



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Proyecto Fin de Carrera

Modelos ARIMA y Función de Transferencia para Previsiones Temporales



ESCUELA TÉCNICA
SUPERIOR DE
INGENIEROS DE
TELECOMUNICACIÓN

Autor : Najib BEKKALI

Director: Jorge Gosálbez Castillo

Curso 2012-2013

Agradecimiento

Primero y antes que nada, dar gracias a Dios, por estar conmigo en cada paso que doy, por fortalecer mi corazón e iluminar mi mente y por haber puesto en mi camino a aquellas personas que han sido mi soporte y compañía durante todo el periodo de estudio.

Agradecer hoy y siempre el apoyo incondicional ofrecido por mis padres durante toda la carrera y la confianza que han depositado en mí. Llegar hasta este punto, no hubiera sido posible sin ellos. Dedico este trabajo, con todo mi cariño a mi padre, sé lo que le hubiera significado este momento. Aunque no estás con nosotros físicamente, estás en nuestros corazones para siempre, te queremos papá.

Mil gracias a la mejor madre del mundo, por su paciencia, por su entrega, por su cariño y por todo....te quiero mamá.

Muchísimas gracias a mis queridas hermanas, que me brindan el apoyo, la alegría y me dan fortaleza necesaria para seguir adelante.

Muchas gracias a mi futura mujer por estar siempre a mi lado y por sus consejos.

Un agradecimiento especial al Profesor Jorge Gosálbez Castillo, por su tiempo, por su apoyo, por haber guiado el desarrollo de este trabajo y llegar a la culminación del mismo.

Muchas gracias a todos los profesores y compañeros que me han acompañado a lo largo de la carrera y me han dado la oportunidad de aprender de ellos. Ha sido un enorme placer compartir estos años con todos ellos.

Muchas gracias.

1.	INTRODUCCIÓN.....	8
1.1	Introducción.....	8
1.2	Objetivos.....	11
1.3	Estructura del trabajo	13
2.	INTRODUCCIÓN A LOS PROCESOS ESTOCÁSTICOS.....	15
2.1	Naturaleza de los procesos estocásticos	15
2.2	Realización de un proceso estocástico. Series temporales.....	16
2.3	Características de un proceso estocástico	18
2.4	Procesos estocásticos estacionarios	18
2.5	Proceso de ruido blanco.....	22
3.	MODELOS ARIMA	23
3.1	Introducción.....	23
3.2	Procesos autorregresivos $AR(p)$	25
❖	Proceso autorregresivo $AR(1)$	25
❖	Proceso autorregresivo $AR(2)$	26
❖	La función de autocorrelación parcial (fap):.....	27
3.3	Procesos de media móviles $MA(q)$	29
❖	Propiedades del proceso $MA(q)$:.....	30
3.4	Procesos $ARMA(p, q)$	32
3.5	Procesos no estacionarios. Modelos $ARIMA(p, d, q)$	32
3.6	Procesos estacionales	33
❖	Estructura $ARIMA$ estacional	34
❖	Características	35
3.7	Modelos $ARIMA(p, d, q)(P, D, Q)$ s generales	35
4.	METODOLOGÍA BOX-JENKINS.....	36
4.1	Introducción.....	36
4.2	Identificación del modelo	37
❖	Identificación de la no estacionariedad. Estacionarización	37
❖	Identificación de la estructura $ARMA(p, q)$ REGULAR	38

❖	Identificación de la estructura ARMA(P, Q) ESTACIONAL.....	38
4.3	Estimación del modelo.....	39
4.4	Validación del modelo	39
4.5	Explotación del modelo: Previsión	40
4.6	Ejemplo ilustrativo	41
❖	Identificación	42
❖	Estimación	44
❖	Validación	45
❖	Explotación del modelo	47
5.	MODELOS DE FUNCIÓN DE TRANSFERENCIA.....	48
5.1	Introducción.....	48
5.2	Modelos de función de transferencia con un único input	48
5.3	Modelos de funciones de transferencia con múltiples inputs	53
5.4	Identificación de la función de transferencia	54
❖	La función de covarianza y correlación cruzada	54
❖	La relación entre la función de correlación cruzada y la función de transferencia	55
❖	El preblanqueo.....	55
❖	Identificación de la función de respuesta al impulso	56
5.5	Verificación de un modelo de función de transferencia.....	58
❖	Contrastes sobre los parámetros.....	58
❖	Contraste sobre los residuos	58
6.	CASO DE ESTUDIO.....	59
6.1	Modelo ARIMA	60
❖	Formulación del modelo.....	60
❖	Identificación del modelo	62
❖	Estimación y validación del modelo	63
6.2	Función de Transferencia-Input Temperatura	66
❖	Formulación del modelo.....	66
❖	Identificación del modelo univariante.....	67
❖	Preblanqueo del input	68
❖	Preblanqueo del output.....	68
❖	Cálculo de la Función de Correlación Cruzada (CCF) entre a_t y β_t	68

❖	Identificación de la función de transferencia	69
❖	Estimación del modelo de función de transferencia	70
❖	Validación y previsiones	71
6.3	Función de Transferencia-Inputs: Temperatura y Precio.....	73
7.	CONCLUSIONES Y LÍNEAS FUTURAS	78
7.1	Resumen	78
7.2	Conclusión.....	78
7.3	Líneas de trabajo futuras	80
8.	BIBLIOGRAFÍA	82

1. INTRODUCCIÓN

1.1 INTRODUCCIÓN

Desde los tiempos más lejanos el hombre ha sentido la necesidad de conocer el futuro, de prever lo que va a suceder. A esta necesidad ha respondido por los métodos más variados y al menos desde la óptica actual, peregrinos. Tales métodos, más próximos a la magia que a la religión, han sido consultados y utilizados por los grandes hombres de la antigüedad, que supeditaban el inicio de nuevas empresas a la existencia de buenos auspicios sobre su futuro

Hoy la situación es distinta, pero más en cuanto a los métodos que en cuanto a la necesidad percibida por el hombre. Así los métodos actuales, con un enfoque absolutamente racionalista y en muchos casos matemático-estadístico, buscan la respuesta la misma pregunta de siempre: ¿cómo será el futuro? También ahora los hombres que rigen empresas e instituciones estudian esas previsiones antes de poner en marcha nuevos proyectos.

Así pues, ¿qué ha cambiado? Ha cambiado la noción de la propia naturaleza del futuro. Que ha pasado de ser algo sometido al libre arbitrio de fuerzas sobrenaturales (idea base de los métodos mágicos) a ser esencialmente el resultado del pasado y del presente, con una componente impredecible (aleatorio es la palabra correcta) más importante cuando más nos adentramos en el futuro.

En qué momento se da tal cambio de concepción es una cuestión de difícil respuesta, más cuanto no es posible hablar de un cambio brusco, sino de una transición larga de uno a otro planteamiento. Ya en siglo XIV, Nicolás Maquiavelo, enfoca la cuestión con una acertada frase: *“Para predecir lo que ha de suceder, hay que observar antes lo que ha ocurrido anteriormente”*. Aquí está la clave de las técnicas que serán estudiadas en este trabajo: el futuro como consecuencia del pasado. En la misma línea, y más recientemente, el Marqués de Halifax afirmaba: *“La mejor cualidad de un profeta es la buena memoria”*. La percepción de tal realidad es de gran importancia, pues de ella se sigue que el análisis y modelización de los hechos pasados va a permitirnos hacer previsiones sobre el futuro.

Hoy en día, La empresa moderna basa su gestión en la habilidad de detectar las oportunidades de desarrollo que brinda el entorno y en la habilidad de anticipar situaciones que puedan construir amenazas a su supervivencia, en esta situación en la que la información se ha convertido en una ventaja competitiva de primera magnitud y

su gestión eficiente en unos de los recursos estratégicos más importantes. Por eso, es necesario que la empresa invierta en herramientas y servicios destinados a transformar en verdadera información sus datos, para poder tomar así las decisiones más rentables con el mínimo grado de incertidumbre posible.

La mayoría de las decisiones que tomamos encierran implícitamente una previsión, es decir, hacemos previsiones para tomar decisiones. Por tanto, es muy importante el nivel de previsión, ya que si no lo hacemos de forma correcta podemos encontrarnos en una situación equivocada para planificar el futuro.

No obstante, esa dependencia del futuro con el pasado no es siempre igual de estrecha, y existen suficientes elementos de incertidumbre como para dar lugar al menos, a dos planteamientos distintos para la previsión: *el proyectivo* y *el prospectivo*.

En el primero de ellos, enfoque proyectivo, se trata de obtener las previsiones en base a la proyección o extrapolación de comportamiento pasados hacia el futuro. Evidentemente para que ese planteamiento sea correcto se ha establecido la hipótesis, quizás implícitamente, de que la situación que en el pasado dio lugar al comportamiento observado seguirá actuando en el futuro. A tal hipótesis se le suele denominar de mantenimiento (o constancia) del entorno. ¿Es razonable establecer esta restricción? En términos absolutos la respuesta es no: la sociedad es un sistema en cambio permanente en el que muy pocos entornos se mantienen estables. Ahora bien, a corto plazo sí que se puede, en muchos casos, asumir que los cambios que experimenta el entorno de un cierto problema que se está estudiando son de poca importancia, y que en conclusión podemos asumir su estabilidad y utilizar planteamientos proyectivos para la realización de previsiones. Conforme el horizonte de previsión se aleja del presente esa hipótesis pierde consistencia y el uso de estas técnicas será menos adecuado. Sólo en el estudio de algunos fenómenos naturales podremos prolongar el plazo de validez de nuestras previsiones, al realizarse en entornos altamente estables (por ejemplo, en astronomía).

El segundo enfoque, denominado prospectivo, parte del hecho arriba comentado del constante cambio a que están sometidos los sistemas sociales y económicos. Ello invalida la hipótesis de constancia del entorno y a la vez pone en tela de juicio la corrección de las previsiones obtenidas por proyección de comportamientos pasados. El principal objetivo de las técnicas prospectivas será la identificación del sentido en que evoluciona el entorno y la definición de la que será la nueva situación que afecte anterior. La inconstancia del entorno afecta especialmente a las previsiones a largo

plazo, y será en este largo plazo donde las técnicas prospectivas encuentran su mayor utilidad.

El uso integrado de ambos planteamientos permite disponer de previsiones basadas en la información histórica, válidas para el inmediato futuro, y uno o varios escenarios alternativos, que describen la situación en un futuro más lejano.

Las técnicas de previsión proyectiva, con que vamos a desarrollar nuestro proyecto, tienen en común el basarse en la información historia para, mediante su análisis, identificar pautas de comportamiento que supondremos siguen siendo válidas en futuro. Requerimos en todos los casos la hipótesis de mantenimiento del entorno y, por poder asegurar validez de ésta sólo a corto plazo, serán técnicas útiles especialmente en las previsiones a corto.

La información que habitualmente manejan las técnicas proyectivas tiene la forma de las llamadas series temporales, lo que es lo mismo un proceso aleatorio en terminología de telecomunicaciones, que son sucesiones de valores referidos a una misma variable (X_t) correspondientes a distintos instantes de tiempo, variable independiente (t).

Para ser más exactos en la definición, un proceso aleatorio es una función de las posibles realizaciones de un experimento, pero, a diferencia de una variable aleatoria (V.A.), dichas realizaciones o resultados del experimento son función del suceso probabilístico, s , y del tiempo, t . De esta forma, a cada realización, s , le asignamos una función temporal $X(t, s)$. A la familia de todas las funciones posibles la denominamos *Proceso Aleatorio (P.A.)* o *Proceso Estocástico*. Por sencillez, a los procesos aleatorios los denominaremos simplemente por X_t . Para aclarar la nomenclatura utilizada, a partir de ahora, y a lo largo de este trabajo se llamará a un proceso aleatorio serie temporal.

Cabe destacar, que una primera división de las técnicas basadas en series temporales, sería entre métodos clásicos y métodos modernos. Los primeros parten de un enfoque determinista, considerando que la serie temporal sigue una pauta de comportamiento fija, en la que las desviaciones observadas son sólo el efecto de hechos puntuales ocurridos en el entorno, sin que esos hechos afecten a la pauta base del modelo. Por este motivo se habla también de métodos deterministas. Los métodos modernos se basan en planteamientos probabilísticos, considerando las series temporales como manifestaciones de proceso estocásticos de una cierta estructura, en la cual esas perturbaciones que el entorno introduce forman parte de la propia estructura de la serie.

En este trabajo, nos centraremos en la metodología *Box-Jenkins*, uno de los de los métodos modernos o estocásticos. Basada en la identificación y posterior explotación del modelo *ARIMA* que mejor refleje el comportamiento pasado de la serie.

A lo largo de los últimos años, las técnicas de pronóstico estadístico se han ido aplicando en muchos sectores, y el sector de la energía no es ajeno a la necesidad de predicción de las variables que le afectan para su funcionamiento diario. Con el fin de tener un control y buena planificación de los sistemas de energía, se utiliza la predicción de la demanda.

Esa necesidad la podemos apreciar también en el sector de telecomunicaciones, una aplicación directa sería por ejemplo la predicción del comportamiento de las redes IP con un doble objetivo. Por un lado, está el objetivo de predecir la carga futura de enlaces individuales o rutas aisladas y así anticipar la aparición de situaciones críticas. Como es de esperar de antemano la aparición de problemas permite solucionarlos antes de que éstos lleguen a producirse. El otro objetivo es establecer una mejor planificación del crecimiento de la infraestructura de red mediante los resultados predictivos que se han obtenido. En este sentido, es más eficiente tratar los datos globalmente, por ejemplo, a partir de las matrices de tráfico.

1.2 OBJETIVOS

Este proyecto final de carrera nació como fruto de una beca de colaboración del Departamento de Estadística e Investigación Operativa Aplicadas y Calidad de la Universidad Politécnica de Valencia. Colaborando en el desarrollo del mismo el Dr. Jorge Gosálbez Castillo, como tutor de la presente memoria.

Este trabajo tiene dos objetivos fundamentales. El primero de ellos, es desarrollar una metodología de modelos de predicción basadas en series temporales mediante la aplicación y desarrollo de herramientas adecuadas, para un caso relacionados con el sector energético, con diferentes tipos y cantidad de datos disponibles, en segundo lugar, realizar una comparación de estas técnicas para la selección del modelo idóneo.

Dentro de los modelos basados en series temporales se distinguen dos modelos: modelo *ARIMA* y modelo de función de transferencia.

Se comenzará utilizando los modelos en los que una variable es explicada utilizando exclusivamente una “exógena”: su propio pasado. Dentro de estos modelos univariantes se desarrollarán suficientemente los conocidos con el nombre de *ARIMA*.

Como procedimiento de análisis estadístico para ajustar a una serie los modelos *ARIMA* emplearemos la metodología *Box-Jenkins*, una metodología de trabajo capaz de extraer todo su potencial, que se basa en un esquema iterativo de trabajo común al todo el método científico, que se esquematiza en *Figura 1.1* de las fases de identificación del modelo a emplear, estimación de sus parámetros, validación de la bondad del modelo y reformulación o en su caso explotación. En los próximos capítulos detallaremos cada una de esas fases del análisis de una serie temporal.

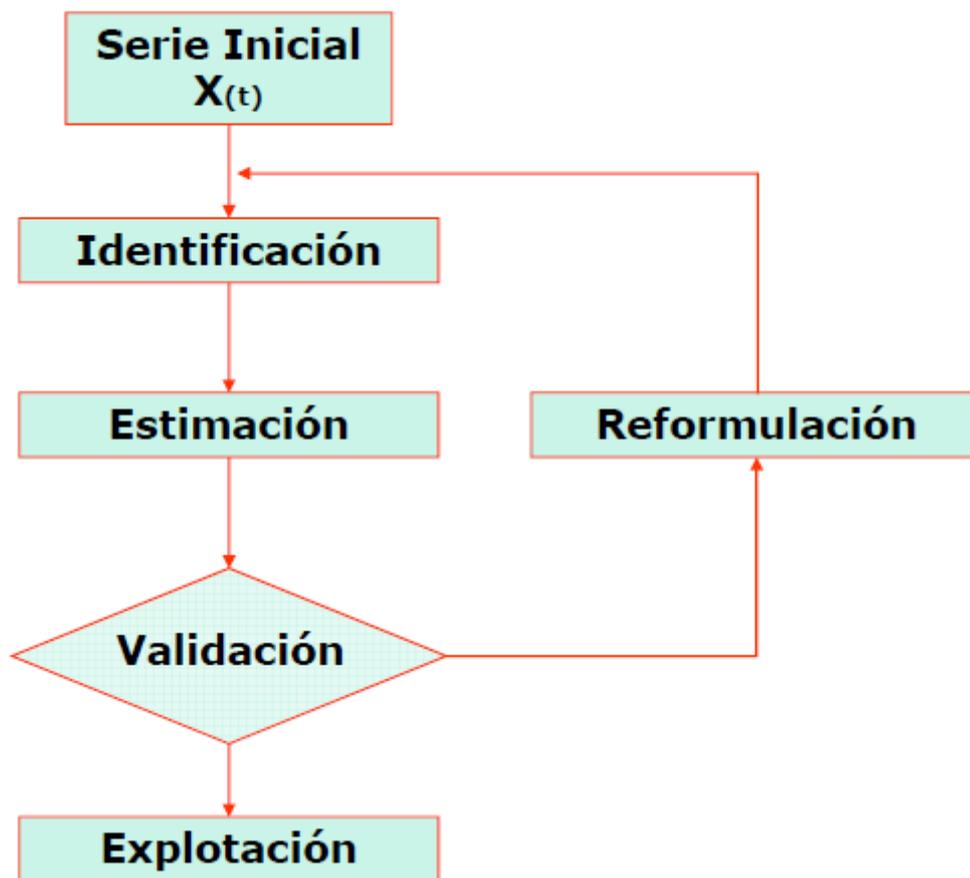


Figura 1.1

Las previsiones univariantes pueden mejorarse incorporando la información de la evolución de otras variables y construyendo modelos que tengan en cuenta esta dependencia. Estos modelos se conocen como *modelos de Regresión Dinámica o de Función de Transferencia*.

En resumen, en los modelos *ARIMA* se consideran la predicción de una serie temporal basada en su propio pasado. En cambio, en los modelos de función de transferencia, vamos a introducir cómo hacer predicciones de una serie temporal cuando

se conoce además de su propio pasado, el pasado y presente de otra serie temporal (variables exógenas) que pueda estar relacionada con la misma, estudiando la relación entre una serie temporal input X_t y una (o más) serie temporal output Y_t .

1.3 ESTRUCTURA DEL TRABAJO

Como se ha comentado anteriormente de la importancia de la predicción, en la actualidad, ya sea en el campo de la medicina, la economía, la tecnología, toda institución o empresa necesita realizar planes para el futuro si desea anticiparse a las situaciones del entorno, en el mercado o entre la sociedad para su supervivencia. La planificación necesita unas previsiones de lo que va a ocurrir en el futuro y la precisión con las que éstas pueden ocurrir es una ventaja competitiva clara de primera magnitud sobre las decisiones, que marcan la diferencia entre el resto de la competencia.

El modelo estadístico utilizado para hacer inferencias sobre el futuro, tiene que tener en cuenta lo ocurrido en el pasado, esto en sí se hace con las técnicas de análisis de series temporales.

Por tanto, con las series temporales trataremos de hacer predicciones sobre esa magnitud, teniendo en cuenta sus características históricas y como ha ido fluctuando las series a lo largo del tiempo.

A lo largo de la beca de colaboración, se han tratado unos cuantos casos, y todos están relacionados con el sector energético, por eso, el objetivo era definir una metodología con la herramienta informática adecuada para ajustar y validar los posibles modelos y así obtener las predicciones correspondientes.

Definiremos una estructura que nos permita, por sus características, cumplir el fin último de predicciones:

En el segundo capítulo se puede encontrar una introducción a los procesos estocásticos, y sus características, donde abordaremos unos conocimientos teóricos básicos para el análisis de los modelos *ARIMA*.

En el tercer y el cuarto capítulo se detalla la descripción de los modelos *ARIMA*, también se explican las herramientas necesarias para realizar un análisis de series temporales basado en Box-Jenkins, con un caso al final del capítulo ilustrando todas las fases necesarias para llevar a cabo unas predicciones idóneas para el modelo validado.

En el quinto capítulo, se explicarán los pasos necesarios para formular y estimar un modelo de función de transferencia para el caso general de varias entradas.

Y para acabar, estudiaremos el caso práctico del consumo de gas natural en el capítulo seis mediante análisis ARIMA y función de transferencia, incluyendo como variables explicativas la temperatura y el precio.

En el capítulo siete se comentarán las conclusiones más relevantes de todo el trabajo subyacentes de la comparación pormenorizada de los resultados.

2. INTRODUCCIÓN A LOS PROCESOS ESTOCÁSTICOS

En este capítulo introducimos los procesos estocásticos como modelo de una serie temporal, conociendo las propiedades más importantes de un proceso estocástico. De este modo, sería importante definir lo que es un proceso estacionario y mostrar las implicaciones estadísticas y temporales que conlleva dicha propiedad, para posteriormente presentar un proceso estacionario muy importante en el manejo de las predicciones: el proceso de ruido blanco.

2.1 NATURALEZA DE LOS PROCESOS ESTOCÁSTICOS

Una variable aleatoria unidimensional puede ser representada como X , y si pasamos a variables n-dimensionales la notación será (X_1, X_2, \dots, X_n) . Consideremos una variable aleatoria multidimensional cuyo término genérico sea X_t , si esta t puede tomar infinitos valores de un conjunto Θ , entonces diremos que $\{X_t\}$ es un proceso estocástico.

Si el conjunto Θ es continuo, por ejemplo R^+ , diremos que X_t es un proceso estocástico de parámetro continuo. Si por el contrario Θ es discreto, por ejemplo N , hablaremos de proceso estocástico de parámetro discreto. Ese parámetro t suele tener una interpretación temporal, con lo cual estaremos hablando de la consideración que le demos al tiempo, continuo o discreto. Por su parte la variable X puede ser discreta o continua, lo cual nos conduce a cuatro tipos de proceso estocástico:

	t DISCRETO	t CONTINUO
VARIABLE X DISCRETA	P.E DISCRETO DE VARIABLE DISCRETA	P.E CONTINUO DE VARIABLE DISCRETA
VARIABLE X CONTINUA	P.E DISCRETO DE VARIABLE CONTINUA	P.E CONTINUO DE VARIABLE CONTINUA

Al igual que en una variable n-dimensional cada componente X_i es a su vez una variable aleatoria, también aquí, para cada t , X_t es una variable aleatoria, con lo que el proceso estocástico podemos interpretarlo como una sucesión de variables aleatorias, cuyas características pueden variar a lo largo del tiempo. Por ejemplo, si consideramos el tiempo discreto y observamos sólo unos pocos valores de t , tendríamos una imagen similar a la de la *Figura 2.1* en la que se representa para cada t la función de densidad de probabilidad correspondiente a X_t . Aunque en la figura se han representado unas funciones de densidad variables, un proceso estocástico no tiene porqué presentar esa

inconstancia en la función de densidad a lo largo del tiempo. Como más adelante se comentará presentan un especial interés aquellos procesos cuyo comportamiento se mantiene constante a lo largo de t .

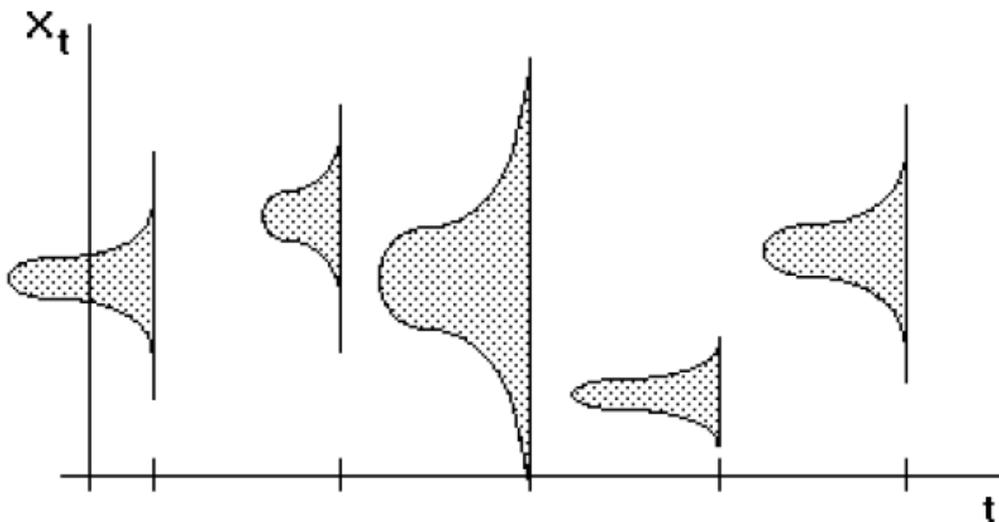


Figura 2.1

2.2 REALIZACIÓN DE UN PROCESO ESTOCÁSTICO. SERIES TEMPORALES

Un concepto que debe ser definido es el de realización: una realización de una experiencia aleatoria es el resultado de una repetición de esa experiencia. Así, en la experiencia aleatoria "lanzar una vez un dado" una realización sería el cinco que nos ha salido en ese único lanzamiento. En ese caso, la realización se reduce a un único número $\{x\}$. En una experiencia n -dimensional tendríamos un grupo de n valores $\{x_1, \dots, x_n\}$ y, siguiendo con la analogía, en un proceso estocástico tendríamos un grupo de infinitos valores $\{x_t\}$, con $t \in \Theta$.

Veamos un ejemplo de realización de un proceso estocástico. Consideremos que estamos interesados en estudiar cómo evoluciona la temperatura en un punto P de un motor durante una hora de funcionamiento del mismo a régimen constante. Podemos considerar que la variable "temperatura instantánea en ese punto P " es un proceso estocástico continuo de parámetro continuo. Si realizamos la experiencia una vez obtendremos una gráfica temporal de la evolución de la temperatura: esa gráfica (o los valores numéricos asociados) es una realización del proceso estocástico. Si repetimos la

experiencia en las mismas condiciones obtendremos otra realización, representada por otra gráfica o por sus correspondientes resultados numéricos. (Figura 2.2)

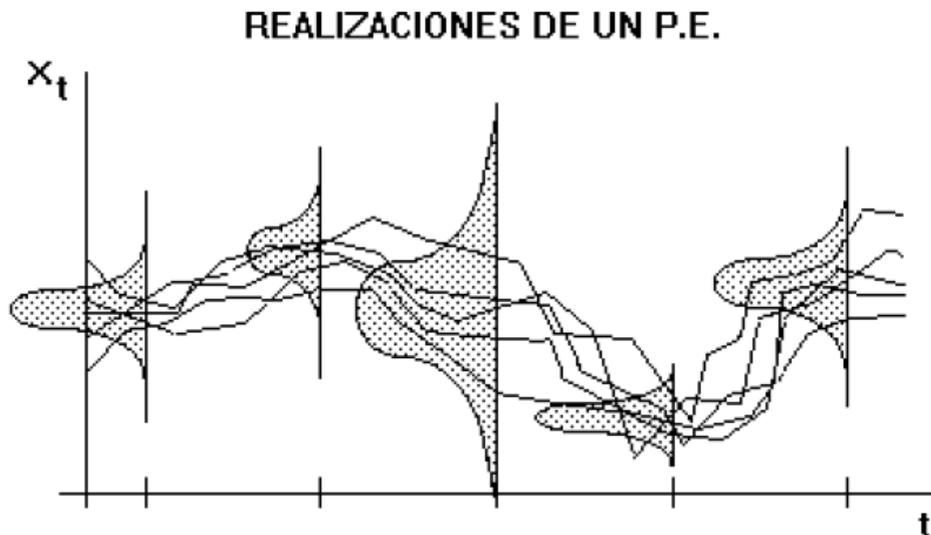


Figura 2.2

En la enumeración anterior se aprecia, por tanto, que la letra mayúscula se reserva siempre para el carácter aleatorio (ya sea el proceso completo o una *V.A.* por él generada), mientras que la letra minúscula indica siempre un carácter determinista (función temporal o amplitud).

Su representación se realiza colocando en el eje de abscisas los valores de t , y en ordenadas el valor de X , con lo que tenemos que una realización de un proceso estocástico es una sucesión de infinitos valores de una cierta variable, a lo largo del tiempo t . Si t tiene una consideración continua x_t será una línea y si t es discreto x_t será una sucesión de puntos, uno para cada.

Por otro lado, una serie temporal es una sucesión de valores de una variable obtenidos secuencialmente durante un tiempo.

Si comparamos estas dos definiciones, vemos que la única diferencia que hay entre ellas radica en que la realización consta de infinitos elementos y la serie temporal de un número limitado. Así, desde el punto de vista de los procesos estocásticos, diremos que una serie temporal es una realización parcial de un proceso estocástico de parámetro tiempo discreto. Con ello, la teoría de los procesos estocásticos será de aplicación a las series temporales.

Los procesos aleatorios se pueden clasificar en dos grupos, en función de cómo se presenta su aleatoriedad:

- P.A. Predecible: Es aquel proceso cuyos valores futuros de cualquier realización pueden predecirse en función de los valores pasados. También se le puede llamar P.A. determinista. Un ejemplo sería el P.A. $X_t = A \cos(\omega_0 t + \theta)$ siendo A , ω_0 y/o θ variables aleatorias.

- P.A No Predecible: También llamado P.A No Determinista. Es aquel proceso que no cumple la condición anterior. Un ejemplo de P.A. no predecible es el ruido blanco que se añade a la señal recibida en cualquier sistema de comunicaciones.

2.3 CARACTERÍSTICAS DE UN PROCESO ESTOCÁSTICO

Puesto que las características del proceso estocástico pueden variar a lo largo de t , deberemos calcular la media para cada instante de tiempo, obteniendo una función de medias cuya expresión es:

$$\mu_t = E[X_t]$$

Podremos también obtener la varianza en cada instante de tiempo:

$$\sigma_t^2 = E[(X_t - E[X_t])^2]$$

Y la covarianza entre valores de x en diferentes instantes:

$$\gamma(t, s) = cov(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)]$$

A partir de esta expresión de la covarianza y de la anterior de la varianza, podemos obtener el coeficiente de correlación entre X_t y X_s :

$$\rho(t, s) = E[X_t X_s] = \frac{\gamma(t, s)}{\sqrt{\sigma_t^2 \sigma_s^2}} = \frac{\gamma(t, s)}{\sigma_t \sigma_s}$$

Por tratarse de covarianza o correlaciones calculados entre valores de un mismo proceso, aunque sea en diferentes instantes de tiempo, se hablará con frecuencia de autocovarianza y autocorrelaciones y de función de autocovarianza y función de autocorrelación.

2.4 PROCESOS ESTOCÁSTICOS ESTACIONARIOS

La estacionalidad es un concepto fundamental en el análisis de series temporales, en una primera aproximación, llamaremos estacionarios a aquellos procesos estocásticos que tengan un comportamiento estadístico constante a lo largo del tiempo. En la práctica del análisis de series con problemas de estacionariedad que afecten a cualquier de sus

parámetros básicos, siendo los que más suelen afectar al proceso de análisis las inconstancias en media y varianza,

Con más precisión, las definiciones de estacionariedad implican una cierta propiedad de las funciones de densidad de probabilidad de orden n : $f_x(x_1, \dots, x_n; t_1, \dots, t_n)$. Sin embargo esta condición es habitualmente demasiado exigente, por lo que se emplea una definición distinta que establece lo que se denomina estacionariedad en sentido amplio, es una condición más relajada que las del primer y segundo orden al depender únicamente de las propiedades de la media y la autocorrelación. Consecuencia de esto es que las funciones de medias y varianzas son constantes:

$$\mu_t = m \quad \sigma_t^2 = \sigma^2 \quad \forall t$$

Y la función de covarianza entre x_t y x_s sólo depende de la diferencia $s-t$, y no de los valores de ambos instantes:

$$\gamma(t, s) = \gamma(s - t) = \lambda(\tau) \quad \text{siendo } \tau = s - t, \quad t < s$$

a este valor de τ se le denomina decalaje entre los instantes t y s .

Consecuencia de este comportamiento de la función de covarianzas es que la función de correlación valdrá:

$$\rho(t, s) = \frac{\gamma(t, s)}{\sqrt{\sigma_t^2 \sigma_s^2}} = \frac{\gamma(t, s)}{\sigma_t^2 \sigma_s^2} = \frac{\gamma(\tau)}{\sigma^2} = \rho(\tau)$$

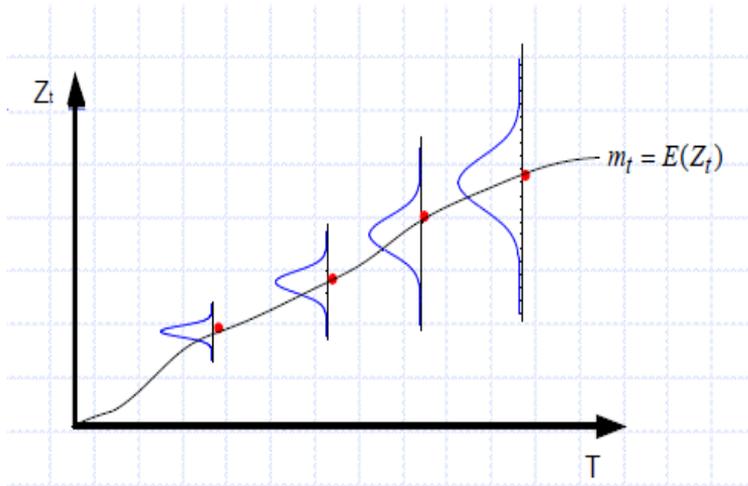
Dependiendo sólo del decalaje. Como además la autocovarianza con de decalaje cero es la varianza:

$$\gamma(t, t) = \gamma(t - t) = \gamma(0) = \sigma^2$$

La expresión anterior quedará:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

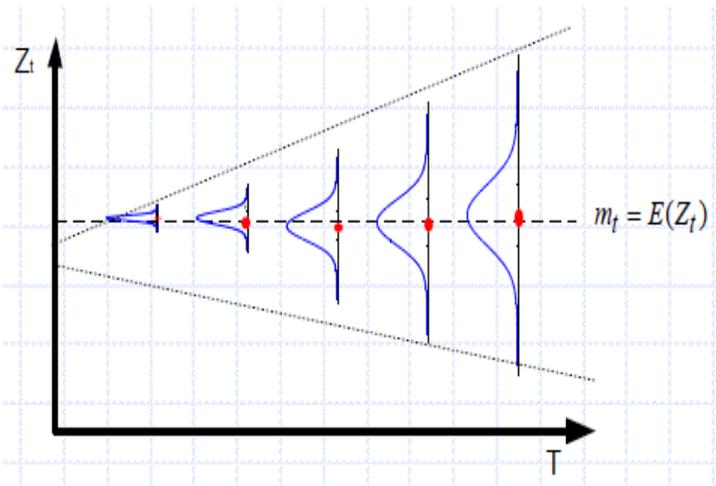
Veremos algunos ejemplos, si un P.E es estacionario o no, a partir de su media y su varianza:



⇒ P.E NO ESTACIONARIO porque:

- Función de varianza No constante
- Función de medias No Constante

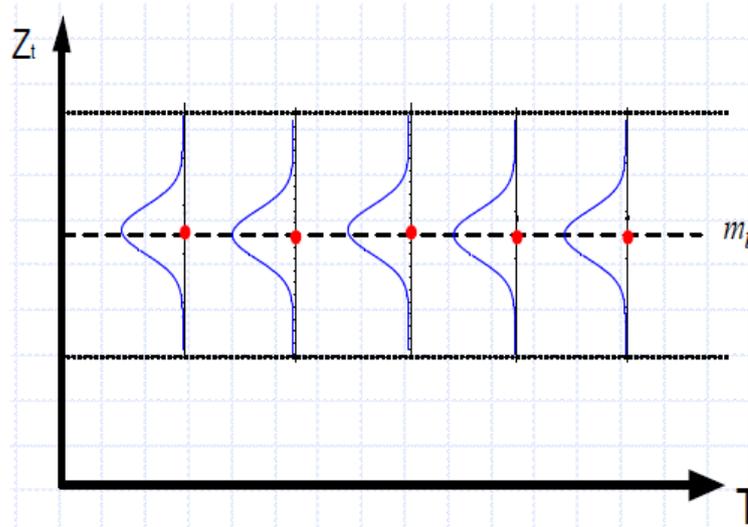
Figura 2.3



⇒ P.E NO ESTACIONARIO porque:

- Función de varianza No constante, a pesar de que la función de medias Constante

Figura 2.4



⇒ P.E ESTACIONARIO porque:

- Función de varianza constante
- Función de medias constante

Figura 2.5

En adelante llamaremos a la función de autocorrelación del proceso la función de autocorrelación simple (*fas*), para diferenciarla de otras funciones ligadas a la autocorrelación que se definirán más adelante.

El gráfico de los valores del coeficiente de autocorrelación frente al decalaje se denomina correlograma y, como veremos más adelante, es la herramienta fundamental en el análisis de series temporales y cabe notar que sólo tiene sentido para series estacionarias pues sólo en ellas el coeficiente de autocorrelación depende exclusivamente del decalaje o retardo.

El cálculo del correlograma se hace mediante el cálculo de los coeficientes de correlación simple entre las siguientes parejas de variables:

El cálculo del correlograma equivale calcular los coeficientes de correlación simple entre las siguientes parejas de variables:

$$\begin{aligned} \rho(1): & (x_1, x_2), (x_2, x_3), (x_3, x_4), \dots, (x_n, x_{n+1}), \dots \\ \rho(2): & (x_1, x_3), (x_2, x_4), (x_3, x_5), \dots, (x_n, x_{n+2}), \dots \\ & \dots \\ \rho(k): & (x_1, x_{k+1}), (x_2, x_{k+2}), (x_3, x_{k+3}), \dots, (x_n, x_{k+n}), \dots \\ & \dots \end{aligned}$$

Cómo se ha comentado anteriormente el correlograma tal como se ha definido sólo tiene sentido para series estacionarias. Y la interpretación de estos coeficientes de autocorrelación es la misma que el coeficiente de correlación ordinaria entre dos variables cualesquiera: toma valores entre -1 y +1, estando en módulo más próximo a la unidad cuanto más se aproxima la relación entre las variables a una recta, indicando el signo si la recta es creciente o decreciente. El coeficiente de autocorrelación correspondiente al decalaje cero, $\rho(0)$, siempre la unidad al recoger la relación, obviamente lineal y exacta, de cada variable consigo misma. (Ver la *Figura 2.6*)

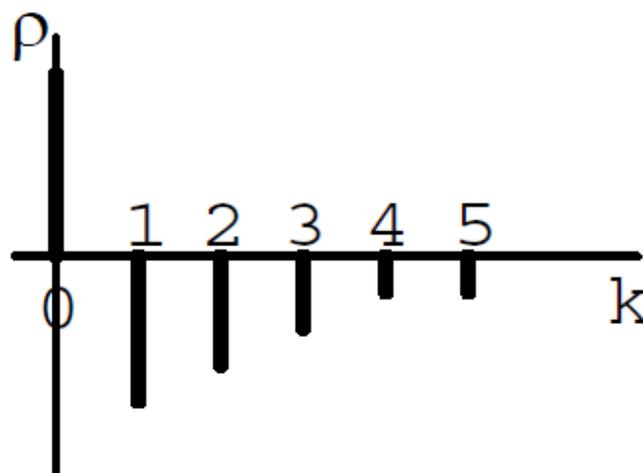


Figura 2.6

2.5 PROCESO DE RUIDO BLANCO

Un proceso estocástico que emplearemos con frecuencia en los apartados siguientes, sobre en la fase de la validación de los modelos es el llamado proceso de ruido blanco. Se trata de un proceso estacionario que representamos habitualmente como a_t y cumple las siguientes características:

1ª.media nula: $E(a_t) = 0 \quad \forall t$

2ª.varianza constante: $var(a_t) = \sigma_a^2 \quad \forall t$

3ª.incorrelacionados: $cov(a_t, a_s) = 0 \quad \forall t \neq s$

Y habitualmente exigiremos también que sigan una distribución normal.

4ª.normales: $\{a_t\} \equiv N(0, \sigma_a) \quad \forall t$

Una de las características principales de un proceso estocástico es su correlograma y en el caso del proceso de ruido blanco no puede ser más sencillo: puesto que las variables correspondientes a distintos instantes están incorrelacionadas, todos los coeficientes de correlación son nulos, y el correlograma no tiene ningún coeficiente de autocorrelación presentado, salvo el que corresponde a retardo cero, que siempre es la unidad.

3. MODELOS *ARIMA*

3.1 INTRODUCCIÓN

Una vez que hayamos definido el concepto de la estacionariedad como una característica principal de los procesos estocásticos, en este capítulo daremos entrada a los modelos univariantes *ARIMA* como técnicas más usuales en el diseño de modelos para la predicción de futuros valores de series temporal, empezando por los modelos simples, tales como el autorregresivos simple (modelos *AR*) y la media móvil (modelos *MA*), hasta llegar a los más complejos; la combinación de ambos modelos sin o con procesos de integración (modelos *ARMA* o *ARIMA*, respectivamente). Sin embargo antes de su presentación se expondrán algunos conceptos necesarios.

La predicción del valor futuro de una serie temporal significa determinar los siguientes valores que tomará la variable que forma la serie temporal a partir del instante presente. Esa determinación del valor futuro de la variable puede realizarse desde dos perspectivas indicadas seguidamente:

1. La predicción de valores futuros se puede realizar por extrapolación de los valores pasados.
2. Se puede establecer un modelo matemático que explique las relaciones entre los valores de la variable en el pasado y utilizar ese modelo para predecir los futuros valores de la variable.

Precisamente nos centraremos en la segunda perspectiva comentada, la de extraer información del pasado. Presentaremos el modelo que mejor explique las relaciones entre los valores pasados de la variable, para ser utilizado posteriormente en la predicción de los valores futuros.

Pero, en muchas ocasiones se dispone de otras informaciones que pueden ser útiles en la determinación de los valores futuros de una variable, de la que disponemos de una serie temporal. La variable que tratamos de predecir puede ser función de otras variables cuyos valores podemos conocer, lo que las hace variables “explicativas” de la variable a predecir. En función de que usemos, o no, esa información con los valores de las posibles variables explicativas, podemos distinguir dos métodos para establecer los valores futuros de la variable a predecir señalados a continuación:

- Métodos de predicción univariante (también conocidos como métodos de series temporales), es el caso de los modelos *ARIMA*: estos métodos pretenden establecer la relación entre los propios valores pasados de la

variable a predecir, una vez establecida la relación, con el modelo se realizan proyecciones hacia el futuro para obtener predicciones de la variable. Matemáticamente, el valor en un futuro de la variable a predecir es una función de sus variables pasados, como indica la ecuación:

$$x_{t+1} = f(x_t, x_{t-1}, \dots, x_{t-d})$$

- Métodos de predicción multivariante (también conocidos como métodos causales) que serían el objetivo de estudio del próximo capítulo: estos métodos pretenden establecer la relación entre los valores pasados de la variable a predecir y los valores, también pasados, de las posibles variables explicativas. Una vez establecida la relación (determinado el modelo) puede ser utilizado para obtener valores futuros de la variable a predecir. La ecuación expresa el valor en un futuro inmediato de la variable a predecir como una función de los valores futuros de otras dos variables (valores futuros, pero conocibles de antemano), u y v .

$$x_{t+1} = f(u_{t+1}, v_{t+1})$$

En el caso más general el método multivariante, el valor futuro de la variable a predecir puede ser función de los valores pasados de la propia variable, del valor futuro de las variables explicativas y de los valores pasados de estas variables, tal y como indica la ecuación:

$$x_{t+1} = f(x_t, x_{t-1}, \dots, x_{t-d}, u_t, u_{t-1}, \dots, u_{t-m}, v_t, v_{t-1}, \dots, v_{t-n})$$

Los modelos *ARIMA* fueron popularizados en los años 70 por *George Box* y *Gwilym Jenkins*. La principal ventaja de esta metodología es que proporciona predicciones óptimas en el plazo inmediato y en el corto plazo. Esto se debe a que la metodología *Box-Jenkins* nos permite elegir entre un amplio rango de distintos modelos según represente mejor el comportamiento de los datos. El sentido de predicciones óptimas significa que ningún modelo univariante puede ofrecer mejores predicciones que un modelo *ARIMA*. Esto no se cumple si ampliamos el modelo *ARIMA* con regresión múltiple o utilizamos metodología multivariante.

La principal desventaja de estos modelos es que la determinación del modelo que mejor se adecua a la serie de datos no es trivial y, por tanto, se requiere que la persona que realice predicciones tenga amplios conocimientos sobre esta metodología.

Antes de empezar con el desarrollo exhaustivo de estos modelos, sería útil definir un elemento que resulta de gran interés es el llamado operador retardo B . Tal operador

actúa sobre un término de un proceso estocástico reduciendo el índice temporal en una unidad:

$$Bx_t = x_{t-1}$$

Aplicando el operador k veces obtenemos:

$$B^k x_t = x_{t-k}$$

3.2 PROCESOS AUTORREGRESIVOS $AR(p)$

Los procesos autorregresivos forman una familia de procesos tales que una observación depende de las observaciones anteriores. Diremos que una serie temporal X_t estacionaria sigue un proceso autorregresivo de orden p si:

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + a_t$$

donde a_t un proceso de ruido blanco, c y ϕ_i son constantes a determinar en el proceso de ajuste. Es decir que una observación estará influida por las p observaciones anteriores de forma directa.

Para simplificar, consideremos la constante del modelo $c = 0$, y aplicando el operador de retardo B , obtenemos:

$$x_t = \phi_1 Bx_t + \phi_2 B^2 x_t + \dots + \phi_p B^p x_t + a_t$$

escribiendo el resultado en forma de polinomio,

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)x_t = a_t$$

que es la manera alternativa de escribir el modelo autorregresivo de orden p en forma de polinomio. Llamando $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$: Polinomio autorregresivo de orden p , obtenemos la notación compacta:

$$\phi_p(B)x_t = a_t$$

Llamaremos ecuación característica del proceso autorregresivo a la ecuación:

$$\phi_p(B) = 0$$

Nótese que para que el proceso $AR(p)$ sea estacionario, las p raíces del polinomio en B , $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ deben estar dentro del círculo unidad.

En la práctica, resulta útil saber identificar los modelos autorregresivos más importantes: $AR(1)$ y $AR(2)$, por eso vamos a dar alguna de sus características.

❖ *Proceso autorregresivo $AR(1)$*

El proceso autorregresivo de primer orden $AR(1)$:

$$x_t = c + \phi x_{t-1} + a_t$$

donde c y $-1 < \phi < 1$ son constantes a determinar y a_t es un proceso de ruido blanco con varianza σ^2 .

- Esperanza y varianza:

$$E[x_t] = \mu = \frac{c}{1 - \phi}$$

Llamando $\tilde{x}_t = x_t - \mu$

$$E[\tilde{x}_t] = \sigma_x^2 = \frac{\sigma^2}{1 - \phi^2}$$

- Función de autocovarianza:

$$\gamma_k = \phi \gamma_{k-1} \quad k = 1, 2, \dots$$

donde $\gamma_0 = \sigma_x^2$. Esta ecuación muestra que, como $|\phi| < 1$, la dependencia entre observaciones se amortigua al aumentar el retardo. En particular:

$$\gamma_1 = \frac{\phi \sigma^2}{1 - \phi^2}$$

- Función de autocorrelación simple (*fas*):

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k \quad k \geq 0$$

lo que muestra que la *fas* de un proceso $AR(1)$ decrece geoméricamente hacia cero.

❖ **Proceso autorregresivo $AR(2)$**

El proceso autorregresivo de segundo orden $AR(2)$:

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t$$

Donde ahora c y ϕ_1 y ϕ_2 son constantes a determinar y a_t es un proceso de ruido blanco con varianza σ^2 .

- Esperanza y varianza:

$$E[x_t] = \mu = \frac{c}{1 - \phi_1 - \phi_2}$$

Llamando $\tilde{x}_t = x_t - \mu$

$$E[\tilde{x}_t] = \sigma_x^2 = \gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)}$$

Para que el proceso sea estacionario esta varianza debe ser positiva, por tanto:

$$-1 < \phi_2 < 1 \quad \phi_1 + \phi_2 < 1 \quad \phi_2 - \phi_1 < 1$$

- Función de autocovarianza:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \quad k \geq 1$$

Particularizando esta ecuación para $k = 1$, como en un proceso estacionario $\gamma_{-1} = \gamma_1$, se obtiene que:

$$\gamma_1 = \frac{\phi_1 \gamma_0}{(1 - \phi_2)}$$

- Función de autocorrelación simple (*fas*):

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad k \geq 1$$

Particularizando para $k = 1$, como en un proceso estacionario $\rho_1 = \rho_{-1}$, obtenemos que:

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

Y particularizando para $k = 2$, y utilizando la ecuación anterior:

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2$$

Lo que da lugar a dos formas de decrecimiento de la función de autocorrelación simple:

1. El decrecimiento de la *fas* será la suma de dos exponenciales y para no entrar en muchos detalles, la forma de la función de autocorrelación dependerá de que los dos factores del polinomio característica del proceso (son reales en este caso) tenga signo igual u opuesto.
2. La *fas* decrecerá de forma sinusoidal en el caso de que los dos factores del polinomio característica del proceso sean complejos.

❖ **La función de autocorrelación parcial (*fap*):**

Determinar el orden de un proceso autorregresivo a partir de su función de autocorrelación simple es difícil. Hemos visto que esta función es una mezcla de decrecimientos exponenciales y sinusoidales, que se amortiguan al avanzar el retardo, y no presenta rasgos fácilmente identificables para determinar el orden del proceso. Para resolver este problema se introduce la función de autocorrelación parcial (*fap*).

Se define el *coeficiente de autocorrelación parcial* de orden k, ρ_k^p como el coeficiente de correlación entre observaciones separadas k periodos, cuando eliminamos de la relación entre las dos variables de dependencia lineal debida a los valores intermedios. Llamaremos *función de autocorrelación parcial (*fap*)* a la representación de los coeficientes de autocorrelación parcial en función del retardo.

El procedimiento del cálculo de los coeficientes de autocorrelación parcial no tiene importancia para nuestro trabajo, ya que se maneja en la modelización y análisis de series temporales un correlograma parecido al correlograma simple, denominado correlograma parcial, facilitado por cualquier software de estadística.

Como norma general, en un proceso $AR(p)$ tendrá los p primeros coeficientes de autocorrelación parcial distintos de cero, y por tanto, el número de coeficientes distintos de cero nos indica el orden del proceso AR . Esta propiedad va a ser clave para identificar el orden de un proceso autorregresivo. Además, el coeficiente de correlación parcial de orden p coincidirá siempre con el parámetro ϕ_p .

Resumimos en la tabla siguiente (*Figura 3.1*) las funciones de autocorrelación simple y parcial de distintos procesos AR .

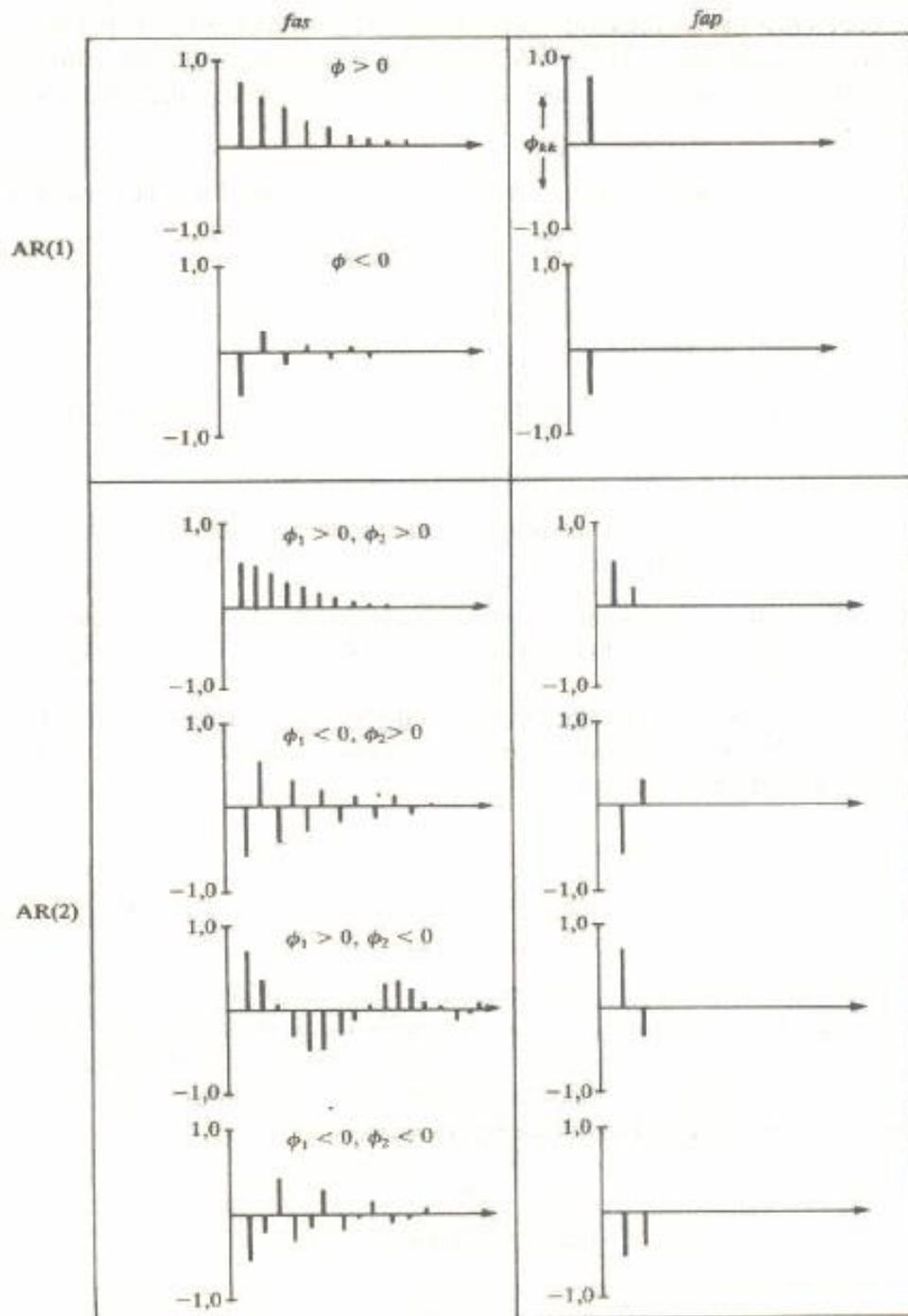


Figura 3.1

3.3 PROCESOS DE MEDIA MÓVILES $MA(q)$

Los procesos autorregresivos que se han introducido anteriormente tienen una característica común: todos ellos tienen memoria larga. La memoria larga aparece en que los valores de la *fas* van decreciendo lentamente. En cambio, hay una nueva

familia de modelos que representa los procesos de memoria corta y se denomina genéricamente *procesos de media móvil*.

Se define el proceso de media móvil de orden q , $MA(q)$ con la siguiente ecuación:

$$x_t = c + a_t - \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}$$

donde a_t un proceso de ruido blanco con varianza σ^2 , c y θ_i son constantes a determinar en el proceso de ajuste.

Este proceso es la suma de los procesos $a_t, \theta_1 a_{t-1}, \dots$ y $\theta_q a_{t-q}$, por tanto, siempre será estacionario, para cualquier valor del parámetro, a diferencia de los procesos AR .

Para simplificar, consideremos la constante del modelo $c = 0$, e introduciendo la notación de operadores:

$$x_t = a_t - \theta_1 B a_t + \theta_2 B^2 a_t + \dots + \theta_q B^q a_t$$

puede escribirse de forma más compacta:

$$x_t = \theta_q(B) a_t$$

donde: $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$

A diferencia de los procesos AR , un proceso $MA(q)$ es siempre estacionario, por ser suma de procesos estacionarios, eso sí, exigimos que las raíces del operador $\theta_q(B) = 0$ sean, en módulo, mayores que la unidad para que el proceso sea *invertible*. Para que un proceso lineal sea útil para la obtención de previsiones debe tener una propiedad, además de la estacionariedad, que es la llamada invertibilidad.

❖ **Propiedades del proceso $MA(q)$:**

Las propiedades de este proceso se obtienen de forma genérica, en este caso no vamos a tratar los procesos más importantes $MA(1)$ y $MA(2)$ por separado.

- Esperanza y varianza:

$$E[x_t] = \mu = c$$

llamando $\tilde{x}_t = x_t - \mu$

$$E[\tilde{x}_t] = \sigma_x^2 = \gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2$$

- Función de autocovarianza:

$$\gamma_k = (-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) \sigma^2 \quad k = 1, \dots, q,$$

$$\gamma_k = 0 \quad k > q,$$

resultando que un proceso $MA(q)$ tiene exactamente los q primeros coeficientes de la función de autocovarianzas distintos de cero.

- Función de autocorrelación simple (*fas*):

$$\rho_k = \frac{\sum_{i=0}^{i=q} \theta_i \theta_{k+i}}{\sum_{i=0}^{i=q} \theta_i^2} \quad k = 1, \dots, q$$

$$\rho_k = 0, \quad k > q$$

donde $\theta_0 = -1$, y $\theta_k = 0$ para $k \geq q + 1$.

Existe una dualidad entre procesos *AR* y *MA*, de manera que la *fap* de un *MA*(*q*) tiene la estructura de la *fas* de un *AR*(*q*), y la *fas* de un *MA*(*q*) tiene la estructura de la *fap* de un *AR*(*q*). La *Figura 3.2* siguiente presenta estas funciones para procesos *MA*(1) y *MA*(2).

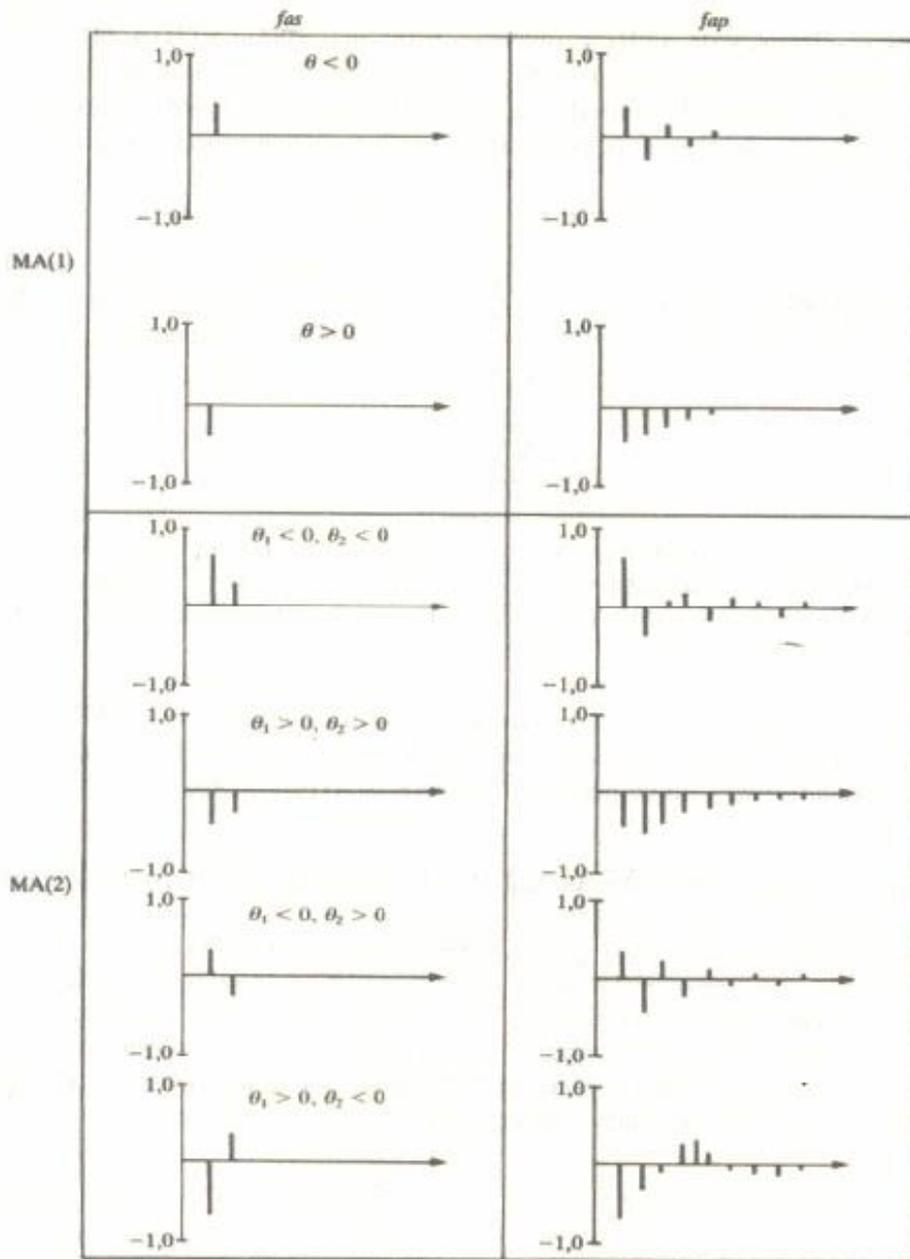


Figura 3.2

3.4 PROCESOS $ARMA(p, q)$

En la práctica se observa que las series no son puras AR o MA , sino que presentan parte AR y parte MA . Los procesos $ARMA(p, q)$ son combinación de estructuras autorregresivas y de media móvil que tienen una parte $AR(p)$ y una parte $MA(q)$.

El proceso $ARMA(p, q)$ será:

$$x_t = c + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

o notación compacta,

$$\phi_p(B)x_t = c + \theta_q(B)a_t$$

El proceso será estacionario si las raíces de $\phi_p(B) = 0$ están fuera del círculo unidad, e invertible si lo están las de $\theta_q(B) = 0$. En la práctica los órdenes p y q suelen ser menores o igual a 2.

La *fas* y la *fap* de los procesos $ARMA$ es el resultado de la superposición de sus propiedades AR y MA : en las *fas* ciertos coeficientes iniciales que dependen del orden de la parte MA y después un decrecimiento dictado por la parte AR . En las *fap* valores iniciales dependientes del orden del AR seguidos del decrecimiento debido a la parte MA . Esta estructura compleja hace que el orden de un proceso $ARMA$ sea difícil de identificar en la práctica.

La tabla siguiente resume estas características.

	<i>fas</i>	<i>fap</i>
$AR(p)$	Muchos coeficientes no nulos	Primeros p no nulos, resto 0
$MA(q)$	Primeros p no nulos, resto 0	Muchos coeficientes no nulos
$ARMA(p, q)$	Muchos coeficientes no nulos	Muchos coeficientes no nulos

3.5 PROCESOS NO ESTACIONARIOS. MODELOS $ARIMA(p, d, q)$

Antes de comenzar con el desarrollo de los modelos $ARIMA$, introducimos un elemento llamado *operador diferencia regular*, siendo B el operador retardo, llamamos operador diferencia regular, ∇ a:

$$\nabla = 1 - B$$

que actúa de manera que, al aplicarlo sobre un término de un proceso estocástico, obtenemos:

$$\nabla x_t = (1 - B)x_t = x_t - x_{t-1}$$

Este operador puede también aplicarse de forma reiterada:

2º orden:
$$\nabla^2 x_t = (1 - B)(1 - B)x_t = x_t - 2x_{t-1} + x_{t-2}$$

orden d :
$$\nabla^d x_t = (1 - B)^d x_t$$

Para el estudio de procesos estocásticos mediante *ARMA* se requiere que estos sean estacionarios, ya que la mayoría de los procesos son no estacionarios. Pero se pueden transformar en estacionarios tomando diferencias, si la no estacionariedad se debe a que la función de medias no es constante en el tiempo. Es decir, con la diferenciación solo conseguimos que la serie sea estacionaria en media.

Diremos que un proceso estocástico es integrado de orden d cuando es necesario diferenciarlo d veces para conseguir un proceso estacionario, en la práctica no suele ser necesario aplicar una diferencia regular superior a 2.

Si el modelo estacionario resultante de la diferenciación es un $ARMA(p, q)$, el modelo original diremos que es un $ARIMA(p, d, q)$, siendo p el orden de la parte autorregresiva, d el orden de diferenciación regular, y q el orden de la parte de medias móviles.

Es decir, que si

$$\nabla^d x_t = ARMA(p, q)$$

entonces

$$x_t = ARIMA(p, d, q)$$

y la expresión del modelo quedaría del siguiente modo:

$$\phi_p(B)\nabla^d x_t = c + \theta_q(B)a_t$$

3.6 PROCESOS ESTACIONALES

Existe un tipo de no estacionariedad peculiar, que es la denominada *estacionalidad*. Los comportamientos estacionales son la repetición periódica regular de pautas de comportamiento en la serie. Ocurre así en multitud de series económicas, por ejemplo de datos mensuales, en las que con periodo un año se repiten comportamientos: máximos en verano y mínimos en invierno para temas relacionados con turismo.

Estos comportamientos entrañan una falta de estacionariedad en la serie, que se puede eliminar tomando diferencias entre observaciones separadas por el periodo estacional. Por tanto, sería necesario introducir un operador nuevo llamado *operador diferencia estacional* ∇_s , siendo B el operador retardo que se define como:

$$\nabla_s = 1 - B^s$$

que actúa de modo similar al operador diferencia regular ∇ , pero de manera que obtenemos el incremento sufrido por el proceso con respecto a s periodos atrás:

$$\nabla_s x_t = (1 - B^s)x_t = x_t - x_{t-s}$$

pero debemos de tener cuidado en no confundir con el operador diferencia regular ∇^s :

$$\nabla_s = 1 - B \neq (1 - B)^s = \nabla^s$$

para el caso de tener que diferenciar estacional D veces:

$$\nabla_s^D x_t = (1 - B^s)^D x_t$$

❖ Estructura ARIMA estacional

Los proceso que presenten pautas cíclicas no estacionarias de comportamiento, pueden ser estacionarizados respecto a ellas por la aplicación, quizás reiterada, del operador diferencia estacional. Al número D de diferencias estacionales que debemos aplicar se le llama orden de diferenciación, rara vez es superior a uno.

Además de esta estructura no estacionaria, un proceso puede incorporar comportamientos cíclicos estables. Tales comportamientos serán modelizados de un modo análogo al empleado en los modelos ARMA no estacionales, mediante una representación autorregresiva, de media móviles o una combinación de ambas. Puesto que refleja la relación entre valores decalados s periodos (ciclo estacional), los polinomios empleados serán polinomios en B^s :

Así un modelo AR puramente estacional sería:

$$(1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps})x_t = a_t$$

o abreviadamente:

$$\Phi_P(B^s)x_t = a_t$$

donde $\Phi_P(B^s)$ es un polinomio en B^s de orden P , y el modelo se denomina $AR(P)_s$, indicando el subíndice s que se trata de un modelo estacional con esa longitud de ciclo.

Analógicamente podrían establecerse los modelos MA y ARMA estacionales:

$$x_t = \Theta_Q(B^s)a_t \quad \text{es un } MA(Q)_s$$

$$\Phi_P(B^s)x_t = \Theta_Q(B^s)a_t \quad \text{es un } ARMA(P, Q)_s$$

Si incluimos la posibilidad de que el modelo estacional fuera no estacionario, requiriendo el uso del operador diferencia estacional D veces para estacionarizarlo, tendremos definido el modelo ARIMA estacionario:

si:
$$\nabla_s^D x_t = ARMA(P, Q)_s$$

entonces:
$$x_t = ARIMA(P, D, Q)_s$$

❖ *Características*

Las características de los modelos *AR*, *MA* y *ARMA* estacionales son similares a los de sus correspondientes no estacionales (llamados regulares). Así, las estructuras de los correlogramas simple y parcial son las mismas, pero se presentan sólo en los retardos múltiplos del ciclo estacional.

Por ejemplo, mientras un modelo $AR(p)$ regular presenta un correlograma parcial con los p primeros coeficientes no nulos, su correspondiente $AR(P)_s$ estacional presenta no nulos los coeficientes de los retardos $s, 2s, \dots, Ps$, es decir los P primeros retardos múltiplos de s , en cambio, en el correlograma simple de este modelo, el decrecimiento mezcla de exponenciales y sinusoidales amortiguadas ubicado en los retardos $s, 2s, \dots$

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

Para concluir este capítulo, buscamos máxima generalidad y poder de representación, debemos considerar el caso, por otro lado muy frecuente, de procesos cuya estructura incorpore comportamientos regulares y estacionales, combinando los dos grupos de modelos *ARIMA* comentados.

Con ello, el modelo más general de la familia de modelos *ARIMA* sería el $ARIMA(p, d, q)(P, D, Q)_s$, cuya expresión es:

Notación compacta: $p, d, q, P, D, Q \leq 2$

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D x_t = c + \theta_q(B)\Theta_Q(B^S)a_t$$

s : Longitud del periodo estacional.

Polinomios	$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$
Parte Regular	$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$
Polinomios	$\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS}$
Parte Estacional	$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS}$

4. METODOLOGÍA BOX-JENKINS

4.1 INTRODUCCIÓN

Después de presentar los modelos generales *ARIMA* en el apartado anterior, en este apartado trataremos de explicar paso a paso el procedimiento estadístico para ajustar a una serie los modelos *ARIMA* según la metodología *Box-Jenkins*. Conociendo las fases más importantes de esta metodología también conoceremos los criterios y herramientas para validar el modelo, con el fin de saber explorarlo y obtener predicciones fiables. Y para acabar, resolvemos uno de los casos tratados durante la beca de colaboración, que ilustra todo el proceso explicado.

Box y Jenkins en su desarrollo de modelos estadísticos para series temporales fijaron distintas fases para su modelado. Básicamente estas fases se resumen en los siguientes pasos:

- La identificación del modelado *ARIMA* adecuado a los datos de la serie: recogida de datos de la serie, representación gráfica, análisis de la estacionariedad, transformaciones previas adecuadas para conseguir la estacionariedad, eliminación de la tendencia si es necesario e identificación efectiva del modelo asociándole la estructura *ARIMA* adecuada.
- Estimación del modelo previamente identificado: cálculo de los estimadores del modelo y residuales.
- Validación del modelo: contrastes para ver si el modelo es adecuado.
- Predicción: selección de los periodos de predicción y cálculo de estadísticos para evaluar la capacidad predictiva.

Las fases a llevar a cabo se resumen en el siguiente esquema (*Figura 4.1*)

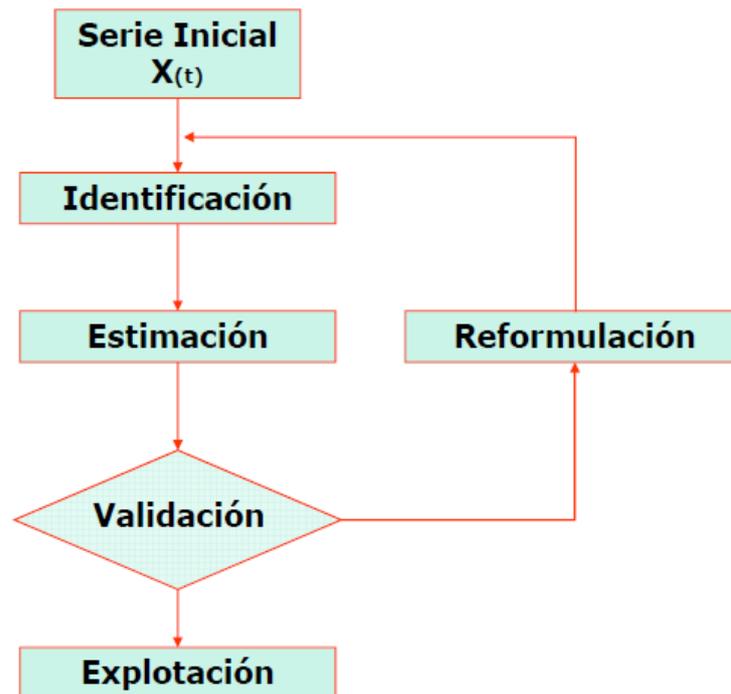


Figura 4.1

4.2 IDENTIFICACIÓN DEL MODELO

❖ *Identificación de la no estacionariedad. Estacionarización*

La fase de identificación del modelo empieza por la identificación de la no estacionariedad de la serie, es decir, comprobar si la serie es estacionaria mediante la observación de la representación gráfica de la serie en función del tiempo, y fijarse en los 3 aspectos siguientes:

- Estacionaria en varianza: ¿serie homocedástica?
- Estacionaria en media: ¿serie con media constante?
- Comportamiento no estacional: ¿serie con comportamiento estacional?

Si la serie temporal presenta no estacionariedad en varianza (varianza no constante), se puede corregir con la *transformación de Box-Cox*. Cuando la relación entre desviación típica y media sea una exponencial de la forma: $\sigma = k \cdot m^\alpha$ la transformación que nos permite estabilizar en sustituir la serie X_t por la transformada:

$$Y_t = \begin{cases} X_t^{1-\alpha} & \alpha \neq 1 \\ \ln X_t & \alpha = 1 \end{cases}$$

La serie así transformada es ya homocedástica. Esta transformación debe ser realizada como primera operación, allí donde sea necesaria, pudiendo pasar a continuación a la siguiente faceta de análisis.

La falta de no estacionariedad en media se detecta mediante la representación gráfica de la serie donde nos indicará si la serie temporal presenta una tendencia, y también con el correlograma simple, que este caso se observa un decrecimiento lento y lineal de los coeficientes de autocorrelación simples (*fas*). Una tendencia lineal será corregida tomando primeras diferencias (diferencia regular), que será el caso más frecuente. Una tendencia no lineal suele llevar en la práctica al uso de dos diferencias como mucho.

Una vez que hayamos tratado la no estacionariedad en varianza y en media, habrá que comprobar si la serie presenta comportamiento estacional, esta falta de estacionariedad se trata mediante la diferencia estacional.

En la práctica, para detectar la estacionalidad puede utilizarse alguno de los caminos:

- El gráfico de la serie da una idea de los posibles periodos estacionales.
- Las funciones de autocorrelación y autocorrelación parcial estimadas también validan los procesos estacionales (*retardos s, 2s, 3s, ...*).
- El periodograma, que es una figura que transforma la serie temporal de su dominio temporal al dominio de las frecuencias, si no hay picos destacables en el periodograma no hay estacionalidad, y si hay picos, cada pico destacable identifica un periodo. A cada amplitud destacable le corresponde una frecuencia cuya inversa es el período estacional.

❖ **Identificación de la estructura $ARMA(p, q)$ REGULAR**

Nos fijaremos en los primeros 6, 8 o 10 coeficientes de autocorrelación de la *fas* y de la *fap*. Con esta información decidiremos que estructura tiene la parte regular contrastado los correlogramas experimentados con los teóricos.

Si la serie tiene, por ejemplo, estructura $ARMA(1,1)$: identificar inicialmente como el modelo más sencillo (por ejemplo $AR(1)$), estimar este modelo y ver si los residuos siguen un modelo $MA(1)$ o son ruido blanco.

❖ **Identificación de la estructura $ARMA(P, Q)$ ESTACIONAL**

Nos fijaremos en los coeficientes de autocorrelación situados sobre el retardo s (longitud del periodo estacional) y múltiplos de s de la *fas* y de la *fap*. Con esta información decidiremos que estructura tiene la parte estacional contrastando los correlogramas experimentados con los teóricos, de forma análoga a la parte regular.

4.3 ESTIMACIÓN DEL MODELO

No es objeto de este trabajo el estudio detallado de los procedimientos de estimación que el uso de la metodología *Box-Jenkins* y los modelos *ARIMA* conllevan, pues su complejidad y nivel de especialización sobrepasa su planteamiento esencialmente aplicado. Ello no obstante, se va a comentar algunos de los puntos relacionados con este proceso, siquiera sea para ilustrar el funcionamiento de las distintas herramientas software disponible.

La estimación del modelo consiste en la identificación del valor de los parámetros del mismo que permite un mejor ajuste entre datos pasados reales y las correspondientes predicciones del modelo, el procedimiento habitual de búsqueda de ese mejor ajuste emplea el criterio de los mínimos cuadrados.

El análisis de la significación de los parámetros: constante y, coeficientes de los polinomios autorregresivo y de medias móviles tanto de la parte regular como la parte estacional se realiza mediante contrastes de hipótesis del tipo:

$$H_0: p_i = 0 \quad vs \quad H_1: p_i \neq 0$$

donde p_i es el parámetro del modelo cuya significación queremos analizar.

Para ello, consideraremos el *p-value* de la tabla de estimación del software utilizado (*SPSS, Statgraphics*).

- Si *p-value* < 0.05, entonces aceptamos como bueno el valor estimado para el parámetro p_i .
- Si *p-value* \geq 0.05, entonces consideramos que el parámetro es igual a 0.

4.4 VALIDACIÓN DEL MODELO

En esta fase hay que comprobar que los residuos del modelo ajustado siguen un proceso de ruido blanco:

- Siguen una distribución Normal,
- De media cero,
- Varianza constante y
- Están incorrelados (independientes)

La diagnosis de los residuos estará basada en:

- Correlograma *fas* y *fap* de la serie residual del modelo ajustado: hay que comprobar que los coeficientes no son significativos para afirmar que residuos incorrelacionados o independientes.

- El test de *Box-Pierce* o test de *Ljung-Box* proporcionan información sobre si el conjunto de los primeros palos de la función de autocorrelación de los residuos son cero. Estos test indican problemas cuando el *p-valor* es bajo, por ejemplo menos que 0.05, cuanto mayor sea, hay más evidencia a favor de que los residuos son ruido blanco.
- Test de rachas: estudia si hay muchos puntos consecutivos por encima o por debajo de la mediana de los residuos.
- Papel probabilístico normal de los residuos: distribución normal y media 0.
- Test para determinar si tanto la media como la varianza son iguales al principio y al final de la serie de residuos.

Cabe señalar que depende de la herramienta informática se emplea un test u otro, por ejemplo el *Statgraphics* usa el contraste de *Box-Pierce* para la incorrelación de los residuos, mientras el *SPSS* chequea que no hay autocorrelación en los residuos mediante el contraste de *Ljung – Box*.

4.5 EXPLOTACIÓN DEL MODELO: PREVISIÓN

Y para acabar, una vez validado el modelo a través fundamentalmente del análisis de los residuos, el siguiente paso es su explotación o uso. Para nosotros el modo más importante de usar un modelo es la obtención de valores futuros de la serie estudiada.

Para obtener esas previsiones haremos evidentemente uso del modelo *ARIMA* que se ha identificado como correcto, proyectándolo hacia el futuro.

El diagrama siguiente resume las fases de la metodología *ARIMA*. (*Figura 4.2*)

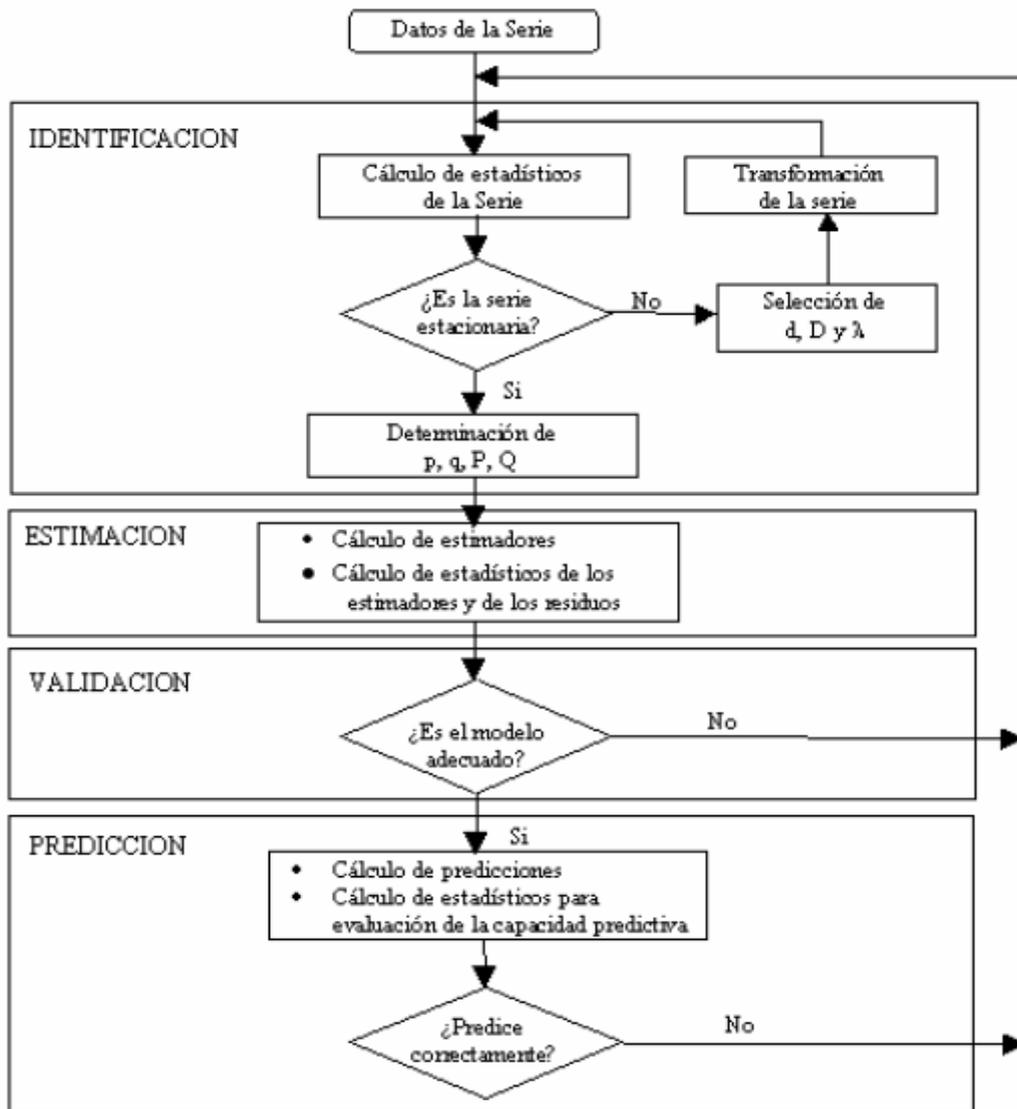


Figura 4.2

4.6 EJEMPLO ILUSTRATIVO

Veamos cómo ajustar a una serie modelo *ARIMA* con el software *Statgraphics*, se trata de una serie que almacena datos mensuales de las acciones en bolsa durante un período de 8 años (96 períodos) de una cadena de supermercados que domina el mercado de alimentos al por menor de un área metropolitana de tamaño medio. Pero a partir del período 73, la cadena de supermercados cambió de propietarios, con este cambio, los dueños apostaron por una campaña publicitaria agresiva que anunciaba su propia línea de producto introducida justo después, el resultado fue un crecimiento rápido y espectacular de sus acciones en bolsa.

Los datos se almacenan en *markets.xls*, y están compuestos por los seis años anteriores a la compra de la compañía y los dos años posteriores a la misma.

El software *Statgraphics* dispone del procedimiento *Métodos Descriptivos* de la opción *Análisis Series Temporales* que permite realizar las tareas de identificación de modelos *ARIMA*. Asimismo *Statgraphics* también dispone del procedimiento de *Predicción*, que permite estimar y realizar predicciones con series temporales a partir de la metodología de *Box-Jenkins*. Para llevar a cabo todo el proceso de modelar nuestra serie a un modelo *ARIMA*, vamos a seguir los pasos explicados anteriormente: identificación, estimación, validación y predicción.

❖ *Identificación*

Comenzamos el proceso de identificación del modelo con la representación gráfica de la serie temporal de acciones en bolsa, para percatarse de la presencia de una posible no estacionariedad debida a la media o a la varianza (no constantes).

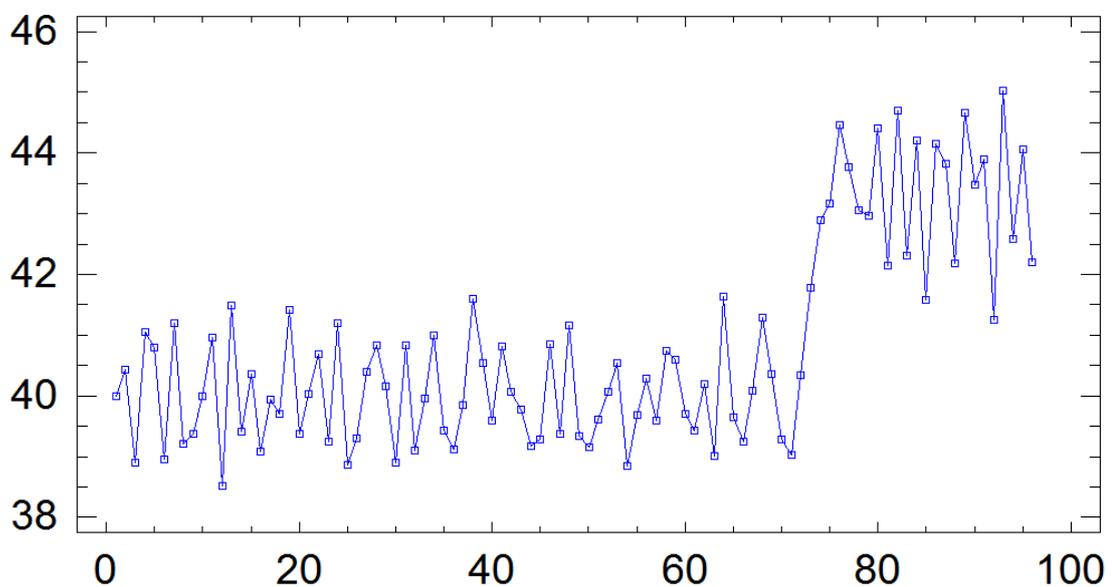


Figura 4.3

La *Figura 4.3* presenta el gráfico de los datos de acciones en bolsa, y muestra claramente un cambio de nivel que experimenta la serie alrededor del período 73 debido a la compra de compañía. El comportamiento post compra sigue igual, pero entorno de una media diferente. Entonces para ajustar la serie correctamente a un modelo *ARIMA* hay que tener en cuenta solo los datos antes del impacto de la compra de la compañía, y después agregamos al modelo validado el efecto de cambio de nivel mediante una o más variables, a este desarrollo le llamamos *análisis de intervención*, lo veremos con más detalles en capítulos posteriores.

Entonces vamos a desarrollar un modelo para la serie antes de la intervención, para esto definimos el período de estimación seleccionado casos, basándose en el rango de tiempo entre 1 y 72.

Escoger un buen modelo *ARIMA* implica observar la serie y decidir si es necesario aplicar una transformación, logarítmica o de raíz cuadrada, para estabilizar la serie y, a continuación observar los gráficos de la función de autocorrelación (*fas*) y de la función de autocorrelación parcial (*fap*) para determinar las ordenes de *ARIMA*.

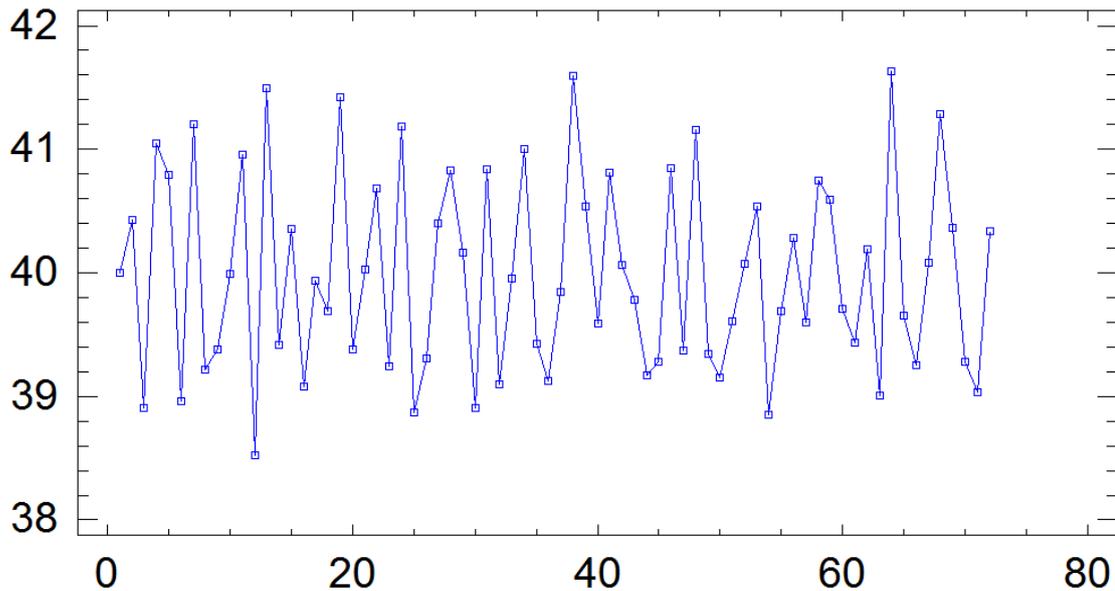


Figura 4.4

El gráfico de *Figura 4.4* muestra a simple vista que no presenta estacionalidad, además tiene una media y varianza constantes, entonces la serie es estacionaria, por lo que no parece necesario aplicar transformaciones a los datos.

A continuación se elige la opción gráfica de las funciones de autocorrelación, que produce el gráfico de *Figura 4.5*, este gráfico presenta la *fas* y las *fap* de la serie, en la *fas* observamos que tiene un primer retardo muy significativo, la *fap* en cambio tiene los tres primeros retardos significativos que van decayendo (decrecimiento exponencial). Por tanto entre los modelos tentativos estaría un proceso de media móvil de orden 1, es decir un *ARIMA(0,0,1)*.

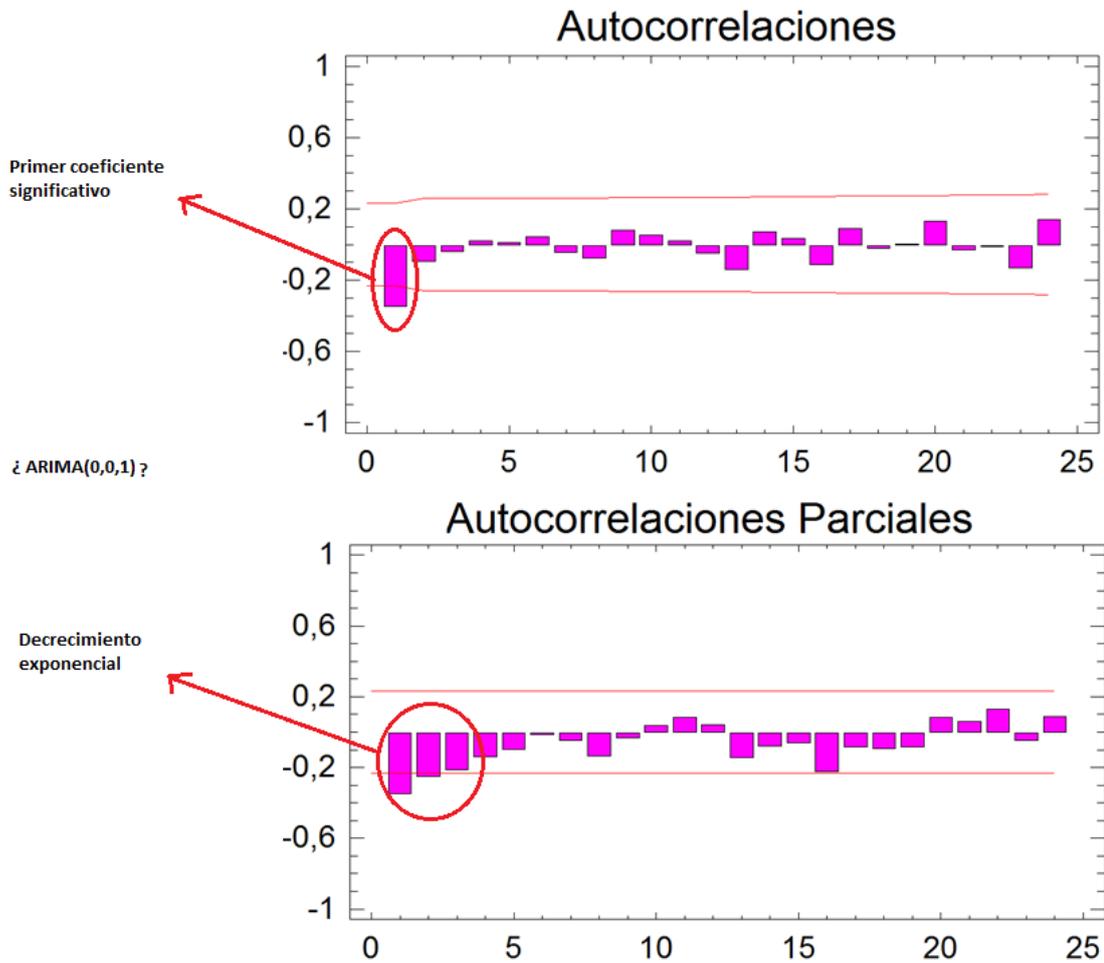


Figura 4.5

❖ **Estimación**

Una vez identificado la serie, es decir, cuando se conoce el modelo $ARIMA(p, d, q)$ que puede seguir, es preciso estimar los parámetros. Ya que el modelo tentativo es $ARIMA(0,0,1)$, equivalente al $MA(1)$.

Es un proceso que sigue la ecuación $x_t = c + a_t - \theta_1 a_{t-1}$, entonces los parámetros que han de estimar son: la constante c y θ_1 , esto lo hace el programa, obteniéndose la tabla de resultados:

Resumen del Modelo ARIMA				
Parámetro	Estimación	Error Estd.	t	P-Valor
MA(1)	0,672149	0,0890377	7,54904	0,000000
Media	39,9878	0,0288022	1388,36	0,000000
Constante	39,9878			

Predicción hacia atrás: si
 Varianza de ruido blanco estimada = 0,49677 con 70 grados de libertad
 Desviación típica del ruido blanco estimado = 0,704819
 Número de iteraciones: 5

Los parámetros son significativos debido a que los p – valores < 0.05 . Por tanto, el modelo queda expresado con un modelo $ARIMA(0,0,1)$ con constante:

❖ Validación

Ahora nos faltaría realizar la validación del modelo comprobando si a_t es ruido blanco o no. Para ello, utilizamos en primer lugar el gráfico de los residuos de la *Figura 4.6*, que muestra un aspecto estacionario en media y varianza. En segundo lugar, la opción gráfica de las funciones de autocorrelación de los residuos de la *Figura 4.7* muestra que las autocorrelaciones residuales se distribuyen aleatoriamente dentro de la franja que las encierra definida por los límites probabilísticos, lo que inicialmente indica que hemos realizado un buen ajuste.

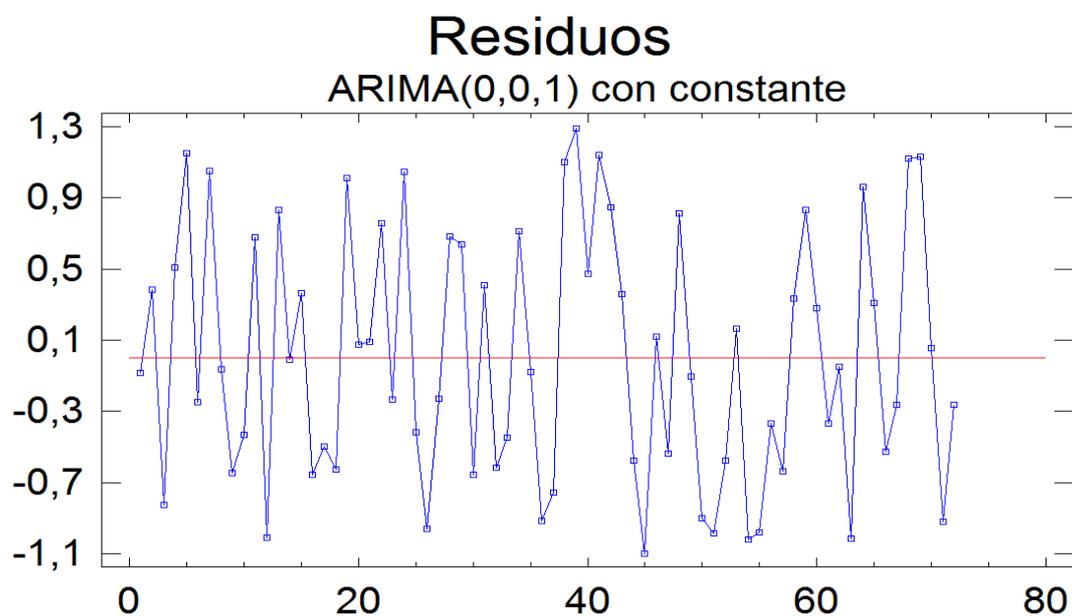
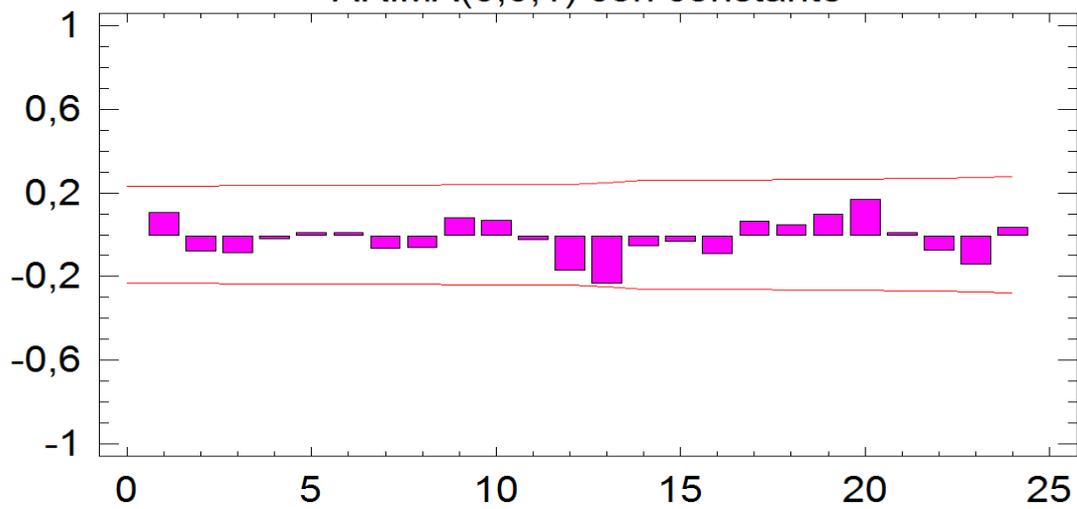


Figura 4.6

Autocorrelaciones Residuales

ARIMA(0,0,1) con constante



Autocorrelaciones Parciales Residuales

ARIMA(0,0,1) con constante

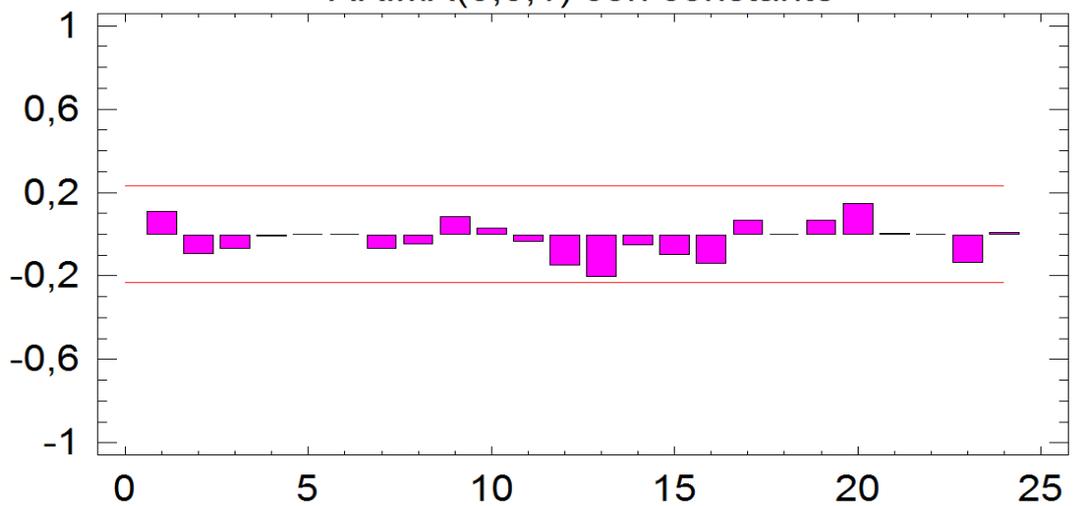


Figura 4.7

No obstante, podemos utilizar el test de *Box-Pierce* que proporciona información bajo la hipótesis nula que los primeros coeficientes de la función de autocorrelación simple de los residuos son cero versus la hipótesis alternativa que son distinto de cero, es decir:

$$H_0: p_i = 0 \quad vs \quad H_1: p_i \neq 0$$

cuya salida es lo siguiente:

Test de Box-Pierce

Test basado en las primeras 24 autocorrelaciones
 Test estadístico de muestra grande = 15,7758
 P-valor = 0,864698

La autocorrelación no son significativos, debido a que el $p - valores > 0.05$. Y para acabar nos queda por comprobar la distribución normal del ruido y su media nula.

En la gráfica siguiente (Figura 4.8), vemos como los residuos se ajustan aproximadamente a una recta diremos que siguen modelo normal, y que para el percentil 50 el valor que nos da se aproxima a 0, con lo cual la media es nula.

Gráfico de Probabilidad Normal

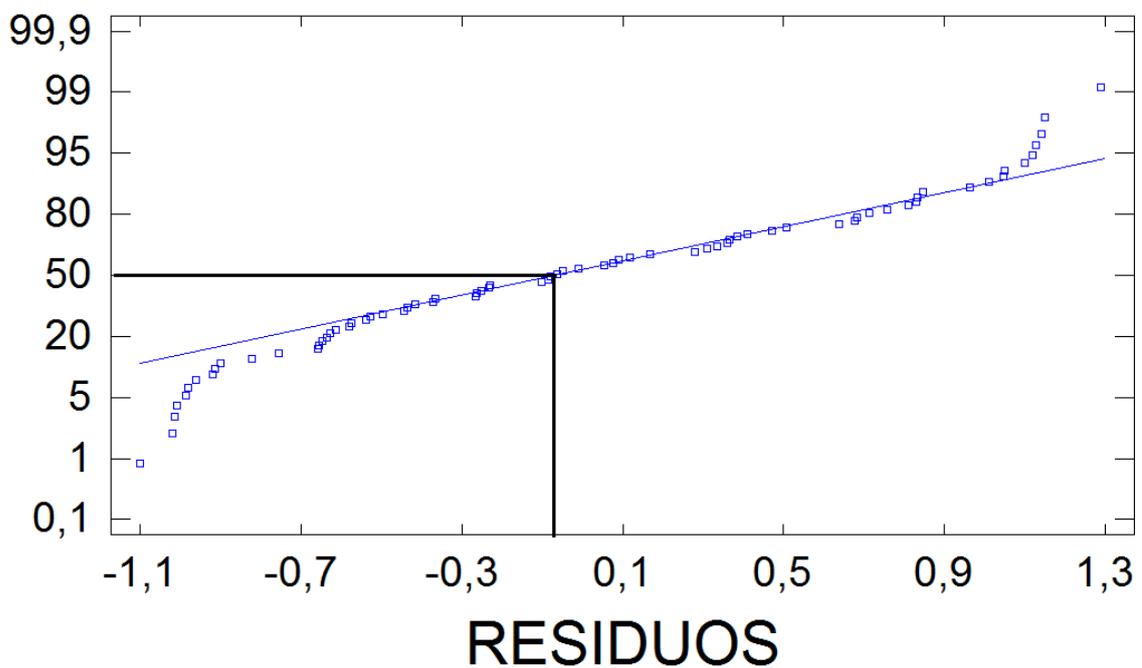


Figura 4.8

Con esto, damos por validado el modelo $ARIMA(0,0,1)$ para la serie de los datos de acciones en bolsa.

❖ **Explotación del modelo**

Nuestra seria temporal sigue un modelo $ARIMA(0,0,1)$, sin diferenciación, ni transformación anteriormente, porque de entrada ya era estacionario, entonces es equivalente a un modelo $MA(1)$, cuya ecuación de predicción será:

$$x_t = 39,9878 + a_t - 0,6721 a_{t-1}$$

donde a_t debe distribuirse como $N(0, \sigma_a^2)$.

5. MODELOS DE FUNCIÓN DE TRANSFERENCIA

5.1 INTRODUCCIÓN

En los modelos de función de transferencia, el objetivo es relacionar dos o más series temporales elaborando modelos causales de predicción. Se considera la forma de relacionar una serie temporal, denominada *output* en función de una u otras series temporales, que se denominan *inputs*. También se considera a priori que existe una causalidad unidireccional desde los inputs hacia el output, desechando la posibilidad de *feedback*.

Estos modelos son muy utilizados en todos los campos científicos para evaluar respuestas dinámicas. Si las variables inputs son controlables, estos modelos permiten simular y evaluar políticas alternativas. Si no lo son, ofrecen la posibilidad de estudiar cómo ciertos “escenarios”, definidos por posibles evoluciones de la variable explicativa, afectan a la variable respuesta.

Además estos modelos son muy útiles para elaborar predicciones, aunque dependerá del intervalo que se tenga entre observaciones.

La construcción de los modelos de Función de Transferencia sigue las mismas etapas que la construcción de los modelos univariantes de series temporales:

Identificación, Estimación, Verificación y Predicción.

5.2 MODELOS DE FUNCIÓN DE TRASFERENCIA CON UN ÚNICO INPUT

El caso más sencillo es el modelo de función de transferencia con un sólo input, que se puede extender sin grandes dificultades a modelos con varios inputs.

En un sistema lineal de un único input y único output, la serie output Y_t y la serie input X_t se relacionan a través de un filtro lineal de la siguiente forma:

$$Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots + N_t = v(B)X_t + N(t) \quad (1)$$

donde $v(B) = v_0 + v_1 B + v_1 B^2 + \dots$ se refiere como función de transferencia del filtro de *Box* y *Jenkins* (1970), y $N(t)$ es el ruido del sistema que es independiente de la serie input X_t . Los coeficientes de $v(B)$ se conocen como *la función de respuesta al impulso*. Para que el sistema (1) sea estable se debe cumplir que una variación finita produzca una variación también finita en el output. Esto es deberá cumplirse que

$$\sum_{j=0}^{\infty} v_j = g \quad (2)$$

siendo g finito. El valor de g representa el cambio total en Y_t motivado por un cambio unitario en X_t mantenido en el tiempo.

Los propósitos de la modelización de función de transferencia son identificar y estimar la función de transferencia $v(B)$ y el modelo del ruido $N(t)$ sobre la base de la información que proporcionan la serie input y la serie output.

El modelo (1) es inestimable, debido a que en él aparece un número infinito de parámetros. Este problema se puede aliviar al expresar la función de transferencia como el cociente de dos polinomios (rationales) finitos:

$$v(B) = \frac{\omega(B)B^b}{\delta(B)} \quad (3)$$

donde $\omega(B) = \omega_0 - \omega_1 B - \dots - \omega_s B^s$, $\delta(B) = 1 - \delta_1 B - \dots - \delta_r B^r$ y b es un parámetro de retardo que representa el tiempo que transcurre antes de que el impulso en la variable input produzca un efecto en la variable output. Para que el sistema sea estable, se asume que las raíces de $\delta(B) = 0$ caen dentro del círculo unidad.

Sustituyendo (3) en (1), se obtiene:

$$Y_t = \frac{\omega(B)B^b}{\delta(B)} X_t + N_t \quad (4)$$

Por otro lado, el término de error no tiene por qué ser necesariamente un ruido blanco. Se puede suponer, con carácter general, que N_t sigue un proceso *ARIMA*, aunque sigue independiente de la variable de input X_t , es decir:

$$N_t = \frac{\theta(B)}{\phi(B)(1-B)^d} a_t \quad (5)$$

con $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ y $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ de manera que todas las raíces de ambos polinomios caen dentro del círculo unidad siendo $(1 - B)^d$ el operador diferencias consecutivas utilizado para inducir estacionariedad y a_t es un ruido blanco.

Sustituyendo (5) en (4), se obtiene:

$$(1 - B)^d Y_t = \frac{\omega(B)B^b}{\delta(B)} (1 - B)^d X_t + \frac{\theta(B)}{\phi(B)} a_t \quad (6)$$

A partir de (6) puede observarse que si el proceso del término de error (6) no es estacionario y, por tanto, debe diferenciarse d veces para conseguir la estacionariedad, este mismo orden de diferenciación recae tanto sobre la variable dependiente u output como sobre la variable explicativa o input.

Sin embargo, en la práctica, para construir un modelo de función de transferencia, se requiere que tanto la variable dependiente (output) como la variable explicativa (input) sean estacionarias, para lo cual obviamente no tiene por qué cumplirse que ambas variables necesiten el mismo orden de diferenciación para lograr su estacionariedad. Por

otra parte, una vez lograda la estacionariedad en ambas variables, el proceso N_t deberá ser un $ARIMA(p, q)$, con lo cual el modelo de función de transferencia se puede escribir como:

$$y_t = \frac{\omega(B)B^b}{\delta(B)} x_t + \frac{\theta(B)}{\phi(B)} a_t \quad (7)$$

donde $y_t = (1 - B)^{d'} Y_t$ y $x_t = (1 - B)^d X_t$, siendo d' el orden de diferenciación de Y_t y d el orden de diferenciación de X_t .

Este modelo de función de transferencia se representa gráficamente como en la *Figura 5.1*.

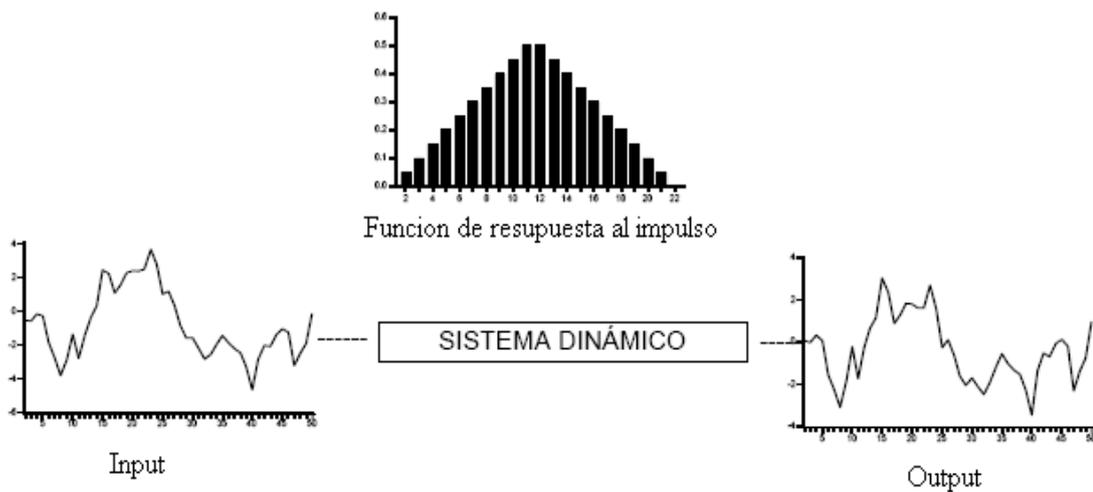


Figura 5.1

Resulta inmediato que, una vez que conozcamos los valores de $\omega(B)$, $\delta(B)$ y b , podremos obtener de inmediato los coeficientes de respuesta del impulso v_j , igualando los coeficientes de B^j de ambos lados de la siguiente ecuación:

$$\delta(B)v(B) = \omega(B) B^b \quad (8)$$

que también puede escribirse como:

$$(1 - \delta_1 B - \dots - \delta_r B^r)(v_0 + v_1 B + v_2 B^2 + \dots) = (\omega_0 - \omega_1 B - \dots - \omega_s B^s) \quad (9)$$

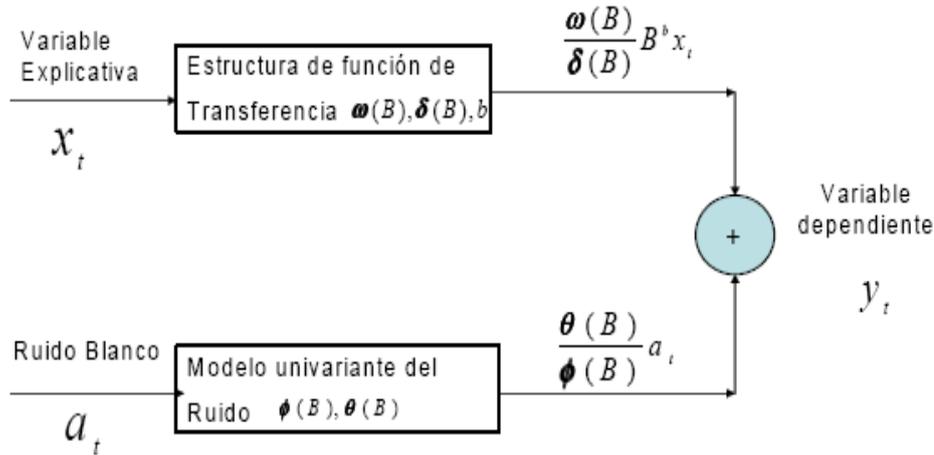


Figura 5.2

A partir de esta expresión, se obtiene:

$$\begin{aligned}
 v_j &= 0 && \text{para } j < b \\
 v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} + w_0 && \text{para } j = b \\
 v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} + w_{j-b} && \text{para } j = b + 1, \dots, b + s \\
 v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} && \text{para } j > b + s
 \end{aligned}$$

Además, las r ponderaciones de respuesta al impulso $v_{b+s}, v_{b+s-1}, \dots, v_{b+s-r+1}$, sirven como valores iniciales para la ecuación en diferencias:

$$\delta(B)v_j = 0 \quad \text{para } j < b \quad (10)$$

Además las r ponderaciones de respuesta al impulso $v_{b+s}, v_{b+s-1}, \dots, v_{b+s-r+1}$, sirven como valores iniciales para la ecuación en diferencias:

De esta forma, las ponderaciones de respuesta al impulso presentan las siguientes características:

- Hay b coeficientes iguales a 0 (v_0, v_1, \dots, v_{b-1}).
- Hay $s - r + 1$ coeficientes ($v_b, v_{b+1}, \dots, v_{b+s-r}$) que no siguen un patrón de comportamiento fijo.
- Hay r coeficientes de respuesta al impulso que sirven como valores iniciales ($v_{b+s-r+1}, v_{b+s-r+2}, \dots, v_{b+s}$)
- Para $j > b + s$, los coeficientes de respuesta al impulso (v_j) siguen el patrón autorregresivo de orden r .

Por lo tanto, si conocemos los coeficientes de respuesta al impulso (v_j), podemos determinar los valores de b, r, s como sigue:

1. El valor de b se determinará teniendo en cuenta el hecho de que $v_j = 0$ para $j < b$ y $v_b \neq 0$.
2. El valor de r se determinará por el patrón de comportamiento de los coeficientes de respuesta al impulso de forma similar a como se identifica el orden p de un modelo *ARIMA* a través de la función de autocorrelación.
3. En cuanto al valor de s . Para un valor dado de b , si $r = 0$ podrá identificarse fácilmente s por cuanto se cumplirá $v_j = 0$ para $j > b + s$. Si $r \neq 0$ el valor de s se encontrara observando cuando comienza a decaer el patrón de la función de respuesta al impulso.

ALGUNAS FUNCIONES DE RESPUESTA AL IMPULSO TÍPICAS

En la práctica, los valores de r y s raramente exceden de 2. Los casos más habituales que suelen encontrarse en variables económicos son:

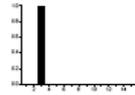
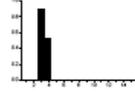
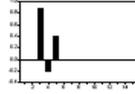
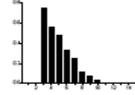
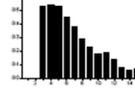
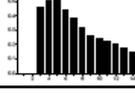
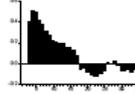
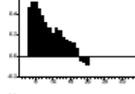
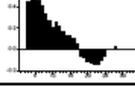
(b, r, s)	Función de Transferencia	Ponderaciones de respuesta al impulso
$r = 0$		
$(3, 0, 0)$	$v(B)x_t = \omega_0 x_{t-3}$	
$(3, 0, 1)$	$v(B)x_t = (\omega_0 - \omega_1 B)x_{t-3}$	
$(3, 0, 2)$	$v(B)x_t = (\omega_0 - \omega_1 B - \omega_2 B^2)x_{t-3}$	
$r = 1$		
$(3, 1, 0)$	$v(B)x_t = \frac{\omega_0}{(1-\delta_1 B)} x_{t-3}$	
$(3, 1, 0)$	$v(B)x_t = \frac{(\omega_0 - \omega_1 B)}{(1-\delta_1 B)} x_{t-3}$	
$(3, 1, 2)$	$v(B)x_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1-\delta_1 B)} x_{t-3}$	
$r = 2$		
$(3, 2, 0)$	$v(B)x_t = \frac{\omega_0}{(1-\delta_1 B - \delta_2 B^2)} x_{t-3}$	
$(3, 2, 1)$	$v(B)x_t = \frac{(\omega_0 - \omega_1 B)}{(1-\delta_1 B - \delta_2 B^2)} x_{t-3}$	
$(3, 2, 2)$	$v(B)x_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1-\delta_1 B - \delta_2 B^2)} x_{t-3}$	

Figura 5.3

5.3 MODELOS DE FUNCIONES DE TRANSFERENCIA CON MÚLTIPLES INPUTS

La formulación puede fácilmente ser extendida para varios inputs, por ejemplo, con dos inputs ($X_{1,t}$ y $X_{2,t}$) tendríamos que:

$$Y_t = v_1(B)X_{1,t} + v_2(B)X_{2,t} = \frac{\omega_1(B)B^{b_1}}{\delta_1(B)}X_{1,t} + \frac{\omega_2(B)B^{b_2}}{\delta_2(B)}X_{2,t}$$

donde:

$$v_i(B) = v_{i,0} + v_{i,1}B + v_{i,2}B^2 + \dots = \frac{\omega_i(B)B^{b_i}}{\delta_i(B)}, \quad i = 1, 2$$

b_i = tiempo muerto para el input $X_{i,t}$ $i = 1, 2$

$\omega_i = \omega_{i,0} - \omega_{i,1}B - \dots - \omega_{i,s_i}B^{s_i}$ $i = 1, 2$

$$\delta_i = 1 - \delta_{i,1}B - \dots - \delta_{i,r_i}B^{r_i} \quad i = 1,2$$

Y extendiendo esta formulación para M inputs, $i = 1,2, \dots, M$, el resultado en forma compacta se escribiría:

$$Y_t = \sum_{i=1}^M v_i(B)X_{i,t} = \sum_{i=1}^M \frac{\omega_i(B)B^{b_i}}{\delta_i(B)} X_{i,t}$$

5.4 IDENTIFICACIÓN DE LA FUNCIÓN DE TRANSFERENCIA

❖ *La función de covarianza y correlación cruzada*

La función de correlación cruzada es una medida muy útil de la fuerza y la dirección de la relación entre dos variables aleatorias. Dados dos procesos estocásticos x_t e y_t para $t = 0, \pm 1, \pm 2, \dots$ con medias μ_x y μ_y , decimos que x_t e y_t son conjuntamente estacionarios si X_t e Y_t son ambos procesos univariantes estacionarios y la covarianza cruzada entre x_t e y_t es una función de la diferencia temporal.

La covarianza cruzada se define:

$$\gamma_{yx}(j) = E[(y_t - \mu_y)(x_{t-j} - \mu_x)] \quad j = 0, \pm 1, \pm 2, \dots$$

o

$$\gamma_{xy}(j) = E[(x_t - \mu_x)(y_{t-j} - \mu_y)] \quad j = 0, \pm 1, \pm 2, \dots$$

Dando valores a j se obtiene la función de covarianza cruzadas.

La función de covarianza cruzada presenta las siguientes características:

Las covarianzas cruzadas no necesitan ser simétricas alrededor de 0

- $\gamma_{yx}(j) \neq \gamma_{yx}(-j)$
- $\gamma_{yx}(j) = \gamma_{xy}(-j)$

Se deben usar series estacionarias ya que las covarianzas cruzadas sólo podrán interpretarse cuando se efectúen sobre series estacionarias.

Dado que suponemos que la variable x puede influir en y pero que la inversa no puede darse, es decir, existe casualidad unidireccional de x hacia y , pero no existe *feedback*, estamos suponiendo de partida que solamente nos interesarán valores no negativos de j en $\gamma_{yx}(j)$.

La función de correlación cruzada se define estandarizando la función de covarianza cruzada, es decir:

$$\rho_{yx}(j) = \frac{\gamma_{yx}(j)}{\sigma_y \sigma_x}$$

siendo σ_y y σ_x las desviación típica de las variables y y x respectivamente.

❖ **La relación entre la función de correlación cruzada y la función de transferencia**

Dado el modelo de función de transferencia:

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \dots + N_t$$

donde se supone que y_t y x_t son estacionarias y que $\mu_x = \mu_y = 0$.

Entonces:

$$\rho_{yx}(j) = \frac{\sigma_x}{\sigma_y} [v_0 \rho_{xx}(j) + v_1 \rho_{xx}(j-1) + \dots + v_j + v_{j+1} \rho_{xx}(1)]$$

A partir de esta expresión se observa que la relación entre la función de correlación cruzada $\rho_{yx}(j)$ y la función de respuesta al impulso aparece claramente contaminada por la estructura de autocorrelación de la serie input x_t .

Sin embargo, si la serie de input fuese un ruido blanco, todas las autocorrelaciones de x_t serían nulas y la expresión anterior se simplificaría considerablemente, obteniéndose:

$$\rho_{yx}(j) = \frac{\sigma_x}{\sigma_y} v_j$$

de manera que

$$v_j = \frac{\sigma_x}{\sigma_y} \rho_{yx}(j)$$

Así, si a_t es un ruido blanco, la función de respuesta al impulso v_j es directamente proporcional a la función de correlación cruzada $\rho_{yx}(j)$.

❖ **El preblanqueo**

Para ello, si escribimos el modelo de función de transferencia como:

$$y_t = v(B)x_t + N_t$$

y suponemos que x_t viene generada por un suceso ARMA del tipo:

$$\phi_x(B) = \theta_x(B)\alpha_t$$

donde $\phi_x(B)$ y $\theta_x(B)$ son los polinomios autorregresivo y de media móvil del proceso *ARMA* para la serie input que cumplen las condiciones de estacionariedad e invertibilidad usuales y α_t es un proceso de ruido blanco. Así, si escribimos α_t :

$$\alpha_t = \frac{\phi_x(B)}{\theta_x(B)} x_t$$

que se denomina serie input preblanqueada. Aplicando la misma transformación de preblanqueo a la serie output, obtenemos:

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)} y_t$$

Si denominamos $\varepsilon_t = \frac{\phi_x(B)}{\theta_x(B)} N_t$, tendríamos:

$$\beta_t = v(B)\alpha_t + \varepsilon_t$$

ecuación a partir de la cual se obtiene fácilmente

$$v_j = \frac{\sigma_\beta}{\sigma_\alpha} \rho_{\beta\alpha}(j)$$

Ya que α_t es un proceso de ruido blanco. Así pues, si conociésemos los elementos poblacionales $\rho_{\beta\alpha}(j)$, σ_β y σ_α , obtendríamos los valores de v_j y, de acuerdo con las características de los coeficientes de la función de respuesta al impulso, podríamos identificar el modelo de transferencia pertinente. En la práctica, al no conocer los valores poblacionales de $\rho_{\beta\alpha}(j)$, σ_β y σ_α , tendremos que estimar estos coeficientes a partir de datos muestrales.

❖ *Identificación de la función de respuesta al impulso*

Consideramos el modelo de función de transferencia con un solo input

$$y_t = v(B)x_{t-b} + N_t = \frac{\omega(B)}{\delta(B)} B^b x_t + N_t$$

con $N_t = \frac{\theta(B)}{\phi(B)} \alpha_t$. La identificación de la función de respuesta al impulso $v(B)$ se realizará siguiendo las siguientes etapas:

- *Etapa 1:* Considerando que y_t y x_t son estacionarias, preblanquear el input de acuerdo con su representación *ARMA*

$$\phi_x(B) = \theta_x(B)\alpha_t \quad \Longrightarrow \quad \alpha_t = \frac{\phi_x(B)}{\theta_x(B)} x_t$$

donde α_t es un ruido blanco con media cero y varianza σ_α^2 .

- *Etapa 2:* Filtrar la serie del output utilizando el mismo filtro que preblanqueó el input, obteniendo:

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)} y_t$$

- *Etapa 3:* Calcular la función de correlación cruzada muestral entre β_t y α_t para estimar v_j de acuerdo con la expresión:

$$\hat{v}_j = \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha} \hat{\rho}_{\beta\alpha}(j)$$

Para analizar qué coeficientes de la función de respuesta al impulso son significativos, se puede utilizar el contraste de significación de las correlaciones cruzadas.

- *Etapa 4:* Identificar r (orden del polinomio $\delta(B)$) y s (orden del polinomio $\omega(B)$) de acuerdo a las reglas descritas anteriormente. Una vez que se escogen b , r y s , se pueden encontrar estimaciones preliminares de los coeficientes de $\delta(B)$ y $\omega(B)$ a partir de sus relaciones con v_j . Así, se obtiene una estimación preliminar de la función de transferencia:

$$\hat{v}(B) = \frac{\hat{\omega}(B)}{\hat{\delta}(B)} B^b$$

Las etapas descritas se realizan suponiendo:

- se procede a identificar $v(B)$ considerando que no hay error, esto es $N_t = 0$.
- Se parte de que las series y_t y x_t son estacionarias.

Una vez identificada la función de transferencia, se puede calcular la serie del ruido estimado:

$$\hat{N}_t = y_t - \hat{v}(B)x_t = y_t - \frac{\hat{\omega}(B)}{\hat{\delta}(B)} B^b x_t$$

identificándose el proceso *ARMA* generador de los residuos mediante sus funciones de autocorrelación *fas* y *fap*. Se obtiene así:

$$\phi(B)\hat{N}_t = \theta(B)a_t$$

conjugando ambas identificaciones, se llega al modelo de función de transferencia:

$$y_t = \frac{\omega(B)}{\delta(B)} B^b x_t + \frac{\theta(B)}{\phi(B)} a_t$$

En la construcción del modelo, se asume que todas las variables y_t , x_t y n_t son estacionarios. Así, para series no estacionarios se tienen que llevar a cabo transformaciones para la estabilización de la varianza y diferenciación de las series para conseguir primero la estacionariedad.

5.5 VERIFICACIÓN DE UN MODELO DE FUNCIÓN DE TRANSFERENCIA

Después de que el modelo haya sido identificado y sus parámetros estimados, antes de aceptarlo conviene efectuar contrastes diagnósticos para comprobar su validez antes de utilizarlo en predicciones. Clasificaremos los contrastes que pueden aplicarse en: sobre los parámetros estimados, y sobre los residuos.

❖ *Contrastes sobre los parámetros*

Debe comprobarse si las raíces de los polinomios autorregresivos cumplen las condiciones de estacionariedad, y que el modelo sea estable mediante las raíces del polinomio $\delta(B) = (1 - \rho_1 B) \dots (1 - \rho_r B^r)$, es decir $|\rho_i| < 1$.

Comprobamos también que las estimaciones de sus parámetros sean significativas de forma similar a los de los modelos *ARIMA* ($p - value < 0.05$).

❖ *Contraste sobre los residuos*

Consisten, como siempre, en comprobar si los residuos son una secuencia de variables aleatorias normales con varianza e independientes. Esto supone realizar contrastes para confirmar:

1. Normalidad.
2. Que su media es cero.
3. Que están incorrelados entre sí.

Estos contrastes son análogos al caso univariante.

6. CASO DE ESTUDIO

En este capítulo se presentan las predicciones obtenidas para el consumo de gas natural al aplicar la metodología *ARIMA* y la metodología de Función de Transferencia (Regresión Dinámica).

Para este estudio, tenemos datos mensuales del consumo de gas natural, en total, son 168 observaciones. Para la construcción del modelo de Función de Transferencia, se analizan las posibles relaciones existentes entre la serie de consumo “ C_t ” y otras variables: climatológicas y económicas. Parece lógico pensar que una mayor subida del precio del gas natural, puede limitar su consumo. Por eso, incluimos para este estudio series temporales adicionales como son: la temperatura “ T_t ” y el precio “ P_t ”.

Cabe apuntar, que el proceso de modelado se va a hacer sobre las 156 primeras observaciones, para hacer predicciones de las 12 últimas muestras, para ver si el modelo estimado se acerca al modelo real.

En primer lugar se determina cual es el modelo *ARIMA* subyacente así como los resultados gráficos numéricos que describen los resultados obtenidos con dicho modelo.

En segundo lugar se determina el modelo de Función de Transferencia obtenido teniendo en cuenta sólo la variable explicativa temperatura, y después incluyendo la variable precio, mostrando los resultados en ambos casos.

Por último, se realiza una comparación de los resultados obtenidos con cada uno de los modelos estudiados.

Para medir la bondad del ajuste, hay varias medidas utilizadas en el modelado de las series temporales, entre ellas elegimos el RMSE como criterio de medida de precisión: Root Mean Squared Error (Raíz cuadrada del error cuadrático medio), es una medida de cuanto se desvía la serie dependiente (serie a modelar) del nivel pronosticado por el modelo, expresado en las mismas unidades que la serie dependiente.

$$RMSE = \sqrt{\frac{\sum_{t=1}^m (\hat{Y}_t - Y_t)^2}{m}}$$

Entre varios modelos alternativos, seleccionamos aquel de menor RMSE. Cabe señalar que el RMSE no sólo sirve para comparar diferentes modelos de ajuste que ya están validados, sino, también para comparar y valorar diferentes configuraciones del mismo modelo durante la fase de identificación.

6.1 MODELO *ARIMA*

Empezamos nuestro estudio con el modelo *ARIMA*, para obtener predicciones fiables, hay que ajustar la serie C_t al que más se aproxime. Los pasos seguidos para la identificación del modelo *ARIMA* son:

1. Formulación del modelo
2. Identificación de términos
3. Estimación de parámetros
4. Validación del modelo
5. Predicción

Cada uno de estos pasos se ha explicado con detalle en el capítulo 4. Para la elaboración de cada uno de estos pasos se ha utilizado el software comercial *SPSS*.

❖ *Formulación del modelo*

El *SPSS* dispone del procedimiento *ARIMA* que permite trabajar y realizar predicciones con series temporales a partir de la metodología de *Box-Jenkins*. Para llevar a cabo las tareas de identificación se utilizan las opciones de gráficos de series temporales. Para comenzar la fase de identificación realizamos una representación gráfica de la serie consumo C_t con el objeto de comprobar su estacionariedad.

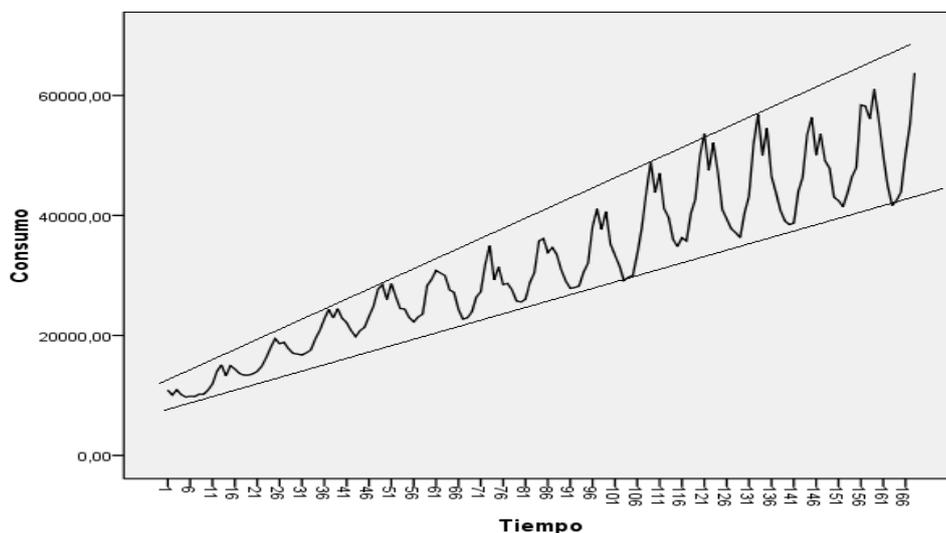


Figura 6.1

La *Figura 6.1* muestra la serie C_t en función del tiempo y se ve claramente que la serie no presenta síntomas de estacionariedad en ninguno de los tres aspectos: varianza, media y estacionalidad.

Para observar mejor la estacionalidad, representamos el periodograma por frecuencia de la serie. Según se indica en la *Figura 6.2*, cuyo segundo pico corresponde a la frecuencia 0,0836, es decir 12 meses.

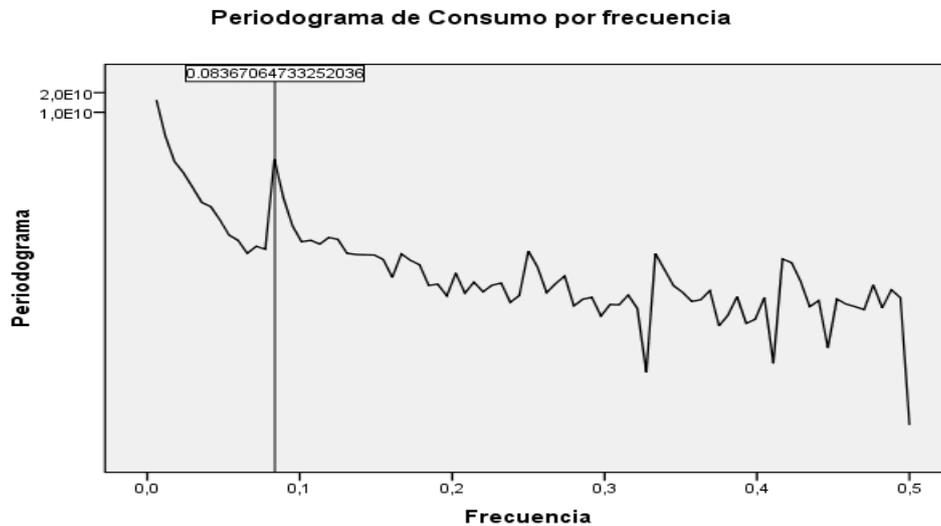


Figura 6.2

La estacionalidad, así como la estacionariedad también pueden detectarse a través de las funciones de autocorrelación estimadas *fas* y *fap* (*ACF* y *ACF* parcial respectivamente).

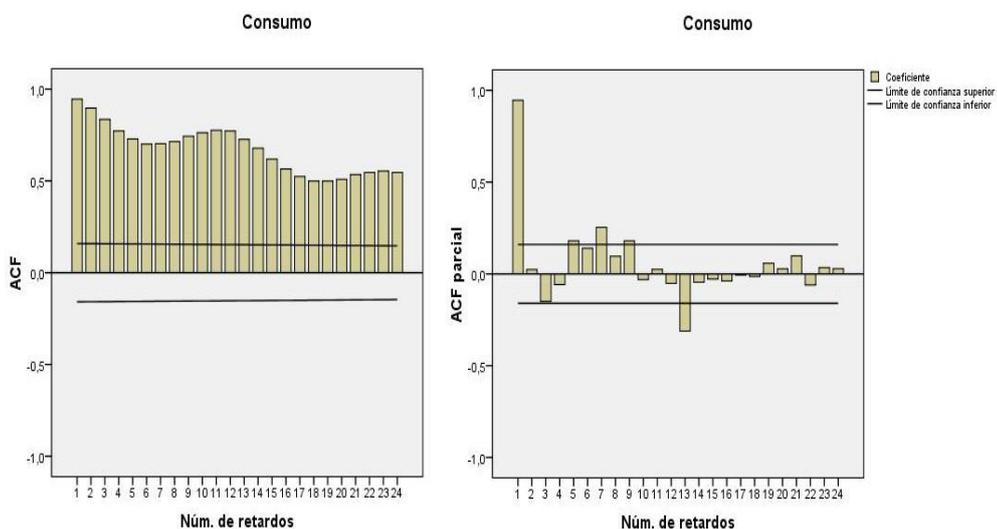


Figura 6.3

Se observa que las funciones de autocorrelación y autocorrelación parcial estimados de la *Figura 6.3* también validan el periodo estacional de la serie. Además, para una cantidad grande de retardos la *ACF* se configura en forma de abanico que completa su ciclo girando sobre el eje de abscisas para una cantidad de retardos igual al periodo estacional. Por otro lado, la *ACF* parcial presenta estructura de coeficientes

significativos para retardos periódicos. La *ACF* y la *ACF* parcial deben considerarse a la vez, pues a veces intercambian sus papeles en el comportamiento estacional. Asimismo, los coeficientes de la *ACF* no decaen rápidamente, lo que indica falta de estacionariedad en media.

Este resultado nos lleva a tomar logaritmos y diferenciar la parte regular ∇ de la serie original con el objetivo de atenuar la falta de estacionariedad en varianza y media. Y para la estacionalidad diferenciamos su parte estacional ∇_{12} . Lo que da como resultado la *Figura 6.4* donde se observa finalmente que la serie transformada es estacionaria por no tener tendencia y una varianza más o menos constante a lo largo del intervalo de tiempo, excepto en algunos valores atípicos presentes.

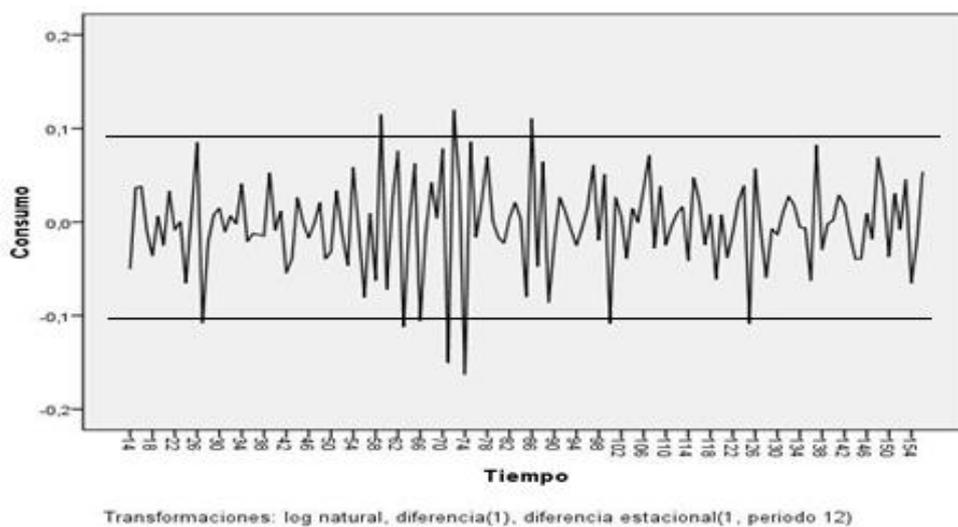


Figura 6.4

❖ *Identificación del modelo*

Una vez conseguida la estacionariedad de la serie, ahora resta por identificar el orden de la parte autorregresiva *AR* y la parte de medias móviles *MA*. Para realizar esta tarea utilizaremos la *ACF* y la *ACF* parcial para las que se obtuvieron la estacionariedad mediante la transformación de logaritmo natural, el operador diferencial regular y estacional. La *Figura 6.5* presenta el correlograma simple y parcial de la serie $\nabla\nabla_{12} \log C_t$. Comenzando con la parte regular, el hecho de que aparezca un coeficiente significativo en la *fas* y un decrecimiento exponencial en los primeros coeficientes de la *fap* sugiere un proceso *MA*(1) para la parte regular. Con la relación a la estacionalidad, existen retardos significativos en 12 y 24 de igual amplitud en la *fap*, en cambio en la *fas* sólo sale uno en el retardo 12. La hipótesis más simple es una

$AR(2)_{12}$. Como conclusión de este análisis, todo indica que la serie se ajusta a primera vista a un modelo de tipo $ARIMA(0,1,1)(2,1,0)_{12}$,

Ya tenemos identificado completamente la serie inicial como un modelo $ARIMA(0,1,1)(2,1,0)_{12}$. Es decir, ya hemos realizado el trabajo más importante en la modelización de una serie temporal mediante la metodología de *Box-Jenkins*. Ahora falta averiguar si hemos acertado en la identificación del modelo mediante las fases de estimación y validación.

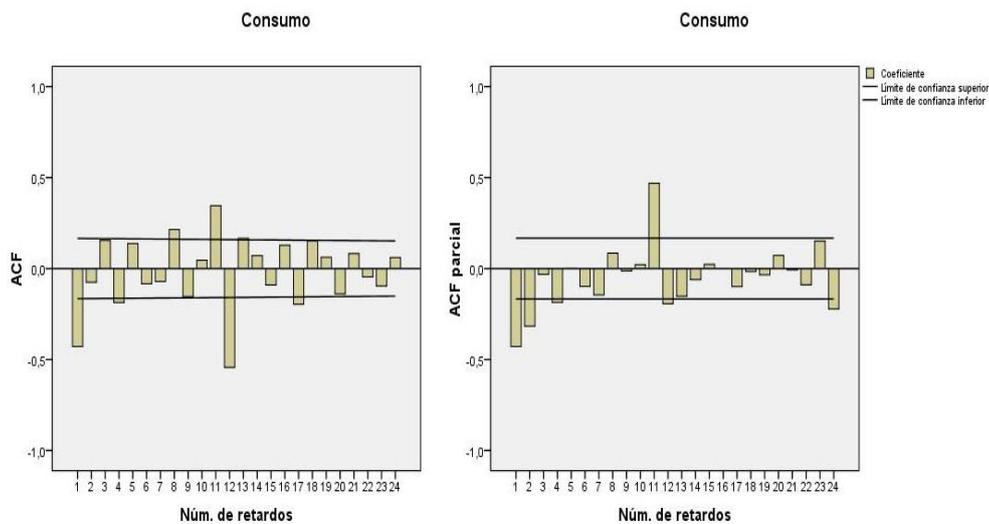


Figura 6.5

❖ **Estimación y validación del modelo**

Una vez identificado un modelo para la serie temporal, *SPSS* dispone del procedimiento *ARIMA* para realizar su estimación y validación así como para obtener predicciones con la serie.

Este procedimiento estima los parámetros del modelo *ARIMA* identificado anteriormente. Para ejecutar el procedimiento, hay que introducir los parámetros p, d y q de la parte regular, y los parámetros P, D y Q de la parte estacional del modelo identificado. En este caso, al ser el modelo $ARIMA(0,1,1)(2,1,0)_{12}$, los parámetros tendrán estos valores:

La parte regular: $p = 0, d = 1, q = 1$

La parte estacional: $P = 2, D = 1, Q = 0$

También hay que marcar la opción *Natural log* porque a la serie original le habíamos aplicado la transformación del logaritmo natural para conseguir la estacionariedad.

La primera estimación de los parámetros la hacemos incluyendo la estimación del término constante, y si sale que no es significativo en el resultado, volvemos a lanzar la estimación sin incluir el término constante. (Ver *Figura 6.6*)

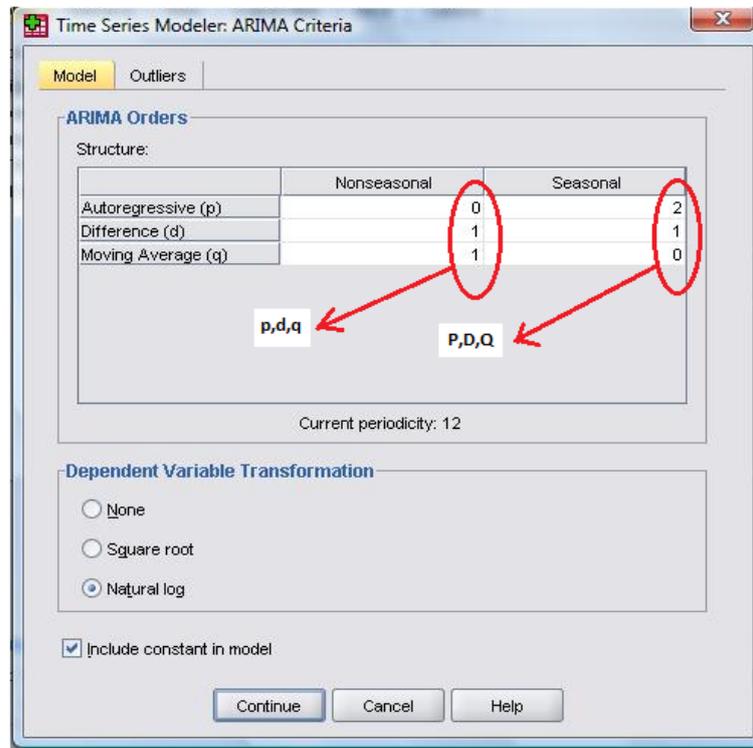


Figura 6.6

Al pulsar *Continue* se obtiene el ajuste a un modelo de *Box-Jenkins ARIMA (0,1,1)(2,1,0)₁₂*.

Parámetros del modelo ARIMA

				Estimación	ET	t	Sig.
Consumo-Model_1	Consumo	Log natural	Constante	-,002	,001	-2,160	,032
			Diferencia	1			
			MA Retardo 1	,446	,079	5,664	,000
			AR, estacional Retardo 1	-,699	,088	-7,905	,000
			Retardo 2	-,316	,087	-3,617	,000
			6 Diferenciación estacional	1			

Figura 6.7

Se observa en la *Figura 6.7* que todos los parámetros son muy significativos (*p* – *value* prácticamente nulos para todos los parámetros), si acaso nos hubiera salido un parámetro no significativo, hubiera sido que el modelo identificado no es el más adecuado, y habría que volver a identificar la serie con otro modelo.

Ahora pasamos a la fase de validación. Para asegurar la validez e idoneidad del modelo y la efectividad de las predicciones, los residuos estimados se deben comportar como un ruido blanco.

Para contrastar la incorrelación de los residuos presentamos la *fas* y la *fap* de los mismos estimados con el modelo diagnosticado.

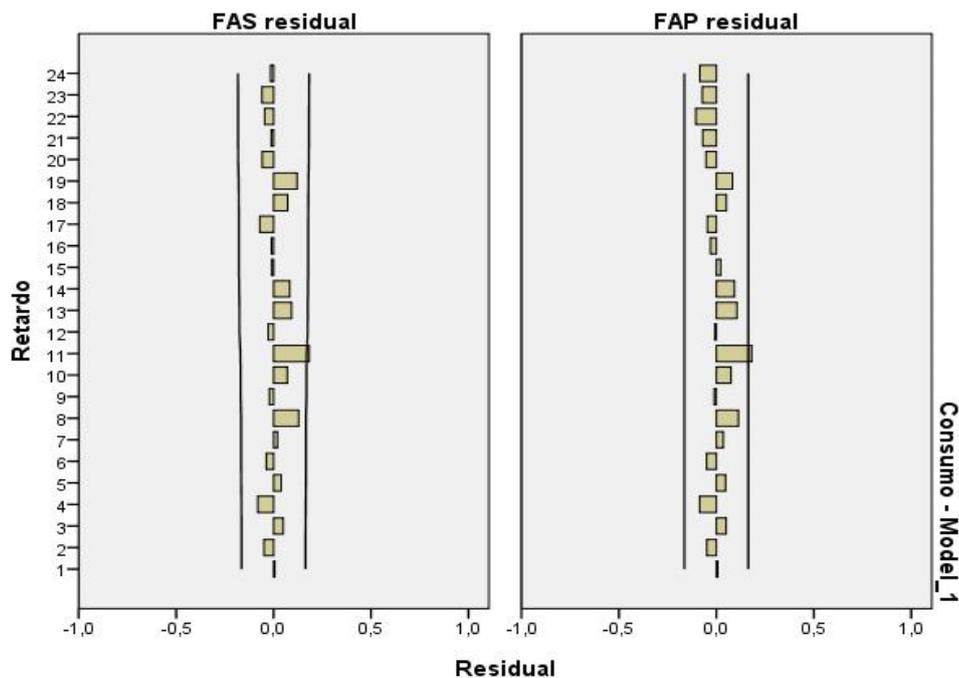


Figura 6.8

La Figura 6.8 presenta el gráfico de las funciones de autocorrelación de los residuos de la serie, se observa que hay algunos valores que superan el intervalo de confianza. El estadístico de *Ljung-Box Q(18)* es de 15,232 cuyo *p-value* es de $0,435 > 0,05$. Con este contraste no encontramos ninguna evidencia de dependencia serial en los residuos, y podemos confirmar que los residuos son incorrelados.

Entonces el modelo quedará validado. El modelo obtenido es:

$$(1 + 0,699B^{12} + 0,316B^{24})\nabla\nabla_{12}\log C_t = -0,02 + (1 - 0,446 B) \alpha_t$$

$$RMSE = 1159,030$$

La Figura 6.9 representa el consumo predicho para el último año (12 predicciones mensuales) junto con el consumo real para dicho año, se ve que se ajustan bastante bien las predicciones a la serie de consumo original.

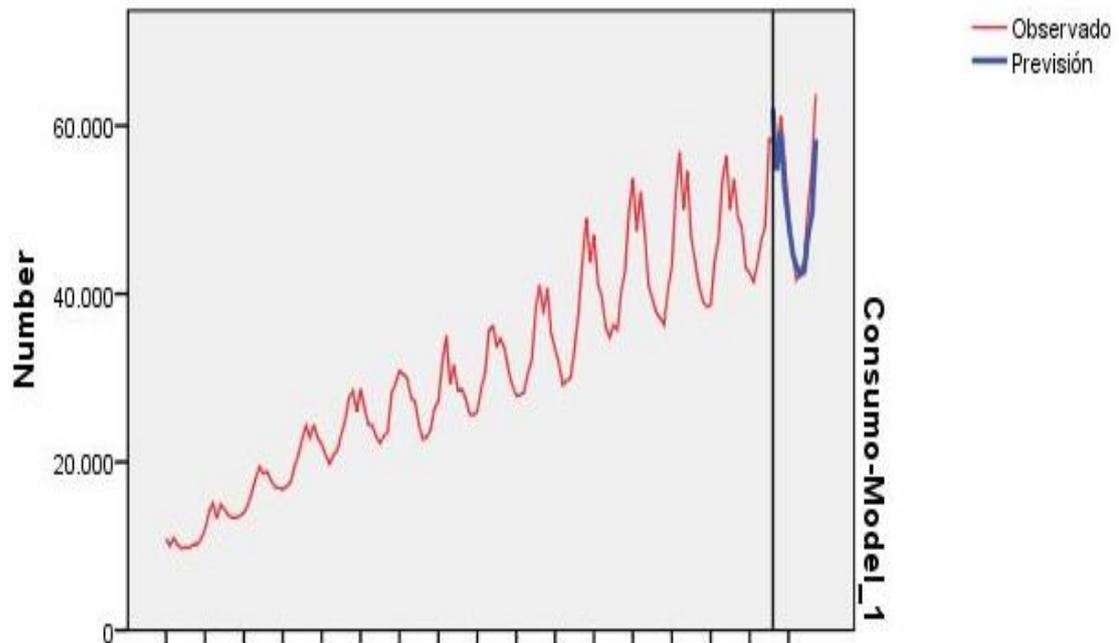


Figura 6.9

6.2 FUNCIÓN DE TRANSFERENCIA-INPUT TEMPERATURA

Para la elaboración del modelo de función de transferencia (Regresión Dinámica) y la obtención de las predicciones se ha utilizado el mismo software *SPSS*.

Para el desarrollo del modelo de regresión dinámica se ha utilizado la temperatura como serie explicativa, ya que ayuda a explicar el comportamiento del consumo del gas natural, con el objetivo de mejorar los resultados obtenidos con el ajuste *ARIMA*.

❖ *Formulación del modelo*

En primer lugar, presentamos la serie Temperatura T_t en función del tiempo para comprobar la correlación de la temperatura hacía el consumo.

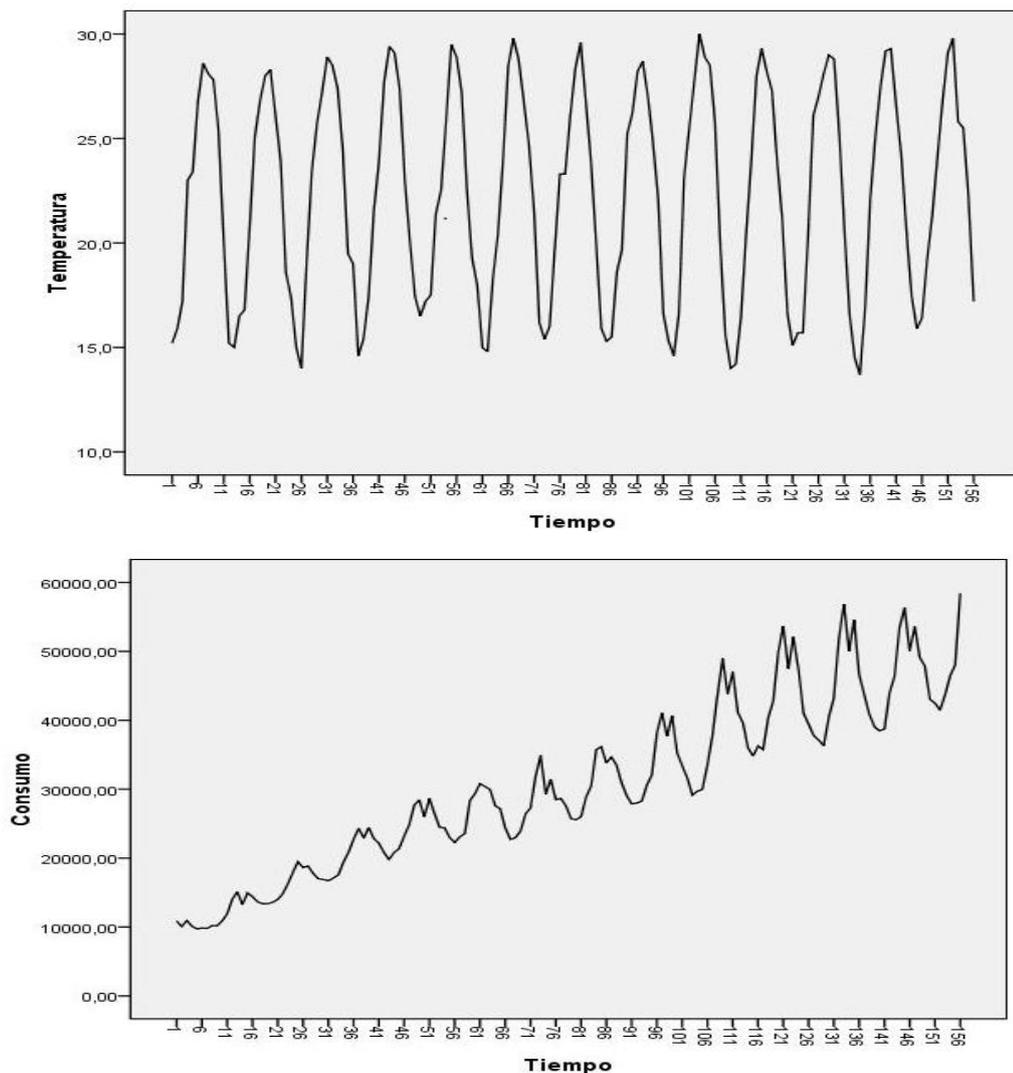


Figura 6.10

Se ve claramente en la *Figura 6.10* que los picos del consumo del gas natural coinciden con las temperaturas más bajas del año.

La serie dependiente C_t se ha transformado a $\nabla\nabla_{12}\log C_t$ para conseguir la estacionariedad, entonces la misma transformación la aplicamos a la serie T_t para dar lugar a la serie $\nabla\nabla_{12}\log T_t$, aunque se aprecia perfectamente en la gráfica que para conseguir la estacionariedad de la serie T_t basta con aplicarle la diferencia estacional ∇_{12} .

❖ *Identificación del modelo univariante*

A partir de las funciones de autocorrelación simple y parcial de esta variable estacionaria y siguiendo el mismo proceso del apartado anterior, se identifica para la serie T_t un proceso $ARIMA(1,1,1)(0,1,1)_{12}$.

El modelo estimado para la variable que consideramos input T_t fue el siguiente:

$$\nabla\nabla_{12}\log T_t = \frac{(1 - 0,997B)(1 - 0,929B^{12})}{(1 - 0,202B)} \alpha_t$$

$$RMSE = 1,103$$

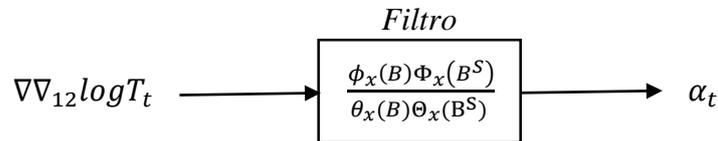
Esto será el modelo que se utilizará para realizar el preblanqueo del output para luego calcular la función de transferencia.

❖ **Preblanqueo del input**

Después de estimar los parámetros del modelo de ajuste para la variable explicativa T_t , toca preblanquear el input de acuerdo con su representación *ARIMA*:

$$\phi_x(B)\Phi_x(B^S)\nabla\nabla_{12}\log T_t = \theta_x(B)\Theta_x(B^S)\alpha_t \implies \alpha_t = \frac{\phi_x(B)\Phi_x(B^S)}{\theta_x(B)\Theta_x(B^S)} x_t$$

Los residuos del modelo α_t se pueden obtener pasando la serie $\nabla_{12}T_t$ por un filtro definido por los parámetros del modelo, este apartado lo hacemos con *MATLAB*.

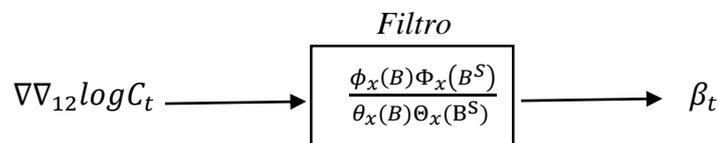


con
$$\frac{\phi_x(B)\Phi_x(B^S)}{\theta_x(B)\Theta_x(B^S)} = \frac{(1-0,202B)}{(1-0,997B)(1-0,929B^{12})}$$

Aplicando el comando *filter* en *MATLAB* pasar la serie T_t estacionaria por el filtro definido por los parámetros del modelo *ARIMA*, conseguimos la serie de residuos α_t .

❖ **Preblanqueo del output**

El siguiente paso es el preblanqueo del output C_t mediante el mismo modelo que el obtenido para el input *ARIMA*(1,1,1)(0,1,1)₁₂, y se obtiene de este modo la serie de residuos β_t , pero en este caso recordamos que la serie $\nabla\nabla_{12}\log C_t$ es la serie estacionaria de C_t después de la transformación.



❖ **Cálculo de la Función de Correlación Cruzada (CCF) entre α_t y β_t**

La correlación cruzada entre α_t y β_t juega un papel importante a la hora de identificar la función de transferencia entre el input y el output. Las funciones de

transferencia que relacionan el consumo C_t con la temperatura T_t , y β_t con α_t son iguales. Calculamos la correlación cruzada entre α_t y β_t en función del retardo k .

Al representar la CCF de β_t con α_t (Figura 6.11) nosotros sólo vamos a considerar que $\rho_{\beta\alpha}(k) = 0$ para $k < 0$ ya que sólo estamos interesados en conocer la relación lineal de T_t hacia C_t . Por ello sólo tendremos en cuenta la parte positiva (retardos positivos) de la CCF.

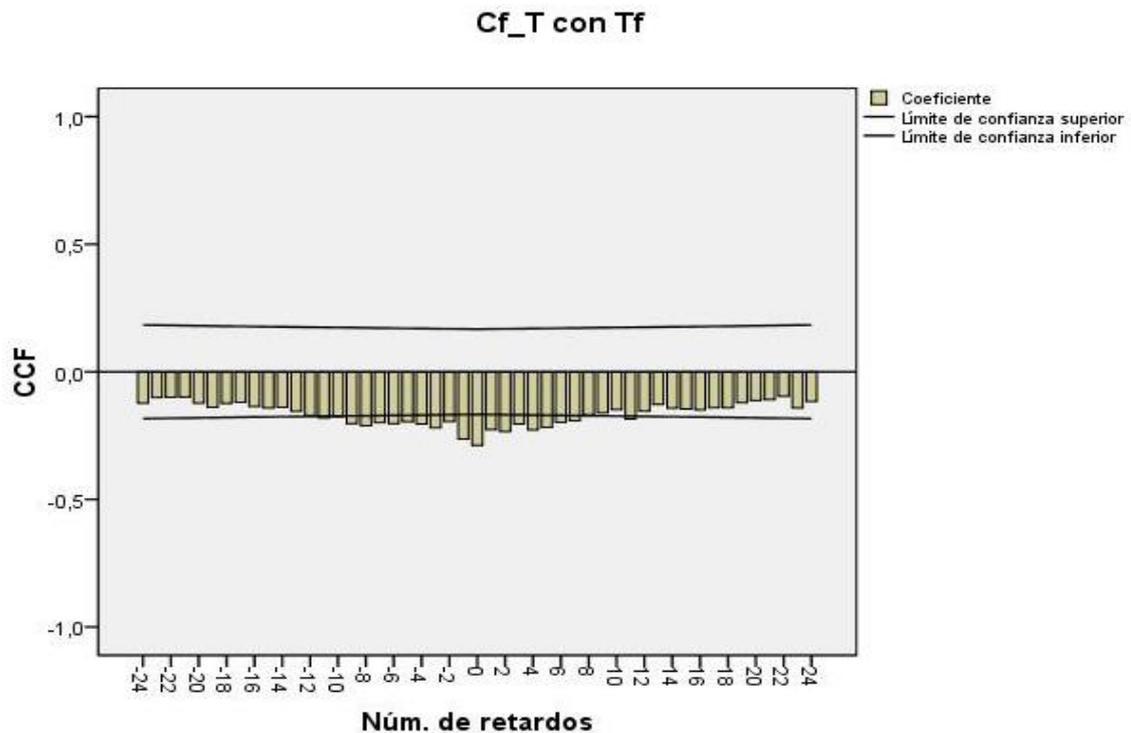


Figura 6.11

❖ *Identificación de la función de transferencia*

La identificación del modelo de función de transferencia consiste en obtener los órdenes r y s de los polinomios del denominador y del numerador respectivamente, así como el retardo inicial b . Los órdenes s y r de los polinomios $\delta(B)$ y $\omega(B)$ se pueden determinar por la forma de la función de correlación cruzada entre α y β a partir del retardo b . En la gráfica de CCF de la Figura 6.12 sale un patrón de comportamiento fijo desde el retardo 0, es decir, un decrecimiento exponencial desde el retardo 0, por lo tanto, podemos determinar los valores de b , r y s según la tabla de la Figura 5.3 del capítulo 5.

$$b = 0, r = 1, s = 0$$

Luego el modelo de la función de transferencia será del tipo:

$$\frac{\omega_0}{1 - \delta_1 B} T_t$$

❖ **Estimación del modelo de función de transferencia**

El siguiente paso será calcular los parámetros ω_0 y δ_1 de función de transferencia después de haber identificado los órdenes de los polinomios. En *SPSS* introducimos los valores de b, r y s :

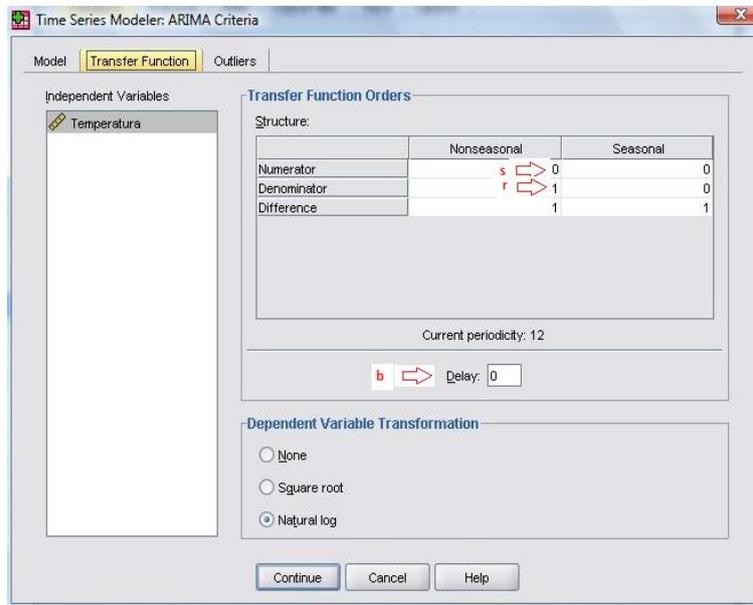


Figura 6.12

Al pulsar *Continue* nos sale la tabla de resultados siguiente:

Parámetros del modelo ARIMA				Estimación	ET	t	Sig.
Consumo-Model_1	Consumo	Log natural	Diferencia	1			
			Diferenciación estacional	1			
	Temperatura	Log natural	Numerador Retardo 0	-,388	,044	-8,880	,000
			Diferencia	1			
			Denominador Retardo 1	,446	,112	3,970	,000
			6 Diferenciación estacional	1			

Con este proceso quedan determinados ω_0 y δ_1 : $\omega_0 = -0,388$ y $\delta_1 = 0,446$

Entonces el modelo de función de transferencia tentativo (por falta de validación) será:

$$\nabla\nabla_{12}\log C_t = -\frac{0,388}{(1 - 0,446B)}\nabla\nabla_{12}\log T_t + N_t$$

Siendo N_t un ruido que no necesariamente blanco y por tanto que podemos modelizar mediante una estructura *ARMA*.

Analizamos la estructura de autocorrelación presente en los residuos N_t mediante la *fas* y la *fap* (Figura 6.13), parece ser que el ruido N_t sigue un modelo *ARIMA*(0,1,1)(0,1,1)₁₂.

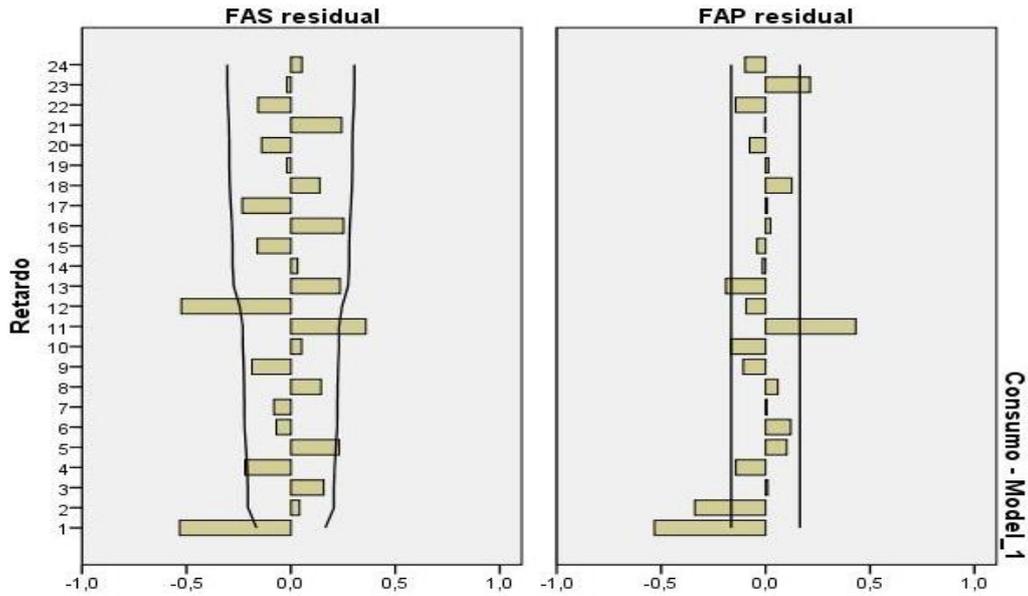


Figura 6.13

Efectivamente, después de ver la tabla de resultados, resulta que todos los parámetros tienen un nivel de significación muy buena ($p_value < 0,05$) incluso la constante. Al volver lanzar la estimación de los parámetros teniendo en cuenta el modelo *ARIMA* del ruido N_t , los parámetros de función de transferencia cambian de valor pero de forma muy ligera.

Parámetros del modelo ARIMA

				Estimación	ET	t	Sig.
Consumo-Model_1	Consumo	Log natural	Constante	-,002	,001	-2,871	,005
			Diferencia	1			
			MA Retardo 1	,480	,078	6,134	,000
			Diferenciación estacional	1			
	Temperatura	Log natural	MA, estacional Retardo 1	,600	,082	7,301	,000
			Numerador Retardo 0	-,360	,037	-9,810	,000
			8 Diferencia	1			
			Denominador Retardo 1	,358	,093	3,854	,000
			10 Diferenciación estacional	1			

El modelo finalmente estimado será:

$$\nabla\nabla_{12}\log C_t = -0,002 - \frac{0,36}{(1 - 0,358B)} \nabla\nabla_{12}\log T_t + (1 - 0,480B)(1 - 0,6B^{12})a_t$$

$$RMSE = 861,951$$

❖ **Validación y previsiones**

Al ver el correlograma de las funciones de correlación de los residuos representado en la *Figura 6.14* y quitando un pico que sale en el retardo 11 debido a valores atípicos, todos los coeficientes están dentro del intervalo de confianza y por tanto se puede dar por válido el modelo de función de transferencia.

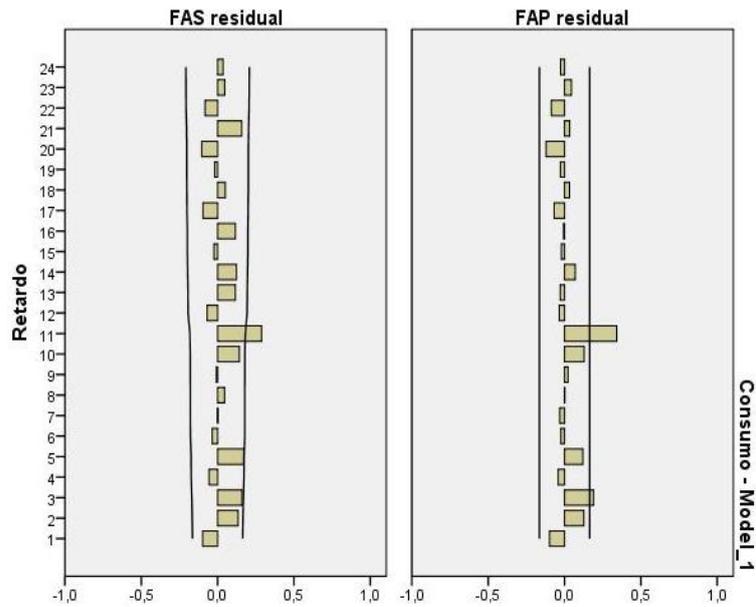


Figura 6.14

Antes de realizar una previsión para el consumo de las 12 últimas muestras y compararlas con sus valores reales, cabe recordar que hay una relación dinámica entre el consumo y la temperatura, entonces primero hay que sacar previsiones de temperatura para las mismas muestras siguiendo el modelo identificado en el preblanqueo del input (Figura 6.15). Y después calcular las predicciones del consumo mediante la función de transferencia utilizando esos valores predichos de temperatura.

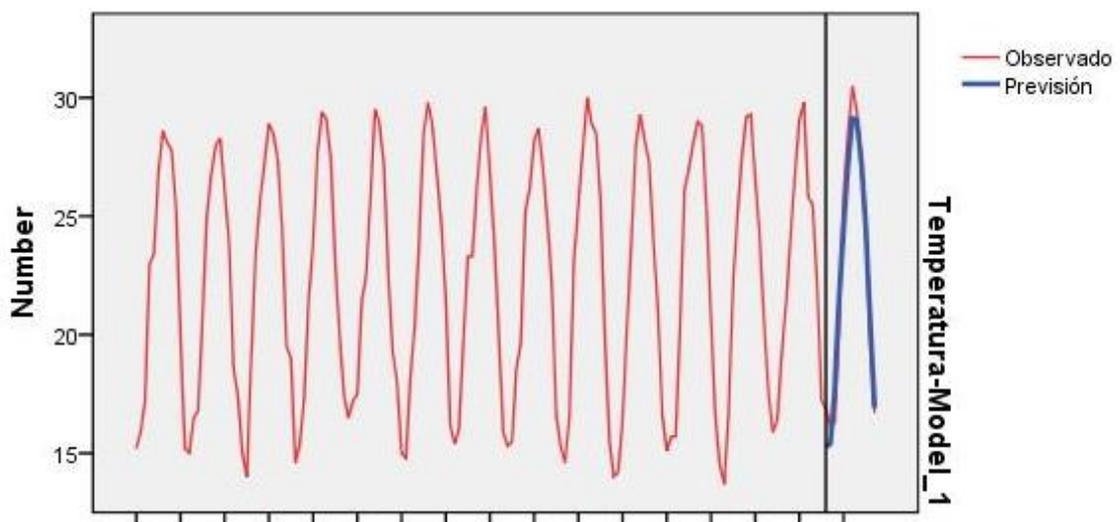
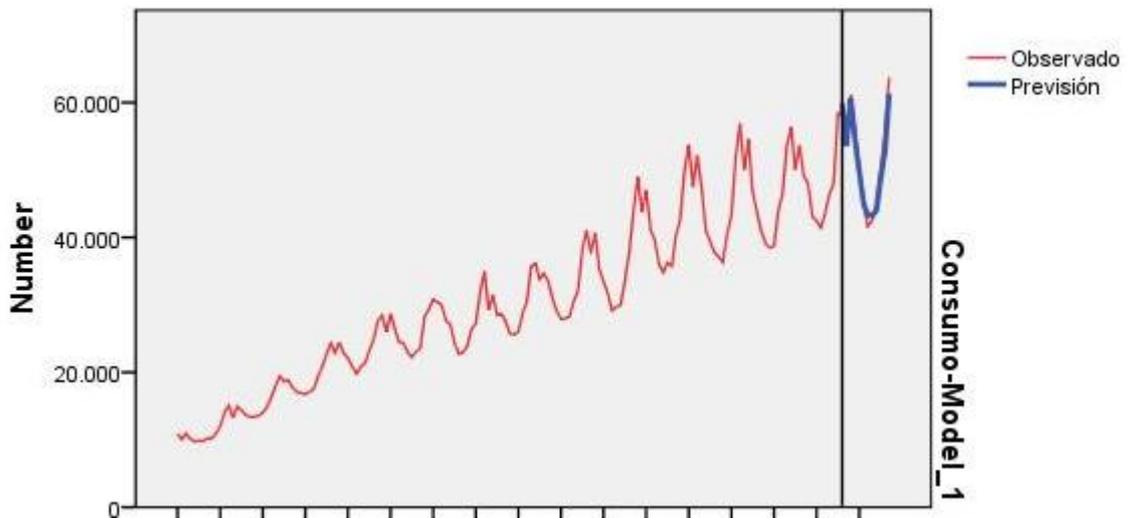


Figura 6.15

Finalmente ya tenemos previsiones del consumo de los últimos 12 meses comparadas con los datos reales, se ve que se ajusta bastante al modelo real.



6.3 FUNCIÓN DE TRANSFERENCIA-INPUTS: TEMPERATURA Y PRECIO

En el apartado anterior vimos como formular la función de transferencia en función de la variable explicativa; la temperatura. En este apartado, la idea sería, manteniendo la misma estructura de la relación dinámica respecto a la temperatura $(b, s, r) = (0,1,0)$ introducir otra variable independiente, es decir, siguiendo los mismos pasos del apartado anterior buscaremos mediante el método de preblanqueo identificar los parámetros (b, s, r) .

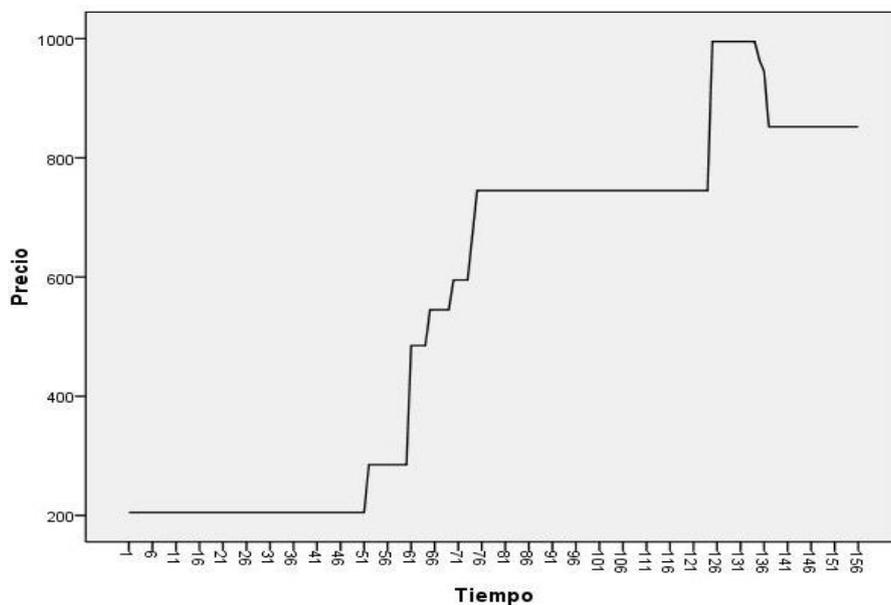


Figura 6.16

Al observar la representación gráfica de la serie precio P_t (la Figura 6.16), se ve claramente que no hay una relación directa entre el precio y el consumo que pueda explicar el comportamiento a lo largo del tiempo del consumo de gas natural, en

cambio, en el caso de la temperatura sí que había una cierta semejanza, de tal manera que los picos del consumo coincidía con las temperaturas más bajas.

Como conclusión, evaluar respuestas dinámicas introduciendo una variable input controlable como es el caso del precio permite simular y evaluar políticas alternativas, es decir, tener un cierto control sobre la variable dependiente u output, en este caso es el consumo. Por ejemplo, en el caso de prever mucha demanda, que no se puede satisfacer, como alternativa estratégica sería subir el precio para atenuar esa demanda prevista y ajustarla a la planificada.

Primero, identificamos el modelo *ARIMA* que se ajusta a la serie P_t . A partir de las funciones de autocorrelación simple y parcial, podemos afirmar que la serie P_t sigue un proceso $ARIMA(0,1,0)(2,1,0)_{12}$.

El modelo estimado para la variable que consideramos input P_t fue el siguiente:

$$\nabla \nabla_{12} \log P_t = \frac{1}{(1 + 0,664B^{12} + 0,321B^{24})} \alpha_t$$

Una vez tenemos estimados los parámetros del modelo *ARIMA* de la serie precio P_t , toca preblanquear el input y el output por un filtro definido por la función:

$$\frac{1}{(1+0,664B^{12}+0,321B^{24})}$$

para conseguir β_t y α_t .

Vamos a representar la correlación cruzada entre β_t y α_t que juega un papel importante la hora de identificar la función de transferencia entre el input y el output.

Los órdenes s y r de los polinomios $\delta(B)$ y $\omega(B)$ se pueden determinar por la forma de la función de correlación cruzada entre β_t y α_t a partir del retardo inicial b .

Viendo la *Figura 6.17* queda claro que el retardo $b = 2$, ya que sólo hay un pico significativo en el retardo 2, y los demás no son significativos, están dentro del intervalo de confianza. Por lo tanto, nuestro caso podría corresponder a un modelo con $b = 2, r = 0$ y $s = 0$. Luego el modelo de la función de transferencia será del tipo: $\omega_0 P_{t-2}$.

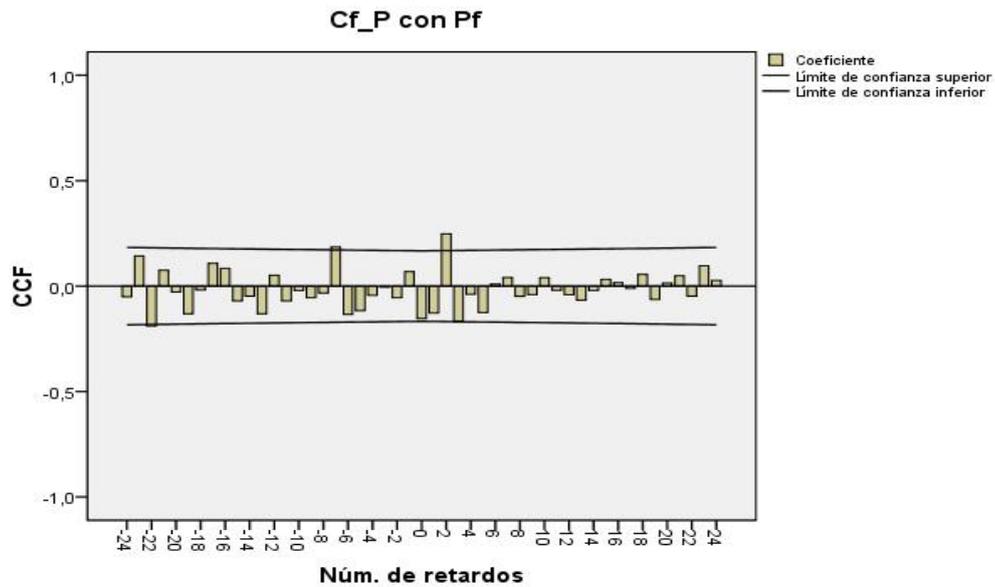


Figura 6.17

El paso siguiente sería rellenar la tabla siguiente de los parámetros (b, s, r) de la variable input P_t .

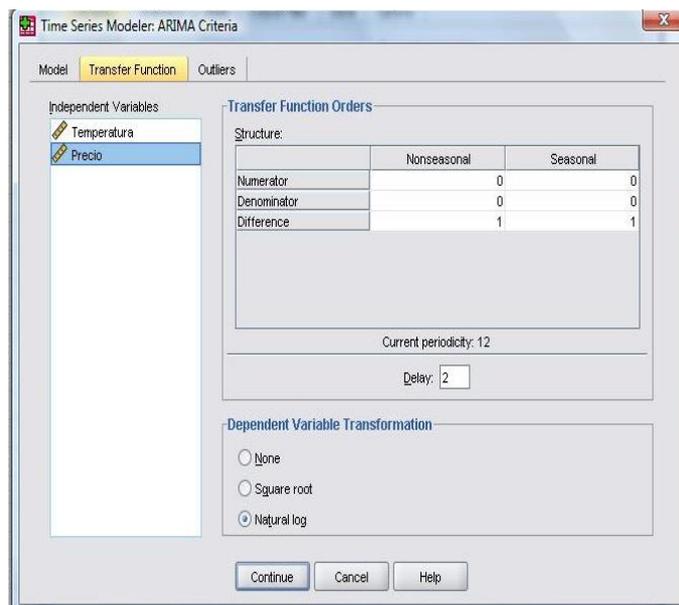


Figura 6.18

Los parámetros (b, r, s) de la variable input Temperatura T_t los mantenemos del apartado anterior. Al pulsar *Continue* conseguimos los siguientes resultados:

Parámetros del modelo ARIMA				Estimación	ET	t	Sig.
Consumo-Model_1	Consumo	Log natural	Constante	-.002	,000	-3,254	,001
			Diferencia	1			
			MA Retardo 1	,531	,078	6,831	,000
			Diferenciación estacional	1			
Temperatura	Log natural	MA, estacional Retardo 1	,593	,083	7,168	,000	
		Numerador Retardo 0	-.365	,036	-10,026	,000	
		8 Diferencia	1				
		Denominador Retardo 1	,368	,090	4,113	,000	
Precio	Log natural	10 Diferenciación estacional	1				
		Retardo	2				
		Numerador Retardo 0	-.068	,028	-2,404	,018	
		12 Diferencia	1				
			Diferenciación estacional	1			

Son resultados satisfactorios, ya que todos los valores son significativos ($p - value < 0.05$).

Analizando la estructura de autocorrelación presente en los residuos mediante la *fas* y la *fap* como es habitual en la fase de la validación del modelo (Figura 6.19), vemos que efectivamente los coeficiente de autocorrelación presentan una estructura de ruido blanco, salvo el pico que sale en el retardo 11, debido a valores atípicos, pero más adelante explicaremos como podemos corregir este anomalía.

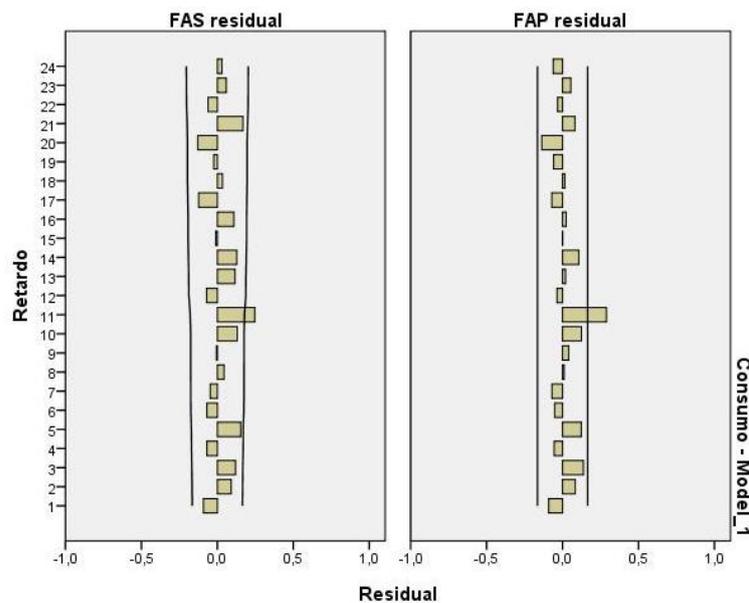


Figura 6.19

Finalmente, el modelo de la función de transferencia con dos inputs queda definido así:

$$\nabla \nabla_{12} \log C_t = -0,002 - \frac{0,365}{(1 - 0,368B)} \nabla \nabla_{12} \log T_t - 0,068 \nabla \nabla_{12} \log P_{t-2} + (1 - 0,531B)(1 - 0,593B^{12})a_t$$

$$RMSE = 855,575$$

Con esta ecuación ya podemos calcular las predicciones para las últimas 12 muestras, o sea, datos mensuales del consumo para el último año, pero en este caso no hace falta calcular primero las predicciones del precio como hicimos con las de temperatura, al ser

una variable controlable y entre en lo planificado para el próximo curso, a pesar de modelizar la serie P_t con un modelo *ARIMA*.

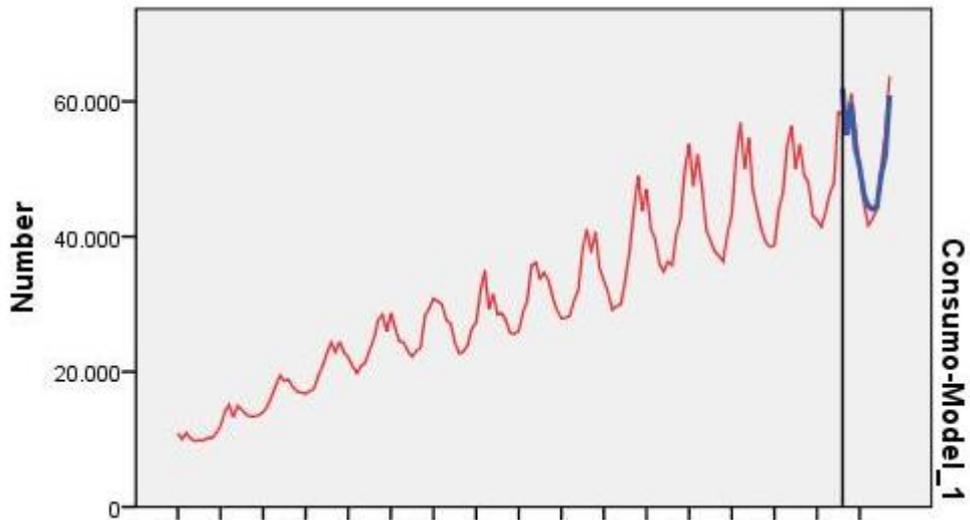


Figura 6.20

7. CONCLUSIONES Y LÍNEAS FUTURAS

En este capítulo se presenta un resumen del trabajo realizado, las conclusiones más relevantes obtenidas, así como las líneas de trabajo futuras que se derivan del mismo.

7.1 Resumen

En este proyecto fin de carrera se han estudiado tres metodologías para la predicción mensual del consumo del gas natural. Las tres se engloban dentro de las técnicas de predicción basadas en series temporales. Estas metodologías son: el modelo *ARIMA*, el modelo de Regresión Dinámica con un input y el modelo de Regresión Dinámica con dos inputs.

En los capítulos 3, 4 y 5 se ha descrito una base teórica de los modelos presentados, el modelo *ARIMA* y el modelo de Regresión Dinámica. Se ha realizado una descripción matemática de cada uno de los modelos y a continuación se ha realizado el estudio del caso utilizando los tres modelos.

7.2 Conclusión

Para comparar los resultados de los tres modelos, se comparan a continuación las predicciones obtenidas al aplicar cada modelo junto con los valores reales del consumo mensual del gas natural durante todo el período de estudio.

A simple vista de la *Figura 7.1* se observa que con el modelo *ARIMA* se obtienen predicciones menos precisas, sobre todo en el período de tiempo dedicado para estimar el modelo (las primeras 156 muestras). Esa comparación resulta más difícil entre los dos modelos de Regresión Dinámica, por eso recurrimos al criterio de medida de precisión *RMSE*.

El *RMSE* facilitado por el software *SPSS* está calculado sobre el período de ajuste, es decir, es el error cuadrático medio entre el valor pronosticado y el valor real hasta la muestra 156. Sería útil también calcular el *RMSE* durante el período de predicción (las últimas 12 muestras) para comparar los tres modelos. En la siguiente tabla resumimos el *RMSE* calculado para cada modelo.

RMSE	Modelo ARIMA	Modelo FT con un input	Modelo FT con dos inputs
RMSE Período ajuste	1159,030	861,951	855,575
RMSE Período predicción	3054,405	2280,352	2168,628

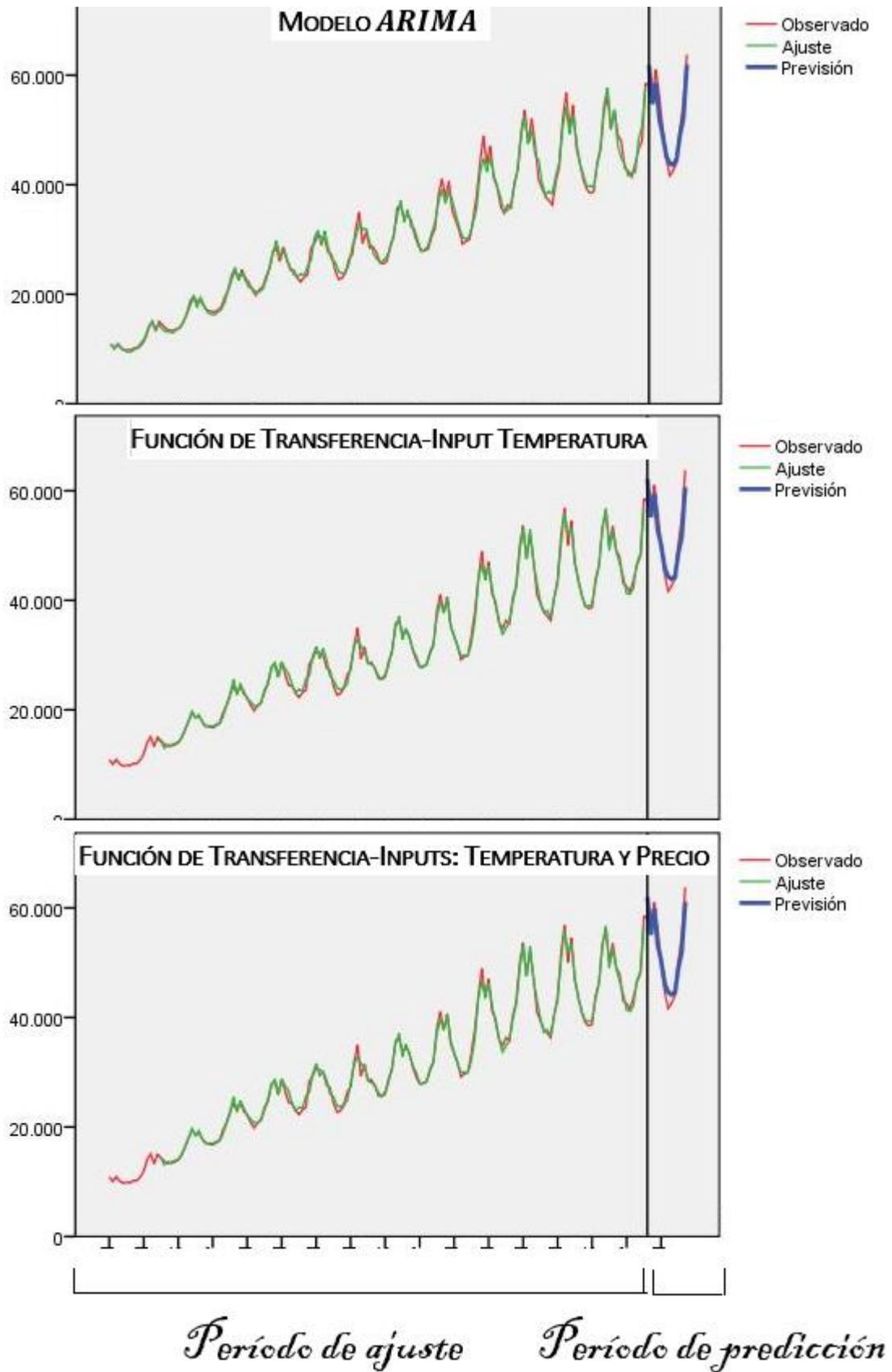


Figura 7.1

Entonces, entre las conclusiones más relevantes obtenidas de este trabajo cabe destacar las siguientes:

El modelo que obtiene mejores predicciones es el modelo de función de transferencia con dos inputs, que presenta en este caso aparte de conseguir buenas prestaciones, la ventaja de tener un cierto control sobre el consumo futuro, ya que la entrada es una serie controlable. El único inconveniente es su proceso de modelado, que es muy laborioso, teniendo en cuenta que la mejora es casi inapreciable respecto al modelo de función de transferencia de un input.

En cambio, el modelo *ARIMA* da resultados aceptables sobre todo teniendo en cuenta su coste computacional que no es tan elevado respecto a otros modelos.

Obviamente, para alcanzar esa mejoría en las previsiones utilizando el modelo de función de transferencia, queda claro que hay que hacer buena labor de ajuste de las variables explicativas. Como comparación, calculamos el *RMSE* del consumo predicho del último año en el caso de tener en cuenta datos pronosticados mediante el modelo *ARIMA* de la temperatura: $RMSE = 2280$ y en el caso de usar datos reales de la temperatura (ajuste perfecto) el $RMSS = 1577,83$, se mejora bastante.

En la práctica, desconocimos los valores de la temperatura para el período de predicción, por no ser variable controlable por eso hay que predecirlos.

7.3 Líneas de trabajo futuras

En este apartado se muestra las posibles líneas de trabajo futuras surgidas del desarrollado del trabajo realizado. Éstas son:

- Las series reales se ven con frecuencia afectadas por sucesos puntuales conocidos. Por ejemplo, una huelga, un año bisiesto, un cambio legal, un accidente o un cambio en una festividad. Si modelamos estos efectos en la serie podemos mejorar la precisión de la estimación de los parámetros y de las previsiones, se denomina *análisis de intervención* a la inclusión en un modelo de series temporales de variables ficticias para representar sucesos que producen efectos deterministas.
- Con mucha frecuencia ocurren en las series reales hechos puntuales que desconocemos. Por ejemplo, la serie puede haber estado sometido a cambios de base, errores de medición, etc. Las observaciones afectadas por estas intervenciones pueden presentar una estructura distinta de las demás y aparecer como datos atípicos, es decir, datos que aparentemente no han sido

generados igual que las demás. En particular, si se comete un error de medida o de transcripción en un dato, este valor no mantendrá la estructura de dependencia del resto de la serie. Es importante ser capaz de identificar estas situaciones desconocidas y separarlas de la dinámica de la serie. En el caso por ejemplo, ha habido valores atípicos, que se manifestaron en la función de autocorrelación de los residuos, que se pueden eliminar para mejorar aún la precisión de las previsiones.

8. BIBLIOGRAFÍA

- ❖ [1]: Análisis de series temporales
Daniel Peña Madrid: Alianza D.L. 2010
- ❖ [2]: Análisis de series temporales
Daniel Peña Madrid: Alianza D.L. 2005
- ❖ [3]: Métodos estadísticos avanzados con SPSS
César Pérez López Madrid: Thomson 2005
- ❖ [4]: Técnicas estadísticas con SPSS 12: aplicaciones al análisis de datos César Pérez López Madrid etc.: Pearson Educación D.L. 2005
- ❖ [5]: Series temporales: técnicas y herramientas
César Pérez López Madrid: Garceta 2011
- ❖ [6]: Análisis de series temporales y técnicas de previsión
Andrés Carrión; Universidad Politécnica de Valencia, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad Valencia: Editorial UPV 2009
- ❖ [7]: Introducción a las señales aleatorias
Departamento de Comunicaciones, Escuela Técnica Superior de Ingenieros de Telecomunicaciones- Servicio de Publicaciones: SPUPV-99.4026
- ❖ [8]: Manual del usuario del sistema del sistema básico de IBM SPSS Statistics 19

Anexo

Datos correspondientes a las series del caso:

Time	Consumo	Temperatura	Precio	Time	Consumo	Temperatura	Precio
1	10.890,85	15,2	205	85	36.153,84	15,3	745
2	10.064,03	15,9	205	86	33.823,48	15,5	745
3	10.937,02	17,2	205	87	34.659,73	18,6	745
4	10.134,52	23,0	205	88	33.488,66	19,7	745
5	9.723,29	23,4	205	89	30.929,37	25,2	745
6	9.861,51	26,8	205	90	29.089,73	26,2	745
7	9.800,81	28,6	205	91	27.864,58	28,2	745
8	10.210,76	28,1	205	92	27.987,55	28,7	745
9	10.176,83	27,8	205	93	28.256,75	27,0	745
10	10.886,21	25,3	205	94	30.573,45	24,8	745
11	11.919,70	20,0	205	95	32.075,97	22,1	745
12	14.051,67	15,2	205	96	38.195,92	16,6	745
13	15.093,05	15,0	205	97	41.080,93	15,3	745
14	13.262,70	16,5	205	98	37.687,38	14,6	745
15	14.952,30	16,8	205	99	40.637,50	16,6	745
16	14.390,79	20,8	205	100	35.233,17	23,1	745
17	13.675,09	25,1	205	101	33.414,19	25,5	745
18	13.374,08	26,8	205	102	31.630,09	27,8	745
19	13.376,34	28,0	205	103	29.143,59	30,0	745
20	13.593,22	28,3	205	104	29.704,45	28,9	745
21	14.004,43	26,1	205	105	29.980,06	28,5	745
22	14.854,28	23,9	205	106	33.578,06	25,7	745
23	16.271,05	18,6	205	107	37.824,87	19,9	745
24	17.970,22	17,4	205	108	43.799,03	15,6	745
25	19.469,86	15,0	205	109	48.960,89	14,0	745
26	18.628,83	14,0	205	110	43.817,43	14,2	745
27	18.853,02	19,0	205	111	47.024,57	16,4	745
28	17.774,51	23,4	205	112	41.137,37	20,4	745
29	17.021,00	25,7	205	113	39.664,03	23,9	745
30	16.897,66	27,2	205	114	36.028,03	28,0	745
31	16.723,50	28,9	205	115	34.811,02	29,3	745
32	17.107,73	28,5	205	116	36.280,80	28,2	745
33	17.586,92	27,4	205	117	35.738,25	27,3	745
34	19.437,98	24,5	205	118	40.353,82	23,9	745
35	20.851,03	19,5	205	119	42.757,96	21,1	745
36	22.743,22	19,0	205	120	49.890,37	16,6	745
37	24.312,66	14,6	205	121	53.669,64	15,1	745
38	22.919,82	15,4	205	122	47.499,20	15,7	745
39	24.445,31	17,4	205	123	52.114,44	15,7	745
40	22.844,76	21,5	205	124	47.406,24	20,4	745
41	22.132,02	23,9	205	125	41.005,22	26,1	995
42	20.805,75	27,7	205	126	39.431,74	27,0	995
43	19.801,66	29,4	205	127	37.792,81	28,1	995
44	20.794,87	29,1	205	128	37.119,80	29,0	995
45	21.359,75	27,4	205	129	36.301,97	28,8	995
46	23.210,20	23,0	205	130	40.436,29	25,3	995
47	24.857,56	20,0	205	131	43.279,42	20,7	995
48	27.683,71	17,4	205	132	51.912,79	16,6	995
49	28.447,72	16,5	205	133	56.848,30	14,5	995

50	25.982,04	17,2	205	134	50.040,26	13,7	995
51	28.656,22	17,5	205	135	54.549,90	16,8	964
52	26.459,92	21,4	285	136	46.617,64	22,1	945
53	24.466,61	22,5	285	137	43.782,82	25,1	852
54	24.393,17	25,9	285	138	40.879,30	27,5	852
55	22.978,80	29,5	285	139	39.099,31	29,2	852
56	22.263,58	28,9	285	140	38.476,13	29,3	852
57	23.086,43	27,2	285	141	38.725,55	26,6	852
58	23.568,65	22,6	285	142	43.947,83	24,2	852
59	28.321,08	19,3	285	143	46.398,30	20,6	852
60	29.348,36	18,0	285	144	53.496,33	17,5	852
61	30.827,67	15,0	485	145	56.337,02	15,9	852
62	30.388,75	14,8	485	146	50.063,04	16,4	852
63	29.966,14	18,2	485	147	53.605,72	19,2	852
64	27.594,86	20,4	485	148	49.112,05	21,3	852
65	27.164,34	24,0	545	149	47.875,57	24,1	852
66	24.366,11	28,5	545	150	43.063,48	26,8	852
67	22.713,57	29,8	545	151	42.476,12	29,1	852
68	22.951,23	28,8	545	152	41.463,24	29,8	852
69	23.896,43	26,8	545	153	43.682,31	25,8	852
70	26.388,93	24,7	595	154	46.447,85	25,5	852
71	27.283,55	21,5	595	155	48.022,00	22,2	852
72	31.874,64	16,2	595	156	58.427,00	17,2	852
73	34.938,74	15,4	595	157	58.243,00	16,9	757
74	29.262,19	16,0	670	158	56.090,00	16,1	757
75	31.440,73	19,7	745	159	61.042,00	16,3	757
76	28.483,92	23,3	745	160	56.027,00	20,3	757
77	28.664,12	23,3	745	161	49.968,00	25,9	757
78	27.574,33	26,1	745	162	44.911,00	28,6	757
79	25.722,60	28,4	745	163	41.654,00	30,5	757
80	25.567,46	29,6	745	164	42.538,00	29,4	757
81	26.027,84	26,8	745	165	43.976,00	27,3	757
82	28.873,43	23,9	745	166	50.249,00	24,6	757
83	30.488,59	20,0	745	167	55.138,00	19,6	757
84	35.728,55	15,9	745	168	63.757,00	16,7	757