



UNIVERSIDAD DE JAÉN
FACULTAD DE CIENCIAS SOCIALES Y JURÍDICAS

Trabajo Fin de Grado

**MODELIZACIÓN DE DATOS
ECONÓMICOS
UTILIZANDO SERIES
TEMPORALES**

Alumna: Miryam Molina del Moral

Junio, 2014

RESUMEN:

La realización de este Trabajo Fin de Grado consiste en conocer la evolución de la actividad productiva de las ramas industriales, excluida la construcción, mediante el estudio del Índice de Producción Industrial (IPI). Durante el desarrollo del trabajo se pretende conocer qué es exactamente este índice, como se calcula y quién es el encargado de su recogida y recopilación de la información, tanto a nivel nacional como de Andalucía, para posteriormente realizar un estudio de series temporales que permita obtener el modelo más apropiado para la serie. Una vez conseguido el modelo, se lleva a cabo uno de los principales objetivos por los que se realiza el análisis econométrico: la predicción de la evolución futura de la serie, la cual es realmente ventajosa al tratarse de un índice reflejo del crecimiento económico de una región. Finalmente, son elaboradas unas conclusiones donde se detallan los resultados obtenidos y son interpretados económicamente.

RESUMEN EN INGLÉS:

The fulfillment of this final degree project is to know the evolution of the production activity of the industrial branches, excluding construction, through the study of the Industrial Production Index (IPI). During the development of the work is meant to know what is exactly this index, how it is calculated and who is the manager of its collection and summary of the information, so much on the national level like of Andalucía, in order to realize a study of temporary series that might give the most appropriate model for the series. Once obtained the model, develops one of the aims for the econometric analysis is performed: the prediction of the future evolution of the series, which is really advantageous to be a reflection rate of economic growth of a region. Finally, some conclusions are elaborated where the obtained results are detailed and interpreted economically.

ÍNDICE:

1. INTRODUCCIÓN	Pág.3
2. ÍNDICE DE PRODUCCIÓN INDUSTRIAL (IPI)	Pág.4
2.1. DEFINICIÓN	Pág.4
2.2. OBJETIVO Y UTILIDADES	Pág.4
2.3. MUESTRA	Pág.5
2.4. RECOGIDA DE LA INFORMACIÓN	Pág.6
2.5. VALIDACIÓN Y DEPURACIÓN DE LOS DATOS	Pág.9
2.6. METODOLOGÍA	Pág.10
3. ÍNDICE DE PRODUCCIÓN INDUSTRIAL DE ANDALUCÍA (IPIAN)	Pág.12
4. APLICACIÓN DE LA METODOLOGÍA BOX-JENKINS AL IPI ANDALUZ	Pág.15
4.1. IDENTIFICACIÓN	Pág.19
4.2. ESTIMACIÓN	Pág.29
4.2.1. Modelo 1	Pág.29
4.2.2. Modelo 2	Pág.30
4.2.3. Modelo 3	Pág.31
4.3. DIAGNÓSIS	Pág.31
4.3.1. Modelo 1	Pág.31
4.3.2. Modelo 2	Pág.35
4.3.3. Modelo 3	Pág.39
4.4. PREDICCIÓN	Pág.39
5. CONCLUSIONES	Pág.43
6. BIBLIOGRAFÍA	Pág.44

1. INTRODUCCIÓN

La elaboración del Trabajo Fin de Grado supone una defensa por parte del alumno de una materia elegida según sus preferencias. En este aspecto, he de mencionar que la decisión por llevar a cabo un análisis econométrico a través de la modelización de series temporales se debe principalmente al interés que presento por este tipo de disciplinas prácticas relacionadas con el cálculo de modelos matemáticos y estadísticos. Asimismo, los conocimientos adquiridos en esta materia son de gran utilidad debido a la importancia que conlleva el análisis, predicción e interpretación de datos económicos. Como estudiante del Grado en Finanzas y Contabilidad, soy consciente de la importancia que supone ser capaz de realizar predicciones para así poder enfrentarnos a situaciones futuras y lograr alcanzar éxito empresarial.

El objetivo fundamental de este trabajo es construir el modelo ARIMA que mejor se ajuste a la variable *Índice de Producción Industrial* (IPI) en Andalucía y que permita hacer predicciones de dicha variable. Para ello, aplicaremos la metodología desarrollada por Box y Jenkins (Box et al., 2008) y utilizaremos el software econométrico *Gretl* (GnuRegression, Econometrics and Time-series Library), libre y de código abierto, disponible en la siguiente dirección:

<http://gretl.sourceforge.net/es.html>

El trabajo está estructurado en dos partes, teórica y práctica, claramente diferenciadas. En la primera de ellas se define la variable económica IPI, cuál es su utilidad y cómo se construye, y se describe brevemente el proceso de recogida y depuración de los datos asociados a dicha variable. Esto favorecerá la interpretación de los resultados obtenidos durante el análisis de la serie temporal. La modelización de la serie temporal conforma la segunda parte del trabajo donde se aplican detalladamente las cuatro fases de la metodología Box-Jenkins, como aplicación de los conocimientos adquiridos en la disciplina econométrica. El análisis se realizará mediante el estudio de gráficos y contrastes de hipótesis, para, finalmente, establecer las conclusiones que culminan el estudio y sintetizan los resultados alcanzados.

2. ÍNDICE DE PRODUCCIÓN INDUSTRIAL

2.1.DEFINICIÓN

El Índice de Producción Industrial (IPI) es un indicador coyuntural que mide la evolución mensual de la actividad productiva de las ramas industriales, excluida la construcción, contenidas en la Clasificación Nacional de Actividades Económicas 2009 (CNAE-2009, Real Decreto 475/2007, de 13 de abril publicado en BOE 102/2007)(Tabla 1).

Concretamente, las ramas industriales consideradas son:

- Industrias extractivas.
- Industrias manufactureras.
- Industrias de producción y distribución de energía eléctrica, agua y gas.

Además, por primera vez para la base 2010, ha sido incluida la división 36: Captación, depuración y distribución de agua del grupo E, y se han completado ramas de actividad que ya estaban recogidas, como la investigación de las clases pertenecientes al grupo 33.1.Reparación de productos mecánicos, maquinaria y equipo. También se han incluido productos nuevos, como las construcciones prefabricadas de madera, vehículos de motor eléctrico y/o híbrido, servicios de impresión y los relacionados con la industria del mueble.

El organismo autónomo de España encargado de calcular este índice desde el año 1975 es el Instituto Nacional de Estadística (INE, <http://www.ine.es/>) difundándose dicha información a escala de las Comunidades Autónomas a partir del 2002, operación estadística que se realiza al amparo de los Planes Estadísticos Nacionales.

2.2.OBJETIVO Y UTILIDADES

Los números índices son utilizados para mostrar cambios durante un periodo comparándose una cantidad (venta, precio, producción, etc.) con el valor correspondiente en algún periodo anterior al que se le conoce como base.

$$\text{Índice} = \frac{\text{Valor del año determinando}}{\text{Valor del año base}} \cdot 100$$

Son números muy versátiles, aplicables a cualquier ciencia o campo de estudio siendo el de mayor utilidad el de la economía, ya que ésta se vale de indicadores económicos para estudiar las situaciones presentes y tratar de predecir las futuras.

En concreto, el IPI tiene como objetivo medir los cambios que se producen en el volumen, en términos de valor añadido, en intervalos mensuales. Por esta razón, se trata de una herramienta analítica ampliamente utilizada para la observación y el análisis de la coyuntura económica, siendo algunas de sus finalidades las expuestas a continuación:

- Conocer el funcionamiento corriente del sistema económico.
- Establecer políticas económicas a largo plazo.
- Analizar el crecimiento y las fluctuaciones de la economía, tanto general como sectorialmente.

El conocimiento del IPI es de esencial importancia debido a que la industria es una de las ramas más significativas en la economía de un país, ya que disponer de una industria fuerte y desarrollada permite fabricar productos manufactureros y exportarlos, invirtiendo sus beneficios y creando puestos de trabajo. Por ello, los países ricos son los más industrializados, los cuales no tienen una excesiva dependencia de las importaciones. Además, la industrialización trae consigo la necesidad de obtener mano de obra especializada, generando personas más educadas, preparadas y cualificadas, es decir, incrementando en capital humano.

2.3.MUESTRA

La información necesaria para el cálculo del IPI se obtiene mediante la información suministrada por una muestra de unos 11.500 establecimientos que, según la Encuesta Industrial Anual de Productos (EIAP), producen un porcentaje significativo de cada bien seleccionado en la cesta de productos, como más representativo de cada clase de la Clasificación Nacional de Actividades Económicas (CNAE), grupos B, C y D (Tabla 1). Esta clasificación estadística asigna un código a cada actividad económica que se puede realizar, de manera que agrupa actividades homogéneas en una misma categoría facilitando así análisis estadísticos e interpretación de los datos. Las unidades que agrupa una clasificación de actividades

son las empresas y establecimientos que tienen actividades comunes. En el concepto de actividades se tienen en cuenta los inputs utilizados el proceso productivo así como el output obtenido.

Tabla 1. Clasificación Nacional de Actividades, CNAE – 2009

Grupo A	AGRICULTURA, GANADERÍA, SILVICULTURA Y PESCA
Grupo B	INDUSTRIAS EXTRACTIVAS
Grupo C	INDUSTRIA MANUFACTURERA
Grupo D	SUMINISTRO DE ENERGÍA ELÉCTRICA, GAS, VAPOR Y AIRE ACONDICIONADO
Grupo E	SUMINISTRO DE AGUA, ACTIVIDADES DE SANEAMIENTO, GESTIÓN DE RESIDUOS Y DESCONTAMINACIÓN
Grupo F	CONSTRUCCIÓN
Grupo G	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACIÓN DE VEHÍCULOS DE MOTOR Y MOTOCICLETAS
Grupo H	TRANSPORTE Y ALMACENAMIENTO
Grupo I	HOSTELERÍA
Grupo J	INFORMACIÓN Y COMUNICACIONES
Grupo K	ACTIVIDADES FINANCIERAS Y DE SEGUROS
Grupo L	ACTIVIDADES INMOBILIARIAS
Grupo M	ACTIVIDADES PROFESIONALES, CIENTÍFICAS Y TÉCNICAS
Grupo N	ACTIVIDADES ADMINISTRATIVAS Y SERVICIOS AUXILIARES
Grupo O	ADMINISTRACIÓN PÚBLICA Y DEFENSA; SEGURIDAD SOCIAL OBLIGATORIA
Grupo P	EDUCACIÓN
Grupo Q	ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES
Grupo R	ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRENIMIENTO
Grupo S	OTROS SERVICIOS
Grupo T	ACTIVIDADES DE LOS HOGARES COMO EMPLEADORES DE PERSONAL DOMÉSTICO
Grupo U	ACTIVIDADES DE ORGANIZACIONES Y ORGANISMOS EXTRATERRITORIALES

Fuente: Clasificación Nacional de Actividades Económicas

2.4.RECOGIDA DE LA INFORMACIÓN

La recolección de la información se realiza a través de las Delegaciones Provinciales y los Servicios Centrales del INE, siendo el sistema de recogida mediante la cumplimentación de unos cuestionarios mensuales por parte de los informantes de los establecimientos.

Los informantes pueden utilizar diversas vías para la devolución de los cuestionarios cumplimentados:

- ❖ Internet, a través del sistema IRIA (Integración de la recogida de información y su administración) (Figura 1). Se trata de una herramienta que permite definir, diseñar, construir y explotar sistemas de recogida de datos del INE por internet tanto para encuestas estructurales o coyunturales dirigidas a hogares o empresas. <https://iria.ine.es/iriaPortal/contents/index.jsf>
- ❖ Cuestionarios electrónicos por email.
- ❖ Vía Postal.
- ❖ Fax.
- ❖ Teléfono. Esta vía se utiliza en el caso de que el informante no haya enviado los datos requeridos dentro del plazo establecido por alguna de las vías anteriores.

En la actualidad, aproximadamente un 70% de los establecimientos optan por la vía de internet (sistema IRIA) por su comodidad y rapidez a la hora de transmitir la información.

Las fases a realizar, así como los plazos en que éstas se llevan a cabo, son las que se muestran en la Tabla 2.

Tabla 2. Fases y plazos para la recogida de información

Tarea a realizar	Fecha
Envío de cuestionarios a las unidades informantes	Última semana del final del mes t
Llegada de cuestionarios a la delegación	A partir del día 5 del mes t+1
Grabación y microdepuración	A partir del día 6 del mes t+1
Recepción de cuestionario en SSCC	Hacia el día 25 del mes t+1
Depuración y cálculo de índices	A partir del día 26 del mes t+1
Publicación de resultados	Según el calendario de disponibilidad de las estadísticas coyunturales del INE, aproximadamente 37 días después del mes de referencia

Fuente: Instituto Nacional de Estadística (INE)

Figura 1. Página web Sistema IRIA

integración de la
recogida de información
y su administración

[Inicio](#) [Contacto](#) [Mapa web](#)

Idioma Español

Portal para la cumplimentación de Encuestas por Internet

A través de esta página podrá cumplimentar de forma ágil y segura algunas de las encuestas que realiza el Instituto Nacional de Estadística. Para entrar a cumplimentar un cuestionario pulse en el botón **Acceso**.

Si antes de entrar a cumplimentar un cuestionario quiere hacer una prueba del funcionamiento del sistema pulse en el botón **Prueba de cumplimentación**.

Encontrará las normas básicas para la navegación por el cuestionario **Normas para la utilización de IRIA**.

También podrá saber por qué ha sido elegido, cuales son sus derechos y obligaciones, y por qué es necesaria su participación **Información General**.

O cuáles son los datos para contactar con las unidades de recogida del INE **Datos de unidades del INE**

El Instituto Nacional de Estadística es consciente del esfuerzo que supone para los informantes participar en una encuesta y le agradece su valiosa colaboración.

Más información en ine.es:

- Colaboradores en Encuestas del INE
- Datos disponibles de las encuestas (INEBASE)
- Cifras INE
- Notas de prensa

Síguenos en:

Cumplimentar encuestas

Acceda a cumplimentar su cuestionario

Acceso

Prueba de cumplimentación

Puede simular la cumplimentación de una encuesta de forma sencilla y rápida sin necesidad de identificarse

Prueba de cumplimentación

Relación de Unidades de Recogida

Sede Central
Paseo de la Castellana, 183
28046 Madrid-ESPAÑA

Direcciones y Teléfonos de las Unidades de Recogida del INE

Enlaces relacionados

2.5.VALIDACIÓN Y DEPURACIÓN DE LOS DATOS

Una buena representación de la realidad del IPI se consigue principalmente con una adecuada recogida de datos. Por esto, la validación y depuración de datos es una fase previa de vital importancia en el cálculo del índice.

Desde la cumplimentación del cuestionario por parte de los informantes se utiliza un sistema de detección de inconsistencias de forma que los usuarios son advertidos y se le requiere que confirmen o modifiquen la información proporcionada. Una vez grabada esta información por el propio personal del INE, donde también existe un sistema de control de grabación, aquellos datos que muestren discrepancias, el personal se pone en contacto con el informante para que confirme o modifique la información proporcionada. Finalmente, una vez compilada toda la información, se lleva a cabo un control de cobertura de la información, al objeto de garantizar la completitud de los datos grabados, así como detectar duplicados y errores de cobertura, falta de respuesta, etc.

En cuanto al proceso de depuración de datos, éste consta de dos fases:

1. **Microdepuración (depuración del microdato):** consiste en validar los datos procedentes de cada establecimiento, atendiendo a la serie histórica proporcionada por el mismo. Este análisis establece comparaciones intermensuales e interanuales.
2. **Macrodepuración:** consiste en corroborar la información agregada, mediante una depuración manual de los errores detectados en las principales variables y una depuración automática del resto de variables.

Por otra parte, la importancia de un año base de referencia para el cálculo del índice hace que sea fundamental la revisión y actualización de los componentes que conforman su cálculo. De esta forma, las últimas publicaciones del IPI son en base 2010, dando así cumplimiento al Reglamento (CE) N° 1165/98 del Consejo sobre las estadísticas coyunturales, donde se establece que los estados miembros deben cambiar la base cada cinco años, concretamente en los terminados en cero y en cinco.

El principal objetivo por el que se realizan estos cambios es conseguir un indicador que se adapte a los cambios producidos en los últimos cinco años en las

distintas actividades del sector industrial, de manera de mida la evolución de dicho sector de forma más precisa, lo que proporcionará un IPI más representativo.

La nueva base, con respecto a la anterior, ha incluido actividades nuevas que:

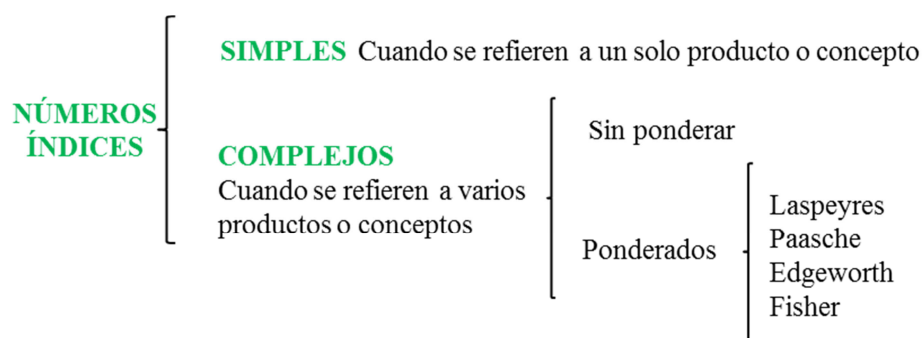
1. Han ampliado el ámbito de aplicación del IPI.
2. Han completado ramas de actividades que ya venían recogidas en las anteriores bases.

Asimismo, en la base 2010 se ha revisado la cesta de productos representativos de cada actividad de la CNAE, eliminando aquellos que aportan escasa representación y añadiendo nuevos productos, siendo un ejemplo de estos los vehículos de motor eléctrico y/o híbridos.

2.6. METODOLOGÍA

El Índice de Producción Industrial se calcula mediante un índice tipo Laspeyres de base fija. Estos índices se tratan de una media aritmética de índices de precios simples que utiliza como ponderaciones el valor de las transacciones realizadas en el periodo base. Se obtiene multiplicando los precios de cada bien en el año dado por las cantidades del año base, dividido por el producto del precio de cada bien en el año base y las cantidades del año base. No cambian los hábitos de consumo, sólo varía el precio.

Figura 2. Esquema de los números índices



Fuente: Elaboración propia

La ventaja de un índice de este tipo, además de su facilidad de cálculo e interpretación, es que permite la comparabilidad de una misma estructura de productos y ponderaciones a lo largo del tiempo que esté en vigor la base y la actividad de las series. Sin embargo, tiene el inconveniente de que la estructura de

ponderaciones pierde vigencia a medida que pasa el tiempo, ya que supone que siempre se adquieren las mismas cantidades que el periodo base y, por ello, es necesario realizar un cambio de base cada cierto tiempo.

El IPI es catalogado como un Índice de Producción por lo que mide la evolución a corto plazo del volumen del output producido por las empresas, en términos de valor añadido. Se define como el ratio entre el volumen de output producido por las empresas en un periodo de tiempo t menos los consumos necesarios para ello, y el volumen de output producido por estas mismas industrias en el periodo base menos los consumos necesarios, es decir,

$${}_0I_Q^t = \frac{\sum_{i=1}^N p_{i,0} q_{i,t} - \sum_{j=1}^{M_t} a_{j,0} \delta_{j,t}}{\sum_{i=1}^N p_{i,0} q_{i,0} - \sum_{j=1}^{M_0} a_{j,0} \delta_{j,0}}$$

Siendo¹:

- N = número de productos producidos
- M_t = número de productos consumidos en el periodo t
- M₀ = cantidad de tipo de productos consumidos en el periodo 0
- ${}_0I_Q^t$ = índice de producción en el periodo t con respecto al periodo base 0
- q_{i,t} = cantidad producida del producto i en el periodo t
- q_{i,0} = cantidad producida del producto i en el periodo 0
- p_{i,0} = precio del producto i en el periodo 0
- δ_{j,t} = cantidad del input j necesario para producir el producto i en el periodo t
- δ_{j,0} = cantidad del input j necesario para producir el producto i en el periodo 0
- a_{j,0} = precio del input j en el periodo 0

Multiplicando y dividiendo por el volumen producido en el periodo base:

$${}_0I_Q^t = \frac{\sum_{i=1}^N p_{i,0} q_{i,t} - \sum_{j=1}^{M_t} a_{j,0} \delta_{j,t}}{\sum_{i=1}^N p_{i,0} q_{i,0} - \sum_{j=1}^{M_0} a_{j,0} \delta_{j,0}} \cdot \frac{p_{i,0} q_{i,0} - \sum_{j=1}^{M_t} a_{j,0} \delta_{j,0}}{p_{i,0} q_{i,0} - \sum_{j=1}^{M_0} a_{j,0} \delta_{j,0}}$$

$${}_0I_Q^t = \sum_{i=1}^N \frac{p_{i,0} q_{i,0} - \sum_{j=1}^{M_t} a_{j,0} \delta_{j,t}}{\sum_{i=1}^N p_{i,0} q_{i,0} - \sum_{j=1}^{M_0} a_{j,0} \delta_{j,0}} \cdot \frac{p_{i,0} q_{i,t} - \sum_{j=1}^{M_t} a_{j,0} \delta_{j,0}}{p_{i,0} q_{i,0} - \sum_{j=1}^{M_0} a_{j,0} \delta_{j,0}}$$

¹ Metodología publicada por el INE, base 2010. <http://www.ine.es/daco/daco43/metoipi10.pdf>

Reordenando términos podemos expresar el primer multiplicando en términos de Valor Añadido Bruto (VAB) en el periodo base, dado que todo está valorado a precios de este periodo:

$${}_0I_Q^t = \sum_{i=1}^N \frac{VAB_{i,0}}{\sum_{i=1}^N VAB_{i,0}} \cdot \frac{p_{i,0}q_{i,t} - \sum_{j=1}^{M_t} a_{j,0}\delta_{j,0}}{p_{i,0}q_{i,0} - \sum_{j=1}^{M_0} a_{j,0}\delta_{j,0}}$$

De esta forma, las ponderaciones del índice quedan expresadas en términos del VAB de cada sector en el año base:

$${}_0I_Q^t = \sum_{i=1}^N W_{i,0} \cdot \frac{p_{i,0}q_{i,t} - \sum_{j=1}^{M_t} a_{j,0}\delta_{j,0}}{p_{i,0}q_{i,0} - \sum_{j=1}^{M_0} a_{j,0}\delta_{j,0}}$$

3. ÍNDICE DE PRODUCCIÓN INDUSTRIAL DE ANDALUCÍA

El Índice de Producción Industrial de Andalucía (IPIAN) tiene por objeto medir la oferta de bienes producidos mensualmente por la industria andaluza, indicando la evolución mensual que experimenta la parte del VAB que tiene su origen en el sector industrial. Comienza a calcularse desde enero de 2006, en base 2005, modificando la base anterior que tenía como periodo de referencia en cuanto a estructura industrial, el año 2000, y como referencia temporal 2002.

A nivel de comunidad autónoma, la información sobre el IPIAN es elaborada con periodicidad mensual por el Instituto de Estadística y Cartografía de Andalucía (IECA) <http://www.juntadeandalucia.es/institutodeestadisticaycartografia/>

Tiene por objeto indicar la evolución que experimenta el valor añadido del sector industrial de Andalucía, en periodos cortos y regulares de tiempo.

Al igual que el IPI nacional, el sector industrial al que hace referencia su cálculo coincide en términos generales con los grupos B, C y D de la CNAE-2009. Sin embargo algunas de estas actividades aquí incluidas entran en conflicto con determinados aspectos metodológicos del IPIAN, por lo que se establece una definición del sector industrial más restrictiva, en la que únicamente se consideran aquellas actividades que siendo utilizadoras de factores de producción, y mediante el empleo de una técnica

determinada, obtienen un volumen de producto no primario cuyas características formales y funcionales son esencialmente diferentes de los factores utilizados.

La exclusión de determinadas actividades incluidas en los grupos anteriormente mencionados atiende a los siguientes criterios:

- Actividades en las que la materia prima no se transforma de manera sustancial (actividades de reparación, de distribución...). Éste sería el caso también de las actividades de reparación y mantenimiento de buques, así como las de material ferroviario; no obstante, se han incluido en el cálculo del índice porque juegan un importante papel dentro de su sector y, en general, de la industria andaluza.
- Actividades artesanales cuyo componente artístico es difícilmente cuantificable (actividades anexas a la industria del mueble, del vestido...).
- Subproductos o productos residuales que dependen totalmente del producto principal en lo que respecta a volumen de producción.

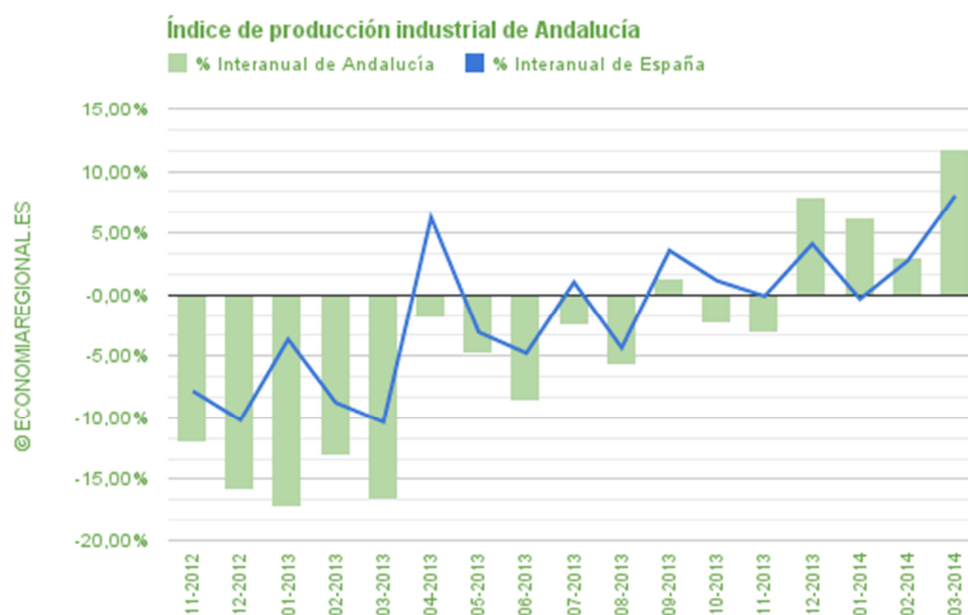
En relación con su ámbito territorial, es de mencionar que el principal indicador de la pertenencia o no de una unidad de producción al sector industrial es el mantenimiento de un establecimiento o centro de trabajo permanente en Andalucía donde se desarrollen actividades incluidas en lo que se ha definido como sector industrial.

Como es natural, si una empresa como unidad informante mantiene algún establecimiento permanente fuera del territorio económico, únicamente se imputará como producción a efectos del IPIAN la correspondiente a los establecimientos situados dentro del mismo.

En cuanto a la formulación para el cálculo del IPIAN, es semejante a la del IPI nacional (Índice de Laspeyres encadenado).

El Instituto de Estadística y Cartografía de Andalucía publica periódicamente información sobre éste índice, siendo su última publicación la referente al mes de abril de 2014: *“La producción industrial andaluza decreció en abril un 0,6% respecto al mismo mes de 2013”* (IECA, 6 de junio de 2014).

Figura 3. Gráfico evolución IPIAN 11/2012 – 03/2014



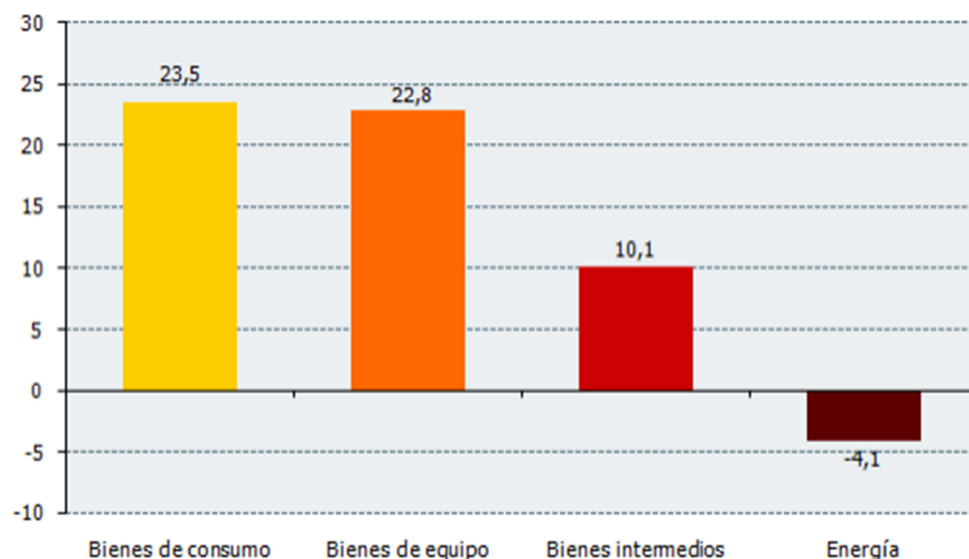
Fuente: Instituto de Estadística y Cartografía de Andalucía.

La Figura 3 muestra el comportamiento de la actividad industrial en España y en Andalucía durante la fase más acentuada de la crisis. Sin embargo, a finales del año 2013 y durante el primer trimestre del año actual, se aprecia una notable mejoría en la comunidad andaluza, probablemente fruto de las mejoras que se están produciendo en el empleo y el intenso esfuerzo realizado para que la internacionalización de las empresas consoliden el consumo interno y las exportaciones.

El crecimiento del IPIAN durante el mes de abril de 2004 es desglosado en función de los diferentes destinos económicos, siendo el incremento de la tasa de variación interanual de un 4.2% en bienes de consumo y 6.1 % en bienes de equipo, así como el decrecimiento de un 7.2% en energía y 0.1% en bienes intermedios.

Corrigiendo los datos del efecto calendario, es decir, teniendo en cuenta la distinta distribución de los días laborables en 2014 respecto a 2013, dicha tasa por destino económico ha crecido y decrecido en los porcentajes que se muestran en la Figura 4.

Figura 4. Índices por destino económico. Tasas de variación interanual corregidas de efecto de calendario. Abril 2014.



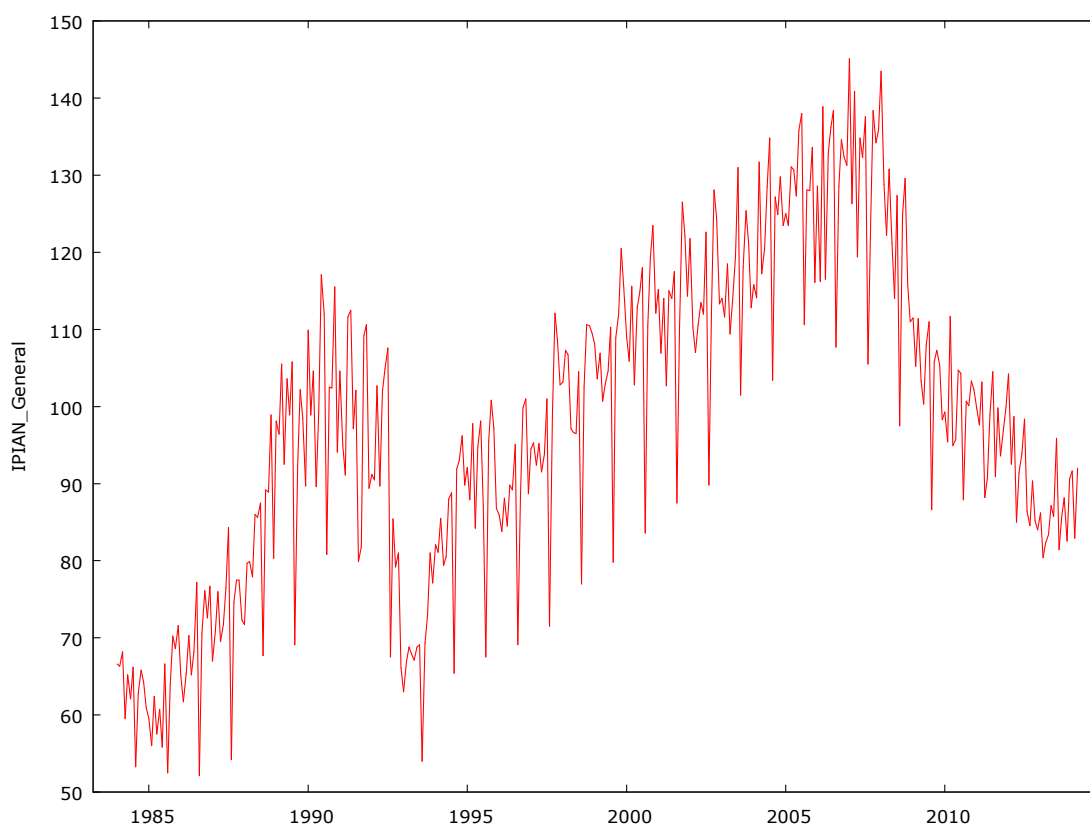
Fuente: Instituto de Estadística y Cartografía de Andalucía.

4. APLICACIÓN DE LA METODOLOGÍA BOX-JENKINS AL IPI ANDALUZ

Tal y como se indicó en la introducción, el objetivo esencial de este trabajo es modelizar la serie temporal de datos del IPI andaluz aplicando la metodología desarrollada por Box y Jenkins y utilizando el software econométrico *Gretl*. Los datos han sido extraídos de la página web del IECA (http://www.juntadeandalucia.es/institutodeestadisticaycartografia/bd/indea/indea_se.jsp?&s=6109).

Esta institución actualiza periódicamente la publicación del índice, pudiéndose consultar y realizar un estudio más específico al ofrecer información desglosada en sectores económicos y destino económico de los bienes. Sin embargo, con el fin de llevar a cabo un estudio más genérico de este índice se han considerado los datos correspondientes al índice general mensual del IPI andaluz con base en 2010 desde enero de 1984 hasta abril de 2014.

Figura 5. Representación gráfica del IPIAN con base 2010 (enero 1984 - abril 2014)



Fuente: Elaboración propia a través de *Gretl*

La representación gráfica de la serie aparece en la Figura 5, donde se aprecian claramente tendencias crecientes y decrecientes en su evolución, dando lugar a caídas sistemáticas en agosto de cada año y repuntes en septiembre, es decir, cada 12 meses, por lo que se trata de una serie con comportamiento estacional. Principalmente, se distinguen tres tramos:

- **Enero 1984 - Agosto 1993:** España firma el Tratado de Adhesión a la Comunidad Económica Europea (12 de junio de 1985), dando lugar a una etapa de crecimiento y expansión debido primordialmente a la liberalización y ampliación de los mercados. Tal desarrollo fue interrumpido con la crisis de 1992-1993, crisis corta que el Gobierno solventó con un plan de estabilización al uso y las tradicionales devaluaciones de la peseta.

Nota de prensa:

“La cifra ya es definitiva. El producto interior bruto (PIB) cayó el año pasado un 1%, en lo que puede considerarse la peor recesión de la economía española

en los últimos 30 años. Una cifra concreta, el 1%, que resume 12 meses de crisis continuada, cientos de miles, de puestos de trabajo destruidos y más de 600.000 nuevos parados”. España sufrió en 1993 la peor recesión económica registrada en los últimos 30 años. El País – 4 de marzo de 1994.

http://elpais.com/diario/1994/03/04/economia/762735606_850215.html

- **Septiembre 1993 – Agosto 2007:** tras la crisis sufrida por la debilidad económica española durante 1992-1993, el país dejó atrás la recesión en el año 1994 con un crecimiento del 2.4% del PIB, llegando a alcanzar una cifra de 2.8% en 1995. Durante estos dos años la tasa de paro bajó hasta el 22% y se crearon 400.000 puestos de trabajo. A partir de 2006, y durante más de una década, se produjo una etapa de prosperidad interrumpida en la economía española con tasas de crecimiento siempre superiores a las de la media europea.
- **Septiembre 2007 – abril 2014:** desde la fecha hasta la actualidad España se halla sumergida en una crisis económica a nivel mundial cuyos orígenes se encuentran en el sector inmobiliario dando lugar al estallido de la burbuja inmobiliaria, la crisis bancaria de 2010 y finalmente, y como consecuencia más visible de la recesión, el aumento vertiginoso del desempleo.

Los cambios bruscos de nivel y de tendencia de la serie hacen necesaria la aplicación de análisis de intervención en la modelización ARIMA. Sin embargo, como dicho estudio supera los objetivos del trabajo fin de grado, hemos optado por modelizar el último tramo de la serie, es decir, a partir de enero de 2008.

La metodología de Box-Jenkins fue desarrollada en 1976, recibiendo su nombre en honor a los estadísticos George Box y Gwilym Jenkins. Consta de cuatro fases con las que se pretende encontrar un modelo matemático que represente de forma plausible el comportamiento de los valores de una serie temporal con el fin de realizar pronósticos lo más acertados posible. Concretamente, las fases son:

1. **Identificación:** Consiste en deducir a partir de los datos qué modelo ARIMA ha podido general los datos de la serie temporal.
2. **Estimación:** Consiste en estimar los parámetros del modelo $ARIMA(p, d, q) \times ARIMA(P, D, Q)_S$.

3. **Diagnosis:** Consiste en comprobar que el modelo verifica las hipótesis de partida haciendo uso de los residuos del modelo estimado. Si no se verifican dichas hipótesis, el modelo no se considera adecuado y debe buscarse otro candidato.
4. **Predicción:** Consiste en utilizar el modelo para obtener valores futuros de la serie.

Un resumen de la metodología Box-Jenkins se resume en el gráfico de la Figura 6.

Figura 6. Fases metodología Box-Jenkins



Fuente: Elaboración propia

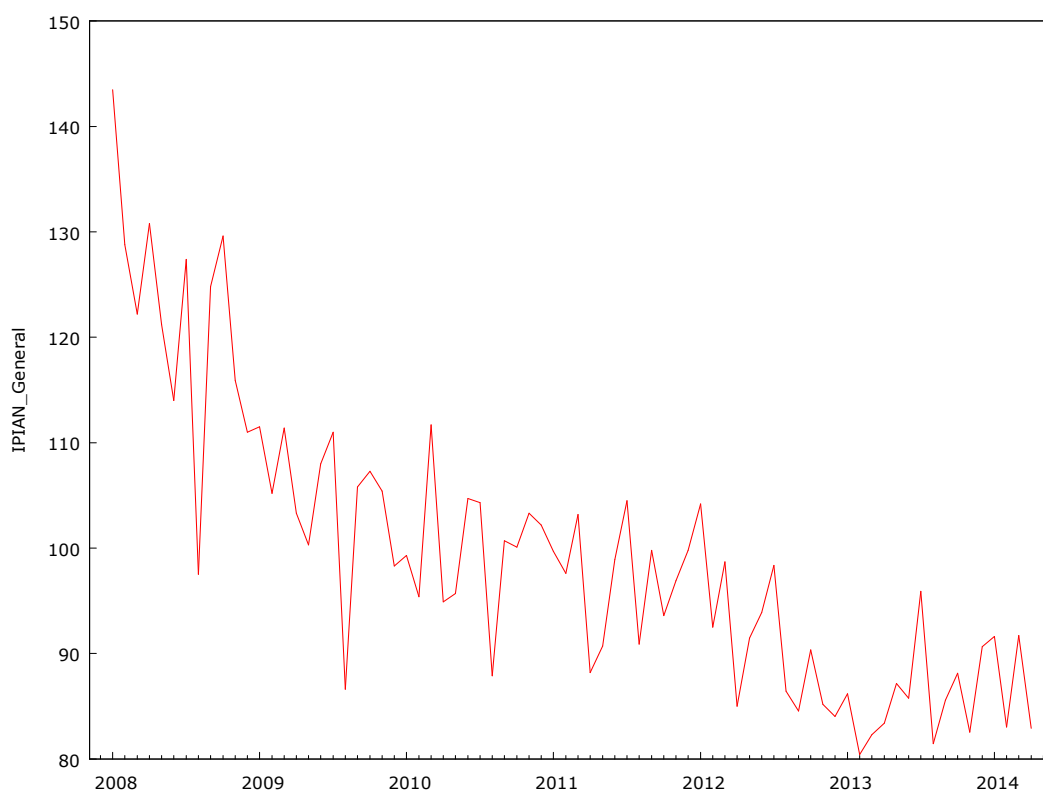
Esta metodología proporciona predicciones sin necesidad de la existencia de ningún tipo de condición previa ya que los modelos se construyen utilizando exclusivamente una variable “exógena”: el pasado de la serie. Además, una vez encontrado el modelo, se pueden efectuar de manera inmediata predicciones y comparaciones entre datos reales y estimados para observaciones pertenecientes al pasado. Sin embargo, además de requerir un elevado número de observaciones, la estimación e interpretación de sus coeficientes es compleja y proporciona peores resultados en previsiones a largo plazo.

4.1. IDENTIFICACIÓN

Como ya ha sido comentado anteriormente, esta primera etapa tiene por finalidad deducir qué modelo se adapta mejor a las características de la serie.

La serie temporal establecida en este rango está compuesta por un total de 76 observaciones y su representación gráfica se detalla en la Figura 7.

Figura 7. Gráfico del IPI andaluz con base 2010 (enero 2008 - abril 2014). Serie original



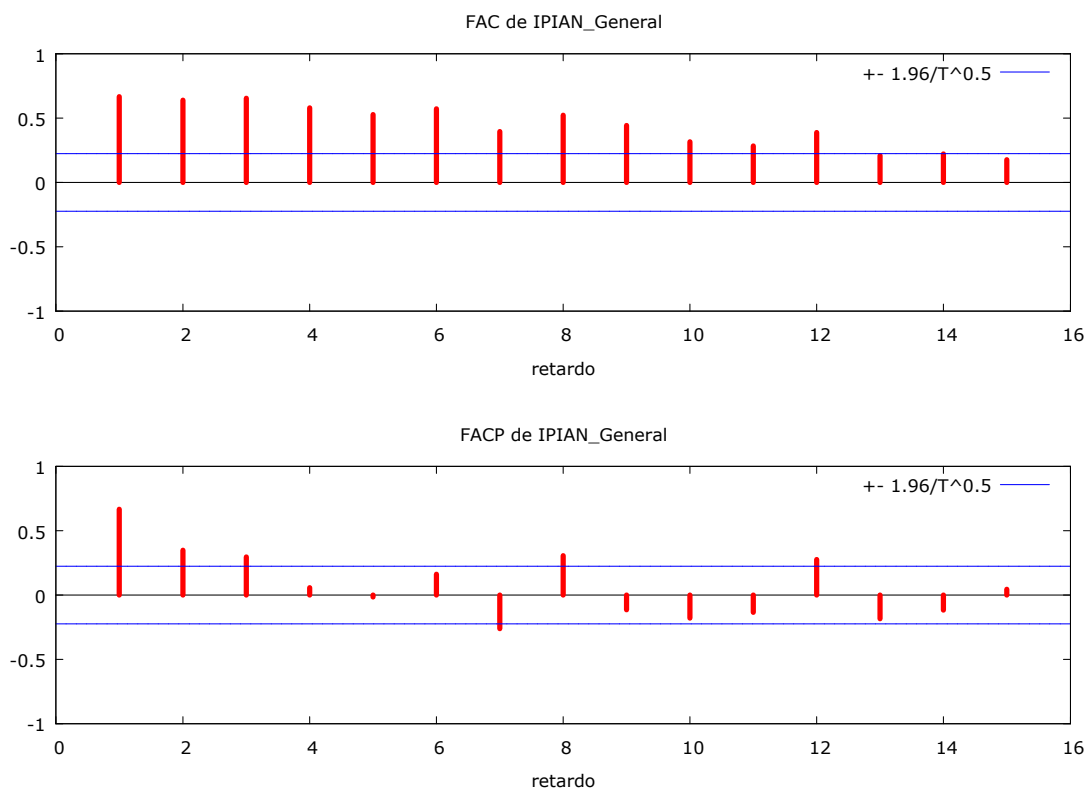
Fuente: Elaboración propia a través de *Gretl*

En este gráfico podemos observar que la serie presenta una clara tendencia decreciente, por tanto, existe evidencia de que la media del proceso subyacente no es constante. Otra forma de detectar este problema es a través de la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP) muestrales (Figura 8), ya que las barras o picos de la FAC decrecen muy lentamente, de forma lineal en lugar de exponencial o sinusoidal, mientras que la FACP posee un pico muy significativo (el 1°).

En cuanto a la varianza, la amplitud de las oscilaciones disminuye ligeramente con el tiempo, por tanto, el gráfico sugiere que la varianza no es constante.

En el gráfico de la serie también se aprecia una componente estacional de periodo 12, ya que el comportamiento se repite cada 12 meses. Esto también puede observarse en la FAC y la FACP muestrales (Figura 8).

Figura 8. Funciones de autocorrelación (FAC) y de autocorrelación parcial (FACP). Serie original



Fuente: Elaboración propia a través de *Gretl*

Para comprobar objetivamente las afirmaciones sobre la falta de estacionalidad realizaremos a continuación los contrastes de hipótesis necesarios.

- Comprobación de estacionariedad en media:

Para confirmar que la media no es constante se realiza un contraste de raíz unitaria, concretamente, el contraste aumentado de Dickey-Fuller, donde existe una raíz unitaria (serie no estacionaria).

H_0 : El modelo posee una raíz unitaria (no es estacionario en media)

H_1 : El modelo no posee una raíz unitaria (es estacionario en media)

Más información al respecto puede encontrarse en:

<http://www.monografias.com/trabajos-pdf4/pruebas-raices-unitarias-eviews/pruebas-raices-unitarias-eviews.pdf>

Los resultados obtenidos mediante *Gretl* son los siguientes:

Contraste aumentado de Dickey-Fuller para IPIAN_General
incluyendo 6 retardos de (1-L)IPIAN_General
(el máximo fue 11, el criterio AIC modificado)
tamaño muestral 76
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 : -0.021
diferencias retardadas: $F(6, 68) = 10.215$ [0.0000]
valor estimado de $(a - 1)$: -0.140983
Estadístico de contraste: $\tau_c(1) = -2.40169$
valor p asintótico 0.1412

con constante y tendencia
modelo: $(1-L)y = b_0 + b_1t + (a-1)y(-1) + \dots + e$
Coef. de autocorrelación de primer orden de e : 0.042
diferencias retardadas: $F(11, 62) = 4.503$ [0.0001]
valor estimado de $(a - 1)$: -0.220725
Estadístico de contraste: $\tau_{ct}(1) = -1.30208$
valor p asintótico 0.8872

Puesto que el p -valor es mayor que 0.05, no se rechaza H_0 al 5% de significación, por lo que la media no es constante, coincidiendo con lo deducido subjetivamente a partir del gráfico.

- Comprobación de estacionariedad en varianza:

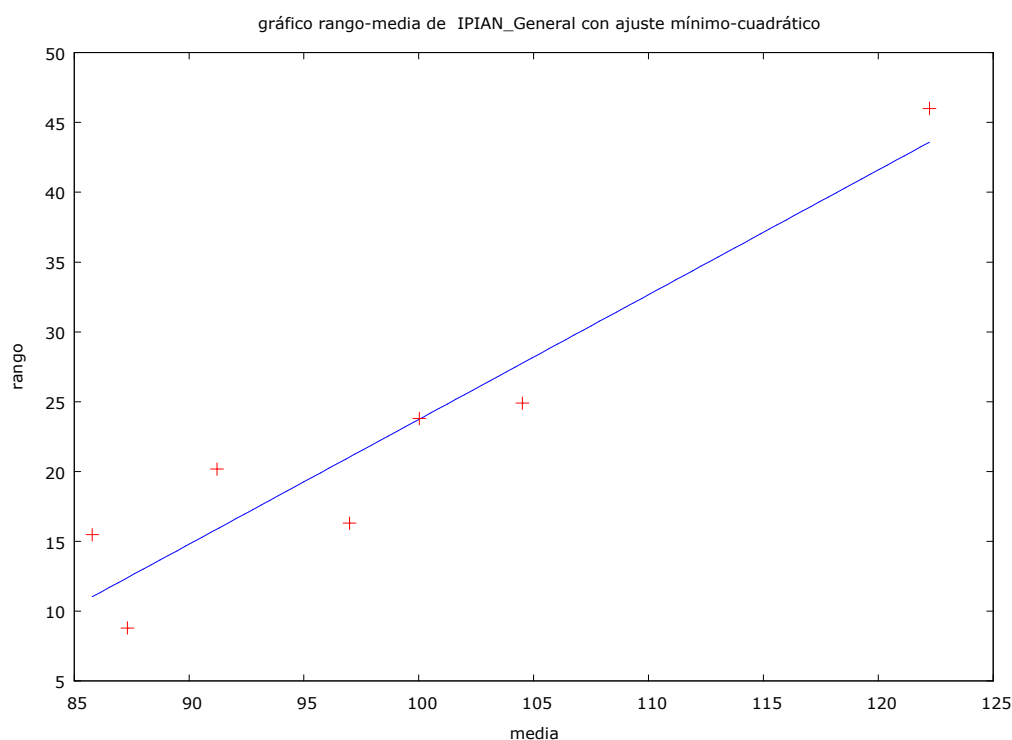
Para confirmar si la varianza es o no constante de manera objetiva se emplea el gráfico rango-media.

El gráfico rango-media se realiza dividiendo el conjunto de observaciones en subconjuntos de igual tamaño (en nuestro caso se trata de 7 subconjuntos de tamaño 12, al tener los datos estructura mensual), calculándose el rango y la media para cada subconjunto. Asimismo, ajusta una recta de regresión del rango con respecto a la media y comprueba si la pendiente de dicha recta es significativa o no.

H_0 : El rango varía en función de la media $\rightarrow$ varianza no es constante (el modelo no es estacionario en varianza)

H_1 : El rango no varía en función de la media $\rightarrow$ varianza es constante (el modelo es estacionario en varianza)

Figura 9. Gráfico rango-media. Serie original



Fuente: Elaboración propia a través de *Gretl*

El gráfico rango-media se muestra en la Figura 9. A continuación se presentan los resultados analíticos que proporciona *Gretl*:

Estadísticos de rango-media para IPIAN_General
utilizando 7 submuestras de tamaño 12

rango	media		
2008:01 - 2008:12	46.0000	122.225	
2009:01 - 2009:12	24.9000	104.508	
2010:01 - 2010:12	23.8000	100.017	
2011:01 - 2011:12	16.3000	96.9833	
2012:01 - 2012:12	20.1800	91.2200	
2013:01 - 2013:12	15.4800	85.7900	
2014:01 - 2014:04	8.80000	87.3200	

pendiente de 'rango' con respecto a 'media' = 0.893274
el valor p para H_0 : pendiente = 0 es 0.00123974

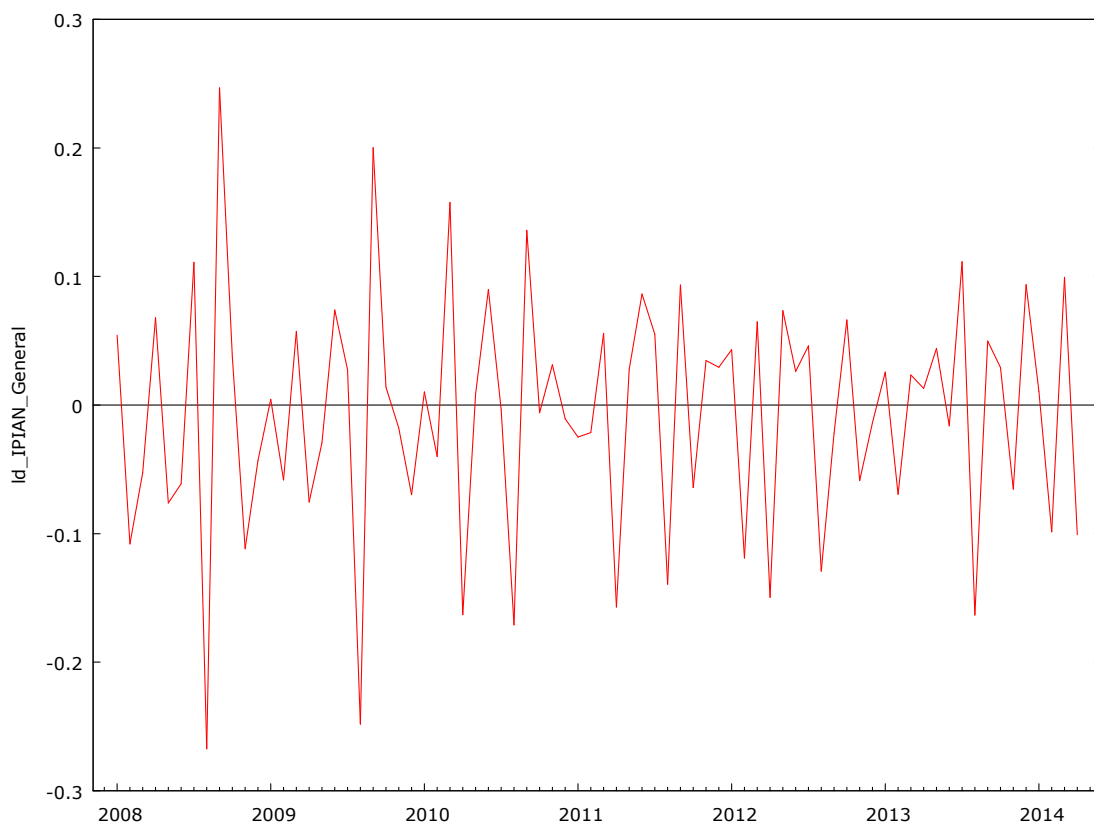
En la serie temporal que estamos analizando, el p-valor es menor que 0.05, lo que indica que hay evidencia para no rechazar H_0 , es decir, la varianza no es constante a medida que varía la media.

Para solucionar el problema de la falta de estacionariedad en media y en varianza, es necesario transformar los datos de la serie original Y_t . Lo más usual es tomar logaritmos y diferenciar regularmente la serie una vez (en ese orden), es decir, se genera una nueva serie Z_t de la forma:

$$Z_t = \ln Y_t - \ln Y_{t-1}$$

La representación gráfica de la nueva serie se muestra en la Figura 10.

Figura 10. Gráfico del IPI andaluz con base 2010 (01/2008 –04/2014). Serie Z_t



Fuente: Elaboración propia a través de *Gretl*

De nuevo se comprueba si los datos de la nueva serie proceden de un proceso estacionario. De manera subjetiva se observa que la serie no presenta tendencia, es decir, fluctúa en torno a una línea horizontal, por tanto no tenemos evidencia de que

la media no sea constante. Asimismo, no se observa forma de embudo ni de altavoz, por lo que la varianza también se considera constante.

Estas afirmaciones deben ser confirmadas de nuevo mediante el test aumentado de Dickey-Fuller y el contraste asociado al gráfico rango-media:

```
Contraste aumentado de Dickey-Fuller para ld_IPIAN_General
incluyendo un retardo de (1-L)ld_IPIAN_General
(el máximo fue 11, el criterio AIC modificado)
tamaño muestral 74
hipótesis nula de raíz unitaria: a = 1
```

```
contraste con constante
modelo: (1-L)y = b0 + (a-1)*y(-1) + ... + e
Coef. de autocorrelación de primer orden de e: -0.134
valor estimado de (a - 1): -2.18933
Estadístico de contraste: tau_c(1) = -11.8404
valor p asintótico 3.461e-025
```

```
con constante y tendencia
modelo: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
Coef. de autocorrelación de primer orden de e: -0.142
valor estimado de (a - 1): -2.21142
Estadístico de contraste: tau_ct(1) = -11.8997
valor p asintótico 5.882e-028
```

Puesto que el p-valor es inferior a 0.05, se rechaza H_0 al 5% de significación, por lo que la media es constante, coincidiendo con lo deducido subjetivamente a partir del gráfico.

El gráfico rango-media proporcionado por Gretl aparece en la Figura 11.

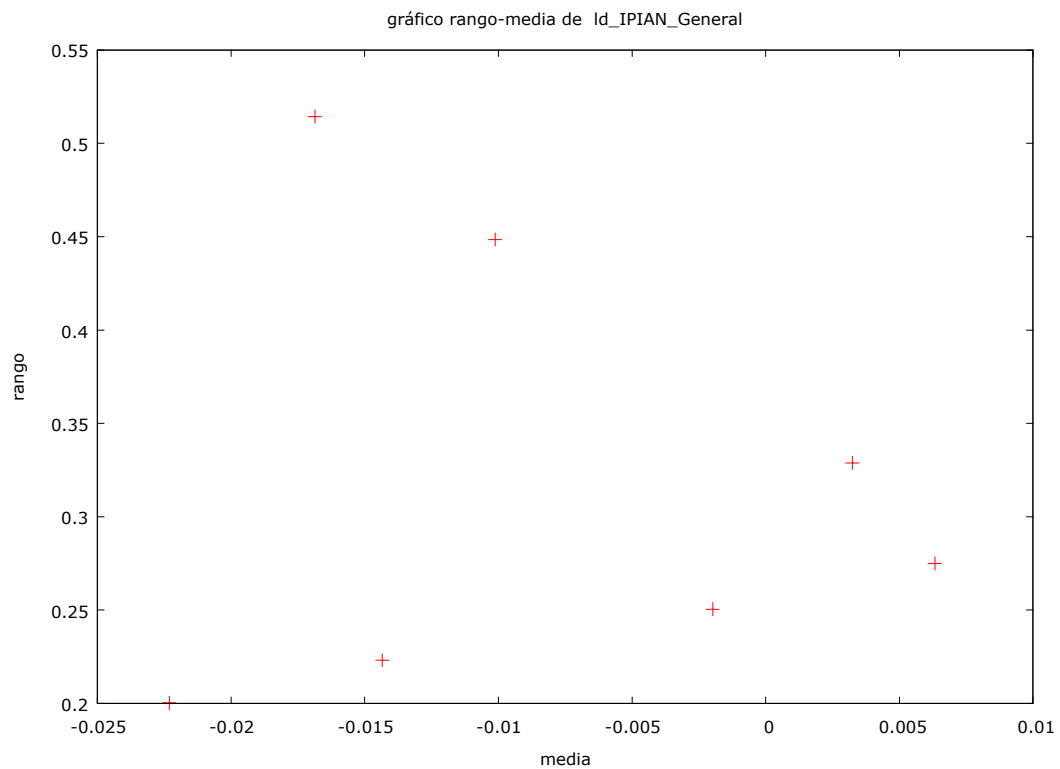
```
Estadísticos de rango-media para ld_IPIAN_General
utilizando 7 submuestras de tamaño 12
```

rango	media		
2008:01 - 2008:12	0.514339	-0.0168658	
2009:01 - 2009:12	0.448481	-0.0101255	
2010:01 - 2010:12	0.328810	0.00324230	
2011:01 - 2011:12	0.250470	-0.00198029	
2012:01 - 2012:12	0.223138	-0.0143428	
2013:01 - 2013:12	0.274984	0.00632926	
2014:01 - 2014:04	0.200305	-0.0223126	

```
pendiente de 'rango' con respecto a 'media' = -0.909412
el valor p para H0: pendiente = 0 es 0.860692
```

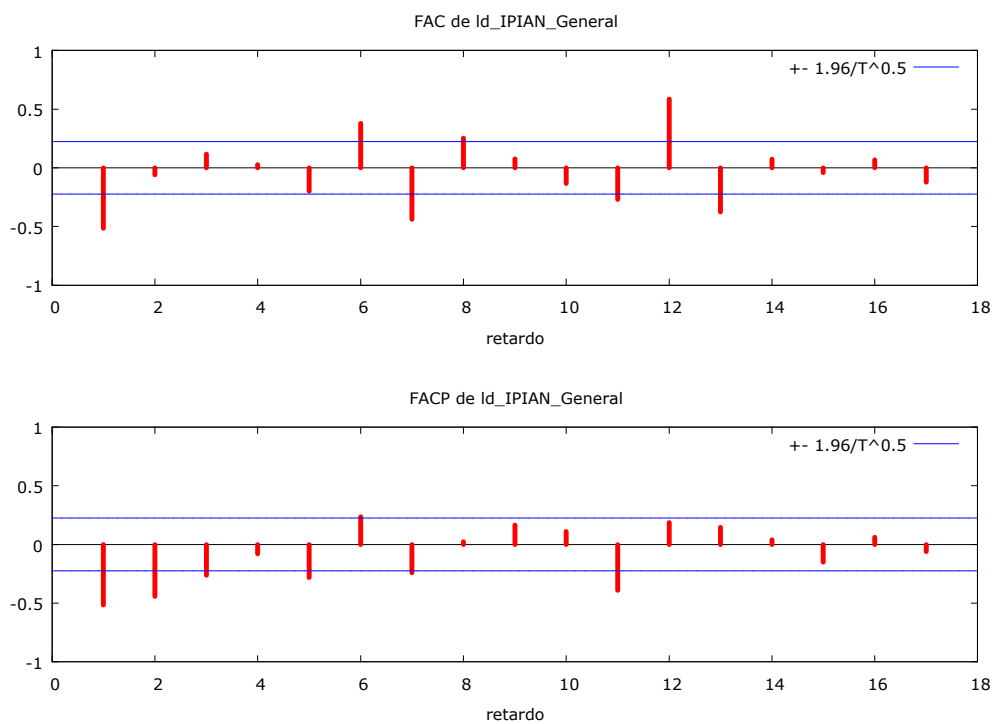
El p-valor correspondiente a este contraste indica que no hay evidencia para rechazar H_0 , con lo que se confirma que la varianza es constante.

Figura 11. Gráfico rango-media. Serie Z_t



Fuente: Elaboración propia a través de *Gretl*

Figura 12. Funciones de autocorrelación (FAC) y de autocorrelación parcial (FACP). Serie Z_t



Sin embargo, se sigue apreciando la componente estacional tanto en el gráfico de la serie transformada como en la FAC y la FACP muestrales asociados a dicha serie (Figuras 10 y 12).

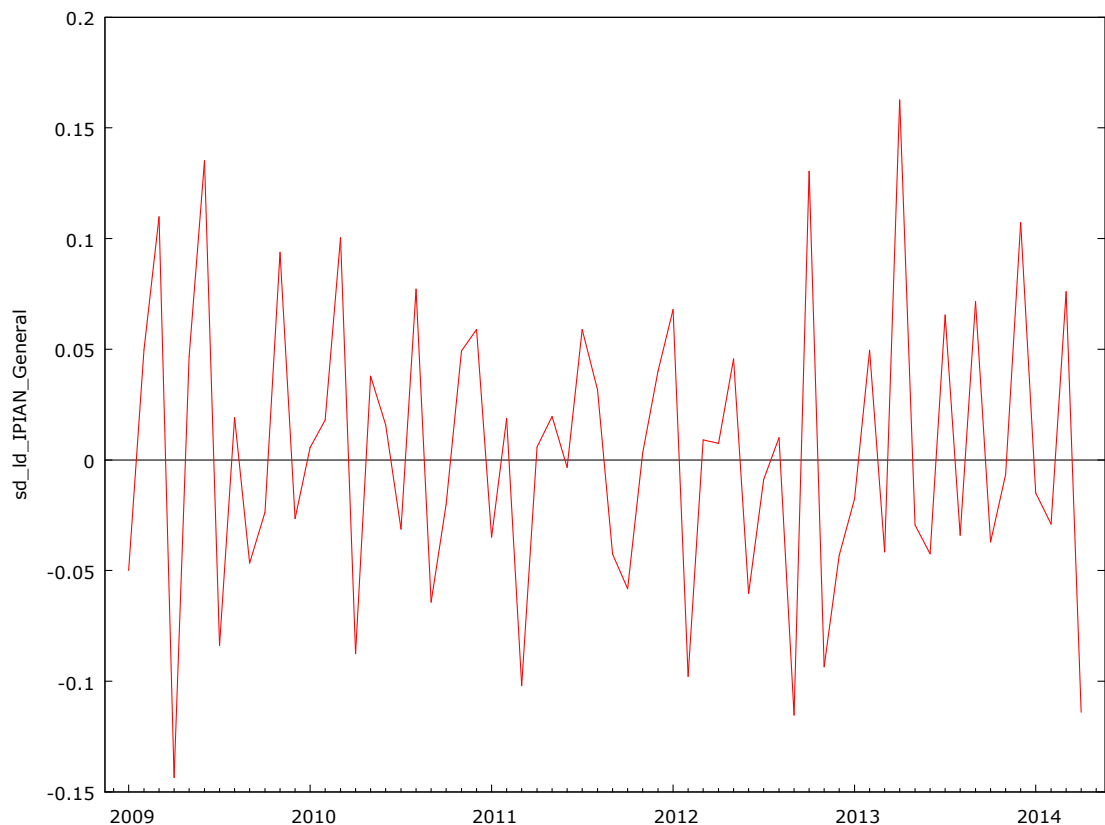
Para filtrar la componente estacional realizamos primeras diferencias estacionales de orden 12, generando una nueva serie que llamamos W_t definida como

$$W_t = Z_t - Z_{t-1} = \ln Y_t - \ln Y_{t-1} - \ln Y_{t-12} + \ln Y_{t-13}$$

En el gráfico de la nueva serie (Figura 13) ya no se aprecia la componente estacional.

Figura 13. Gráfico del IPI andaluz con base 2010 (enero 2008 - abril 2014).

Serie W_t



Fuente: Elaboración propia a través de *Gretl*

Al igual que el gráfico anterior (Figura 10), no presenta tendencias crecientes ni decrecientes deduciendo de este modo que la media es constante. En relación con la varianza, no se observa forma de embudo ni de altavoz, por tanto, la varianza es constante.

Una vez más, desarrollamos los contrastes para comprobar la estacionariedad en media y varianza:

Los resultados del contraste de Dickey – Fuller calculado que proporciona el programa *Gretlson*:

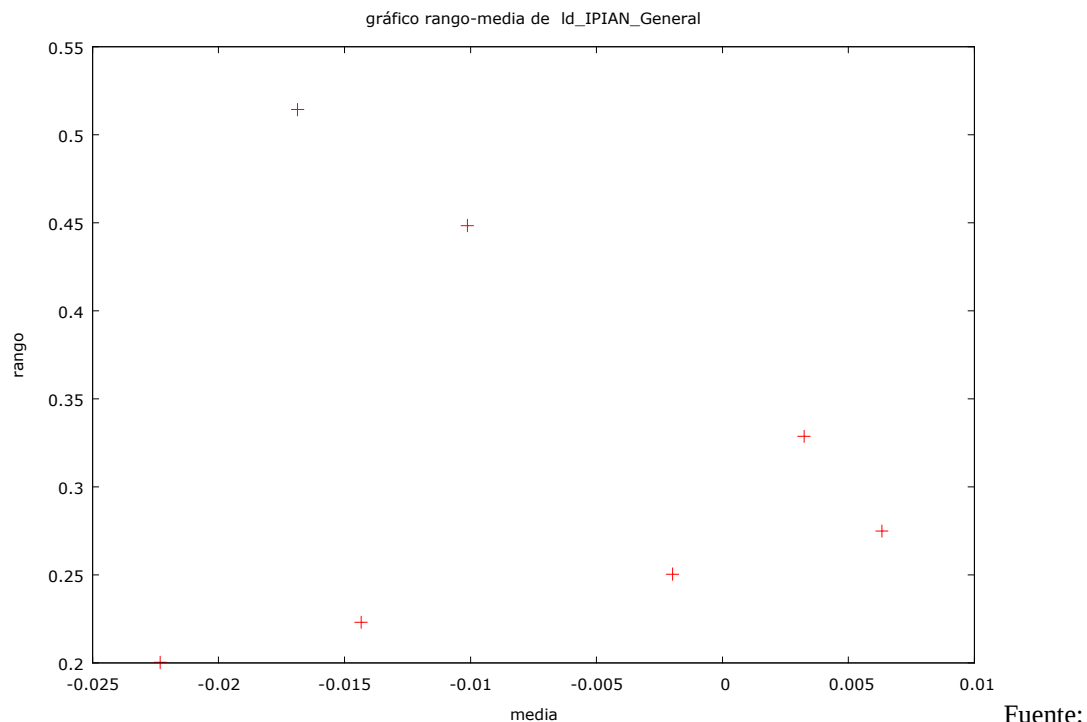
```
Contraste aumentado de Dickey-Fuller para sd_ld_IPIAN_General
incluyendo un retardo de (1-L)sd_ld_IPIAN_General
(el máximo fue 1, el criterio AIC modificado)
tamaño muestral 62
hipótesis nula de raíz unitaria:  $\alpha = 1$ 

contraste con constante
modelo:  $(1-L)y = b_0 + (\alpha-1)y(-1) + \dots + e$ 
Coef. de autocorrelación de primer orden de e: -0.019
valor estimado de  $(\alpha - 1)$ : -1.94169
Estadístico de contraste:  $\tau_{\alpha}(1) = -9.31284$ 
valor p asintótico 5.049e-017

con constante y tendencia
modelo:  $(1-L)y = b_0 + b_1t + (\alpha-1)y(-1) + \dots + e$ 
Coef. de autocorrelación de primer orden de e: -0.015
valor estimado de  $(\alpha - 1)$ : -1.94352
Estadístico de contraste:  $\tau_{\alpha}(1) = -9.25491$ 
valor p asintótico 6.858e-017
```

Puesto que el p-valor es inferior a 0.05, se rechaza H_0 al 5% de significación, por lo que la media es constante, coincidiendo con lo deducido subjetivamente a partir del gráfico.

Figura 14. Gráfico rango-media. Serie W_t



Elaboración propia a través de *Gretl*

El gráfico rango-media de la serie W_t aparece en la Figura 14 y los resultados analíticos proporcionados por Gretl son:

Estadísticos de rango-media para sd_ld_IPIAN_General
utilizando 6 submuestras de tamaño 12

rango	media		
2009:01 - 2009:12	0.278710	0.00674025	
2010:01 - 2010:12	0.187977	0.0133678	
2011:01 - 2011:12	0.160852	-0.00522260	
2012:01 - 2012:12	0.245659	-0.0123625	
2013:01 - 2013:12	0.204996	0.0206720	
2014:01 - 2014:04	0.189987	-0.0204309	

pendiente de 'rango' con respecto a 'media' = 0.229415
el valor p para H_0 : pendiente = 0 es 0.8743

El p-valor correspondiente a este contraste indica que no hay evidencia para rechazar H_0 , con lo que confirma que la varianza es constante.

En resumen, la serie es estacionaria tanto en media como en varianza, por tanto, pasamos a determinar el orden de la parte autorregresiva y de la de medias móviles del proceso ARMA que se considere que haya podido generar la serie estacionaria, tanto en su parte regular como en su parte estacional.

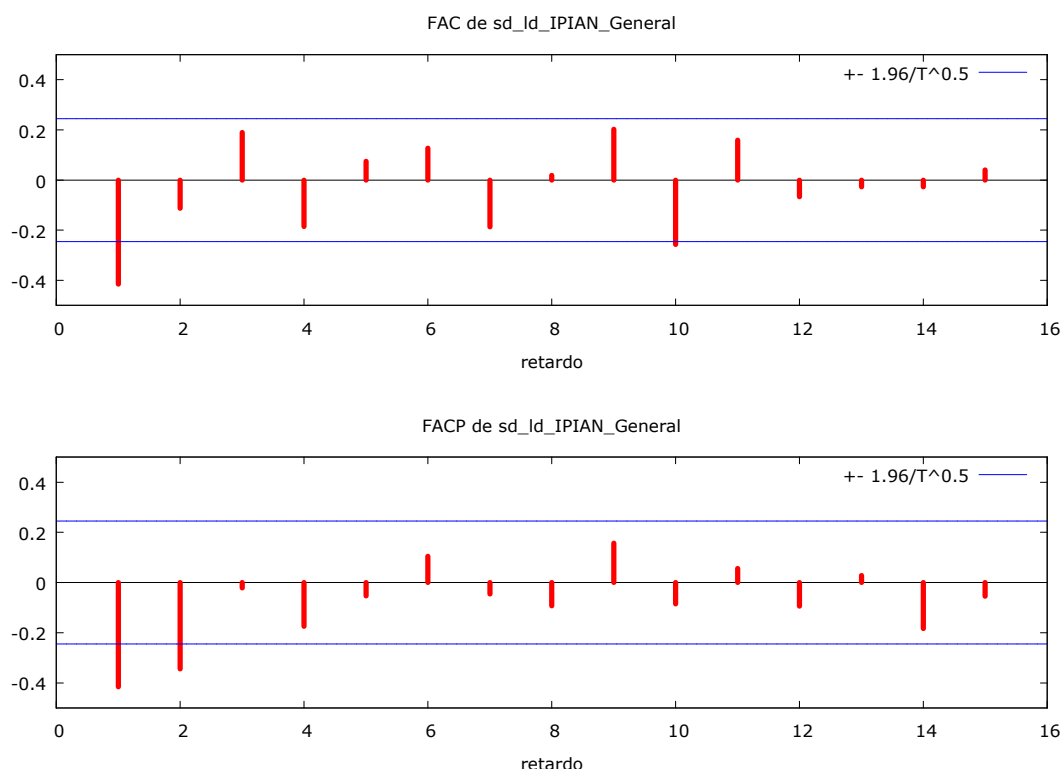
Para ello, se utilizarán las características de la FAC y de la FACP estimadas (Figura 15).

Se distingue que los dos primeros retardos de la FAC y de la FACP estimadas son significativos al 1%, 5% y 10%, al igual que lo es el segundo retardo de la FACP estimada. Gráficamente también puede apreciarse, ya que los picos correspondientes a estos retardos superan las bandas de confianza al 5% de significación.

Esta información permite identificar los siguientes candidatos para la parte regular: un modelo $MA(1)$ y un modelo $AR(2)$, mientras que para la parte estacional se propone un proceso de ruido blanco (PRB). En consecuencia, los candidatos son:

1. $ARIMA(0,1,1) \times ARIMA(0,1,0)_{12}$
2. $ARIMA(2,1,0) \times ARIMA(0,1,0)_{12}$
3. $ARIMA(2,1,1) \times ARIMA(0,1,0)_{12}$

Figura 15. Funciones de autocorrelación (FAC) y de autocorrelación parcial (FACP). Serie W_t



Fuente: Elaboración propia a través de *Gretl*

4.2. ESTIMACIÓN

4.2.1. MODELO 1

Una vez identificado el modelo, se procede a la estimación de sus parámetros. Para ello se asume que los errores tienen distribución normal y que el proceso es estacionario e invertible. El método de estimación utilizado es máxima verosimilitud exacta.

Especificando al programa *Gretl* los órdenes del modelo más simple, es decir, $ARIMA(0,1,1) \times ARIMA(0,1,0)_{12}$, se obtienen los siguientes resultados:

Modelo 1: ARIMA, usando las observaciones 2009:02-2014:04 (T = 63)				
Variable dependiente: (1-L)(1-Ls) l_IPIAN_General				
Desviaciones típicas basadas en la matriz de productos externos				
	Coefficiente	Desv. Típica	Z	Valor p
Const	0.00393215	0.00287341	1.3685	0.17117
theta_1	-0.603974	0.10281	-5.8747	<0.00001 ***
Media de la vble. dep.	0.003913	D.T. de la vble. dep.	0.066377	
media innovaciones	0.001311	D.T. innovaciones	0.056223	
Log-verosimilitud	91.72151	Criterio de Akaike	-177.4430	

Criterio de Schwarz -171.0136 Crit. de Hannan-Quinn -174.9143

	<i>Real</i>	<i>Imaginaria</i>	<i>Módulo</i>	<i>Frecuencia</i>
MA				
Raíz 1	1.6557	0.0000	1.6557	0.0000

Así, la expresión del modelo estimado sería:

$$\widehat{W}_t = 0.00393215 - 0.603974\hat{\varepsilon}_{t-1}$$

Teniendo en cuenta la relación entre W_t e Y_t y deshaciendo los cambios, se tiene la siguiente expresión del modelo estimado:

$$\hat{Y}_t = \exp(0.00393215 - 0.603974\hat{\varepsilon}_{t-1} + \ln\widehat{Y}_{t-1} + \ln\widehat{Y}_{t-12} - \ln\widehat{Y}_{t-13})$$

4.2.2. MODELO 2

Los resultados obtenidos mediante Gretl para el segundo candidato son:

Modelo 2: ARIMA, usando las observaciones 2009:02-2014:04 (T = 63)
Variable dependiente: (1-L)(1-Ls) l_IPIAN_General

Desviaciones típicas basadas en la matriz de productos externos

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>Z</i>	<i>Valor p</i>	
Const	0.00394396	0.00366167	1.0771	0.28144	
phi_1	-0.583841	0.120513	-4.8446	<0.00001	***
phi_2	-0.359959	0.121432	-2.9643	0.00303	***

Media de la vble. dep.	0.003913	D.T. de la vble. dep.	0.066377
media innovaciones	0.001181	D.T. innovaciones	0.055929
Log-verosimilitud	92.03764	Criterio de Akaike	-176.0753
Criterio de Schwarz	-167.5027	Crit. de Hannan-Quinn	-172.7037

	<i>Real</i>	<i>Imaginaria</i>	<i>Módulo</i>	<i>Frecuencia</i>
AR				
Raíz 1	-0.8110	-1.4562	1.6668	-0.3309
Raíz 2	-0.8110	1.4562	1.6668	0.3309

Por tanto, la expresión del modelo estimado sería:

$$\widehat{W}_t = 0.00766627 - 0.583841\widehat{W}_{t-1} - 0.359959\widehat{W}_{t-2}$$

Teniendo en cuenta la relación entre W_t e Y_t y deshaciendo los cambios, se tiene la siguiente expresión del modelo estimado::

$$\begin{aligned} \hat{Y}_t = \exp(& 0.0076627 + 0.416159\ln\widehat{Y}_{t-1} - 0.223882\ln\widehat{Y}_{t-2} + 0.359959\ln\widehat{Y}_{t-3} \\ & + \ln\widehat{Y}_{t-12} - 0.416159\ln\widehat{Y}_{t-13} - 0.223882\ln\widehat{Y}_{t-14} \\ & - 0.359959\ln\widehat{Y}_{t-15}) \end{aligned}$$

4.2.3. MODELO 3

Los resultados obtenidos por *Gretl* para el modelo más complejo, es decir, $ARIMA(2,1,1) \times ARIMA(0,1,0)_{12}$ son:

Modelo 3: ARIMA, usando las observaciones 2009:02-2014:04 (T = 63)				
Variable dependiente: (1-L)(1-Ls) l_IPIAN_General				
Desviaciones típicas basadas en la matriz de productos externos				
	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>Z</i>	<i>Valor p</i>
Const	0.00397621	0.00314469	1.2644	0.20608
phi_1	-0.271941	0.316728	-0.8586	0.39056
phi_2	-0.210502	0.202034	-1.0419	0.29745
theta_1	-0.347273	0.317154	-1.0950	0.27353
Media de la vble. dep.	0.003913	D.T. de la vble. dep.	0.066377	
media innovaciones	0.001249	D.T. innovaciones	0.055811	
Log-verosimilitud	92.16225	Criterio de Akaike	-174.3245	
Criterio de Schwarz	-163.6088	Crit. de Hannan-Quinn	-170.1100	
	<i>Real</i>	<i>Imaginaria</i>	<i>Módulo</i>	<i>Frecuencia</i>
AR				
Raíz 1	-0.6459	-2.0817	2.1796	-0.2979
Raíz 2	-0.6459	2.0817	2.1796	0.2979
MA				
Raíz 1	2.8796	0.0000	2.8796	0.0000

Así, la expresión del modelo estimado sería:

$$\widehat{W}_t = 0.0058945 - 0.271941\widehat{W}_{t-1} - 0.210502\widehat{W}_{t-2} - 0.347273\hat{\varepsilon}_{t-1}$$

De nuevo, teniendo en cuenta la relación entre W_t e Y_t y deshaciendo los cambios, se tiene la siguiente expresión del modelo estimado:

$$\begin{aligned}\widehat{Y}_t = \exp(&0.00589450468 + 0.728059\widehat{\ln Y}_{t-1} - 0.061439\widehat{\ln Y}_{t-2} \\ &+ 0.210502\widehat{\ln Y}_{t-3} + \widehat{\ln Y}_{t-12} - 0.728059\widehat{\ln Y}_{t-13} \\ &- 0.061439\widehat{\ln Y}_{t-14} - 0.210502\widehat{\ln Y}_{t-15} - 0.347273\hat{\varepsilon}_{t-1})\end{aligned}$$

4.3. DIAGNOSIS

4.3.1. MODELO 1

- Análisis de los coeficientes

En el estudio de la significación de los coeficientes del modelo 1 se aprecia que la constante no es significativa. Por tanto, debemos de modificar el modelo

eliminando la constante, siendo el modelo 1 obtenido el que se detalla a continuación:

Modelo 1: ARIMA, usando las observaciones 2009:02-2014:04 (T = 63)					
Variable dependiente: (1-L)(1-Ls) l_IPIAN_General					
Desviaciones típicas basadas en la matriz de productos externos					
	Coeficiente	Desv. Típica	Z	Valor p	
theta_1	-0.571362	0.105618	-5.4097	<0.00001	***
Media de la vble. dep.	0.003913	D.T. de la vble. dep.		0.066377	
media innovaciones	0.010205	D.T. innovaciones		0.057022	
Log-verosimilitud	90.86090	Criterio de Akaike		-177.7218	
Criterio de Schwarz	-173.4355	Crit. de Hannan-Quinn		-176.0360	
	Real	Imaginaria	Módulo	Frecuencia	
MA					
Raíz 1	1.7502	0.0000	1.7502	0.0000	

Ahora, la expresión del modelo estimado es:

$$\hat{Y}_t = \exp(-0.571362\hat{\varepsilon}_{t-1} + \widehat{\ln Y}_{t-1} + \widehat{\ln Y}_{t-12} - \widehat{\ln Y}_{t-13})$$

Al analizar las propiedades de estacionariedad e invertibilidad del modelo 1 comprobamos que se cumplen, ya que la pendiente es menor que 1 en valor absoluto o, equivalentemente, la raíz del correspondiente polinomio en retardos es mayor que 1 en valor absoluto (1.7502).

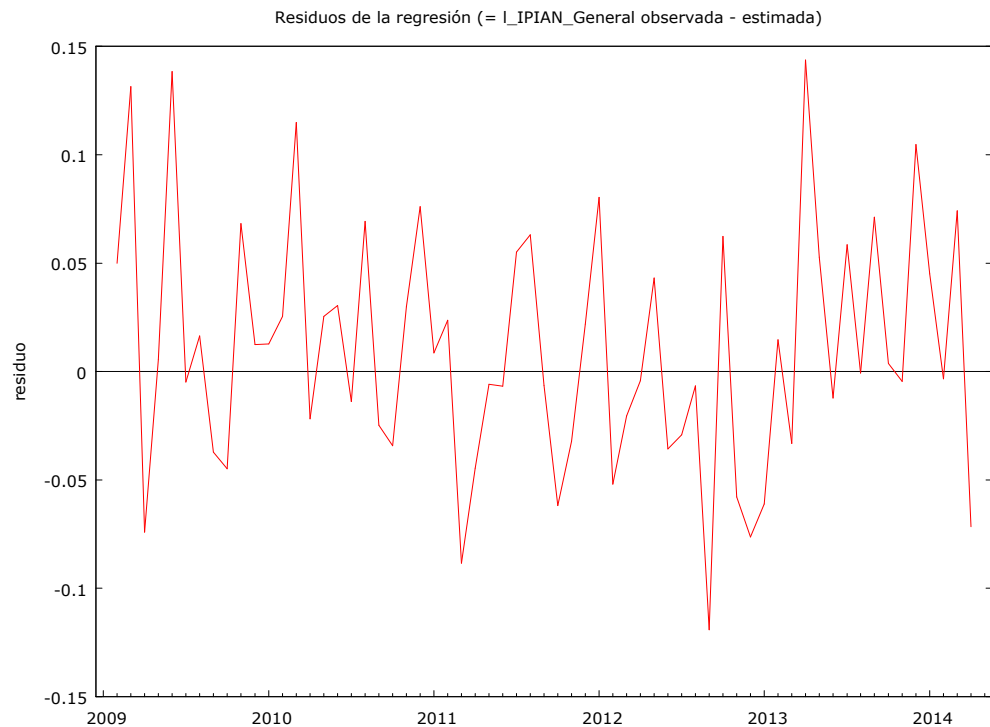
- Análisis de los residuos

La última parte del diagnóstico del modelo se lleva a cabo mediante el análisis de los residuos del modelo estimado, donde se comprueba si estos proceden de un PRB. Para ello, realizaremos tres comprobaciones:

❖ *Comprobación de aleatoriedad, media cero y varianza constante a través del gráfico de los residuos.*

La Figura 16 confirma que los residuos no presentan tendencia y que, por tanto, la media es constante y está situada en torno al cero. Del mismo modo, la varianza también se considera constante al presentar una amplitud de las oscilaciones de manera aleatoria. Por tanto, no se incumple la hipótesis de homocedasticidad.

Figura 16. Gráfico de residuos del modelo 1



Fuente: Elaboración propia a través de *Gretl*

❖ *Comprobación de ausencia de autocorrelación mediante la FAC y FACP estimadas a partir de los residuos y el contraste de Ljung-Box.*

El Test de Ljung-Box indica si existe dependencia entre los m primeros residuos estimados, es decir, si estos residuos presentan correlación. El contraste de hipótesis correspondiente es el siguiente:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$$

$$H_1: \rho_k \neq 0, \text{ para algún } k = 1, \dots, m$$

En la última fila de los resultados proporcionados por *Gretl* aparecen los p-valores asociados a dicho contraste para $m = 1, \dots, 15$:

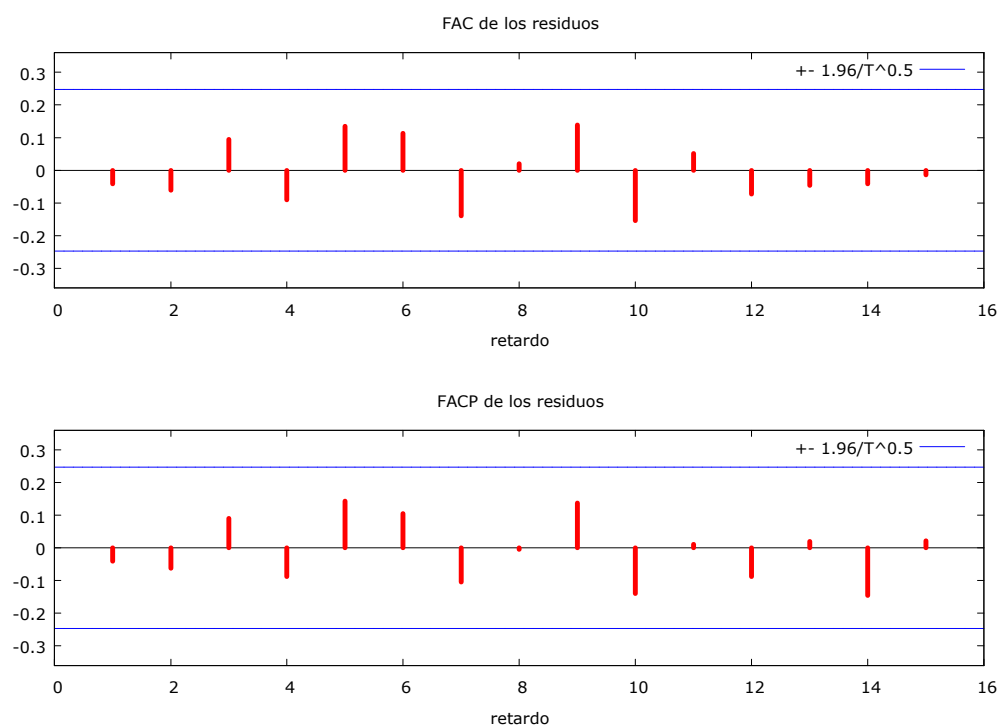
Función de autocorrelación de los residuos

RETARDO	FAC	FACP	Estad-Q.	[valor p]
1	-0.0413	-0.0413	0.1127	[0.737]
2	-0.0609	-0.0627	0.3615	[0.835]
3	0.0945	0.0898	0.9712	[0.808]
4	-0.0904	-0.0879	1.5389	[0.820]
5	0.1353	0.1432	2.8305	[0.726]
6	0.1134	0.1051	3.7549	[0.710]

7	-0.1399	-0.1044	5.1855	[0.637]
8	0.0205	-0.0046	5.2168	[0.734]
9	0.1390	0.1373	6.6817	[0.670]
10	-0.1544	-0.1400	8.5230	[0.578]
11	0.0512	0.0112	8.7293	[0.647]
12	-0.0730	-0.0881	9.1566	[0.690]
13	-0.0466	0.0196	9.3346	[0.747]
14	-0.0414	-0.1457	9.4777	[0.799]
15	-0.0144	0.0214	9.4955	[0.850]

En este caso, todos los p-valores asociados son mayores que 0.05 por lo que no puede rechazarse H_0 al 5% de significación. Esto implica que no hay evidencia de que se incumpla la hipótesis de ausencia de autocorrelación de los errores de manera conjunta y que todos los coeficientes de autocorrelación son nulos lo que se puede apreciar en el gráfico de la FAC y de la FACP estimadas de los residuos (Figura 17) donde ningún pico supera las bandas de confianza.

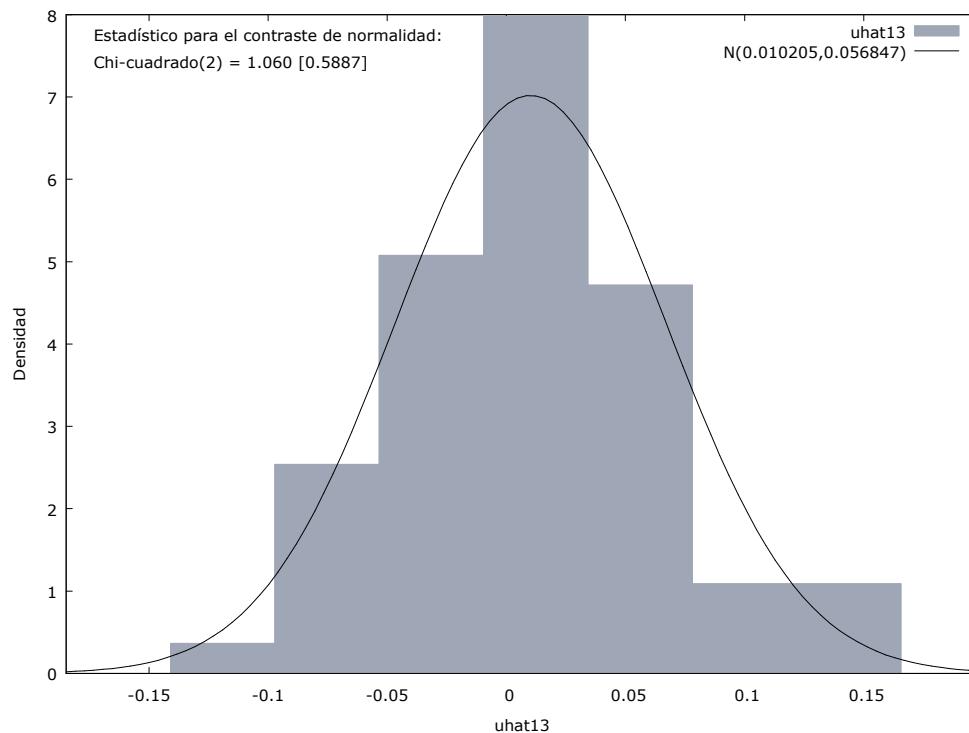
Figura 17. FAC y FACP estimadas a partir de los residuos del Modelo 1



Fuente: Elaboración propia a través de *Gretl*

❖ *Comprobación de la normalidad de los residuos mediante un contraste de normalidad:*

Figura 18. Gráfico normalidad de los residuos del Modelo 1



Fuente: Elaboración propia a través de *Gretl*

El contraste de normalidad se plantea como:

H_0 : los residuos proceden de una v. a. con distribución normal

H_1 : los residuos no proceden de una v. a. con distribución normal

El p-valor asociado al contraste es 0.5887, mayor que cualquier nivel de significación considerado, con lo que no se rechaza H_0 , es decir, los residuos proceden de un variable con distribución normal.

En resumen, el modelo verifica todos los controles de diagnóstico y puede emplearse para obtener predicciones de valores futuros de la serie.

4.3.2. MODELO 2

- Análisis de los coeficientes:

Al igual que en el modelo 1, la constante no es significativa. Sin embargo, el resto de parámetros del modelo sí lo son al 1%, 5% y 10% de significación. En este caso, debemos de modificar el modelo eliminando la constante, siendo el modelo 2 obtenido el que se detalla a continuación:

Modelo 2: ARIMA, usando las observaciones 2009:02-2014:04 (T = 63)

Variable dependiente: (1-L)(1-Ls) l_IPIAN_General

Desviaciones típicas basadas en la matriz de productos externos

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>Z</i>	<i>Valor p</i>	
phi_1	-0.570425	0.120732	-4.7247	<0.00001	***
phi_2	-0.349979	0.121771	-2.8741	0.00405	***
Media de la vble. dep.	0.003913	D.T. de la vble. dep.	0.066377		
media innovaciones	0.008640	D.T. innovaciones	0.056445		
Log-verosimilitud	91.47036	Criterio de Akaike	-176.9407		
Criterio de Schwarz	-170.5113	Crit. de Hannan-Quinn	-174.4120		

	<i>Real</i>	<i>Imaginaria</i>	<i>Módulo</i>	<i>Frecuencia</i>
AR				
Raíz 1	-0.8149	-1.4809	1.6904	-0.3301
Raíz 2	-0.8149	1.4809	1.6904	0.3301

En consecuencia, la expresión del modelo estimado sería:

$$\hat{Y}_t = \exp(0.429575\hat{\ln Y}_{t-1} - 0.2204464\hat{\ln Y}_{t-2} + 0.349979\hat{\ln Y}_{t-3} + \hat{\ln Y}_{t-12} - 0.429575\hat{\ln Y}_{t-13} - 0.220446\hat{\ln Y}_{t-14} - 0.3499794\hat{\ln Y}_{t-15})$$

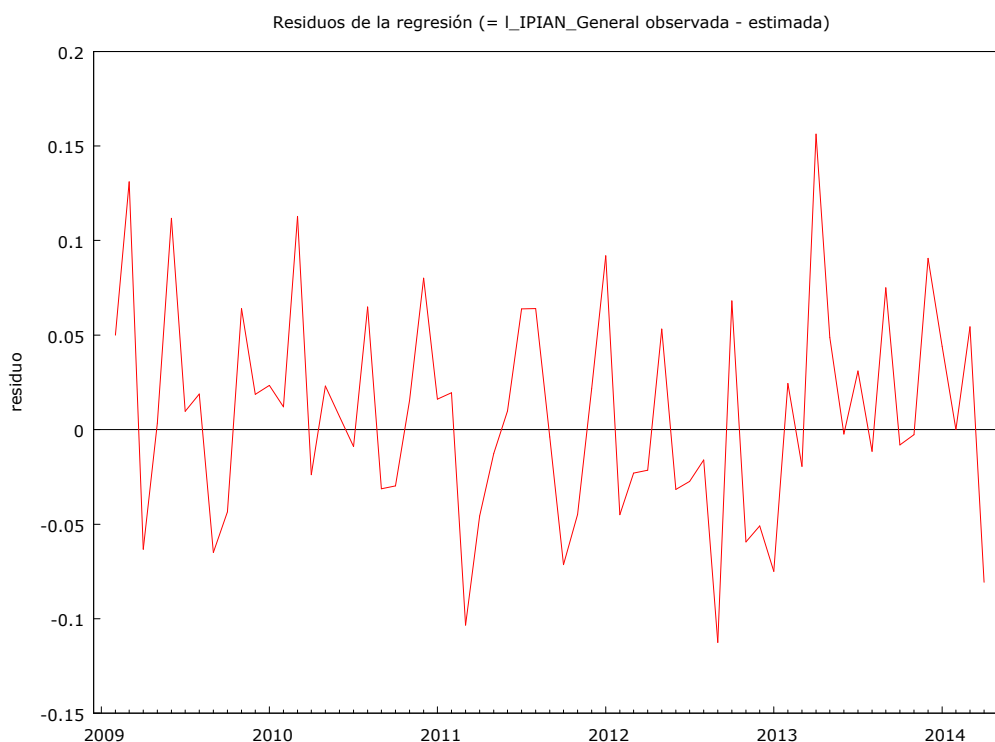
En segundo lugar, se analiza si el modelo estimado verifica las propiedades de estacionariedad e invertibilidad. Estas características se cumplen al ser las dos raíces del polinomio de retardos mayores que 1 en módulo (1.6904).

- Análisis de los residuos:

- ❖ *Comprobación de aleatoriedad, media cero y varianza constante a través del gráfico de los residuos:*

El gráfico de la Figura 19 confirma que los residuos proceden de una variable aleatoria, no presentan tendencia y fluctúan en torno al 0. Por tanto, la media es constante y vale cero. Del mismo modo, la varianza también se considera constante al no observarse forma de embudo ni de altavoz. Por tanto, no se incumple la hipótesis de homocedasticidad.

Figura 19. Gráfico de residuos del modelo 2



Fuente: Elaboración propia a través de *Gretl*

❖ *Comprobación de ausencia de autocorrelación mediante la FAC y la FACP estimadas a partir de los residuos y el contraste de Ljung-Box.*

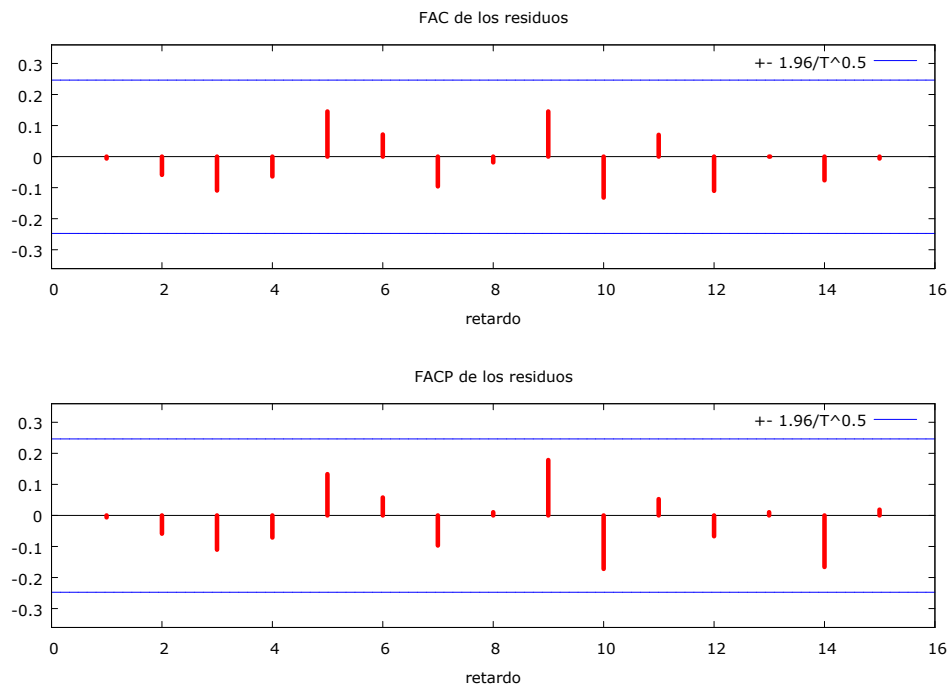
Función de autocorrelación de los residuos

RETARDO	FAC	FACP	Estad-Q. [valor p]	
1	-0.0057	-0.0057	0.0021	[0.963]
2	-0.0586	-0.0586	0.2324	[0.890]
3	-0.1087	-0.1098	1.0389	[0.792]
4	-0.0639	-0.0704	1.3223	[0.858]
5	0.1456	0.1332	2.8183	[0.728]
6	0.0719	0.0582	3.1897	[0.785]
7	-0.0957	-0.0969	3.8600	[0.796]
8	-0.0177	0.0109	3.8835	[0.867]
9	0.1454	0.1781	5.4871	[0.790]
10	-0.1313	-0.1715	6.8185	[0.742]
11	0.0709	0.0531	7.2143	[0.781]
12	-0.1103	-0.0667	8.1918	[0.770]
13	0.0001	0.0101	8.1918	[0.831]
14	-0.0754	-0.1654	8.6665	[0.852]
15	-0.0056	0.0186	8.6692	[0.894]

En todos los casos, el p-valor asociado es mayor que 0.05 por lo que no puede rechazarse H_0 al 5% de significación. Esto implica que no hay evidencia de que se incumpla la hipótesis de ausencia de autocorrelación de

los errores de manera conjunta y que todos los coeficientes de autocorrelación son nulos lo que se puede apreciar en el gráfico de la FAC y de la FACP estimadas de los residuos donde ningún pico supera las bandas de confianza.

Figura 20. FAC y FACP estimadas a partir de los residuos del Modelo 2



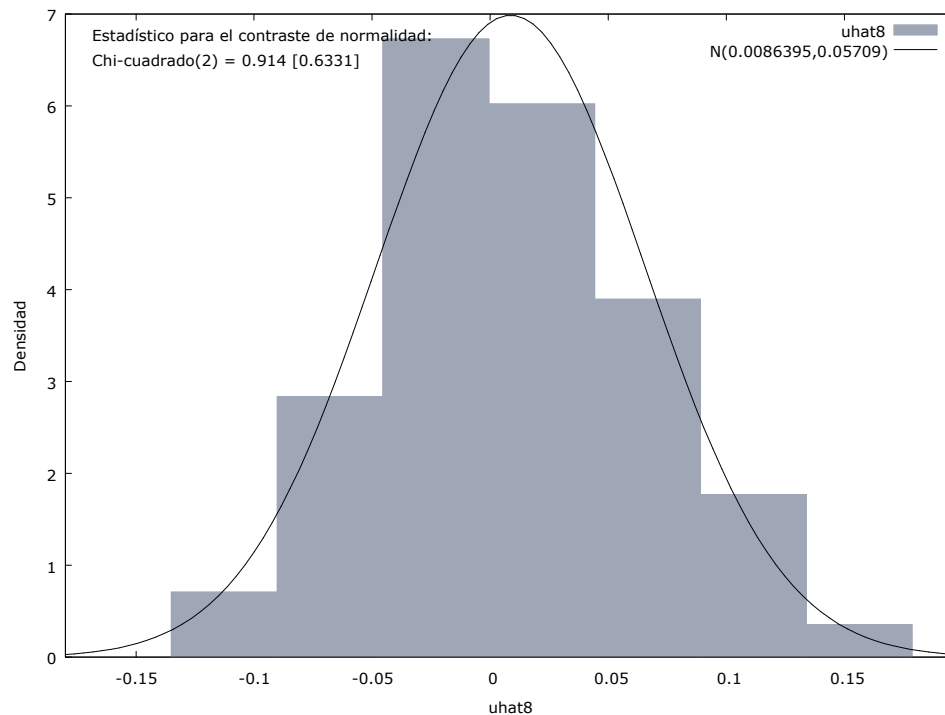
Fuente: Elaboración propia a través de *Gretl*

❖ *Comprobación de la normalidad de los residuos mediante un contraste de normalidad:*

Tal y como se observa en la Figura 21, el p-valor asociado al contraste es 0.6331, mayor que cualquier nivel de significación considerado, con lo que no se rechaza H_0 , es decir, los residuos proceden de una variable aleatoria con distribución normal.

En resumen, el modelo verifica todos los controles de diagnóstico y puede emplearse para obtener predicciones de valores futuros de la serie.

Figura 21. Gráfico de normalidad de los residuos del Modelo 1



Fuente: Elaboración propia a través de *Gretl*

4.3.3. MODELO 3

- Análisis de los coeficientes:

En los resultados obtenidos anteriormente se aprecia que la constante y las pendientes no son significativas ni al 1%, ni al 5%, ni al 10% de significación. Por tanto, con este análisis es suficiente para desechar el modelo.

4.4. PREDICCIÓN

Una vez aplicadas las tres primeras fases de la metodología Box-Jenkins, podemos concluir que los modelos 1 y 2 son válidos para realizar predicciones futuras de la serie temporal del IPIAN. Sin embargo, para decidir cuál de los dos modelos es preferible acudimos a los criterios de selección de modelos que *Gretl* incluye en la propia ventana de cada uno de estos modelos (Modelo 1 y Modelo 2): Criterio de Akaike, criterio de Schwarz y el criterio de Hannan-Quinn. Seleccionaremos aquel modelo que presente menor valor de estos criterios.

CRITERIOS	MODELO 1	MODELO 2
Criterio de Akaike	-177.7218	-176.9407
Criterio de Schwarz	-173.4355	-170.5113
Criterio de Hannan-Quinn	-176.0360	-174.4120

Los tres criterios anteriores confirman que el modelo 1 es preferible al modelo 2, por tanto, el modelo estimado adecuado para la predicción futura de valores de la serie del IPI andaluz es:

$$ARIMA(0,1,1) \times ARIMA(0,1,0)_{12}$$

cuya ecuación estimada es:

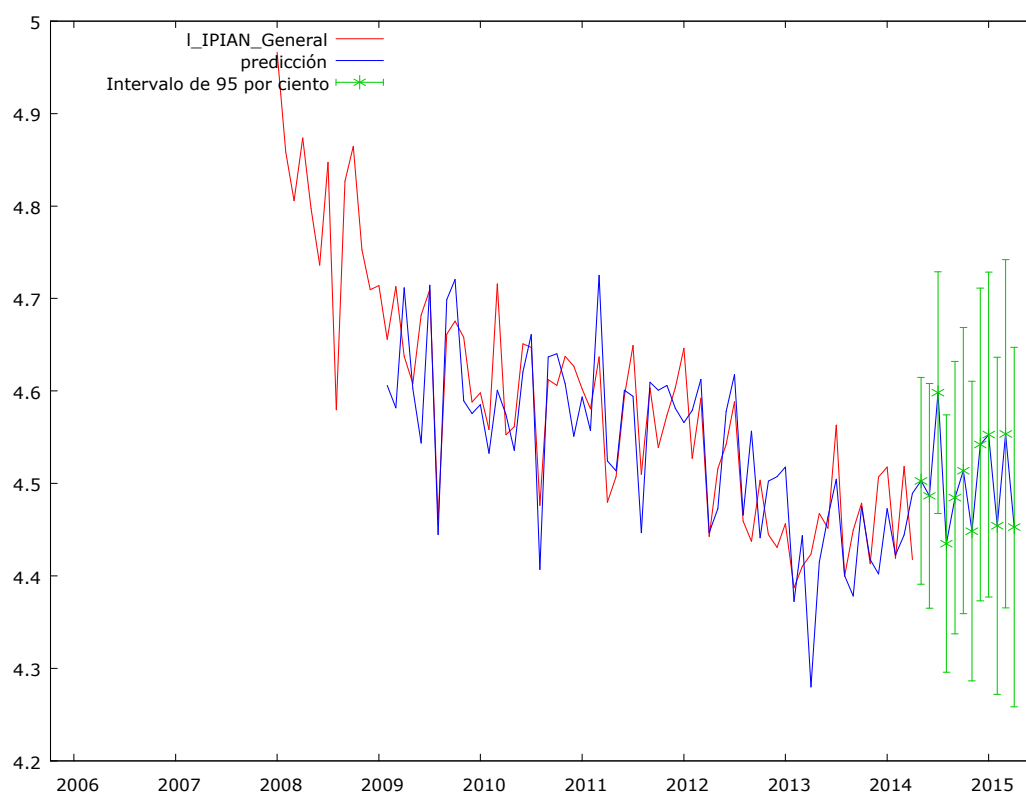
$$\hat{Y}_t = \exp(-0.571362\hat{\varepsilon}_{t-1} + \ln\hat{Y}_{t-1} + \ln\hat{Y}_{t-12} - \ln\hat{Y}_{t-13})$$

Elegido el modelo preferido entre los modelos candidatos, comenzamos a realizar las predicciones de la serie del IPI andaluz. Con el apoyo de *Gretl*, llevamos a cabo un análisis de predicción automática, con un intervalo de confianza del 95%, y una duración de un año posterior al último valor obtenido. La representación gráfica puede contemplarse en la Figura 22 y los resultados analíticos son los siguientes:

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

Observaciones	Predicción	Desv. Típica	Intervalo de 95%
2014:05	4.50284	0.0570224	(4.39108, 4.61460)
2014:06	4.48665	0.0620400	(4.36505, 4.60825)
2014:07	4.59830	0.0666811	(4.46761, 4.72899)
2014:08	4.43496	0.0710196	(4.29577, 4.57416)
2014:09	4.48466	0.0751079	(4.33746, 4.63187)
2014:10	4.51391	0.0789848	(4.35910, 4.66872)
2014:11	4.44850	0.0826802	(4.28645, 4.61055)
2014:12	4.54210	0.0862173	(4.37312, 4.71108)
2015:01	4.55285	0.0896149	(4.37721, 4.72850)
2015:02	4.45430	0.0928883	(4.27224, 4.63636)
2015:03	4.55373	0.0960502	(4.36547, 4.74198)
2015:04	4.45285	0.0991113	(4.25860, 4.64711)

Figura 22. Gráfico de la predicción del IPI andaluz, expresado en logaritmos

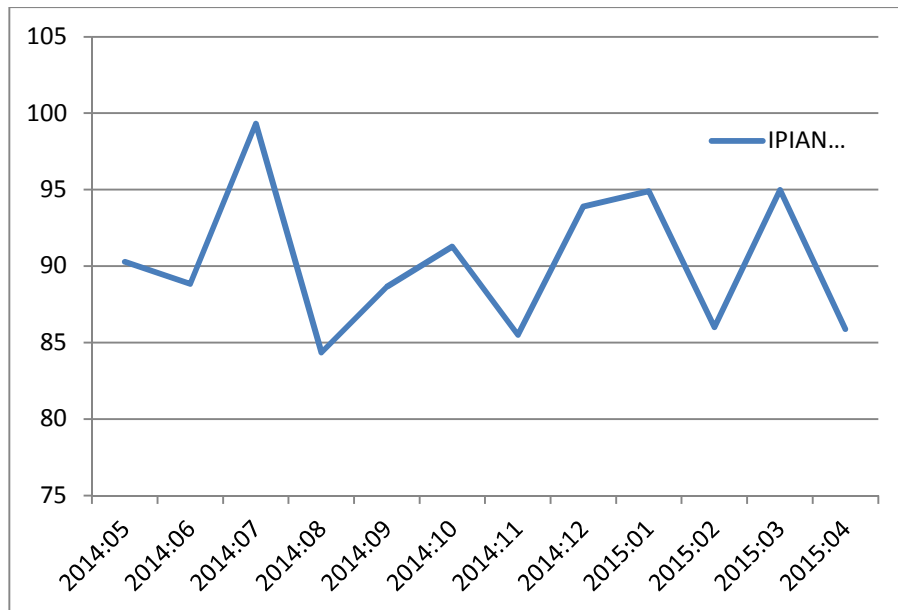


Fuente: Elaboración propia a través de *Gretl*

Sin embargo, debido a la transformación de la serie, estas cifras son logaritmos de los valores reales, es decir, es necesario modificar los datos realizando la operación inversa. Así, las predicciones obtenidas son las representadas gráficamente en la Figura 23 y mostradas numéricamente en la siguiente tabla:

Observaciones	IPI Pronosticado
2014:05	90,27314332
2014:06	88,82338856
2014:07	99,31533598
2014:08	84,34875087
2014:09	88,64680577
2014:10	91,27801874
2014:11	85,49859987
2014:12	93,88775753
2015:01	94,90249537
2015:02	85,99593262
2015:03	94,98604632
2015:04	85,87132888

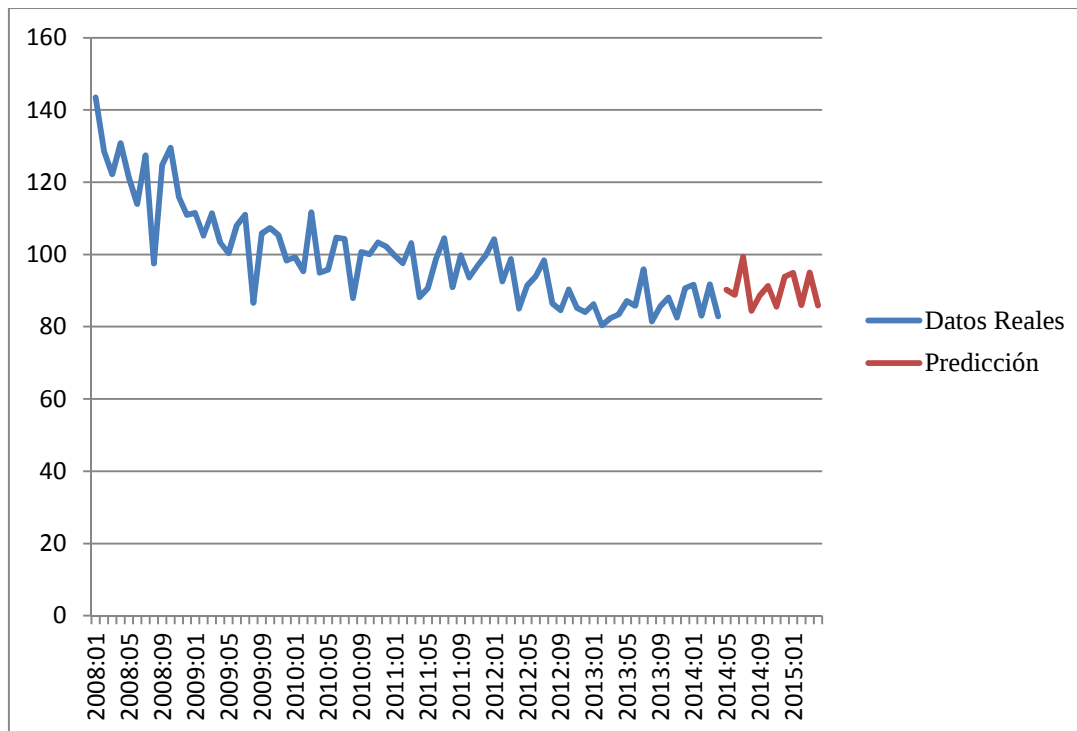
Figura 23. Gráfico de la predicción del IPI andaluz (mayo de 2014 - abril de 2015)



Fuente: Elaboración propia a través de *Microsoft Excel 2010*

La representación gráfica tanto de los valores originales de la serie como de los predichos por Gretl se puede observar en la Figura 24.

Figura 24. Gráfico de la predicción del IPI andaluz (enero de 2008 - abril de 2015).



Fuente: Elaboración propia a través de *Microsoft Excel 2010*

5. CONCLUSIONES

Como cumplimiento del estudio econométrico presentado, haré unos comentarios sobre los resultados obtenidos tras el análisis de la serie temporal, así como una pequeña opinión personal respecto a la elaboración del mismo.

La modelización de la serie del IPI andaluz, y más en concreto la última fase de la metodología Box-Jenkins, la predicción, da pie a unas observaciones dignas de ser comentadas y analizadas como concluyentes de este trabajo fin de grado. Por un lado, al observar la evolución de la serie gráficamente, es de esperar que el componente estacional que lo caracteriza vuelva a reflejarse en los próximos meses con una bajada del índice durante el mes de agosto de 2014. Sin embargo, sin un estudio más allá de la propia imaginación y deducción, existe una gran incertidumbre y riesgo al tomar decisiones que afecten al futuro próximo, lo que implica que los resultados obtenidos a través de la modelización de la serie proporcionen cierta seguridad para la ayuda a la toma de decisiones. En este sentido, los resultados obtenidos en la serie analizada durante la elaboración del trabajo, prevén una evolución semejante a la obtenida hasta el momento e incluso puede apreciarse un ínfimo crecimiento del índice, y por consiguiente, en la economía española. Este incremento, al considerarse único durante los últimos años de crisis podía ser considerado por aquellos analistas económicos más optimistas, junto con otros factores que deben ser analizados, como una señal de prosperidad en la economía española. No obstante, no hay que olvidar que todo futuro es incierto y que, por lo tanto, todas las decisiones futuras deben ser tomadas con cautela.

Finalmente, he de comentar que personalmente la elaboración de este trabajo me ha aportado como autora del mismo no solo una adquisición y puesta en práctica de unos conocimientos econométricos de vital importancia en relación con mis estudios cursados, sino además la experiencia de elaborar un proyecto de gran consideración que sin duda me será útil en situaciones futuras.

6.BIBLIOGRAFÍA

Antúñez Irgoin, C. H. (2010). “*Pruebas de raíces unitarias en EViews*”. Disponible online: <http://www.monografias.com/trabajos-pdf4/pruebas-raices-unitarias-evIEWS/pruebas-raices-unitarias-evIEWS.pdf>

Box, G. E. P., Jenkins, G. M. y Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. John Wiley and Sons: Hoboken, New Jersey, 4ª edición.

Casillas, C. (2010). “¿Por qué es tan importante el índice de producción industrial de un país?”. *Revista Inversión & Finanzas*. Disponible online: http://www.finanzas.com/noticias/economia/2010-03-08/255276_por-importante-indice-produccion-industrial.html

De Arce, R. y Mahía, R. (2011). *Modelos ARIMA*. Disponible online: http://www.uam.es/personal_pdi/economicas/anadelsur/pdf/Box-Jenkins.PDF

Economía Regional Andaluza (2014). “Índice de Producción Industrial de Andalucía. Marzo 2014”. Disponible online: <http://www.economiaregional.es/2014/05/indice-de-produccion-industrial-de.html>

Espasa, A. y Cancelo, J.R. (1993). *Métodos cuantitativos para el análisis de la coyuntura Económica*. Alianza Editorial, Madrid

García Luarces, P. (2011). “España firma el tratado de adhesión a la CEE, hoy Unión Europea”. Disponible online: <http://blogs.libertaddigital.com/almanaque-de-la-historia-de-espana/1985-espana-firma-el-tratado-de-adhesion-a-la-cee-hoy-union-europea-9797/>

González Enciso, A. y Matés Barco, J.M. (2007). *Historia Económica de España*. Editorial Ariel S.A., 1ª Edición, pp. 859-878.

Instituto de Estadística y Cartografía de Andalucía – Junta de Andalucía (6 de junio de 2014). “NOTA DE PRENSA - Índice de Producción Industrial de Andalucía. Abril 2014”. Disponible online:

<http://www.juntadeandalucia.es/institutodeestadisticaycartografia/NotasDePrensa/pdf/PIANotaprensa.pdf>

Instituto Nacional de Estadística (2010). “Índice de Producción Industrial (IPI), Base 2010, Metodología.”. Disponible online: <http://www.ine.es/daco/daco43/metoipi10.pdf>

Jiménez, J. F., Gázquez, J. C. y Sánchez, R. (2006). “La capacidad predictiva en los métodos Box-Jenkins y Holt-Winters: una aplicación al sector turístico”. *Revista Europea de Dirección y Economía de Empresa*, vol.15, núm. 3, pp. 187-191.

Mauricio, J. 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>

Page, D. (2013). “Todas las crisis de España: de la Edad Media al banco malo”. *Revista Expansión*. Disponible online: <http://www.expansion.com/2013/05/04/economia/1367690914.html>

Peña Sánchez de Rivera, D. (2010). *Análisis de series temporales*. Alianza Editorial D.L.: Madrid.

Sánchez, L. (2012). “El INE aborda la segunda fase del proyecto IRIA”. Disponible online: <http://www.computing.es/gestion/noticias/1059357001401/ine-aborda-segunda-fase-proyecto-iria.1.html>

Wei, W. W. S. (2006). *Time series analysis: univariate and multivariate methods*. Addison-Wesley: Redwood City.