



UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado

ANÁLISIS DE LA EVOLUCIÓN DEL DESEMPLEO EN ESPAÑA COMO SERIE TEMPORAL

Alumno: Ana Quesada Cazalla

Junio, 2015

RESUMEN

El objetivo perseguido con la realización de este Trabajo de Fin de Grado es conocer la evolución que ha experimentado el desempleo en España a lo largo de la última década para poder realizar predicciones en un futuro próximo. Y para ello, se empieza explicando el funcionamiento del Mercado de Trabajo, quien lo compone y cuáles son las fuentes que nos proporcionan la mejor información sobre él. A continuación, se elabora una introducción sobre Series Temporales que será necesaria conocer para, posteriormente, llevar a cabo un análisis econométrico de series temporales sobre el desempleo, con el que conseguiremos determinar un modelo que describa de una manera próxima a la realidad su comportamiento futuro. Finalmente, se elaboran las conclusiones en las cuales son detallados e interpretados los resultados conseguidos con el análisis.

ABSTRACT

The objective pursued with the development of this Final Degree Project is to know the evolution that the unemployment has experienced in Spain over the past decade to make predictions in the near future. And for that, we start explaining the functioning of the Labour Market, who composes it and what are the sources that give us the best information about it. Then, we will elaborate an introduction to time series that will later be useful to know to carry out an econometric analysis of time series on unemployment, to determine a model that describes its future behavior of a way close to reality. Finally, we will elaborate the conclusions where we will expose and explain the results that have been obtained in the analysis.

ÍNDICE

RESUMEN	1
1. INTRODUCCIÓN	3
2. EL MERCADO DE TRABAJO	4
2.1. Encuesta de Población Activa (EPA)	5
2.2. Servicio Público de Empleo Estatal (SEPE)	8
2.3. Diferencias entre ambos métodos	9
2.4. El mercado de trabajo en la actualidad.....	10
3. SERIES TEMPORALES	11
3.1. Concepto de serie temporal	11
3.2. Componentes de una serie temporal.....	11
3.3. Tipos de metodologías de análisis.....	15
3.4. Clasificación descriptiva de series temporales.....	17
3.5. Metodología ARIMA. El enfoque Box-Jenkins.....	18
4. APLICACIÓN DE LA METODOLOGÍA BOX-JENKINS	23
4.1. Identificación.....	25
4.2. Estimación del Modelo.....	31
4.3. Examen de diagnóstico.....	33
4.4. Pronóstico.....	39
5. CONCLUSIONES	41
ANEXO	42
BIBLIOGRAFÍA.....	45
RECURSOS WEB.....	47

1. INTRODUCCIÓN

Durante los últimos años en España, el desempleo ha sido uno de los grandes problemas que se han experimentado en nuestro país, ocasionando tanto desestabilidad en el mercado de trabajo como rechazo y exclusión social en las personas que son objeto del mismo.

La evolución del desempleo queda reflejada con la información que proporciona el Instituto Nacional de Estadística en la Encuesta de Población Activa (EPA) que elabora de cada trimestre. En el último periodo del año 2005, la tasa de paro en España se encontraba en el 10,17 por ciento, y actualmente, diez años más tarde, se sitúa en un 23,78 por ciento, es decir, ha sufrido un notable incremento, que nos hace apreciar, de forma clara, el fuerte impacto que está teniendo la actual crisis económica en España.

La elaboración de este trabajo constará de dos partes diferenciadas. Por un lado, una parte teórica donde trataré de hacer una introducción acerca del Mercado de trabajo y las Series Temporales, hablando sobre sus aspectos básicos, características, etc. Y por otro lado, una parte práctica, en la cual procederé a la estimación de un modelo econométrico que me ayude a obtener un pronóstico sobre la futura evolución del desempleo en España. Para ello, necesitaré la ayuda del software econométrico Gretl (Gnu Regression, Econometrics and Time-series Library), totalmente libre y disponible en la dirección web: <http://gretl.sourceforge.net>

Para llevar a cabo esta parte práctica sobre el análisis de la evolución del desempleo en España voy a utilizar la metodología ARIMA, para que a través de variaciones y regresiones de datos estadísticos pueda llegar a encontrar patrones para una predicción con vistas hacia el futuro. Se trata de un modelo donde las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes.

2. EL MERCADO DE TRABAJO

Cuando hablamos del mercado de trabajo, sólo podemos referirnos a la parte de la población que legalmente puede ser retribuida, por lo tanto, podría definirse como el entorno económico en el que conviven las personas que buscan un empleo y las empresas que ofrecen dicho empleo.

Para la teoría económica, el mercado de trabajo debe analizarse de la misma forma que es analizado el mercado de cualquier otro bien. Con ayuda de las siguientes gráficas (Figura 1), podemos ver de forma clara que si se produce un exceso de oferta de empleo (Figura 1.2), el salario cae automáticamente y si, por el contrario, lo que tenemos es un exceso de demanda (Figura 1.1), podemos decir que se produce un aumento del salario real.

Figura 1. Desequilibrios en el mercado laboral

Figura 1.1

Desplazamiento de la demanda de empleo

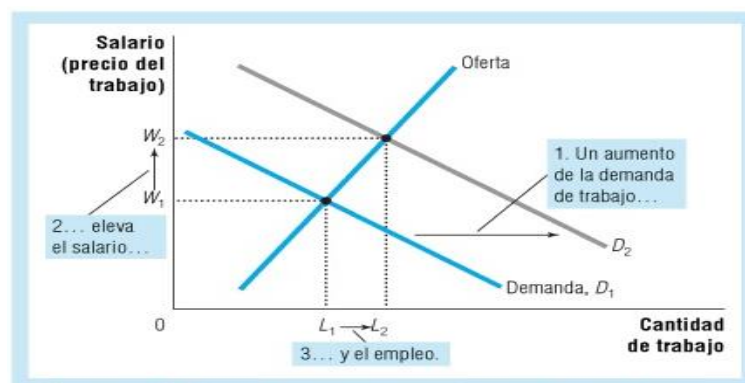


Figura 1.2

Desplazamiento de la oferta de empleo

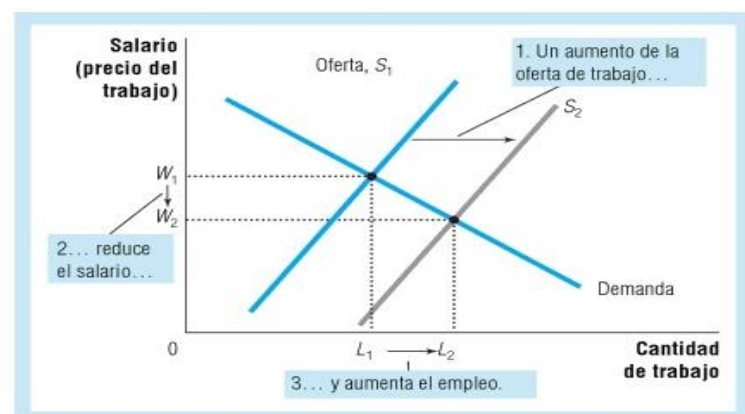
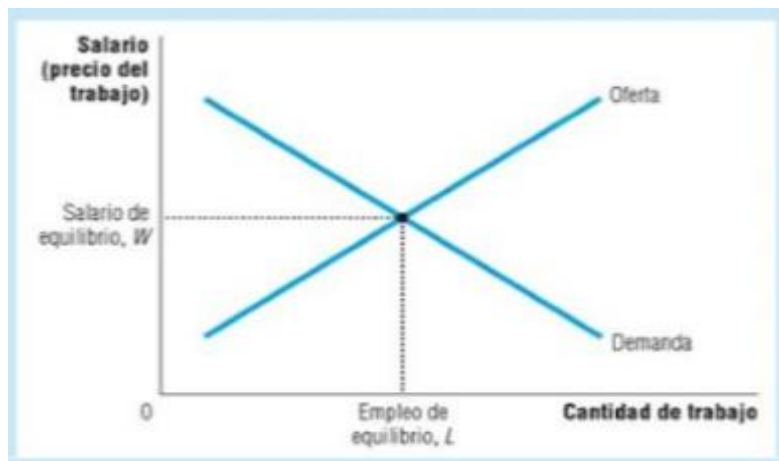


Figura 2. Equilibrio del mercado de trabajo



Fuente: Principios de Economía (N. Gregory Mankiw)

Por otro lado, también podemos llegar a un punto de equilibrio en el mercado de trabajo, que se produce cuando las curvas de oferta y demanda se cruzan. Es decir, existe un punto de equilibrio en el que se obtiene el salario para el cual la cantidad de personas dispuestas a trabajar, coincide con la cantidad de empresas que quieren contratar (*Figura 2*).

El objetivo de este trabajo en definitiva es analizar el comportamiento del desempleo, por lo que se van a presentar las dos fuentes de datos que se suelen considerar: la Encuesta de Población Activa y el Servicio Público de Empleo Estatal.

2.1. Encuesta de Población Activa (EPA)

Según establece el propio Instituto Nacional de Estadística, “la Encuesta de Población Activa (EPA) es una investigación por muestreo, continua y de periodicidad trimestral, dirigida a las viviendas familiares, que se realiza desde 1964 y cuyo objetivo principal es obtener datos de la fuerza de trabajo y de sus diversas categorías (ocupados, parados), así como de la población ajena al mercado laboral (inactivos)”. A continuación, se van a definir los aspectos básicos que se analizan en ella para llevar a cabo el estudio del mercado de trabajo español:

- Población Activa

La población activa está formada por “personas de 16 o más años que, durante la semana de referencia, suministran mano de obra para la producción de bienes y

servicios o están disponibles y en condiciones de incorporarse a dicha producción”.
(INE, *Nota de prensa*, 23 de Abril 2015)

- Población Ocupada

Por contra, la población ocupada son “personas de 16 o más años que durante la semana de referencia han estado trabajando durante al menos una hora, a cambio de una retribución (salario, jornal, beneficio empresarial,...) en dinero o especie”. Y además también son ocupados “quienes teniendo trabajo han estado temporalmente ausentes del mismo por enfermedad, vacaciones, etcétera”. (INE, *Nota de prensa*, 23 de Abril 2015)

- Población Parada o Desempleada

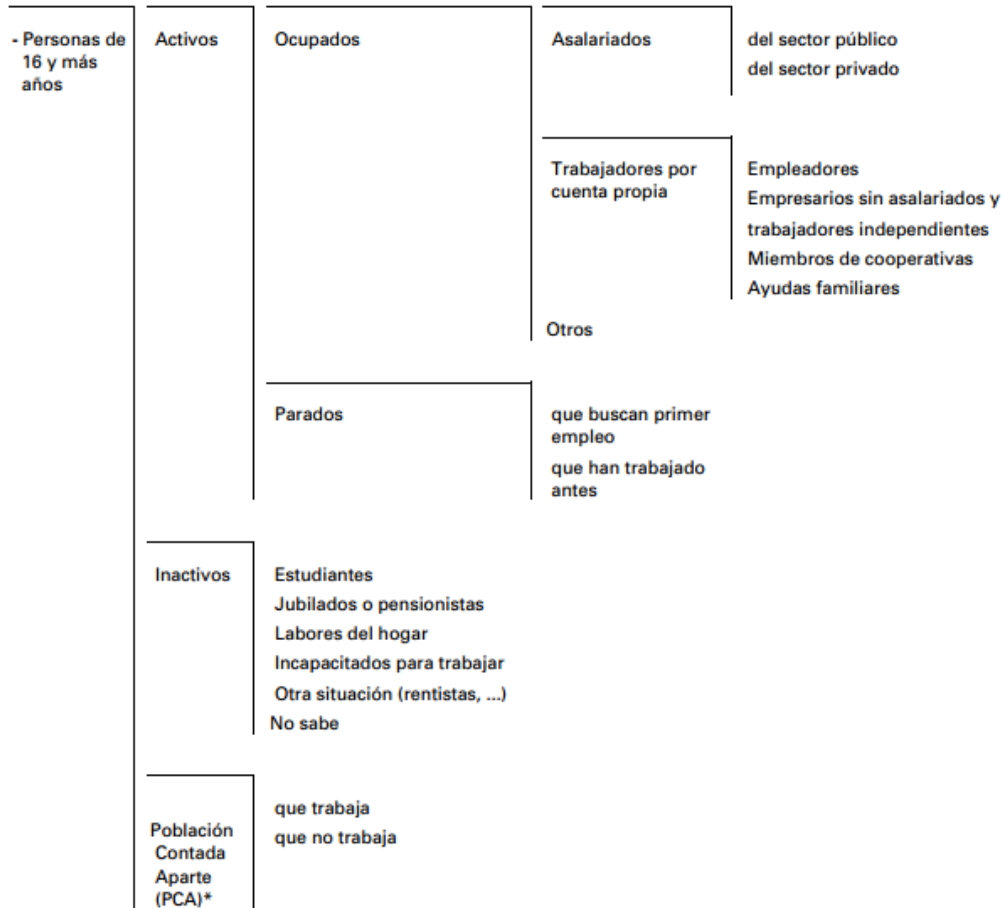
La población parada o desempleada son aquellas “personas de 16 o más años que durante la semana de referencia han estado sin trabajo, disponibles para trabajar y buscando activamente empleo”. “Quienes ya han encontrado un trabajo y están a la espera de incorporarse a él” también forman parte de esta clasificación de la población. (INE, *Nota de prensa*, 23 de Abril 2015)

- Población Inactiva

Por último, se considera población inactiva a aquella parte de la población con edad de trabajar que no queda incluida en ninguna de las categorías anteriores. (INE, *Nota de prensa*, 23 de Abril 2015)

Figura 3. Esquema de clasificación de la población

- Menores de 16 años



Fuente: INE

Para realizar el análisis de la actividad económica de la población, el Instituto Nacional de Estadística (INE), obtiene los datos necesarios a través de la realización de encuestas personales y telefónicas, lo que facilita la estimación de las siguientes tasas:

- Tasa de Actividad

La tasa de actividad es calculada como el cociente entre el número total de personas activas y la población de 16 y más años.

$$T_{(actividad)} = \text{Activos} / \text{Población de 16 o más años}$$

- **Tasa de Ocupación**

La tasa de ocupación se calcula como el cociente entre el número total de personas ocupadas y la población con edad de trabajar.

$$T_{(Ocupación)} = \text{Ocupados} / \text{Población de 16 o más años}$$

- **Tasa de Paro**

La tasa de paro se obtiene como el cociente entre el número total de personas demandantes de empleo y el número de personas económicamente activas.

$$T_{(Paro)} = \text{Demandantes de empleo} / \text{Personas económicamente activas}$$

2.2. Servicio Público de Empleo Estatal (SEPE)

Por otro lado, además del informe que nos ofrece el INE, tenemos acceso a otro gran indicador de la situación laboral en España, el Servicio Público de Empleo Estatal. Éste es un órgano autónomo que forma parte del Ministerio de Empleo y Seguridad Social desde donde se promueven, diseñan y desarrollan medidas para el empleo, como explica la propia organización y cuya ejecución es totalmente descentralizada y se ajusta de forma flexible a las distintas realidades territoriales.

Según el propio Servicio Público de Empleo Estatal expone, sus principales funciones se pueden resumir en lo siguiente:

- “Planificar e impulsar propuestas de políticas de empleo centradas en las necesidades de las personas y de las empresas.” (SEPE, 5 de Mayo 2015)
- “Gestionar las prestaciones por desempleo, haciendo efectivo el derecho de las personas desempleadas a protección.” (SEPE, 5 de Mayo 2015)
- “Realizar investigaciones, estudios y análisis, de ámbito estatal, sobre la situación del mercado de trabajo y las medidas para mejorarlo.” (SEPE, 5 de Mayo 2015)

El Servicio Público de Empleo Estatal, por lo tanto, nos ayuda a disponer de una información completa, ya que es obtenida con la colaboración de los diferentes

Servicios Públicos Autonómicos, y veraz referente a la situación del mercado de trabajo para facilitar la inserción laboral de la población, además de ayudar al gobierno vigente a tomar importantes decisiones sobre las políticas de empleo que se deben implantar y los métodos de protección que se deben poner en marcha para las personas desempleadas.

2.3. Diferencias entre ambos métodos

Sin duda, ambos métodos podemos decir que ofrecen datos totalmente válidos, pero siempre tenemos que tener en cuenta que no miden las mismas cosas ni utilizan los mismos métodos para el registro de dichos datos, luego, esto nos lleva a creer en la existencia de importantes diferencias entre los resultados que nos ofrece por un lado, el Instituto Nacional de Estadística y por otro, el Servicio Público de Empleo Estatal.

Para llevar a cabo la recogida de la información que se va a necesitar para realizar el estudio de la situación laboral, el INE, en la Encuesta de Población Activa utiliza la vía telefónica para recopilar todos los datos necesarios cada trimestre, puesto que este método le da la posibilidad de obtener un gran número de datos, una información muy completa, pero que no podemos dejar de considerar un muestreo. Sin embargo, el Servicio Público de Empleo Estatal cuenta con unas oficinas en las cuales llevan a cabo un registro mensual en el que las personas que se encuentran desempleadas pueden inscribirse aunque, al no ser un acto obligatorio, no todas las personas demandantes de empleo lo hacen, debido a lo que los expertos denominan “el efecto desánimo”, generándose de esta forma y en buena medida la diferencia de los resultados de ambos informes.

En cuanto a la fiabilidad que nos ofrecen estos dos métodos, los expertos dicen que para acercarnos a una estimación óptima del desempleo del país, la Encuesta de Población Activa es una fuente mucho más fiable que el paro registrado. Además, si prestamos atención a la evolución que han tenido ambos métodos en estos últimos años, en los cuales el aumento del desempleo ha sido importante, podemos ver que la EPA viene mostrando un número mayor de parados, lo que nos lleva a pensar que utiliza un método más eficiente a la hora de obtener la información frente al SEPE.

2.4. El mercado de trabajo en la actualidad

En España, durante las últimas décadas, el mercado de trabajo ha ido evolucionando al ritmo de la economía, es decir, cuando la economía afrontaba una etapa de expansión, el mercado de trabajo se veía afectado positivamente y en el caso contrario, cuando el país atravesaba una etapa recesiva, el mercado de trabajo también veía ralentizado su crecimiento. Pero quizás, lo más destacable de la evolución de nuestro mercado de trabajo puede ser el cambio estructural que éste ha experimentado en cuanto a la incorporación de la mujer al mundo laboral tras el incremento de su nivel de formación profesional.

El catastrófico porcentaje (54 por ciento) de desempleo juvenil de finales del pasado año, según la EPA, es otro de los aspectos en el que debemos centrar nuestra atención ya que, siendo ésta la generación mejor preparada académicamente de la historia, no tiene acceso a un trabajo digno y sólo puede acceder a la precariedad laboral.

Por último, observando el nivel de cualificación de los parados actualmente, podemos ver que el paro afecta de una forma más intensa y duradera a las personas con un grado de formación menor. Aunque hay que tener en cuenta que éste no es un factor determinante para el acceso a un empleo, pero una mayor cualificación sí que puede asegurar una mayor permanencia en el mismo.

3. SERIES TEMPORALES

3.1. *Concepto de serie temporal*

Una serie temporal puede definirse como una sucesión de observaciones cuantitativas a lo largo de un determinado periodo de tiempo, ordenadas de forma cronológica y pueden estar referidas a una determinada característica (serie univariante) o a varias características (serie multivariante). Para realizar el análisis de una serie temporal podemos prestar atención a cualquiera de las distintas ramas de la ciencia, puesto que en todas ellas existen datos que pueden ser analizados como series temporales. Como ejemplos de ello, podríamos observar desde el precio del alquiler de la vivienda o el índice de precio del petróleo hasta la tasa de mortalidad infantil por año o la temperatura media mensual.

Con el análisis de una serie temporal se pretende poder predecir una pauta de comportamiento para adelantarnos a su evolución con vistas al futuro, teniendo siempre en cuenta que se mantendrán las mismas condiciones aunque nos enfrentemos al transcurso del tiempo. Por lo tanto, este análisis se ocupará de aquellas series en las que conociendo los valores pasados, no podamos vaticinar con claridad y a simple vista un comportamiento futuro de la variable, lo que se conoce como serie temporal no determinista o aleatoria.

Evidentemente, la predicción que obtengamos sobre el valor futuro de nuestra variable no podemos considerarla totalmente exacta, pero sí que será muy próxima a la realidad, ya que se estudiará una regularidad en los datos a la hora de estimar el modelo de comportamiento de la variable.

3.2. *Componentes de una serie temporal*

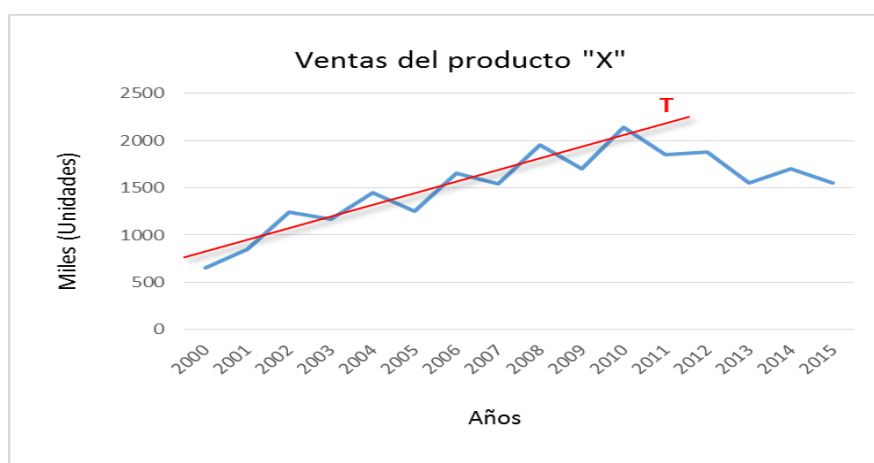
Desde el punto de vista clásico de análisis de series temporales, podemos decir que cualquier serie es la suma de varios componentes, entre ellos: la tendencia (T), variaciones estacionales (VE), variaciones cíclicas (V) y por último, variaciones accidentales (R). Este no es más que un procedimiento que nos hace el estudio mucho más fácil. No siempre que se analiza una serie de tiempo y se procede a su descomposición, existen cada uno de los componentes.

○ **Tendencia (T)**

Esta componente puede ser definida como el comportamiento que es adoptado por la serie a largo plazo, es decir, refleja su evolución. Hará falta que dispongamos de un número elevado de observaciones durante muchos años para que podamos comprobar si existe un patrón de comportamiento en las mismas, de manera que los datos a lo largo del tiempo crezcan, decrezcan, o simplemente exista cierta estabilidad.

Si observamos el ejemplo de la Figura 4, podemos observar que la tendencia de la serie a lo largo de todos los años, es prácticamente una línea recta con pendiente positiva, aunque el ritmo de crecimiento no solo se reduce al final del periodo, sino que a partir del año 2010 parece tener lugar un incipiente cambio de tendencia. De manera que tenemos una serie creciente para casi todo el periodo analizado.

Figura 4. Ejemplo Tendencia de una Serie Temporal



Fuente: Elaboración propia con datos inventados

○ **Variaciones estacionales (VE)**

Las variaciones estacionales en una serie temporal son las oscilaciones que se observan en periodos anuales o inferiores a un año y que se repiten de forma periódica a lo largo del tiempo. Pueden ser muchas las causas de estas variaciones, entre ellas: de tipo climatológico, factores culturales (navidad, vacaciones, etc.). Además, si los datos son recogidos de forma diaria también puede afectar en el comportamiento de algunas series.

Figura 5. Ejemplo Variaciones estacionales de una Serie Temporal



Fuente: Elaboración propia con datos inventados

En la Figura 5 podemos ver una pauta de estacionalidad claramente definida: las ventas de helado desciende notablemente en los meses de invierno y en los meses de verano estas llegan a niveles mucho más elevados. Por lo tanto, éste sería un claro ejemplo de serie de tiempo con componente estacional.

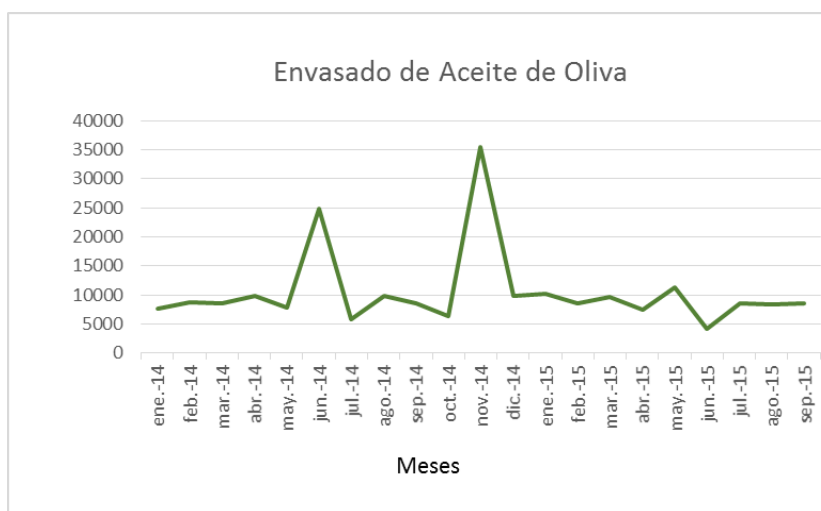
- **Variaciones cíclicas (C)**

Esta es una componente de la serie que muestra las oscilaciones que experimentan las observaciones y que se repiten a lo largo de distintos periodos o ciclos. La principal causa para que esto ocurra es la alternancia de etapas expansivas y recesivas en la actividad económica. Para que la componente cíclica pueda ser reconocida en una serie, el número de observaciones debe aumentar a la vez que aumenta el número de años que consideramos un ciclo completo.

- **Variaciones accidentales (R)**

Las variaciones accidentales o residuales recogen un comportamiento errático de la variable que se produce como resultado de fenómenos que no se pueden controlar. Estos fenómenos pueden ser distinta naturaleza, por ejemplo pueden afectar a una serie de tiempo las convocatorias de huelgas, inundaciones, un pedido excepcional que se produce en una empresa, etc.

Figura 6. Ejemplo Variaciones accidentales de una Serie Temporal



Fuente: Elaboración propia con datos inventados

En la gráfica que muestra la Figura 6 se observa la existencia de dos importantes oscilaciones en la trayectoria normal de producto envasado cuya causa podríamos atribuirle a un pedido extraordinario tanto en el mes de junio de 2014 como en el mes de octubre del mismo año.

La forma en la que estos cuatro componentes interactúan da como resultado a la propia serie temporal. Existen diversas formas de interacción entre ellas, pero generalmente se ha optado por las dos siguientes: el esquema aditivo y el esquema multiplicativo. No obstante, a veces se procede a utilizar una combinación de ambos.

Esquema Aditivo: $y_t = T_t + VE_t + C_t + R_t$

Esquema Multiplicativo: $y_t = (T_t)(VE_t)(C_t)(R_t)$

Como podemos ver en la ecuación resultante del esquema aditivo, éste interpreta que las componentes de la serie son totalmente independientes entre sí, mientras que el esquema multiplicativo, las relaciona de forma que la serie que se obtiene es el resultado de la interrelación de todos y cada uno de sus componentes.

Elegir uno de estos métodos de interacción es solo el principio de un largo camino para llegar a obtener un modelo que explique de forma apropiada la evolución de la serie

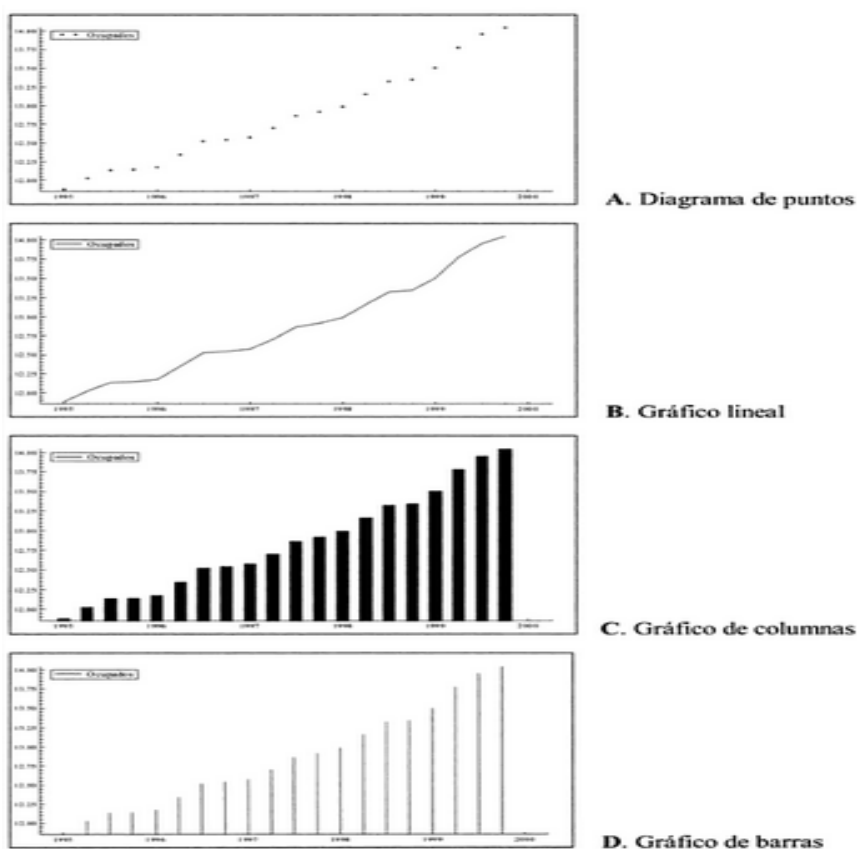
estudiada. Aunque para obtener dicho modelo, antes tenemos que identificar la tendencia determinista que posee la serie, a través de la cual conoceremos la existencia de las reglas matemáticas que nos permitirán obtener unos resultados futuros precisos.

El único componente que no va a mostrarnos una trayectoria regular de que podamos utilizar para la formulación del modelo es el componente accidental o residual, debido a su naturaleza errática. Sólo podremos identificarlo una vez que hayamos eliminado los demás componentes de la serie.

3.3. Tipos de metodologías de análisis

Lo primero que tenemos que hacer para llevar a cabo el análisis de una serie temporal, una vez finalizada la recogida de los datos sujetos a estudio, es la representación gráfica de dicha serie para visualizar de una forma clara su evolución temporal.

Figura 7. Representaciones gráficas de series temporales



Fuente: “Análisis de series temporales económicas II”

Hernández Alonso, José (2007)

Una vez hecha la representación gráfica, disponemos de tres métodos de análisis para realizar un diagnóstico sobre la existencia de regularidades en el comportamiento de la serie, lo que nos llevaría a continuar con nuestro análisis, o por el contrario, si no se encontrara una pauta de actuación, pasaríamos a la interrupción del mismo.

El profesor de la Universidad Complutense de Madrid, José Hernández Alonso en su libro “Análisis de series temporales económicas II”, nos habla de los tres tipos o procedimientos metodológicos de los que disponemos para el análisis de series temporales: análisis clásico, análisis causal y por último, el conocido como el análisis moderno de series temporales.

1. Método clásico:

Bajo esta metodología, se pretende descomponer una serie temporal en sus componentes, anteriormente desarrolladas: tendencia, variaciones cíclicas, variaciones estacionales y variaciones accidentales y así poder extrapolarlos hacia el futuro, para finalmente prever la evolución de la variable.

2. Método causal:

Con este método, explicaremos la evolución futura de la variable por medio de una ecuación con la que podemos relacionarla con otras variables que influyen directamente sobre ella. Por lo tanto, conociendo los valores que tendrán las variables causa, también conoceremos los valores futuros de variable que estudiamos.

3. Análisis moderno de series temporales:

Modelo de análisis que fue diseñado por Box-Jenkins (1970) y que intenta expresar la evolución de la variable sujeta a estudio a partir de sus propios valores pasados. En definitiva, si conseguimos conocer la pauta de comportamiento que ha venido siguiendo la variable que tratamos de estudiar, podemos predecir el comportamiento futuro de la misma.

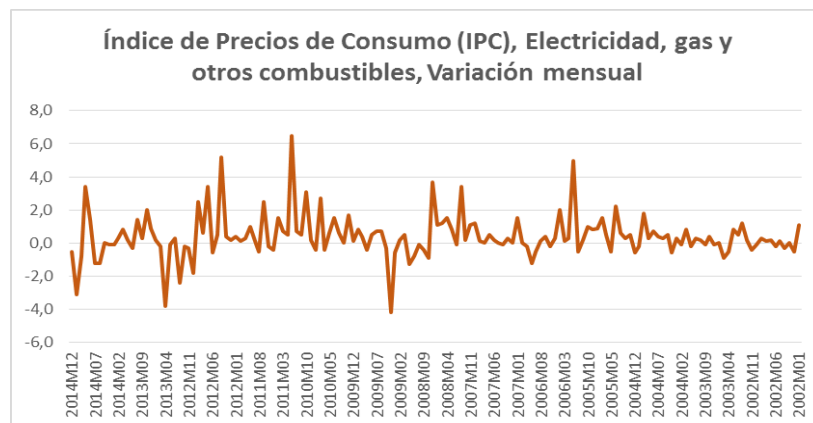
3.4. Clasificación descriptiva de series temporales

Para realizar la clasificación descriptiva de una serie de tiempo tenemos dos opciones en las que incluirla: series de tiempo estacionarias o series de tiempo no estacionarias.

○ Series estacionarias

Se considera estacionaria una serie temporal cuando existe un equilibrio constante tanto en la media como en la varianza de los datos durante el transcurso del tiempo, es decir, no muestra tendencia en su trayectoria y además adopta una apariencia plana a lo largo de la misma.

Figura 8. Serie temporal estacionaria



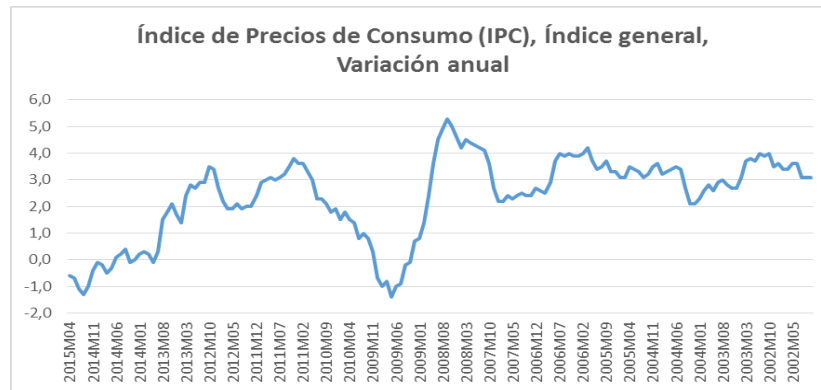
Fuente: Elaboración propia con datos del INE

Observación: La serie que nos muestra el movimiento del Índice de Precio de Consumo referido a la electricidad, gas y otros combustibles de los último años, refleja un comportamiento irregular pero con un cierto patrón de regularidad. Las observaciones presentan un equilibrio alrededor de la media con un valor constante.

○ Series no estacionarias:

Son aquellas series en las que los valores de la media y/o varianza no son constantes a largo plazo. Gráficamente en este tipo de series se aprecia de forma clara una tendencia creciente o decreciente a lo largo del tiempo, por lo tanto, es imposible que exista oscilación alguna sobre un determinado valor estable.

Figura 9. Serie temporal no estacionaria



Fuente: Elaboración propia con datos del INE

Observación: Esta serie representa los movimientos del Índice de Precios de Consumo general que refleja un comportamiento totalmente irregular a lo largo del tiempo. Podemos ver como la tendencia de la serie es ligeramente creciente a largo plazo, por lo tanto, el valor de la media queda lejos de ser constante.

3.5. Metodología ARIMA. El enfoque Box-Jenkins

En este apartado, profundizaremos sobre la metodología ARIMA que fue desarrollada en la década de los 70 por G.E.P. Box, profesor de Estadística de la Universidad de Wisconsin y por G.M. Jenkins, profesor de Ingeniería de Sistemas de la Universidad de Lancaster, consiguiendo la herramienta de pronóstico más utilizada para el análisis de series de tiempo desde entonces.

El enfoque Box-Jenkins para los Modelos Autorregresivos Integrados de Medias Móviles cuyas siglas en inglés dan nombre a la famosa metodología ARIMA, distingue tres fases importantes: identificar, estimar y diagnosticar modelos de series temporales para los cuales es imprescindible que las características que se observan en la serie perduren en el tiempo. Con la metodología ARIMA, son los propios datos pasados de la variable que estamos estudiando los que nos van a llevar a una predicción futura del comportamiento de la misma.

Definición y conceptos básicos de los modelos ARIMA

- Procesos estocásticos

Aunque la predicción del comportamiento futuro que finalmente se obtiene con el análisis de series temporales no puede ser considerada totalmente exacta, tampoco podemos creer que es un comportamiento totalmente aleatorio, es decir, deben existir regularidades en el comportamiento de la serie de las que podamos extraer una estructura probabilística. De este modo, podemos definir un proceso estocástico, como una serie consecutiva de variables aleatorias que tienen asociada una distribución de probabilidad.

- El ruido blanco

Podemos definir el ruido blanco o también denominado proceso puramente aleatorio como una serie consecutiva en el tiempo de variables aleatorias cuya media es cero, su varianza es constante y covarianzas nulas. Puesto que su media y su varianza son constantes a lo largo del tiempo, podemos decir que estas son series de tiempo estacionarias.

La metodología Box-Jenkins nos lleva a la reflexión de que a través de dos modelos básicos: los Modelos Autorregresivos (AR) y los Modelos de Medias Móviles (MA), podemos finalmente obtener el modelo que perseguimos.

- Los Modelos Autorregresivos (AR)

Podemos definir un Modelo Autorregresivo como aquel modelo en el que el valor de la variable en un momento t determinado es igual a su valor en ciertos momentos anteriores $t-1, t-2, \dots, t-p$ más una alteración aleatoria (ruido blanco). Seguidamente de la abreviatura de los modelos autorregresivos se debe señalar el orden que sigue el modelo: AR(1), AR(2),... etc. Este nos muestra las N observaciones de retraso que intervienen en la serie.

Un Modelo Autorregresivo de orden p o AR(p) puede ser expresado de la siguiente forma:

$$Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \dots + \emptyset_p Y_{t-p} + \varepsilon_t$$

Así en un modelo AR(1), “el valor de pronóstico de Y en el periodo t es simplemente alguna proporción ϕ de su valor en el periodo t-1 más una perturbación aleatoria en el tiempo t.” *Gujarati y Porter (2010, Pág. 775)*

- Los Modelos de Medias Móviles (MA)

Denominamos Modelo de Media Móvil a aquel modelo que es capaz de dar un valor a la variable que estamos estudiando a través de un término independiente y una combinación lineal de alteraciones aleatorias. La expresión genérica de dicho modelo sería la siguiente:

$$Y_t = \mu + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q}$$

“donde μ es una constante y ε , al igual que antes, es el término de error estocástico de ruido blanco. Aquí, Y en el periodo t es igual a una constante más un promedio móvil de los términos de error presente y pasado. Así, en este caso decimos que Y sigue un proceso de promedios móviles de orden q”. *Gujarati y Porter (2010, Pág. 776)*

- Los Modelos Autorregresivos de Medias Móviles (ARMA)

Un Modelo Autorregresivo de Medias Móviles se obtiene a través de una conjunción de los dos modelos anteriormente expuestos, AR(p) y MA(q). Por lo tanto, tras esta combinación, las siglas de este modelo quedarían como ARMA (p, q), donde p sería el orden de la parte del modelo autorregresivo y q sería el orden de la parte del modelo de medias móviles.

En general, la expresión de un Modelo Autorregresivo de Medias Móviles “ARMA (p, q)” quedaría de la siguiente forma:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} + \mu$$

- Los Modelos Autorregresivos Integrados de Media Móvil (ARIMA)

Todos los modelos de series temporales que venimos explicando tienen algo en común, y es que suponen que dichas series son estacionarias, es decir, su media y varianza son constantes y su covarianza es invariante a lo largo del transcurso del tiempo. Pero podemos encontrarnos con series no estacionarias, también llamadas integradas, que tendremos que transformar en estacionarias para poder llevar a cabo el análisis.

“Si una serie de tiempo es integrada de orden 1 [es decir, si es $I(1)$], sus primeras diferencias son $I(0)$, es decir, estacionarias. En forma similar, si una serie de tiempo es $I(2)$, sus segundas diferencias son $I(0)$. En general, si una serie de tiempo es $I(d)$, después de diferenciarla d veces se obtiene una serie $I(0)$.” *Gujarati y Porter (2010, Pág. 776)*

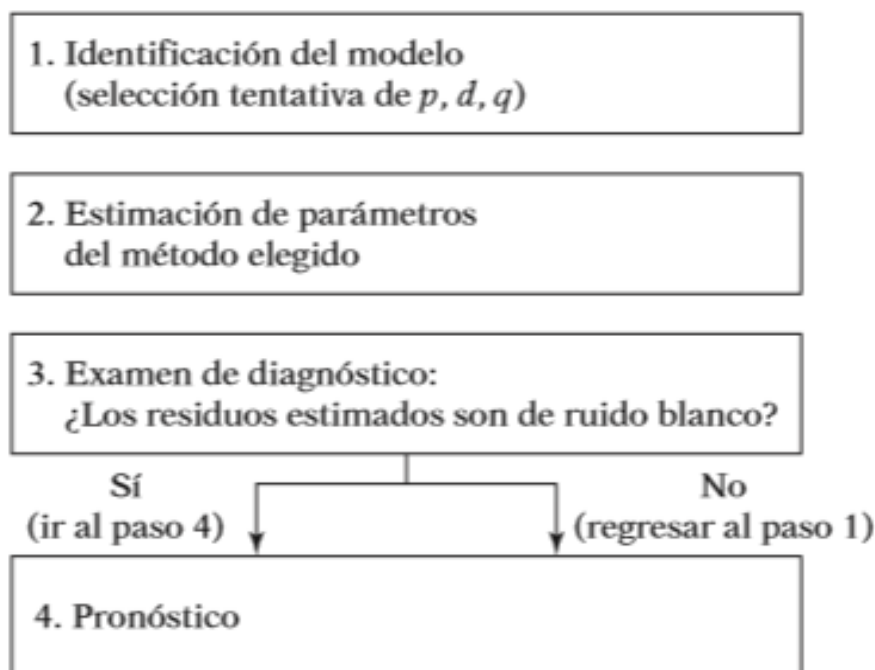
Por lo tanto, si tenemos que transformar una serie no estacionaria, a través de diferenciarla d veces, en estacionaria para después poder aplicar el modelo ARMA (p , q), podemos decir que la serie primaria de la que disponíamos es ARIMA (p , d , q), es decir, es una serie temporal autorregresiva integrada de promedios móviles.

A partir de ahora, el problema se encuentra en identificar qué tipo de modelo sigue la serie que estamos estudiando y en cuyo caso, conocer los valores de p , d y q correspondientes. La metodología Box-Jenkins da solución a este problema a través de un proceso que consta de cuatro fases: identificación, estimación, examen de diagnóstico y por último el pronóstico.

I. Identificación. En esta primera fase, lo que se persigue es determinar qué modelo o modelos de series temporales son más adecuados para representar el comportamiento de la serie que estamos estudiando. Para seguir adelante con este proceso, necesitamos que la serie temporal utilizada sea estacionaria, por lo tanto, si no es así, habría que diferenciarla tantas veces como fuese necesario hasta conseguirlo. Y a partir de ahí, procedemos a observar la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP) muestrales, cuyo gráfico cambia en función no sólo de la clase de modelo que estemos estudiando, sino también de su grado de parametrización (*ANEXO*).

- II. Estimación.** Una vez hallados los valores de p , d y q en la fase anterior, ahora en esta fase tenemos que determinar los parámetros de los términos autorregresivos y de promedios móviles que hemos incorporado en el modelo.
- III. Examen de diagnóstico.** Una vez que hemos seleccionado un modelo ARIMA y estimado sus parámetros, tenemos que asegurarnos de que el modelo seleccionado se ajusta a los datos de una manera próxima a la realidad, ya que puede existir algún otro modelo ARIMA que se ajuste también. Para saber si el modelo que hemos seleccionado es el correcto, tenemos que hacer una serie de comprobaciones, y entre ellas, tenemos que ver si los residuos estimados con este modelo son de ruido blanco, si es así, tenemos el modelo correcto, si no es así, tenemos que volver a comenzar el proceso.
- IV. Pronóstico.** Una vez que tenemos la certeza de que hemos seleccionado el modelo adecuado, empezamos a realizar un pronóstico, es decir, predecimos el comportamiento futuro de la variable estudiada.

Figura 10. Metodología Box-Jenkins



Fuente: Damodar N. Gujarati y Dawn C. Porter
(Econometría 5ª Edición)

4. APLICACIÓN DE LA METODOLOGÍA BOX-JENKINS

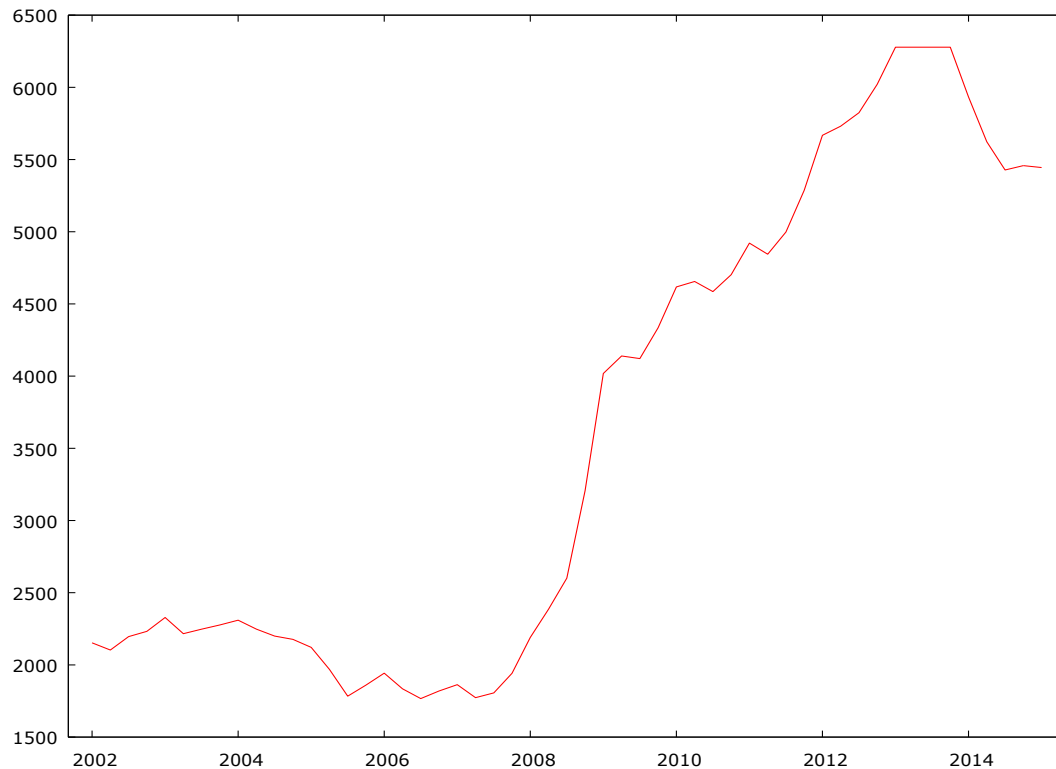
Una vez explicada la teoría sobre la metodología Box-Jenkins y los modelos ARIMA, pasamos a desarrollar la parte práctica, en la cual, se va a realizar un análisis econométrico sobre la evolución de los valores que el desempleo ha experimentado en España desde el año 2002 hasta 2015 y además se elaborará una predicción futura del comportamiento de dicha variable. Los datos que se muestran en la Tabla 1, se encuentran en valor absoluto y han sido recogidos de la página web oficial del Instituto Nacional de Estadística (INE).

Tabla 1. Datos empleados en el análisis práctico

Total Datos Desempleo en España 2002-2015 (Miles de personas)				
	Trimestre I	Trimestre II	Trimestre III	Trimestre IV
2002	2.152,8	2.103,3	2.196,0	2.232,4
2003	2.328,5	2.216,0	2.247,5	2.276,7
2004	2.309,8	2.247,6	2.199,8	2.176,9
2005	2.121,3	1.969,1	1.783,5	1.860,3
2006	1.942,8	1.834,4	1.766,9	1.819,4
2007	1.863,2	1.773,2	1.806,2	1.942,0
2008	2.190,5	2.385,7	2.600,7	3.206,8
2009	4.018,2	4.139,6	4.121,4	4.335,0
2010	4.617,7	4.655,3	4.585,4	4.702,2
2011	4.921,2	4.844,2	4.998,0	5.287,3
2012	5.667,9	5.731,0	5.824,2	6.021,0
2013	6.278,2	6.278,2	6.278,2	6.278,2
2014	5.933,3	5.622,9	5.427,7	5.457,7
2015	5.444,6			

Fuente: Elaboración propia con datos del INE

Figura 11. Representación gráfica del desempleo en España durante el periodo 2002-2015



Fuente: Elaboración Propia con Gretl

Observando la Figura 11, podemos detectar de forma clara el cambio estructural de tendencia que experimenta el desempleo en España, a finales del año 2007. Se puede ver como el desempleo, al principio de la última década, obtenía valores muy bajos con cierto equilibrio y como a partir del comienzo de la crisis económica que está atravesando nuestro país, el desempleo empieza a tomar valores cada vez más altos. Por lo tanto, el análisis que se va a desarrollar, sólo se va a centrar en el periodo que dista entre el tercer trimestre de 2007 hasta el primero de 2015, puesto que son los años que suponen una mayor problemática.

Para llevar a cabo nuestro análisis de modelos ARIMA, vamos a utilizar “GRETl” que es un software totalmente gratuito que podemos encontrar en Internet (en la dirección web: <http://gretl.sourceforge.net>) y del cual nos vamos a servir para realizar todos los cálculos necesarios.

A modo de resumen, trataremos de analizar la trayectoria del desempleo en España a lo largo del periodo que dista entre el tercer trimestre de 2007 y el primer trimestre de 2015, y por último, haremos un pronóstico de la evolución que sufrirá el desempleo hasta finales de 2016.

Comenzamos con la primera fase de la metodología Box-Jenkins:

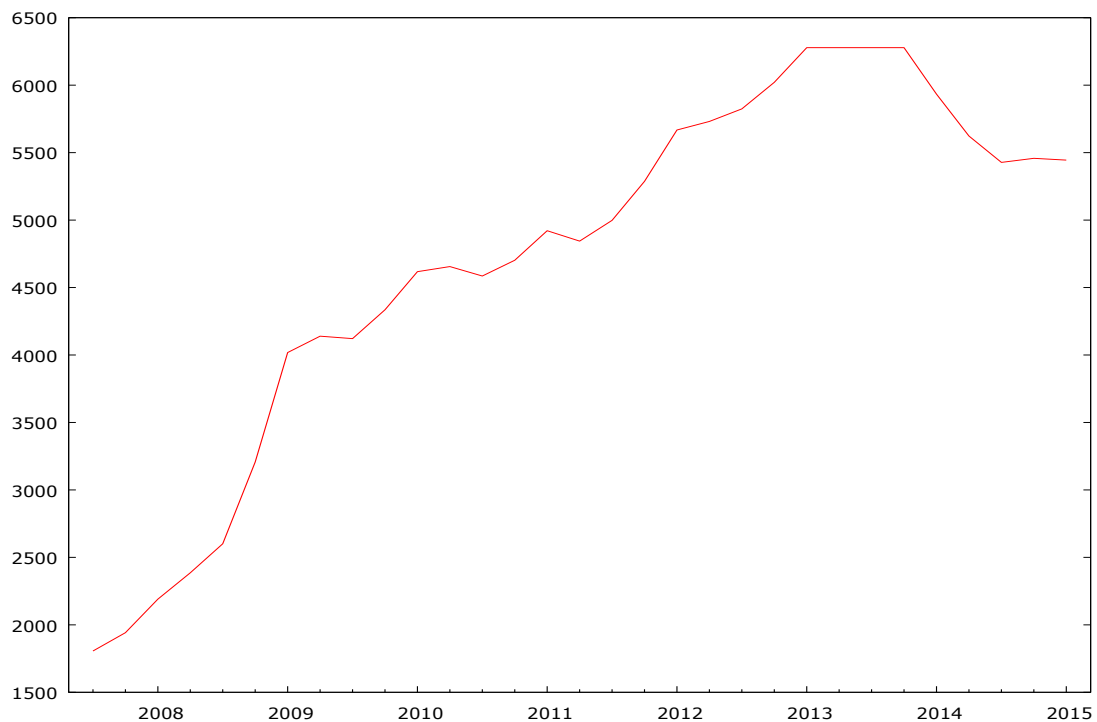
4.1. Identificación

El objetivo de esta primera fase, como ya hemos explicado previamente, es identificar los valores de p , d y q , es decir, vamos a intentar predecir el modelo que mejor se va a ajustar a nuestra serie de tiempo. Para ello, vamos a utilizar la función de autocorrelación y la función de autocorrelación parcial muestrales que nos ayudarán a conocer qué modelo o modelos se adaptan mejor a las características de la serie.

Nuestra serie temporal consta de 31 observaciones que van a quedar reflejadas en la siguiente representación gráfica (Figura 12):

Figura 12. Representación gráfica del desempleo en España (2007-2015)

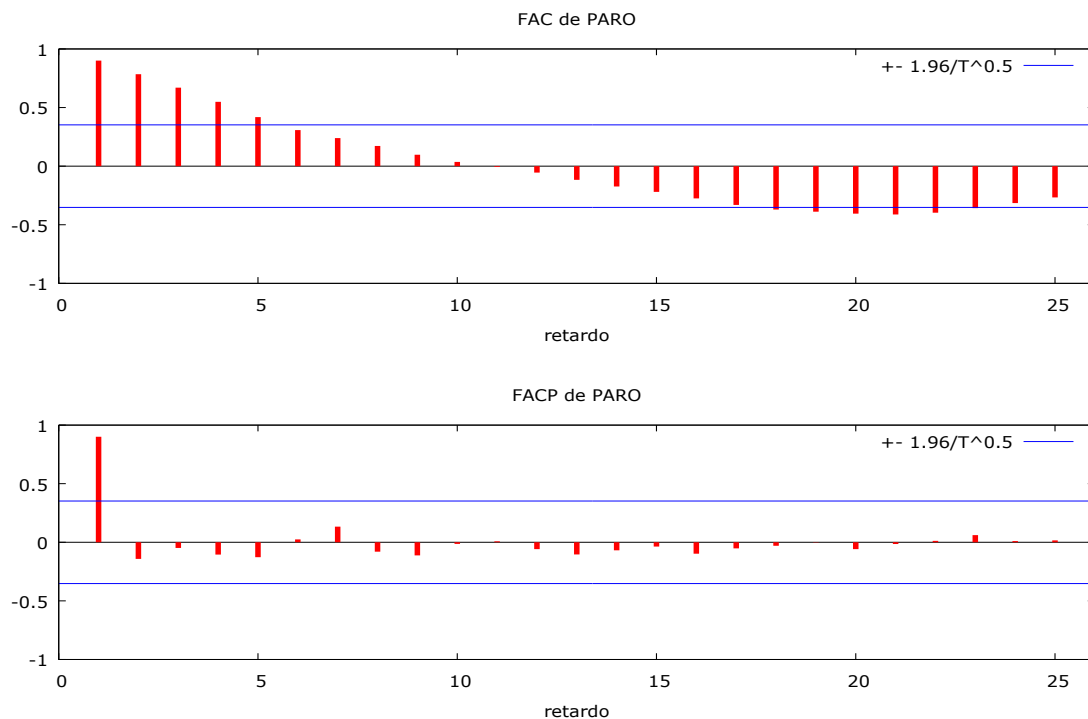
Serie original



Fuente: Elaboración Propia con Gretl

En el gráfico que presenta la Figura 12, podemos ver claramente que existe una tendencia creciente en la serie, por lo tanto, tenemos evidencia para decir que la media del proceso no es constante (serie no estacionaria). También podemos demostrarlo observando la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP) muestrales (Figura 13), puesto que se ve como la FAC decrece muy lentamente y de forma lineal y además, la FACP nos muestra un pico muy significativo.

Figura 13. Funciones de autocorrelación (FAC) y autocorrelación parcial (FACP)
Serie Original



Fuente: Elaboración Propia con Gretl

Para demostrar que la serie que estamos analizando no es estacionaria, vamos a realizar los contrastes de hipótesis que pueden ayudarnos a clarificarlo.

- Para comprobar la existencia de estacionalidad en media, aplicamos el contraste aumentado de Dickey-Fuller:

$$\left\{ \begin{array}{l} H_0: \text{Existe una raíz unitaria (la serie no es estacionaria en media)} \\ H_1: \text{No existe raíz unitaria (la serie es estacionaria en media)} \end{array} \right.$$

El resultado del contraste es el siguiente:

Contraste aumentado de Dickey-Fuller para PARO

Incluyendo un retardo de $(1-L)$ PARO

Tamaño muestral 31

Hipótesis nula de raíz unitaria: $a = 1$

Contraste con constante

Modelo: $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$

Coef. de autocorrelación de primer orden de e: 0.084

Valor estimado de $(a - 1)$: -0.0560235

Estadístico de contraste: $\tau_c(1) = -2.37663$

Valor p asintótico 0.1484

Con constante y tendencia

Modelo: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$

Coef. de autocorrelación de primer orden de e: 0.078

Valor estimado de $(a - 1)$: -0.0727037

Estadístico de contraste: $\tau_{ct}(1) = -1.11169$

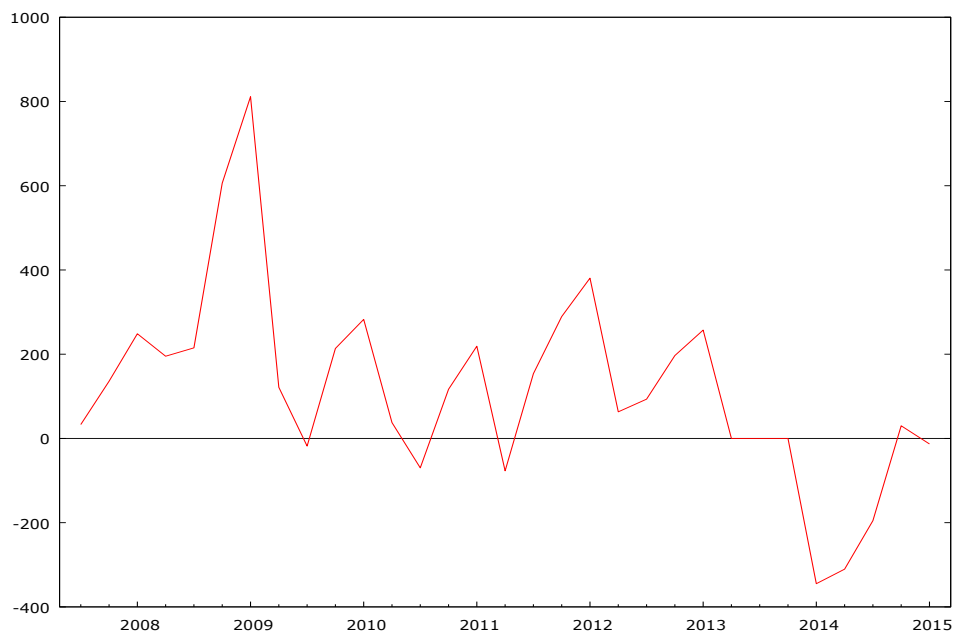
Valor p asintótico 0.9258

Observando el resultado obtenido, vemos que el p-valor es mayor que 0.05, por lo tanto no se rechaza H_0 al 5% de significación, luego la media no es constante, tal y como habíamos podido deducir observando el gráfico.

Una vez visto que la serie no es estacionaria en media, no hace falta que comprobemos la existencia de estacionalidad en varianza, puesto que de igual forma tendremos que aplicar las primeras diferencias regulares de la serie para seguir adelante con el análisis, es decir, generaremos una serie nueva que tendrá la siguiente forma:

$$W_t = Y_t - Y_{t-1}$$

Figura 14. Representación gráfica del desempleo en primeras diferencias regulares



Fuente: Elaboración Propia con Gretl

Una vez más, tenemos que proceder a realizar la comprobación por medio de los contrastes de los que disponemos (Dickey-Fuller y Gráfico rango-media) para saber a ciencia cierta si esta nueva serie puede considerarse estacionaria, tanto en media como en varianza. Sólo observando la representación gráfica (Figura 14), y de forma totalmente subjetiva, podemos ver que la serie no tiene una tendencia creciente ni decreciente, luego podemos decir que la serie a simple vista parece estacionaria en media. Además, también podemos observar que si dividimos la serie en periodos de tiempo iguales, observamos que las oscilaciones que experimenta tienen una amplitud similar, por lo tanto, también podemos considerarla estacionaria en varianza.

Procedemos a la realización del contraste Dickey-Fuller:

$$\begin{cases} H_0: \text{Existe una raíz unitaria (la serie no es estacionaria en media)} \\ H_1: \text{No existe raíz unitaria (la serie es estacionaria en media)} \end{cases}$$

El resultado obtenido mediante Gretl, es el siguiente:

Contraste aumentado de Dickey-Fuller para d_PARO
Incluyendo un retardo de (1-L)d_PARO

Análisis de la evolución temporal del desempleo en España

Tamaño muestral 31

Hipótesis nula de raíz unitaria: $a = 1$

Contraste con constante

Modelo: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$

Coef. de autocorrelación de primer orden de e : 0.099

Valor estimado de $(a - 1)$: -0.563065

Estadístico de contraste: $\tau_c(1) = -3.26797$

Valor p asintótico 0.01641

Con constante y tendencia

Modelo: $(1-L)y = b_0 + b_1*t + (a-1)y(-1) + \dots + e$

Coef. de autocorrelación de primer orden de e : 0.057

Valor estimado de $(a - 1)$: -0.766006

Estadístico de contraste: $\tau_{ct}(1) = -4.31305$

Valor p asintótico 0.002941

En esta ocasión, podemos observar que el p-valor es inferior a 0,05, por lo tanto podemos rechazar H_0 con un 95% de confianza, así que podemos concluir que la media es constante, tal y como habíamos supuesto sólo observando el gráfico.

Realizamos el Gráfico rango-media (Figura 15) para comprobar la estacionalidad en varianza, cuyo resultado es el siguiente:

$$\left\{ \begin{array}{l} H_0: \text{El rango no varía en función de la media (la varianza es constante)} \\ H_1: \text{El rango varía en función de la media (la varianza no es constante)} \end{array} \right.$$

Estadísticos de rango-media para d_PARO

Utilizando 5 submuestras de tamaño 6

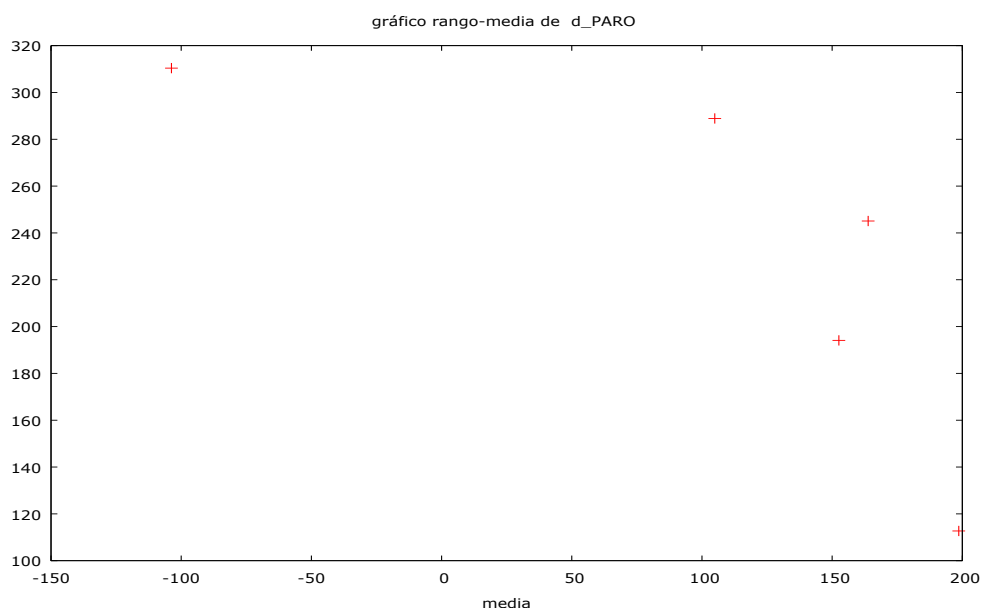
	Rango	Media
2007:3 - 2008:4	112.700	198.625
2009:1 - 2010:2	245.100	163.825
2010:3 - 2011:4	288.900	104.925
2012:1 - 2013:2	194.100	152.575
2013:3 - 2015:1	310.400	-103.740

Pendiente de 'rango' con respecto a 'media' = -0.492484

El valor p para H_0 : pendiente = 0 es 0.146565

Con este resultado, podemos ver que el p-valor es inferior a 0.05, lo que nos lleva a pensar que no podemos rechazar H_0 , luego la varianza también puede considerarse constante.

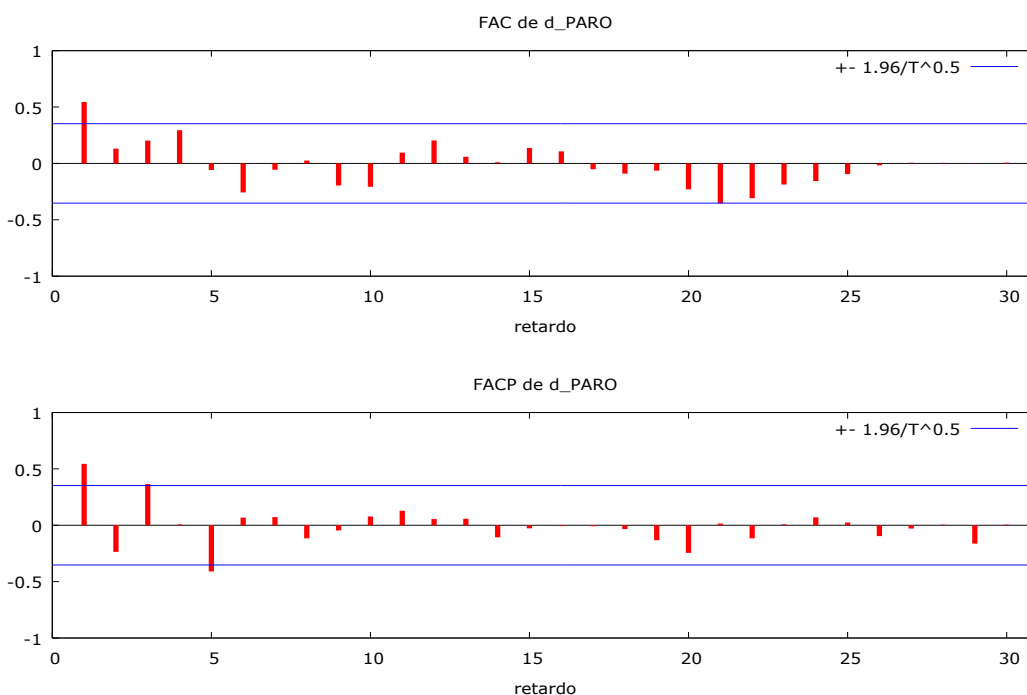
Figura 15. Gráfico rango-media de la serie en primeras diferencias



Fuente: Elaboración Propia con Gretl

Ya que hemos comprobado que la nueva serie es estacionaria tanto en media como en varianza, podemos realizar su correlograma y proceder a la identificación del modelo.

Figura 16. Correlograma serie en primeras diferencias regulares



Fuente: Elaboración Propia con Gretl

Observando el gráfico de la función de autocorrelación y autocorrelación parcial que aparece en la Figura 16, podemos ver que el primer retardo de ambas funciones es significativo, es decir, los picos que corresponden a dichos retardos sobrepasan las bandas de confianza. Comparando esta información con los correlogramas que representan a los diferentes modelos que existen, podemos llegar a pensar que estaríamos ante un MA (1), o quizás ante un ARMA (1,1).

1. Modelo Integrado de Medias Móviles, $ARIMA(0,1,1)$

2. Modelo Autorregresivo Integrado de Medias Móviles, $ARIMA(1,1,1)$

Pasaremos a realizar la estimación de ambos modelos, los cuales serán comparados para definitivamente quedarnos con el que nos proporcione una información más próxima a la realidad acerca del comportamiento de nuestra serie temporal sometida al análisis.

4.2. Estimación del Modelo

Una vez que se han identificado los posibles modelos que pueden ser útiles para ilustrarnos el comportamiento de la serie estudiada, se procede a estimar los parámetros que los componen.

- Modelo 1

Modelo 1: ARIMA, usando las observaciones 2007:3-2015:1 (T = 31)

Estimado usando el filtro de Kalman (MV exacta)

Variable dependiente: (1-L) PARO

Desviaciones típicas basadas en la matriz de productos externos

	Coefficiente	Desv. Típica	z	Valor p	
const	114.358	51.5329	2.219	0.0265	**
theta_1	0.613598	0.149462	4.105	4.04e-05	***
Media de la vble. dep.	118.4323	D.T. de la vble. dep.	230.5938		
Media innovaciones	0.552928	D.T. innovaciones	182.3260		
Log-verosimilitud	-205.6030	Criterio de Akaike	417.2060		
Criterio de Schwarz	421.5079	Crit. de Hannan-Quinn	418.6083		

	Real	Imaginaria	Módulo	Frecuencia
MA				
Raíz 1	-1.6297	0.0000	1.6297	0.5000

La expresión del modelo estimado quedaría de la siguiente manera:

$$W_t = 114.358 + 0.613598\varepsilon_{t-1}$$

La ecuación del Modelo 1 que hemos obtenido, representa que el comportamiento futuro del desempleo en España depende de irregularidades y fluctuaciones aleatorias correspondientes al trimestre inmediatamente anterior.

- Modelo 2

Modelo 2: ARIMA, usando las observaciones 2007:3-2015:1 (T = 31)

Estimado usando el filtro de Kalman (MV exacta)

Variable dependiente: (1-L) PARO

Desviaciones típicas basadas en la matriz de productos externos

	Coeficiente	Desv. Típica	z	Valor p	
const	112.059	60.9067	1.840	0.0658	*
phi_1	0.244436	0.266638	0.9167	0.3593	
theta_1	0.494016	0.247253	1.998	0.0457	**
Media de la vble. dep.	118.4323	D.T. de la vble. dep.	230.5938		
Media innovaciones	0.876024	D.T. innovaciones	179.0232		
Log-verosimilitud	-205.0848	Criterio de Akaike	418.1695		
Criterio de Schwarz	423.9055	Crit. de Hannan-Quinn	420.0393		
	Real Imaginaria		Módulo	Frecuencia	
AR					
	Raíz	1	4.0911	0.0000	4.0911 0.0000
MA					
	Raíz	1	-2.0242	0.0000	2.0242 0.5000

La expresión del modelo estimado quedaría de la siguiente manera:

$$W_t = 112.059 + 0.244436 W_{t-1} + 0.494016\varepsilon_{t-1}$$

La ecuación del Modelo 2 que hemos obtenido, representa que el comportamiento futuro del desempleo en España depende de irregularidades y fluctuaciones aleatorias correspondientes al trimestre inmediatamente anterior, pero además también depende del comportamiento que haya tenido el desempleo en el trimestre anterior.

4.3. Examen de diagnóstico

Para llevar a cabo el examen de diagnóstico, tenemos que realizar el análisis de los coeficientes del modelo y en el análisis de los residuos, que serán los que nos den la información necesaria para saber si el modelo quedará validado o si por el contrario tendremos que volver a comenzar el análisis.

- Modelo 1

Modelo 1: ARIMA, usando las observaciones 2007:3-2015:1 (T = 31)

Estimado usando el filtro de Kalman (MV exacta)

Variable dependiente: (1-L) PARO

Desviaciones típicas basadas en la matriz de productos externos

	Coeficiente	Desv. Típica	z	Valor p	
Const	114.358	51.5329	2.219	0.0265	**
theta_1	0.613598	0.149462	4.105	4.04e-05	***
Media de la vble. dep.	118.4323	D.T. de la vble. dep.	230.5938		
Media innovaciones	0.552928	D.T. innovaciones	182.3260		
Log-verosimilitud	-205.6030	Criterio de Akaike	417.2060		
Criterio de Schwarz	421.5079	Crit. de Hannan-Quinn	418.6083		
	Real Imaginaria		Módulo	Frecuencia	
MA					
	Raíz	1	-1.6297	0.0000	1.6297 0.5000

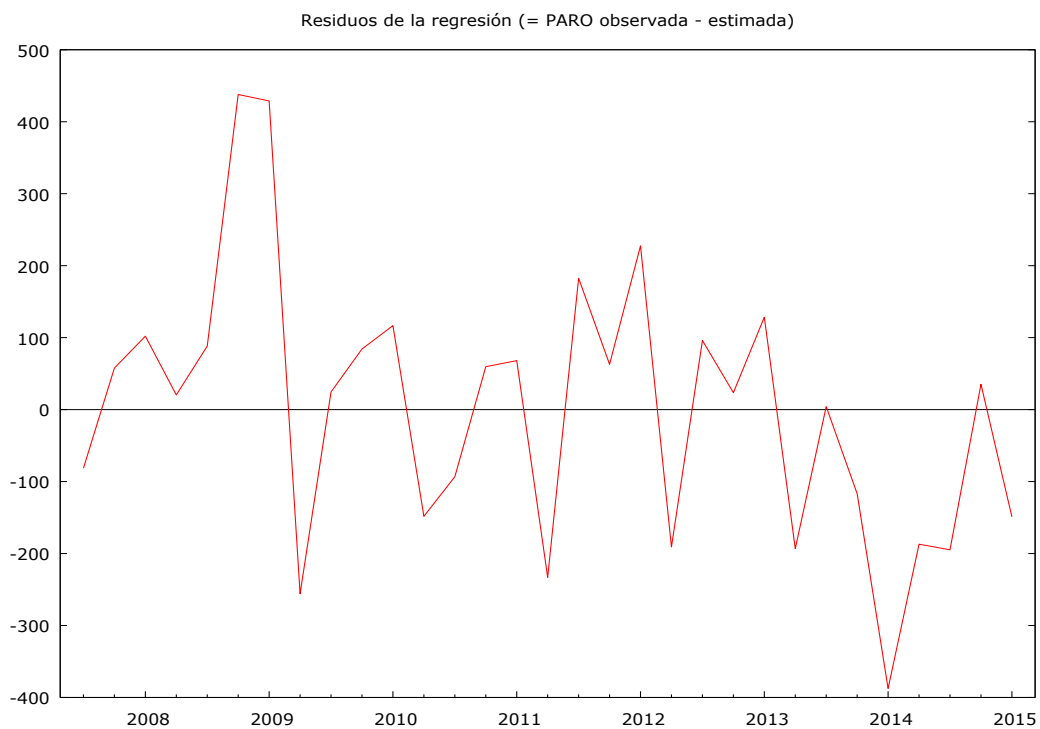
Respecto al análisis de coeficientes, si observamos el modelo estimado en la fase de identificación, podemos apreciar que todos los parámetros del modelo son significativos, por lo tanto, no tenemos que modificar ningún aspecto del modelo que hemos considerado.

Además, si analizamos las propiedades de estacionariedad e invertibilidad del *Modelo 1* podemos comprobar que éstas se cumplen, puesto que la raíz del correspondiente polinomio en retardos está por encima de 1 en valor absoluto, o lo que es lo mismo, queda fuera del círculo de radio unitario.

Ahora pasamos al análisis de los residuos del modelo estimado, donde podremos comprobar si estos provienen de un proceso de ruido blanco.

El gráfico de los residuos del *Modelo 1* (Figura 17), demuestra que éstos oscilan alrededor de cero, luego la media se considera constante. Por otro lado, podemos observar la aleatoriedad en la amplitud de las oscilaciones que recorren el comportamiento de la serie, luego podemos considerar que la varianza también es constante.

Figura 17. Gráfico de los residuos del Modelo 1



Fuente: Elaboración Propia con Gretl

Los resultados que obtenemos del análisis de los residuos para un total de 15 retardos, serían los siguientes:

Función de autocorrelación de los residuos

RETARDO	FAC	FACP	Estad-Q. [Valor p]	
1	0.1443	0.1443		
2	0.0673	0.0474	0.8702	[0.351]
3	0.0484	0.0330	0.9557	[0.620]

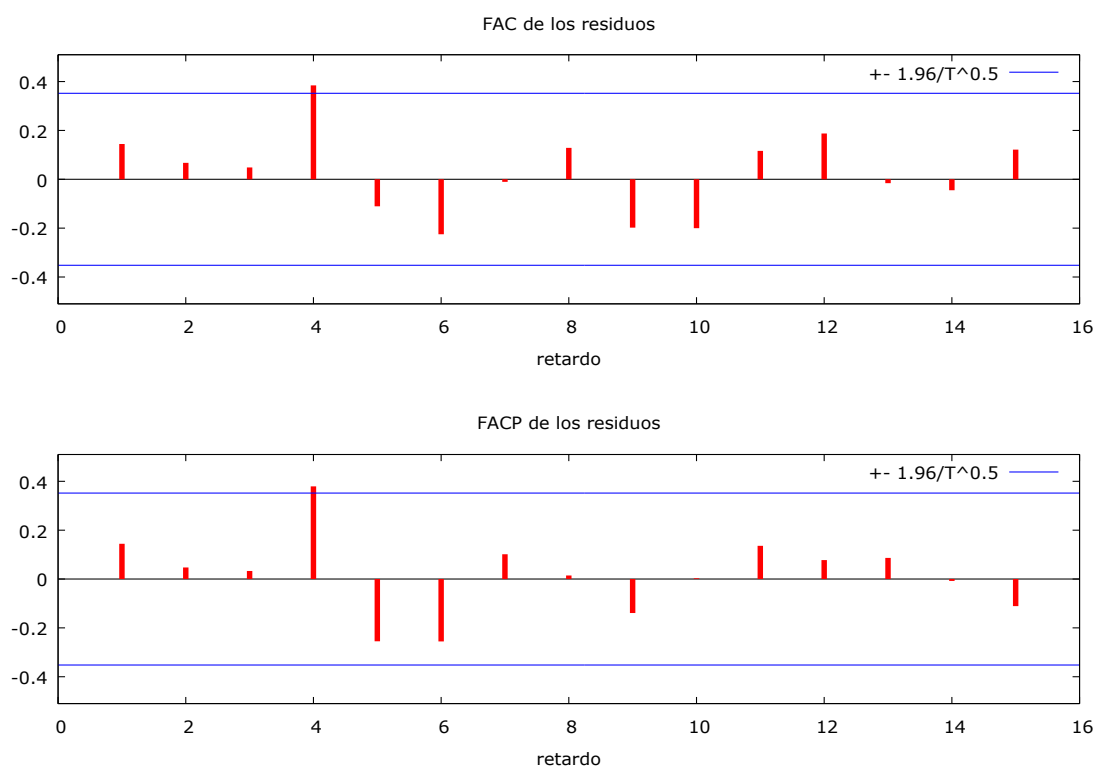
Análisis de la evolución temporal del desempleo en España

4	0.3845 **	0.3794 **	6.5566	[0.087]
5	-0.1107	-0.2549	7.0384	[0.134]
6	-0.2251	-0.2553	9.1123	[0.105]
7	-0.0105	0.1015	9.1170	[0.167]
8	0.1287	0.0145	9.8534	[0.197]
9	-0.1980	-0.1391	11.6756	[0.166]
10	-0.2003	0.0031	13.6292	[0.136]
11	0.1162	0.1360	14.3198	[0.159]
12	0.1875	0.0776	16.2127	[0.133]
13	-0.0160	0.0864	16.2273	[0.181]
14	-0.0449	-0.0077	16.3487	[0.231]
15	0.1214	-0.1109	17.2916	[0.241]

Observando las dos últimas columnas de estos resultados, vemos el valor del estadístico Ljung-Box y su correspondiente p-valor, sobre los cuales podemos destacar que sólo en el tercer retardo se sobrepasa el 5% de significación, por lo tanto, no tenemos evidencia para rechazar la hipótesis de que existe ruido blanco.

Como consecuencia, podemos decir que los residuos son la imagen empírica de un proceso de ruido blanco y por tanto, el modelo quedará validado.

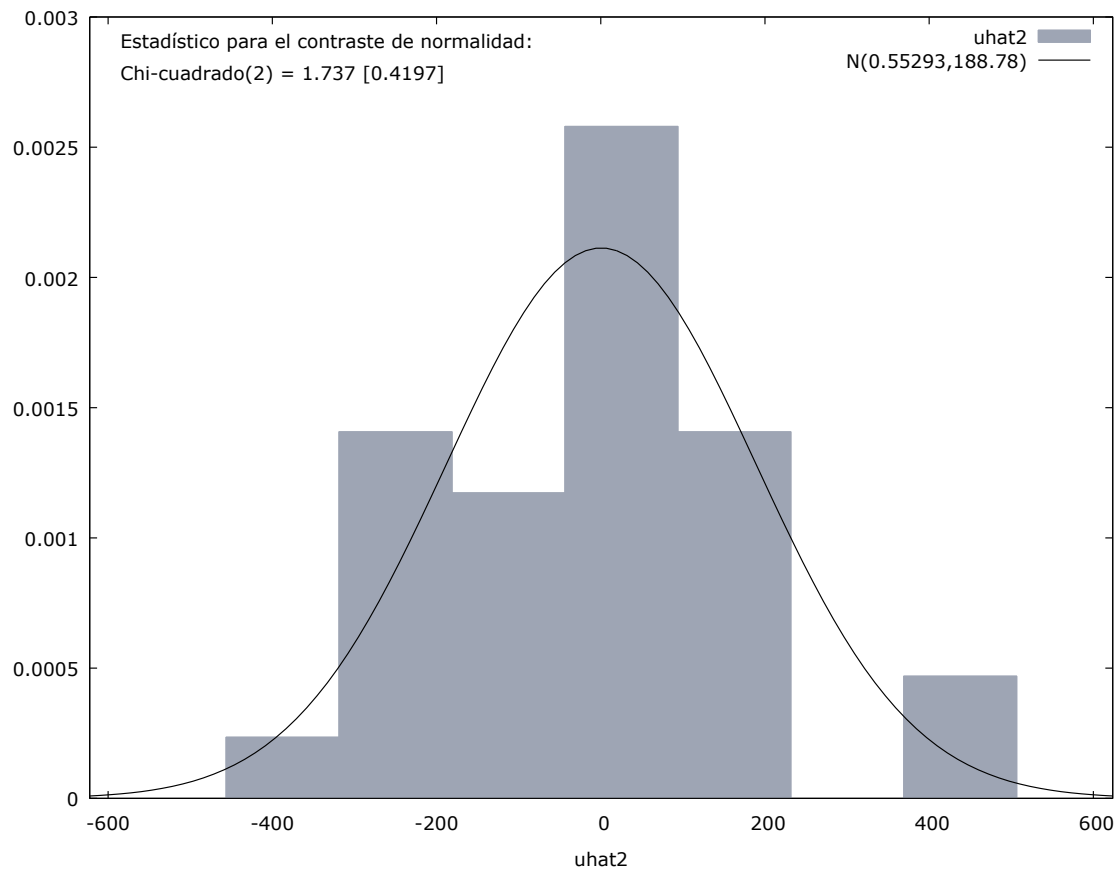
Figura 18. Correlograma de los residuos del Modelo 1



Fuente: Elaboración Propia con Gretl

Y por último, para confirmar la validez del *Modelo 1* tenemos que estudiar la normalidad de los residuos:

Figura 19. Gráfico normalidad de los residuos del Modelo 1



Fuente: Elaboración Propia con Gretl

El contraste de normalidad lo planteamos de la siguiente manera:

$$\begin{cases} H_0: \text{los residuos proceden de una variable aleatoria con distribución normal} \\ H_1: \text{los residuos no proceden de una variable aleatoria con distribución normal} \end{cases}$$

El resultado que hemos obtenido mediante Gretl, es el siguiente:

Distribución de frecuencias para uhat2, observaciones 23-53
Número de cajas = 7, media = 0.552928, desv.típ.=188.781

Intervalo	punto medio	frecuencia	rel	acum.
-----------	-------------	------------	-----	-------

Análisis de la evolución temporal del desempleo en España

< -318.75	-387.52	1	3.23%	3.23%	*
-318.75 - -181.19	-249.97	6	19.35%	22.58%	*****
-181.19 - -43.634	-112.41	5	16.13%	38.71%	*****
-43.634 - 93.922	25.144	11	35.48%	74.19%	*****
93.922 - 231.48	162.70	6	19.35%	93.55%	*****
231.48 - 369.03	300.26	0	0.00%	93.55%	
>= 369.03	437.81	2	6.45%	100.00%	**

Contraste de la hipótesis nula de distribución normal:

Chi-cuadrado (2) = 1.737 con valor p 0.41966

El p-valor que obtenemos del contraste es 0.41966, superior en cualquier caso a los niveles de significación que podemos considerar, por lo que no rechazamos H_0 , es decir, los residuos proceden de un variable aleatoria con distribución normal.

En definitiva, todos los controles de diagnósticos a los que hemos sometido al *Modelo 1* han sido superados por éste, luego el modelo queda validado y por consiguiente podemos utilizarlo para obtener una predicción del comportamiento futuro de la serie.

- Modelo 2

El Modelo 2 queda descartado al comprobar que el parámetro correspondiente a la parte autorregresiva no es significativo, por lo que eliminando dicho término nos quedaría el Modelo 1. Además, si analizamos los valores de los siguientes criterios:

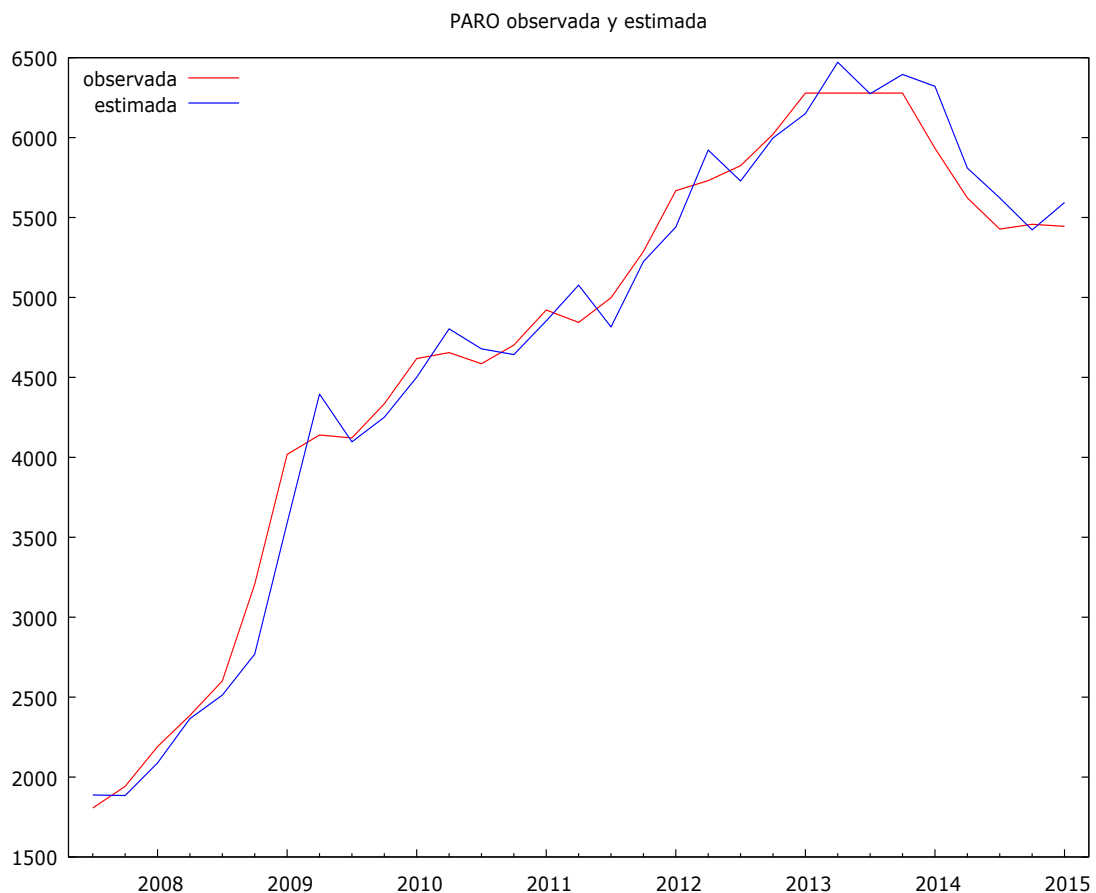
Tabla 2. Criterios de selección modelos ARIMA

CRITERIOS DE SELECCIÓN DEL MODELO			
Modelo 1		Modelo 2	
Criterio de Akaike	417.2060	Criterio de Akaike	418.1695
Criterio de Schwarz	421.5079	Criterio de Schwarz	423.9055
Crit. de Hannan-Quinn	418.6083	Crit. de Hannan-Quinn	420.0393

Fuente: Elaboración Propia

Observamos que los valores que obtenemos para el Criterio de Akaike, el Criterio de Schwartz y el Criterio de Hannan-Quinn del Modelo 1 son inferiores con respecto a los del Modelo 2, por lo tanto, podemos concluir que el primer modelo es preferible, es decir, el modelo adecuado para la predicción que vamos a realizar es el de Medias Móviles.

Figura 20. Gráfico desempleo real y estimado



Fuente: Elaboración Propia con Gretl

En la Figura 20, queda reflejado el buen ajuste que existe entre el comportamiento real que ha tenido el desempleo en España en los últimos años, y lo que el modelo que hemos estimado muestra, por lo tanto, esto nos hace pensar que la predicción que vamos obtener para el resto de 2015 y el año 2016 con los valores que hemos analizado, puede ser cercana a la realidad.

4.4. Pronóstico

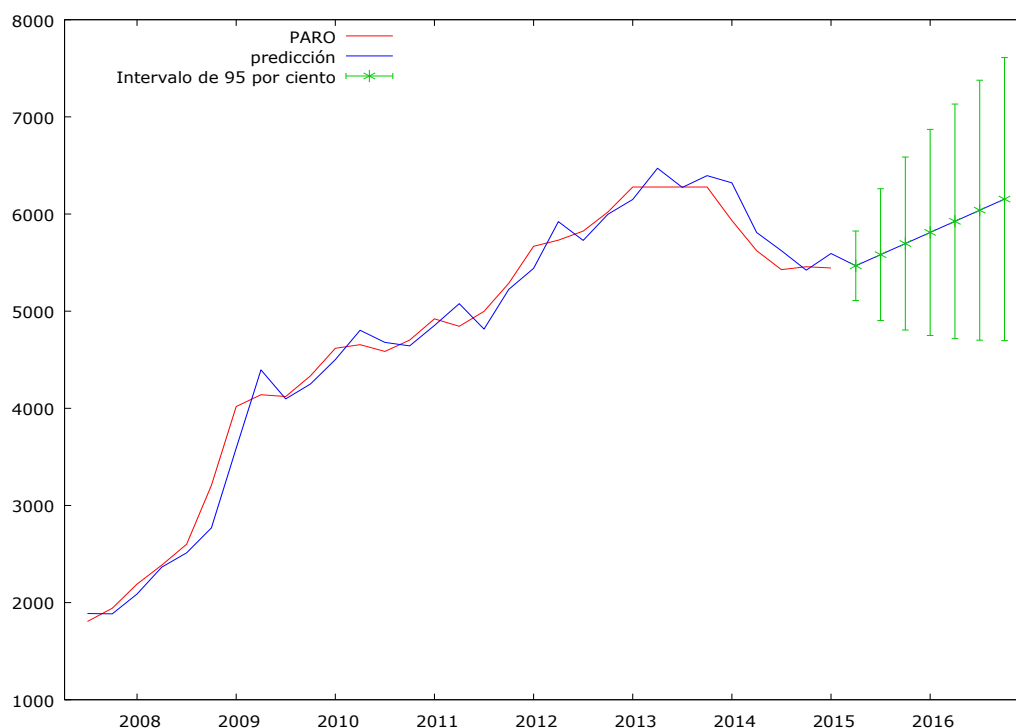
Puesto que ya hemos seleccionado el modelo adecuado y este ha sido validado a través de los criterios correspondientes, ya podemos llevar a cabo el pronóstico futuro sobre el comportamiento del desempleo en España durante el periodo seleccionado.

Con la ayuda de Gretl, realizamos un análisis de predicción automática, con un intervalo de confianza del 95%, y una duración de 7 trimestres, es decir, hasta el último trimestre de 2016.

La representación gráfica la podemos contemplar en la Figura 21 y los resultados analíticos que Gretl nos proporciona, son los siguientes:

Para intervalos de confianza 95%, $z(0.025) = 1.96$			
Observaciones	Predicción	Desv. Típica	Intervalo de 95%
2015:2	5467.50	182.326	(5110.15, 5824.85)
2015:3	5581.86	346.117	(4903.48, 6260.24)
2015:4	5696.22	454.259	(4805.89, 6586.55)
2016:1	5810.57	541.207	(4749.83, 6871.32)
2016:2	5924.93	616.003	(4717.59, 7132.28)
2016:3	6039.29	682.652	(4701.32, 7377.26)
2016:4	6153.65	743.349	(4696.71, 7610.59)

Figura 21. Gráfico predicción 2015-2016



Fuente: Elaboración Propia con Gretl

A continuación, se procede al desarrollo del pronóstico resultante sobre el comportamiento futuro de la serie temporal que hemos sometido a este análisis econométrico.

Teniendo en cuenta que la ecuación del modelo que hemos elegido corresponde a la ecuación de una recta, podemos observar en el gráfico anterior (Figura 21), que la predicción sobre el comportamiento futuro de nuestra variable, viene dada por una recta con pendiente creciente, lo que indica que el desempleo en España se espera que aumente a lo largo de este año 2015 y el próximo año 2016. Aun así, no debemos olvidar que este resultado ha podido estar influenciado por el hecho de no haber sometido al análisis los primeros valores de la serie, es decir, el resultado que ofrece el modelo que hemos estimado puede estar motivado por la tendencia creciente que experimentan los valores del desempleo durante el periodo analizado, sin tener en cuenta su trayectoria previa.

Por último, se debe tener en cuenta que aunque conociésemos con toda seguridad los valores exactos de los coeficientes que forman el modelo, lo que es prácticamente

imposible que se de en la práctica, la realización de un pronóstico futuro trae consigo siempre cierto grado de incertidumbre y, por tanto, un margen de error en la predicción que no se puede eliminar.

5. CONCLUSIONES

Para concluir el estudio realizado sobre la evolución del desempleo en España a lo largo de los últimos años, es necesario hacer algunos comentarios sobre el resultado conseguido y además expresar una breve opinión personal acerca de lo que me ha supuesto a nivel personal la elaboración de este Trabajo de Fin de Grado.

Realizado el análisis de series temporales que se pretendía, hemos obtenido como resultado que el desempleo en España a lo largo del 2016 se espera que siga aumentando ligeramente sus cifras, por lo tanto, la situación económica de nuestro país puede considerarse crítica. Aunque también hemos observado que éste puede ser un resultado influenciado por la tendencia que experimentan los valores que forman la muestra analizada.

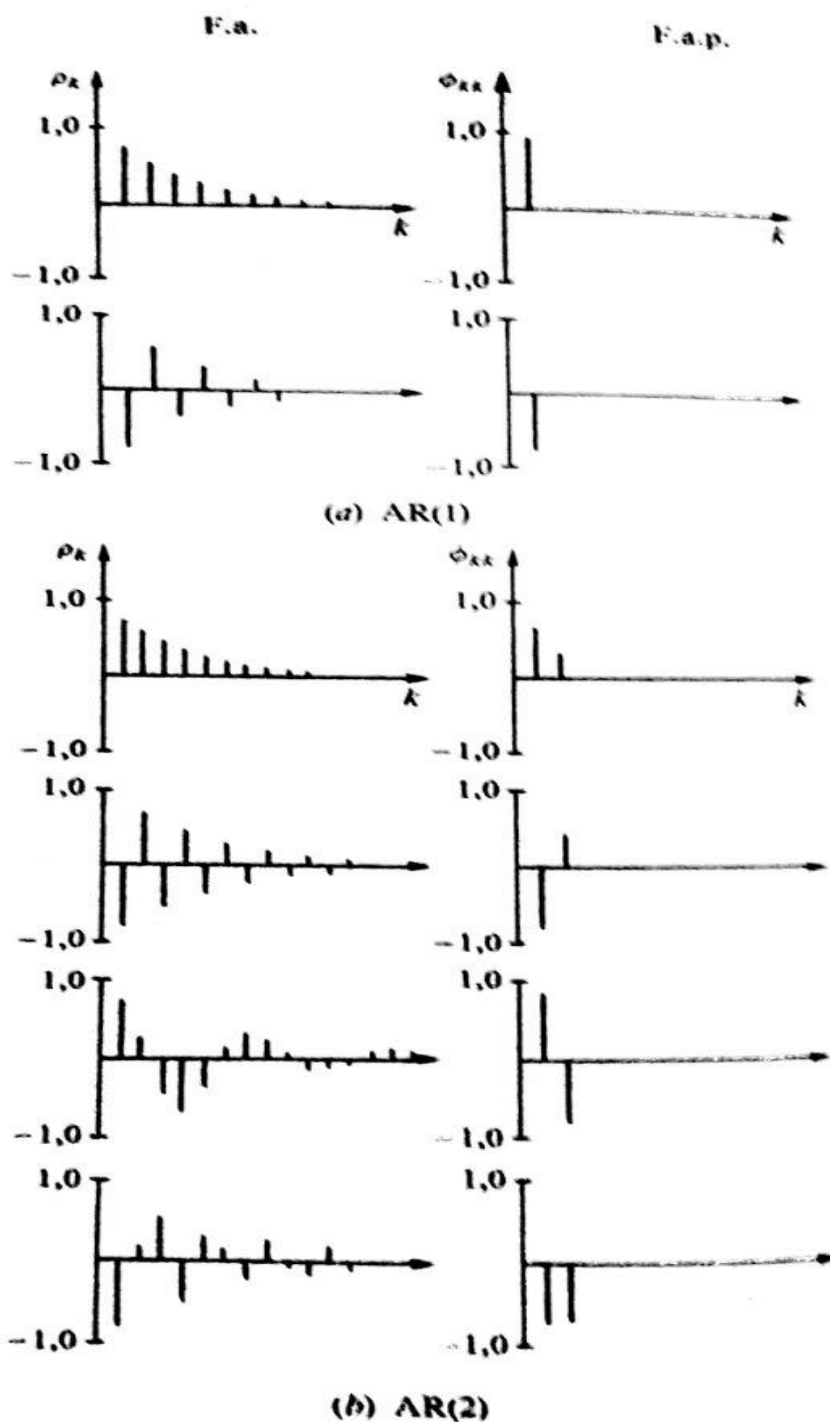
En España, el efecto del desempleo sobre la economía ha sido tan fuerte que con el tiempo hemos vuelto a ser un país de emigrantes, pero ahora con un perfil de emigrante joven y cualificado sin oportunidades de progresar profesionalmente, lo que hace que se pierda toda una inversión, que una vez se realizó sobre dichos jóvenes, con vistas a conseguir en un futuro un país más próspero.

Aun así, no debemos olvidar que el futuro no es algo cierto, sino que conlleva un importante grado de incertidumbre, por lo tanto, se deben tomar decisiones para cambiar la situación actual y mejorar, pero siempre deben ser tomadas con astucia y cautela.

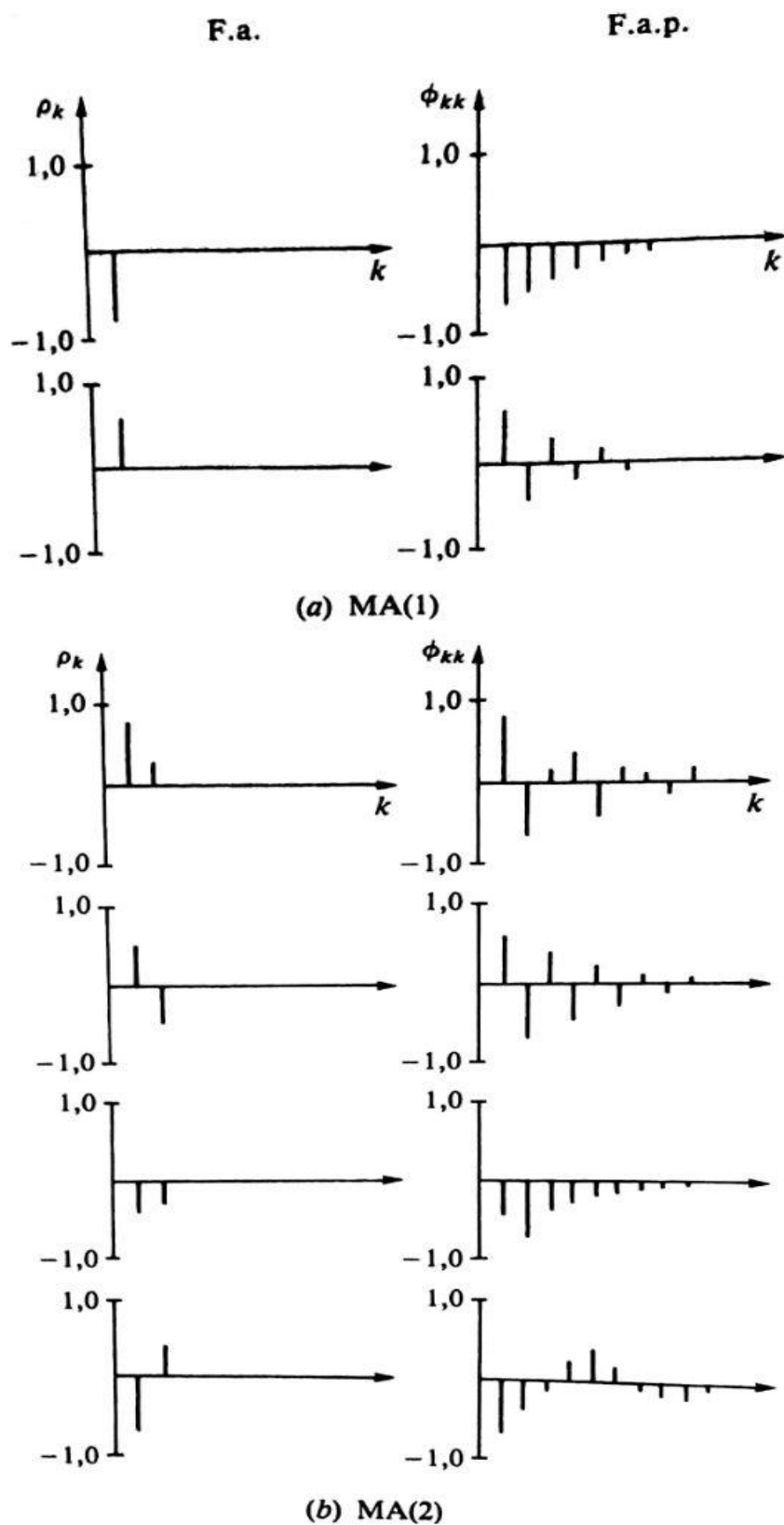
Finalmente, tengo que mencionar que a nivel personal la realización de este proyecto me ha supuesto un reto, pero a la vez me ha ayudado a desarrollar en profundidad los conocimientos ya adquiridos durante mis estudios cursados sobre economía, pero también sobre el área de econometría, que sin duda alguna me serán de utilidad en mi futuro profesional.

ANEXO

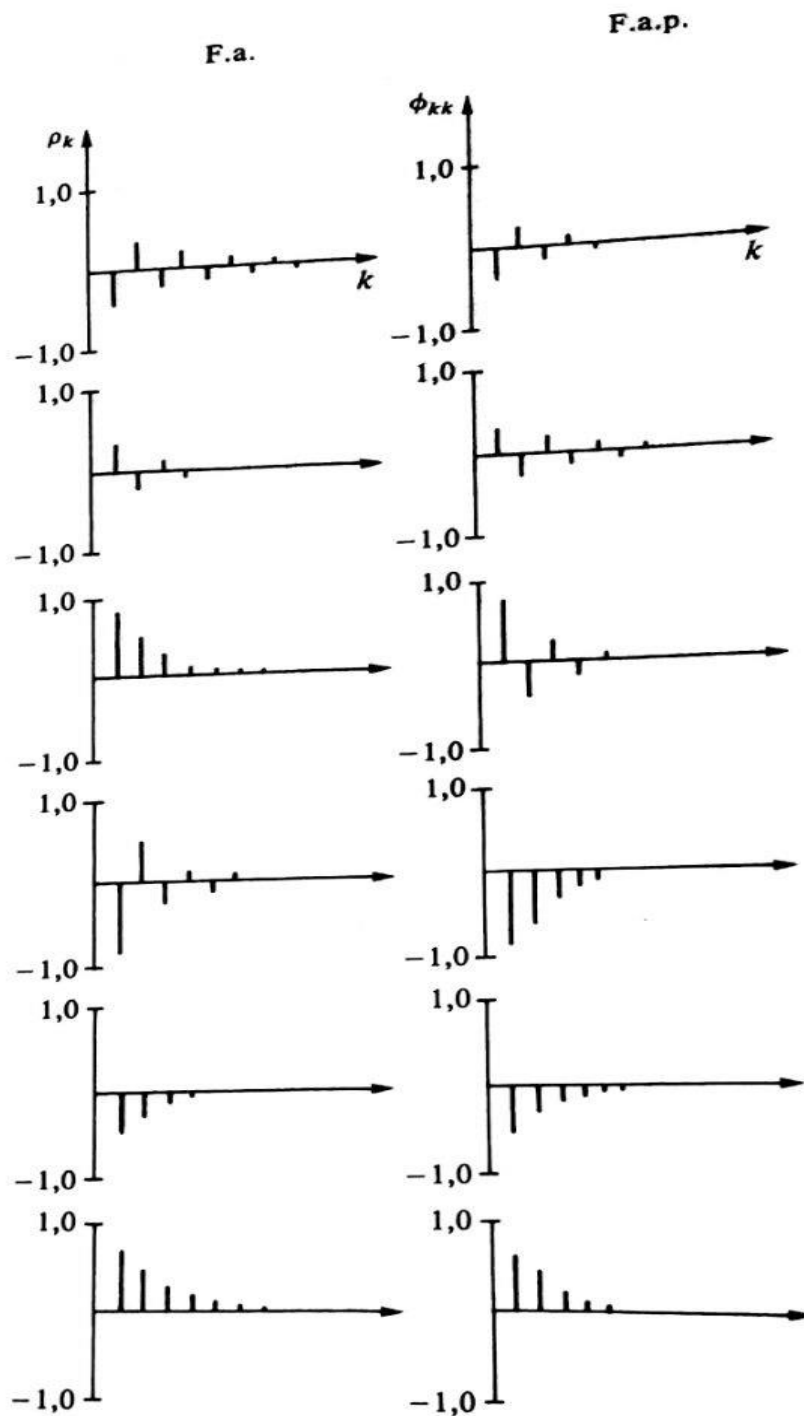
Formas que puede adoptar el Correlograma en los modelos ARIMA.



F.a. y f.a.p. para procesos AR(1) y AR(2).



F.a. y f.a.p. para procesos MA(1) y MA(2).



F.a. y f.a.p. de procesos ARMA(1, 1).

Fuente: "Econometría: Series temporales y predicción" escrito por Jose M^a Otero (1993, pág. 230, 235 y 239)

BIBLIOGRAFÍA

- Cáceres, José J., (2007) “*Conceptos básicos de estadística para ciencias sociales*”, Publicaciones Delta Univ., S.L.
- De Arce, R. y Mahía, R. (2011). *Modelos ARIMA*. Disponible online:

http://www.uam.es/personal_pdi/economicas/anadelsur/pdf/Box-Jenkins.PDF

- Del Arce, R. y Mahía, R., “*Modelos ARIMA: Definiciones básicas*” Disponible online:

[https://www.uam.es/personal_pdi/.../rmc/.../ARIMA\(1\)_Conceptos.doc](https://www.uam.es/personal_pdi/.../rmc/.../ARIMA(1)_Conceptos.doc)

- Fernández, A. (2003), “*Análisis de series de tiempo*” Disponible online:

http://www.ccee.edu.uy/ensenian/catectr/material/arima_I.pdf

- González, P., (2009), “*Análisis de series de tiempo: Modelos ARIMA*” Disponible online:

<https://addi.ehu.es/bitstream/10810/12492/1/04-09gon.pdf>

- Gujarati, Damodar N. y Porter, Dawn C., (2010) “*Econometría*” (5ª Edición), McGraw Hill
- Hernández, J., (2007) “*Análisis de series temporales económicas II*”, ESIC Editorial
- Matilla, M., Pérez, P. y Sanz, B., (2013), “*Econometría y Predicción*”, McGraw Hill
- Mauricio, José A., (2007). “*Introducción al Análisis de Series Temporales*”. Universidad Complutense de Madrid, pp. 1-8. Disponible online:

<https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf>

- Novales, A. (1993) “*Econometría*”, Segunda Edición, McGraw Hill

- Otero, Jose M^a., (1993), *“Econometría, Series temporales y Predicción”*, Editorial AC
- Pérez, P., (2011), *“Tema 9: Modelos de Series Temporales”*. Universidad Rey Juan Carlos. Disponible online:

<http://www.centrokepler.net/pdfs/Aula%20Virtual/Universidad/Econometria/Tema%209.pdf>

- Peña, D. (2005) *“Análisis de series temporales”*, Alianza Editorial.
- Requeijo, J., Iranzo, Juan E., Pedrosa, M., Salido, J., Izquierdo, G., Moral, J., Martínez de Dios, J., Barquín, R. y Arranz, N., (2005) *“Economía Española”*, Publicaciones Delta Univ., S.L.
- Rosales, Luis A., (2002), *“Series de Tiempo Estacionarias. Capítulo 15.”* Disponible online:

<https://econometriaii.files.wordpress.com/2010/01/series-de-tiempo-cap-15-16.pdf>

- Sánchez, J., (2003), *“Introducción a la Estadística Empresarial. Capítulo 4. Series Temporales.”*. Disponible online:

<http://www.eumed.net/cursecon/libreria/2004/jsf/4.pdf>

- Uriel, E. y Peiró, A. (2000) *“Introducción al análisis de series temporales”*, S.A. Alfa Centauro

RECURSOS WEB

- Blog de Economía y Finanzas:
<http://www.elblogsalmon.com/economia/que-dice-la-teoria-economica-sobre-el-desempleo> (22/04/2015)
- Blog sobre el Mercado Laboral:
<http://www.laboral-social.com/aspectos-teoricos-mercado-trabajo.html>
(22/04/2015)
- Sitio web del grupo eumed.net (Universidad de Málaga):
<http://www.eumed.net/ce/2007b/jajj.htm> (23/04/2015)
- Encuesta de Población Activa (2005):
<http://www.ine.es/daco/daco43/resumetepa.pdf> (25/04/2015)
- Encuesta de Población activa (Primer trimestre 2015):
<http://www.ine.es/daco/daco42/daco4211/epa0115.pdf> (23/04/2015)
- Periódico “EL PAÍS” (versión online):
http://economia.elpais.com/economia/2015/04/23/empleo/1429771389_288615.html (25/04/2015)
- Periódico “EL PAÍS” (versión online):
http://economia.elpais.com/economia/2014/08/14/actualidad/1408036451_169670.html (25/04/2015)
- Diario “La Nueva España”:
<http://www.lne.es/noticias-hoy/diferencias-epa-inem.html> (05/05/2015)
- Informe del Servicio Público de Empleo Estatal:
https://www.sepe.es/contenidos/que_es_el_sepe/estadisticas/datos_avance/datos_datos_2015/AV_SISPE_1504.pdf (05/05/2015)
- Sitio web del Servicio Público de Empleo Estatal:
https://www.sepe.es/contenidos/que_es_el_sepe/quienes_somos/quienes_somos.html (05/05/2015)
- Arellano, M. (2001): "Introducción al Análisis Clásico de Series de Tiempo", [en línea] Estadística:
<http://www.5campus.com/leccion/seriest> (08/05/2015)