predicciones las compararé más adelante con las del análisis multivariante.

En el apartado de análisis multivariante, he analizado conjuntamente las tasas de desempleo de España, Italia y Francia. Como dije antes, no he incluido en este análisis la tasa de desempleo de Alemania debido a que no es homogénea con el resto por la necesidad de ser transformada mediante logaritmos.

Mediante el análisis de los criterios de información y el test de razón de verosimilitudes, se obtiene que el número óptimo de retardos a considerar en la realización de este modelo es 3. De esta manera, he realizado un modelo VAR (3) con el que he podido predecir los valores futuros de las variables incluidas en el modelo. Estos valores están recogidos en las tablas 4.2.3, 4.2.4 y 4.2.5. A continuación, compararé estas predicciones con las obtenidas en el apartado univariante.

Las predicciones de los diversos modelos analizados, en líneas generales, siguen una misma tónica descendente para todas las variables estudiadas. De todos modos, queda pendiente el análisis comparativo de las diferentes series incluyendo sus respectivas predicciones. Para ello, en primer lugar, es necesario comparar las predicciones realizadas con el análisis univariante y multivariante.

En primer lugar, para las predicciones de la tasa de desempleo de España (tablas 4.1.1.10 y 4.2.4), los estadísticos de análisis de predicciones presentan valores superiores para las predicciones realizadas mediante el modelo VAR (3) salvo en la proporción de sesgo, que es ligeramente inferior. Por ello, las mejores predicciones corresponderían, en principio, al modelo univariante. Los estadísticos de comparación están resumidos en la Tabla 5.1.

Tabla 5.1: comparación de los estadísticos de evaluación de predicciones para la tasa de desempleo de España.

Estadístico	Univariante	Multivariante
Porcentaje de error absoluto medio	0,9239	1,0064
U de Theil	1,3229	1,4866
Proporción de sesgo	0,99798	0,99324

En segundo lugar, las predicciones de la tasa de desempleo de Alemania (Tabla 4.1.3.12) solo ha sido posible obtenerlas mediante el análisis univariante de la serie debido a que esta serie no es homogénea con el resto.

En tercer lugar, las predicciones de la tasa de desempleo de Francia (Tablas 4.1.2.11 y 4.2.5) muestran unos estadísticos bastante más bajos en el modelo univariante que en el multivariante, salvo la proporción de sesgo, que es notablemente menor en el modelo multivariante. Debido a que tanto el porcentaje de error absoluto medio como la U de Theil son bastante menores en el modelo univariante, considero que estas predicciones son mejores que las del modelo VAR (3). Los estadísticos de comparación están resumidos en la Tabla 5.2.

Tabla 5.2: comparación de los estadísticos de evaluación de predicciones para la tasa de desempleo de Francia.

Estadístico	Univariante	Multivariante
Porcentaje de error absoluto medio	0,30893	0,47246
U de Theil	0,0037663	0,40759
Proporción de sesgo	0,5075	0,0037033

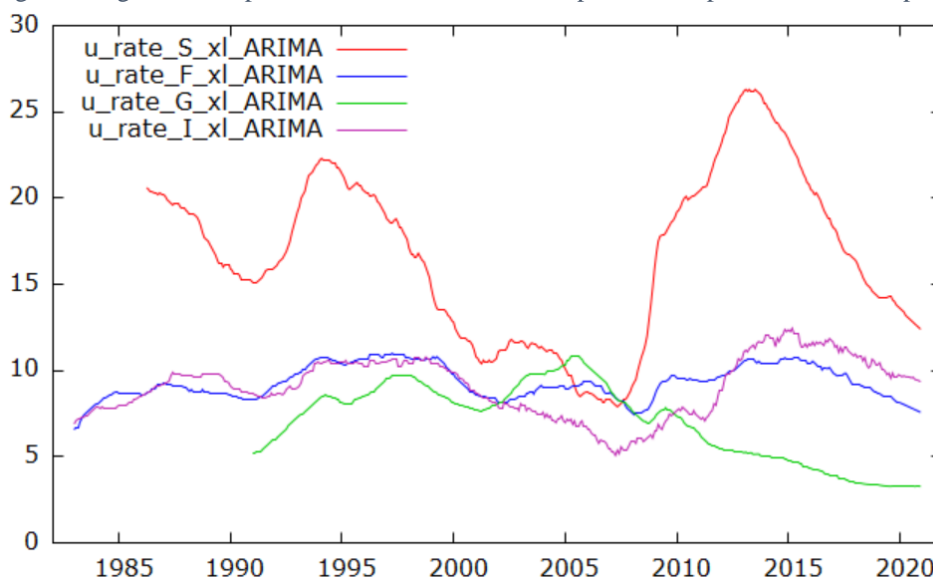
Finalmente, en cuanto a las predicciones de la tasa de desempleo de Italia (Tablas 4.1.4.11 y 4.2.3), estas muestran unos estadísticos notablemente mejores para el caso univariante. Por ello, considero que, en principio, las predicciones para esta variable según el modelo univariante realizado serían mejores que las del modelo VAR (3). Los estadísticos de comparación están resumidos en la Tabla 5.3.

Tabla 5.3: comparación de los estadísticos de evaluación de predicciones para la tasa de desempleo de Italia.

Estadístico	Univariante	Multivariante
Porcentaje de error absoluto medio	0,2616	0,65917
U de Theil	0,11132	0,35889
Proporción de sesgo	0,75944	0,83762

Tras estas comparaciones, incorporo las predicciones seleccionadas a la base de datos, con el fin de comparar todas las series temporales. Con estas incorporadas, las cuatro series temporales pueden estudiarse en un único gráfico comparativo (Figura 5.1) con el fin de analizar la evolución de todas las series.

Figura 5.1: gráfico comparativo de todas las series temporales con predicciones incorporadas.



En este gráfico se puede ver que, en líneas generales, si contamos desde 1991 (momento a partir del cual se disponen de datos mensuales para Alemania), puede que sí que exista una convergencia a la baja en este indicador, aunque no es algo que se pueda afirmar con rotundidad a simple vista.

Es muy notable cómo fluctúa la tasa de desempleo de España en comparación con la de los otros países, pero, contando las predicciones, sus valores disminuyen a una velocidad mayor que las otras tasas de desempleo.

Las tasas de desempleo de Italia y de Francia, a lo largo del tiempo, son dos series temporales bastante similares en cuanto a valores, salvo por un momento de divergencia entre 2005 y 2010. Su tendencia futura parece ser decreciente, aunque en menor medida que la tasa de desempleo de España.

Finalmente, la tasa de desempleo de Alemania es la que actualmente toma valores menores y parece estar lejos del resto. De todos modos, puede observarse que las predicciones hacen que la tendencia actual de esta serie se aplane.

La combinación de una tendencia a la baja rápida para el caso de España (serie de valores más elevados), de unas tendencias a la baja más lentas para los casos de Italia y Francia (valores medios) y de una tendencia más o menos plana para el caso de Alemania (valores más bajos) podría indicar la existencia de convergencia en la actualidad.

Estas conclusiones las he obtenido en base a la realización de un estudio previo a la expansión del COVID-19. Actualmente, la tendencia que llevaban las tasas de desempleo de la mayor parte de los países del mundo ha cambiado drásticamente debido a los grandes incrementos de las tasas de desempleo que estamos viviendo actualmente. Las predicciones de este trabajo, por tanto, sirven como aproximación de lo que habría ocurrido este año en caso de no haberse dado esta pandemia a nivel global.

De todos modos, es de esperar que con la pandemia haya asimetrías en la recuperación económica, ya que, en principio, la pandemia ha afectado más a países como España e Italia que a países como Alemania. Es probable que, por un tiempo, estemos ante un escenario de divergencia hasta que los datos comiencen a estabilizarse.

6. BIBLIOGRAFÍA

DATOS

Federal Reserve Economic Data:

<https://fred.stlouisfed.org/series/LMUNRRTTDEQ156S> que contiene los datos mensuales de la tasa de desempleo de Alemania (Consultada el 25 de abril de 2020).

<https://fred.stlouisfed.org/series/LRHUTTTTESM156S> que contiene los datos mensuales de la tasa de desempleo de España (Consultada el 25 de abril de 2020).

<https://fred.stlouisfed.org/series/LRHUTTTTFRQ156S> que contiene los datos mensuales de la tasa de desempleo de Francia (Consultada el 25 de abril de 2020).

<https://fred.stlouisfed.org/series/LRHUTTTTITQ156S> que contiene los datos mensuales de la tasa de desempleo de Italia (Consultada el 25 de abril de 2020).

Instituto Nacional de Estadística (INE):

<https://ine.es/> página correspondiente al Instituto Nacional de Estadística español.

Institut National de la Statistique et des Études Économiques (INSÉÉ):

<https://www.insee.fr/fr/accueil> página correspondiente al Instituto Nacional de Estadística francés.

Statistisches Bundesamt (SB):

https://www.destatis.de/DE/Home/_inhalt.html página correspondiente al Instituto Nacional de Estadística alemán.

Istituto Nazionale di Statistica (ISTAT):

<https://www.istat.it/> página correspondiente con el Instituto Nacional de Estadística italiano.

REFERENCIAS BIBLIOGRÁFICAS

Box, G.E.P.; Jenkins, G.M. (1976): *Time Series Analysis: Forecasting and Control*, (2ª ed.), Ed. Holden-Day, San Francisco.

Cottrell, A.; Luchetti, R. (2020): Gretl User's Guide, <http://gretl.sourceforge.net/gretl-help/gretl-guide.pdf>

European Comission (2020): *Joint Employment Report*, Directorate-General for Employment, Social Affairs and Inclusion.

Pérez, R. y López, A. J. (2019): *Econometría Aplicada con Gretl*, Creative Commons, ISBN 978-84-09-13601-8, <https://sites.google.com/view/econometriaaplicadacongretl>

Wooldridge, J. M. (2014): *Introduction to Econometrics*, Europe, Middle East and Africa Edition, Cengage Learning, Hampshire.

PÁGINAS WEB CONSULTADAS

Comisión Europea (CE): <https://ec.europa.eu/social/main.jsp?langId=en&catId=101> que contiene información sobre la Estrategia Europea de Empleo (Consultado el 24 de abril de 2020).

https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/european-semester/european-semester-timeline/national-reform-programmes-and-stability-convergence-programmes_es que contiene información sobre los Programas Nacionales de Reforma y Programas de Estabilidad o Convergencia (Consultado el 25 de abril de 2020).

https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/european-semester_en#map que contiene información sobre el Semestre Europeo (Consultado el 26 de abril de 2020).

Consejo Europeo:

<https://www.consilium.europa.eu/es/policies/european-semester/> que contiene información sobre el Semestre Europeo (Consultado e