

UNIVERSIDAD AUTÓNOMA
METROPOLITANA-IZTAPALAPA



DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍAS

ANÁLISIS DE SERIES DE TIEMPO UNIVARIADAS Y
MULTIVARIADAS MEDIANTE LOS MODELOS
ARIMA Y MVAR

TESIS

PARA OBTENER EL GRADO DE:
Maestro en Ciencias
(Matemáticas Aplicadas e Industriales)

PRESENTA:
Margarita Reyes Flores

ASESOR:
Dr. Gabriel Escarela Pérez

JURADO CALIFICADOR:

Presidente: Juan Ruiz de Chávez

Secretario: Gabriel Núñez Antonio

Vocal: Gabriel Escarela Pérez

Vocal: Flaviano Godínez Jaimes

CIUDAD DE MÉXICO, DICIEMBRE DE 2016

Índice general

Resumen	V
Introducción	VII
1. Series de tiempo univariadas	1
1.1. Ejemplos	3
2. Procesos estocásticos	7
2.1. Procesos estocásticos estacionarios	8
2.2. Proceso de ruido blanco	13
2.3. Procesos lineales	15
3. Modelos ARIMA	17
3.1. Introducción	17
3.2. Modelos de medias móviles	17
3.3. Modelos autorregresivos	23
3.3.1. Proceso AR(1)	24
3.3.2. Proceso AR(2)	26
3.3.3. Proceso AR(p)	30
3.4. Modelos autorregresivos de medias móviles	31
3.5. Modelos autorregresivos integrados de medias móviles	33
3.6. Otros modelos	36
4. Metodología de Box & Jenkins	39
4.1. Identificación	39
4.1.1. Identificación de la estructura no estacionaria	39
4.1.2. Identificación del modelo ARMA	41

4.2. Estimación de los parámetros	41
4.2.1. Método de momentos	42
4.2.2. Método de máxima verosimilitud	42
4.2.3. Estimación por mínimos cuadrados no condicionales	42
4.2.4. Estimación por mínimos cuadrados condicionales	43
4.2.5. Estimación no lineal	43
4.3. Validación del modelo	43
4.3.1. Análisis de los residuales	43
4.3.2. Método backcasting o backforecasting	44
4.3.3. Reformulación del modelo	45
5. Ejemplos de la metodología de Box & Jenkins	47
5.1. Introducción	47
5.2. Serie QQQ	49
5.2.1. Identificación de la estructura no estacionaria	49
5.2.2. Identificación del modelo ARMA	53
5.2.3. Validación del modelo	53
5.3. Serie MSFT	56
5.3.1. Identificación de la estructura no estacionaria	56
5.3.2. Identificación del modelo ARMA	60
5.3.3. Validación del modelo	61
6. Series de tiempo multivariadas	65
6.1. Introducción	65
6.1.1. Series de tiempo multivariadas estacionarias	65
6.2. Matrices de correlaciones y covarianzas cruzadas	66
6.2.1. Dependencia lineal	67
6.2.2. Matriz de correlaciones cruzadas muestrales	68
6.3. Modelos vectoriales autorregresivos	69
6.3.1. Proceso VAR(1)	69
6.3.2. Proceso VAR(p)	71
6.3.3. Ajuste de un modelo VAR	72
6.4. Modelos vectoriales de medias móviles	74

6.4.1. Proceso VMA(q)	74
6.4.2. Ajuste de un modelo VMA	75
6.5. Modelos vectoriales autorregresivos y de medias móviles	76
6.5.1. Proceso VARMA(p,q)	76
6.5.2. Modelos marginales de los componentes	78
6.6. Ajuste de la serie bivariada (QQQ,MSFT)	79
7. Modelo MVAR	85
7.1. Introducción	85
7.2. Modelos de mezclas	86
7.2.1. Aplicaciones de los modelos de mezclas	86
7.2.2. Número de componentes	87
7.2.3. Estimación de los parámetros	87
7.3. Modelo MAR	88
7.4. Modelo MVAR	89
7.4.1. Supuestos del modelo	89
7.4.2. Definición	91
7.4.3. Condiciones de estacionaridad	92
7.4.4. Estimación de los parámetros	93
7.4.5. Elección del modelo	94
7.4.6. Validación del modelo	95
8. Ajuste del MVAR a una serie bivariada	97
8.1. Introducción	97
8.2. Elección del modelo MVAR	99
8.3. Eliminación de parámetros no significativos	100
8.4. Validación del modelo	105
9. Conclusiones	111
Anexos	115
A. Función de log verosimilitud del modelo MVAR(2,3;1,1,1) en R	117
Bibliografía	121

Resumen

En el presente trabajo se introduce al lector a la teoría de series de tiempo univariadas y multivariadas haciendo un repaso de los modelos ARIMA, de la metodología de Box & Jenkins y de los modelos VARIMA. También se implementa un modelo multivariado de series de tiempo para el estudio de los precios de acciones incluídas en un portafolio de inversiones que es una combinación de los modelos de mezclas y de los modelos AR para series multivariadas. Este modelo, llamado MVAR (Mixture Vector Autoregressive Model), consiste en una mezcla de componentes gaussianos vectoriales autoregresivos. Para estimar los parámetros se utilizó el método de máxima verosimilitud y el cálculo se realizó mediante la función `nlm` del software estadístico R que usa un algoritmo de tipo Newton. Los datos analizados corresponden a los precios de las acciones del Índice Nasdaq 100 y de Microsoft Corporation.

Introducción

Una gran cantidad de información acerca de las características económicas y financieras, tanto de individuos como de empresas o países, se recopila con fines de análisis para poder planear y tomar decisiones. Ejemplos de esto son los precios de las acciones de una compañía tomados diariamente, el total mensual de exportaciones, el promedio mensual de ingresos, las ganancias anuales de una compañía, etc. Al registro metódico de mediciones u observaciones a lo largo del tiempo, efectuado en periodos fijos, se le conoce como serie de tiempo. No sólo en el ámbito económico es de gran utilidad el análisis de las series de tiempo ya que también existen en una gran variedad de campos como las ciencias físicas, particularmente en meteorología, ciencia marina o geofísica. Algunos ejemplos son las mediciones diarias de la lluvia o la temperatura del aire medida en horas, días o meses. El análisis de las series de tiempo también es aplicable a la mercadotecnia. Las variables observadas podrían incluir la cantidad semanal, mensual o anual de ventas de algún producto, los montos facturados, los costos de publicidad, etc. Con el estudio de las series de tiempo es posible pronosticar la cantidad de ventas para elaborar un plan de producción o examinar la relación existente entre las ventas y otras series de tiempo tales como gastos en publicidad. En demografía son de mucha ayuda las series de tiempo. Por ejemplo en el estudio del crecimiento de las poblaciones surgen series de tiempo formadas por conteos de personas que viven en ellas o por conteos de nacimientos ocurridos en ellas y en estos casos es posible predecir la cantidad de pobladores o la cantidad de nacimientos en la población dentro de 10 o 20 años.

Además, en los campos antes mencionados surge la necesidad de incluir en los análisis no una sino varias series de tiempo. Por ejemplo, en el ámbito financiero, la globalización económica y la comunicación mediante Internet han acelerado la integración de los mercados financieros mundiales en años recientes. Los movimientos en los precios de un mercado pueden extenderse fácilmente a otro mercado. Por ésta razón los mercados financieros son más dependientes entre ellos y los analistas deben considerarlos conjuntamente para entender mejor la estructura dinámica del mercado global. Así, bajo ciertas circunstancias, un mercado puede liderar otro mercado, mientras que en otras circunstancias puede ser al revés. En consecuencia, el conocimiento de cómo los

mercados están interrelacionados es de gran importancia en finanzas. De la misma forma, para un inversionista o una institución financiera con múltiples bienes, las relaciones dinámicas entre el rendimiento de los activos juegan un rol importante en la toma de decisiones. Un ejemplo del uso de series multivariadas en Finanzas se tiene en Broda y Paoletta (2009), donde se optimiza un portafolio de inversiones compuesto por 30 series de tiempo.

Lo anterior pone de manifiesto que la teoría de las series de tiempo es de gran relevancia ya que mediante ella es posible describir y pronosticar los valores de variables unidimensionales o multidimensionales a lo largo del tiempo, y por lo tanto constituye un área importante de la estadística. Los modelos ARIMA de Box & Jenkins y la metodología que dichos autores desarrollaron para la construcción de este tipo de modelos han sido utilizados ampliamente para pronosticar el comportamiento de series de tiempo ya que mediante ellos se obtienen buenas aproximaciones. Incluso el modelo ARIMA se ha generalizado para describir y pronosticar series de tiempo multivariadas dando paso al modelo VARIMA. Sin embargo, a pesar del respaldo teórico de los modelos ya mencionados, los analistas continúan buscando nuevos modelos que mejoren los pronósticos. Ejemplo de ello es el modelo multipredictor autorregresivo para series de tiempo (MATS) de Martin (1992) para describir los saltos multimodales en series de tiempo, el modelo de mezclas Gaussiano (GMTD) de Le, Martin, y Raftery (1996) para modelar estiramientos, estallidos y valores atípicos en series de tiempo, el modelo MAR de Wong y Li (2000) que es una generalización del modelo GMTD cuya distribución predictiva puede ser también multimodal, el modelo MVAR (Mixture Vector Autoregressive Model) de P. W. Fong, Li, Yau, y Wong (2007) que consiste en una mezcla de componentes vectoriales Gaussianos autorregresivos, entre otros.

El primer objetivo de ésta tesis es introducir al lector a la teoría de series de tiempo univariadas y multivariadas haciendo un repaso de los modelos ARIMA, de la metodología de Box & Jenkins y de los modelos VARIMA. El segundo objetivo es implementar un modelo multivariado de series de tiempo para el estudio de los precios de acciones incluidas en un portafolio de inversiones que corresponden a los precios de las acciones del Índice Nasdaq 100 y de Microsoft Corporation.

Para cumplir con los objetivos establecidos, la tesis está estructurada de la siguiente forma. El Capítulo 1 presenta conceptos básicos y ejemplos de series de tiempo univariadas. El Capítulo 2 presenta conceptos básicos y ejemplos de algunos procesos estocásticos íntimamente relacionados con las series de tiempo. El Capítulo 3 está dedicado al estudio de los modelos ARIMA. En el Capítulo 4 se expone la metodología de Box & Jenkins, que se refiere a la metodología estadística para la modelización de una serie de tiempo univariada mediante un mo-

delo ARIMA. El Capítulo 5 contiene la aplicación de la metodología de Box & Jenkins a las series univariadas de los precios de las acciones del Índice Nasdaq 100 y de Microsoft Corporation. El Capítulo 6 está dedicado al estudio de los modelos VARIMA que son el equivalente de los modelos ARIMA para series de tiempo multivariadas. En el Capítulo 7 se define el modelo MVAR y se establecen las condiciones de estacionaridad, la estimación de sus parámetros y la elección del modelo. El Capítulo 8 contiene el ajuste del modelo MVAR a la serie bivariada compuesta por las series de tiempo de los precios de las acciones del Índice Nasdaq 100 y de Microsoft Corporation. El último capítulo presenta las conclusiones de los resultados obtenidos en la tesis.

Capítulo 1

Series de tiempo univariadas

Intuitivamente una **serie de tiempo** es un conjunto de datos respuestas que son el resultado de observar o medir algún fenómeno durante cierto periodo de tiempo, en donde las observaciones contiguas pueden ser dependientes. Al igual que en una muestra aleatoria, su estudio puede llevarse a cabo con herramientas estadísticas descriptivas e inferenciales. El enfoque descriptivo es fundamental en el análisis de las series de tiempo univariadas, siendo de gran importancia las gráficas que contienen en el eje x al tiempo y en el eje y a los datos, ya que dándoles un vistazo se tiene idea de su comportamiento en ese periodo. Por otro lado, la forma clásica en que se lleva a cabo el análisis inferencial es la de descomposición de series que consiste en identificar tres componentes: la *tendencia*, que representa su comportamiento a largo del tiempo, la *estacionalidad*, que representa los comportamientos que se repiten en ciertos periodos de tiempo, y la *irregularidad*, que representa los movimientos aleatorios.

Las series de tiempo se pueden clasificar según nuestra capacidad de observación del tiempo. Si el fenómeno en estudio sólo se puede observar en ciertos instantes se dice que la serie es *discreta*. Por el contrario, cuando es posible observar el fenómeno en cualquier momento se dice que la serie es *continua*. Cabe destacar que la naturaleza discreta o continua de una serie de tiempo depende, como ya se dijo, de la capacidad de observación del tiempo y no de la naturaleza discreta o continua de la respuesta. Usualmente las respuestas (discretas o continuas) son observadas en intervalos iguales como horas, días, meses, años, etc. Un ejemplo de una serie de tiempo discreta generada por una variable continua es aquella que contiene mediciones de temperatura, humedad o contaminantes en el ambiente tomadas por ejemplo cada día, mientras que un ejemplo de una serie de tiempo discreta generada por una variable discreta es la que está formada por conteos de cierto suceso en intervalos de tiempo determinados como el número de muertes ocurridas en un hospital, número de accidentes

ocurridos en cierta carretera o número de piezas defectuosas de un determinado artículo en una tienda tomados por ejemplo cada mes.

El objetivo fundamental del estudio de las series de tiempo es hacer pronósticos sobre los valores que tomará su variable generadora. Para lograr la estimación de sus valores futuros primero es necesario describirlas a través de un análisis inicial; es decir, sintetizar la información contenida en ellas a través de su gráfica y usando herramientas de la estadística descriptiva. También se requiere explicarlas, es decir, contar con un modelo matemático del que éstas puedan ser una muestra. El pronóstico de una serie se realiza extrayendo toda la información contenida y, basándose en el comportamiento de sus datos históricos, se puede conjeturar su futuro resultado. Sin embargo, existen series *deterministas*, es decir series que se pueden predecir con exactitud tales como las generadas por las leyes de la Física clásica newtoniana. Se dice que una serie es *aleatoria* o *estocástica* cuando su futuro está parcialmente determinado por su pasado. Al contrario de las series deterministas, las series aleatorias no se pueden pronosticar con exactitud y lo que queda es obtener la distribución de probabilidad de los valores futuros pero condicionada por los valores pasados. También existen series de tiempo con los dos componentes, el determinista y el aleatorio, de manera que su predicción también implica algún grado de incertidumbre.

Como ya se explicó antes, en las series de tiempo existen dos tipos principales de variaciones: la tendencia y la periodicidad. La primera se refiere al cambio del nivel a lo largo del tiempo tomando en cuenta el número de observaciones disponibles y la segunda se refiere a comportamientos repetitivos, por ejemplo las gráficas de series compuestas por mediciones de temperatura muestran variaciones que se repiten anualmente debido a que la temperatura baja en invierno y sube en verano haciendo que éste patrón se repita cada año. Cuando una serie de tiempo muestra alguno de los comportamientos anteriores y el interés del estudio es más bien ajustar un modelo entonces es recomendable transformarla para obtener una serie nueva sin tendencia ni periodicidad aplicando por ejemplo logaritmo o raíz cuadrada a los datos. Si existe alguna tendencia y la varianza se incrementa a la par del nivel, por ejemplo si la desviación estándar es directamente proporcional a la media, se recomienda una transformación logarítmica la cual eliminará dicha tendencia. Si la varianza cambia a lo largo del tiempo sin que exista alguna tendencia entonces no es necesaria una transformación. Otro caso en el que se requiere aplicar una transformación es cuando existe alguna tendencia y periodicidad pero el tamaño del efecto de ésta última aumenta junto con el nivel medio. En contraste con lo anterior, existen series llamadas *estacionarias* tales que su gráfica varía alrededor de un valor constante y no muestra cambios sistemáticos en la varianza ni comportamientos periódicos. Por lo tanto las transformaciones se utilizan para convertir en estacionarias a las series que no lo son.

Todo lo anterior resalta la importancia que tiene la gráfica de la serie en el análisis estadístico descriptivo. Por lo tanto el primer paso en el estudio de las series de tiempo es dibujar su gráfica para observar características como la tendencia, la periodicidad, la existencia de discontinuidades o de observaciones atípicas. A continuación se muestran algunos ejemplos de series de tiempo.

1.1. Ejemplos

Ejemplo 1.1. La Figura 1.1 muestra los precios de las Letras del Tesoro en el mercado de Chicago durante 100 días hábiles consecutivos en 1981. Es posible notar una marcada tendencia decreciente de los precios con el tiempo y no parece haber periodicidad.

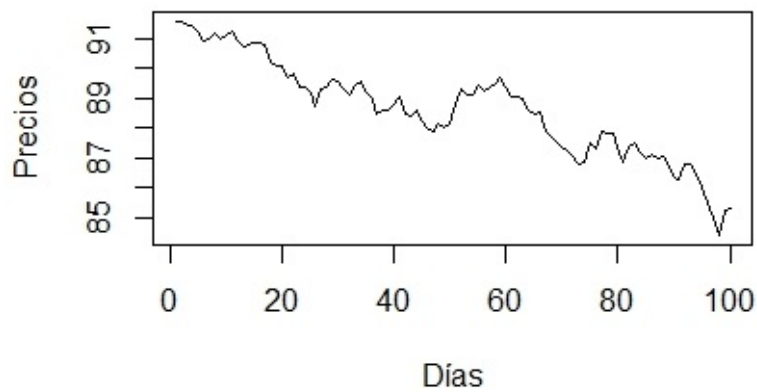


Figura 1.1: Precios de las Letras del Tesoro en el mercado de Chicago tomados diariamente durante cien días hábiles consecutivos en 1981

Ejemplo 1.2. La Figura 1.2 muestra la serie de tiempo de la producción de cerveza mensual en Australia desde enero de 1991 hasta agosto de 1995. La serie no exhibe tendencia pero sí estacionalidad ya que en los últimos dos o tres meses del año siempre aumenta la producción de cerveza.

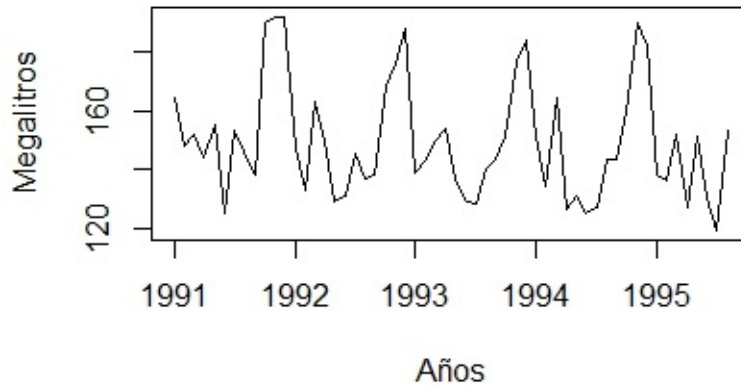


Figura 1.2: Producción de cerveza mensual en Australia medida en megalitros de enero de 1991 a agosto de 1995

Ejemplo 1.3. La Figura 1.3 muestra las ventas mensuales del “producto C” de una gran compañía petrolera durante cuatro meses. Aunque no se aprecia alguna tendencia en la gráfica, sí cierta estacionalidad.

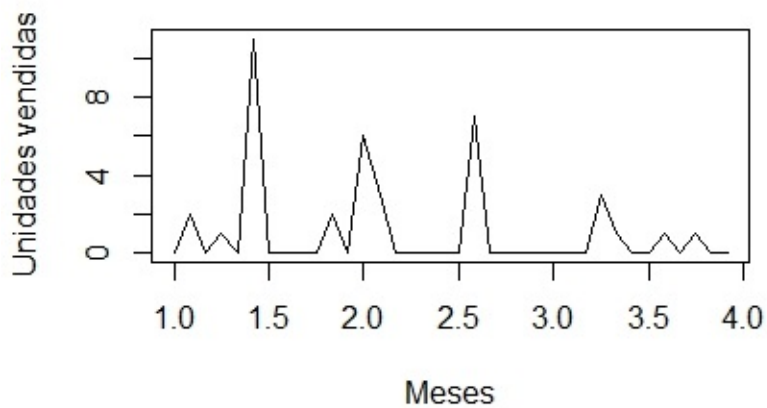


Figura 1.3: Unidades vendidas mensualmente del “producto C” durante cuatro meses

Ejemplo 1.4. La Figura 1.4 muestra la producción de electricidad mensual en Australia desde marzo de 1956 hasta agosto de 1995. Notoriamente la tendencia se incrementa cada año junto con la variación de los datos y existe un patrón estacional muy marcado.

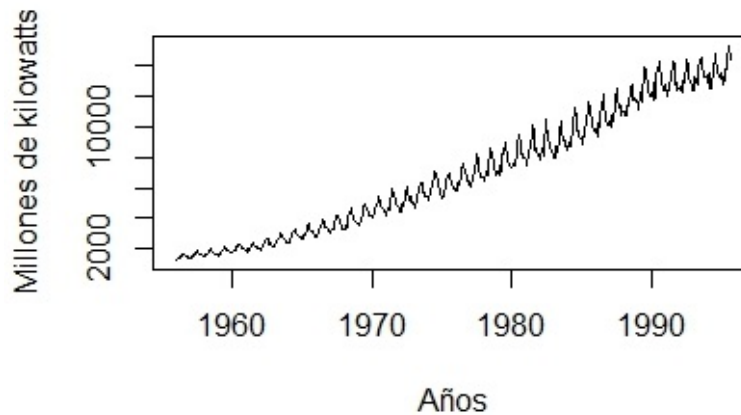


Figura 1.4: Producción mensual de electricidad en Australia desde marzo de 1956 hasta agosto de 1995

Ejemplo 1.5. La Figura 1.5 muestra la producción mensual de ladrillos de arcilla en Australia desde marzo de 1956 hasta septiembre de 1994. Es muy notoria la tendencia creciente. Además se aprecia un comportamiento estacional y mucha variabilidad en los datos.

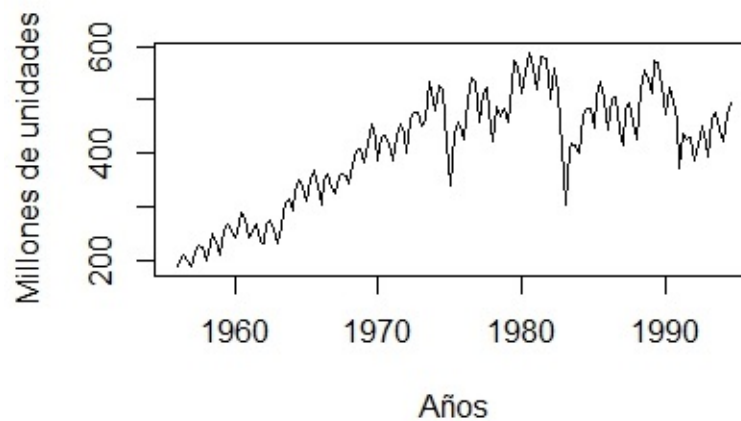


Figura 1.5: Producción mensual de ladrillos de arcilla en Australia de marzo de 1956 a septiembre de 1994

Capítulo 2

Procesos estocásticos

El contexto matemático sobre el cual se definirán las series de tiempo es el de *proceso estocástico*. En este trabajo se supondrá que una observación de una serie de tiempo en el instante t es un valor elegido al azar de entre los posibles valores que puede tomar una variable aleatoria en ese mismo instante. En consecuencia, una serie de tiempo con n observaciones es una muestra de un conjunto de n variables aleatorias ordenadas en el tiempo.

Definición 2.1. *Un proceso estocástico $\{X_t\} = \{X_t : t \in T\}$ es una colección de variables aleatorias que están definidas sobre el mismo espacio muestral, están indexadas por el conjunto T y asumen valores en un mismo conjunto $S \subseteq \mathbb{R}$. El conjunto T es llamado **conjunto de índices** de X y representa al tiempo. El conjunto S es llamado **espacio de estados** de X . Si T es un subconjunto de \mathbb{Z} entonces X es llamado **proceso estocástico en tiempo discreto** o simplemente **proceso estocástico discreto** y si T es un subintervalo de \mathbb{R} entonces X se llama **proceso estocástico en tiempo continuo**, o simplemente **proceso estocástico continuo**.*

Dada la definición anterior resulta natural pensar en una serie de tiempo $x = \{x_t\}$ como una parte de una realización del proceso estocástico $\{X_t\} = \{X_t : t \in T\}$. De aquí que si $\{X_t\}$ es un proceso estocástico discreto o continuo entonces x es una serie de tiempo también discreta o continua. En este trabajo se tratarán series de tiempo discretas cuyas observaciones se tomaron en intervalos de tiempos iguales y para referirse a ellas se utilizarán letras mayúsculas con la letra t como subíndice. Para indicar que se tienen n observaciones sucesivas de una serie de tiempo se escribirá $\{X_1, X_2, \dots, X_t, \dots, X_n\}$ entendiendo que son datos tomados de forma equidistante en los momentos $t_0 + 1h, t_0 + 2h, \dots, t_0 + th, \dots, t_0 + nh$ donde t_0 es algún punto en el tiempo y h es la longitud del intervalo de tiempo en el que se tomó la siguiente observación. Sin embargo, ya que no interesa definir t_0 ni h , en este trabajo se utilizará la notación corta $\{X_n\}$ para referirse a las n observaciones sucesivas de la serie.

Al igual que una variable aleatoria, un proceso estocástico con n variables aleatorias está caracterizado por su función de densidad conjunta $f(X_1, \dots, X_n)$. Tratándose de una serie de tiempo, el proceso estocástico que está asociado a ella está constituido de variables aleatorias que de alguna manera están correlacionadas entre sí, por lo que no se puede hablar de independencia, y como consecuencia no existe una manera directa de encontrar la función de densidad conjunta a partir de las funciones de densidad individuales usando el hecho de que $f(X_1, \dots, X_n) = f(X_1) \cdots f(X_n)$ si X_1, \dots, X_n son variables aleatorias independientes. Por lo tanto, con el fin de lograr caracterizar a las series de tiempo, en lugar de encontrar la función de densidad conjunta se buscan expresiones para sus medias, varianzas y covarianzas ya que éstas medidas logran resumir las características de la distribución de las series. En particular las funciones de autocovarianzas y autocorrelación miden la dependencia lineal que existe entre las variables aleatorias que forman el proceso.

Definición 2.2. Sea $\{X_t\}$ un proceso estocástico con $E(X_t^2)$ finita para todo t .

- La **función de media** de $\{X_t\}$ es

$$\mu_t = E(X_t).$$

- La **función de varianza** de $\{X_t\}$ es

$$\sigma_t^2 = \text{Var}(X_t) = E[(X_t - \mu_t)^2].$$

- La **función de autocovarianza** de $\{X_t\}$ es

$$\gamma_{t_1, t_2} = \text{Cov}(X_{t_1}, X_{t_2}) = E[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})].$$

- La **función de autocorrelación** de $\{X_t\}$, también llamada **función de autocorrelación simple** denotada por **FAS**, es

$$\rho_{t_1, t_2} = \frac{\gamma_{t_1, t_2}}{\sqrt{\sigma_{t_1}^2 \sigma_{t_2}^2}}.$$

2.1. Procesos estocásticos estacionarios

Como ya se dijo antes no se puede encontrar directamente la función de distribución de un proceso estocástico por lo que para estimarla es necesario estudiar una gran cantidad de sus realizaciones. Sin embargo, existen fenómenos que no son estables y que no se pueden repetir a lo largo del tiempo. Un ejemplo es la venta de cierto artículo en un centro comercial donde la serie de tiempo obtenida a partir de los conteos de esas ventas es la única realización del proceso. En estos casos, para calcular la media y varianza de las variables aleatorias

que conforman el proceso estocástico a partir de su evolución a lo largo del tiempo se requiere suponer que algunas de sus características, como la distribución en cada instante, son invariantes en el tiempo. El concepto de estacionaridad surge de las restricciones que deben imponerse a un proceso estocástico para determinar sus características a partir de una realización.

Definición 2.3. Sea $\{X_t\}$ un proceso estocástico con $E[X_t^2]$ finita para todo t . $\{X_t\}$ es **estrictamente estacionario o estacionario en sentido estricto** si

$$f(X_{t_1}, \dots, X_{t_n}) = f(X_{t_1+k}, \dots, X_{t_n+k})$$

para toda $k > 0$, $n \geq 1$, $t_i, t_i + k$ en $\{0, \pm 1, \pm 2, \dots\}$ e $i = 1, \dots, n$.

Un proceso estocástico estrictamente estacionario se comporta de manera muy parecida en cualquier periodo de tiempo, es decir, la gráfica de n observaciones contiguas es casi igual a la gráfica de otras n observaciones contiguas pero ahora tomadas k periodos hacia atrás o hacia adelante. Existen otros conceptos de estacionaridad basados en las funciones de media y autocovarianza.

Definición 2.4. Sea $\{X_t\}$ un proceso estocástico con $E[X_t^2]$ finita para todo t . $\{X_t\}$ es **estacionario de primer orden o estacionario en media** si

$$\mu_t = \mu_{t+k}$$

para todo t y $k > 0$, es decir si μ_t es independiente de t .

Definición 2.5. Sea $\{X_t\}$ un proceso estocástico con $E[X_t^2]$ finita para todo t . $\{X_t\}$ es **estacionario de segundo orden o débilmente estacionario** si

i) es estacionario de primer orden y

ii) $\gamma_{r,t} = \gamma_{r+k,t+k}$ para todo $r, t, r+k, t+k$ en $\{0, \pm 1, \pm 2, \dots\}$, es decir si $\gamma_{t,t+k}$ es independiente de t para toda k .

Corolario 2.1. Sea $\{X_t\}$ un proceso estocástico con $E[X_t^2]$ finita para toda t . Si $\{X_t\}$ es estrictamente estacionario entonces es débilmente estacionario.

La demostración del corolario anterior se puede encontrar en (Guerrero Guzmán, 2003, pág. 18). Los procesos estocásticos estacionarios de primer y segundo orden no exhiben tendencia, es decir, sus observaciones suben y bajan alrededor de un nivel fijo. El recíproco del corolario no siempre es cierto ya que si un proceso es débilmente estacionario no necesariamente es estrictamente estacionario. Sin embargo, en el caso de que la distribución conjunta del proceso estocástico sea la Normal entonces la estacionaridad débil sí implica la estacionaridad fuerte haciendo equivalentes los dos conceptos. Debido a esto, es común suponer que la distribución

de las variables aleatorias asociadas al proceso estocástico es la Normal (Shumway y Stoffer, 2010, pág. 23). En este caso, basta con conocer la media y la función de autocovarianza para tener la estacionaridad estricta y así obtener la función de distribución conjunta de $\{X_{t_0}, \dots, X_{t_0+n}\}$ para cualquier t_0 .

Se llama **incremento** al proceso estocástico que resulta de la resta de dos de sus variables aleatorias consecutivas, es decir, $\{W_t\} = \{X_t - X_{t-1}\}$. Si $\{X_t\}$ es un proceso estocástico estacionario entonces el proceso $\{W_t\}$ también es estacionario. Se llama **retraso** a la distancia entre un par de variables aleatorias de un proceso estocástico. Ya que la covarianza entre dos variables de un proceso estacionario sólo depende de la distancia entre ellas, es posible reducir la notación $\gamma_{t,t+k}$ a γ_k ya que sólo hace referencia al retraso k . Lo mismo ocurre con la función de autocorrelación la cual se denotará por $\rho_k = \gamma_k/\gamma_0$. Como resultado se tiene que las funciones de covarianza y autocorrelación de los procesos estacionarios son simétricas respecto al cero ya que si X_t y X_{t+k} son dos de sus variables la covarianza entre ellas es igual a la covarianza entre X_t y X_{t-k} , por lo que $\gamma_k = \gamma_{-k}$ y por lo tanto $\rho_k = \rho_{-k}$; además, se cumple que $\gamma_0 = Var(X_t) \geq 0$, $\rho_0 = 1$ y $|\rho_k| \leq 1$. La demostración de la última desigualdad puede encontrarse en (González Velasco y del Puerto García, 2009, pág. 57).

Definición 2.6. Sea $\{X_t\}$ un proceso estocástico estacionario tal que $E[X_t] = 0$. La **función de autocorrelación parcial** de $\{X_t\}$ en el retraso k denotada por **FAP** es

$$\alpha_k = \begin{cases} \rho_1 & \text{si } k = 1 \\ \frac{Cov(X_t - \tilde{X}_t, X_{t+k} - \tilde{X}_{t+k})}{\sqrt{Var(X_t - \tilde{X}_t)}\sqrt{Var(X_{t+k} - \tilde{X}_{t+k})}} & \text{si } k \neq 1 \end{cases}$$

donde

$$\tilde{X}_t = \beta_1 X_{t+k-1} + \beta_2 X_{t+k-2} + \dots + \beta_{k-1} X_{t+1},$$

$$\tilde{X}_{t+k} = \delta_1 X_{t+k-1} + \delta_2 X_{t+k-2} + \dots + \delta_{k-1} X_{t+1}$$

y las constantes $\{\beta_i\}$ y $\{\delta_i\}$, $i = 1, 2, \dots, k-1$ se obtienen minimizando las expresiones $E[(X_t - \tilde{X}_t)^2]$ y $E[(X_{t+k} - \tilde{X}_{t+k})^2]$ respectivamente.

El coeficiente de autocorrelación parcial de orden k mide la relación lineal existente entre observaciones separadas k periodos sin considerar el efecto de las variables intermedias. Ésta función, al tratarse de procesos estacionarios, solo depende de la distancia entre las variables, no de su posición en el tiempo, y es simétrica respecto al cero. Es posible dar un algoritmo iterativo para encontrar la FAP.

Sean $\phi_{k,j} = a_j$, $j = 1, \dots, k$ y $k = 1, 2, \dots$ los coeficientes que minimizan

$$E[(X_{k+1} - a_1 X_k - \dots - a_k X_1)^2] = \gamma_0 - 2 \sum_{i=1}^k a_i \gamma_i + \sum_{i=1}^k a_i^2 \gamma_0 + 2 \sum_{i,j=1, \dots, k-1, i < j} a_i a_j \gamma_{j-i}. \quad (2.1)$$

Derivando respecto a a_i e igualando a cero se obtiene el sistema de ecuaciones

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \vdots \\ \phi_{k,k} \end{pmatrix}.$$

Multiplicándolo por γ_0^{-1} se obtiene la ecuación equivalente

$$\boldsymbol{\phi}_k = \mathbf{R}_k^{-1} \boldsymbol{\rho}_k \quad (2.2)$$

donde $\boldsymbol{\phi}_k = (\phi_{k,1}, \phi_{k,2}, \dots, \phi_{k,k})^t$, $\mathbf{R}_k = (\rho_{i-j})_{i,j=1}^k$ y $\boldsymbol{\rho}_k = (\rho_1, \dots, \rho_k)^t$. El valor α_k de la FAP en el retraso k se obtiene resolviendo el sistema (2.2) mediante la regla de Cramer (González Velasco y del Puerto García, 2009, pág. 95). La solución es

$$\alpha_k = \phi_{k,k} = \frac{\rho_k - \hat{\boldsymbol{\rho}}_{k-1}^t \mathbf{R}_{k-1}^{-1} \boldsymbol{\rho}_{k-1}}{1 - \hat{\boldsymbol{\rho}}_{k-1}^t \mathbf{R}_{k-1}^{-1} \hat{\boldsymbol{\rho}}_{k-1}}$$

con $\hat{\boldsymbol{\rho}} = (\rho_k, \rho_{k-1}, \dots, \rho_1)^t$. Desarrollando la expresión anterior se obtiene el siguiente algoritmo iterativo que calcula la FAP sin necesidad de encontrar la matriz \mathbf{R}_{k-1}^{-1} :

- Sea $\phi_{0,0} = 0$.

- Para $n \geq 1$,

$$\phi_{n,n} = \frac{\rho_n - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho_{n-k}}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho_k}. \quad (2.3)$$

- Para $n \geq 2$, $\phi_{n,k} = \phi_{n-1,k} - \phi_{n,n} \phi_{n-1,n-k}$ con $k = 1, 2, \dots, n-1$.

Generalmente al inicio del análisis de una serie de tiempo no se cuenta con un modelo matemático que la describa. Para encontrarlo es necesario determinar el grado de dependencia entre los datos mediante la función de autocorrelación muestral.

Definición 2.7. Sea $\{X_t\}$ un proceso estocástico estacionario con media μ , varianza $\sigma^2 = \gamma_0$ y covarianzas $\{\gamma_k, k = 1, 2, \dots\}$. Sea $\{X_n\}$ la serie de tiempo relativa al proceso anterior.

- La **media muestral** de $\{X_n\}$ es

$$\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t.$$

- La **función de autocovarianzas muestral** de $\{X_n\}$ es

$$\bar{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_{t+k} - \bar{X}_n)(X_t - \bar{X}_n) \quad (2.4)$$

donde $0 \leq k < n$.

- La función de autocorrelación muestral de $\{X_n\}$ es

$$\bar{\rho}_k = \frac{\bar{\gamma}_k}{\bar{\gamma}_0} \quad (2.5)$$

donde $0 \leq k < n$.

- Supongamos que $\mu = 0$. La función de autocorrelación parcial muestral de $\{X_n\}$ es

$$\bar{\alpha}_k = \bar{\phi}_{k,k}$$

con $1 \leq k < n$ y los $\bar{\phi}_{k,k}$ se determinan mediante la Ecuación (2.2) donde los ρ_j se reemplazan por sus estimaciones siempre que $X_i \neq X_j$.

Es importante notar que si $k \geq n$ las estimaciones anteriores de las autocovarianzas y autocorrelaciones carecen de sentido. Una sugerencia para obtener una buena estimación de ρ_k es que el tamaño de la serie n sea mayor a 50 y que k tome valores menores a $n/4$ (Box, Jenkins, y Reinsel, 1994, pág. 32).

En general las series de tiempo con las que se trabaja en la práctica no son estacionarias ya que presentan ciertas características de irregularidad como la tendencia que se puede observar en la gráfica de la serie cuando ésta no se mantiene fluctuando alrededor de cierta línea paralela al eje x llamada nivel medio. Otra característica que aparece comúnmente es la periodicidad que se refiere a comportamientos repetitivos en intervalos de tiempo observados en la gráfica de la serie. Sin embargo es posible convertir estas series en estacionarias, es decir, en series sin tendencia ni periodicidad y que su gráfica siempre suba y baje alrededor de un nivel medio fijo sin comportamientos repetidos en ciertos periodos. Se llama **correlograma** a la gráfica que en el eje x contiene los retrasos k y en el eje y los valores de la función $|\bar{\rho}_k|$. Esta gráfica es utilizada como un indicador de la no estacionaridad: el correlograma de una serie con alguna tendencia decae lentamente a medida que k crece mientras que el de una serie con periodicidad imita este comportamiento. A continuación se presentan ejemplos de la FAS y FAP de una serie de tiempo.

Ejemplo 2.1. La Figura 2.1 representa la función de autocorrelación simple de la serie del Ejemplo 1.2 de megalitros de cerveza mensual producida en Australia. La estacionalidad de la serie se refleja en el marcado comportamiento senoidal de los valores de la FAS. La Figura 2.2 representa la función de autocorrelación parcial de la misma serie. En ella se observa que la mayoría de las autocorrelaciones parciales son negativas y que cuatro de ellas son significativas.

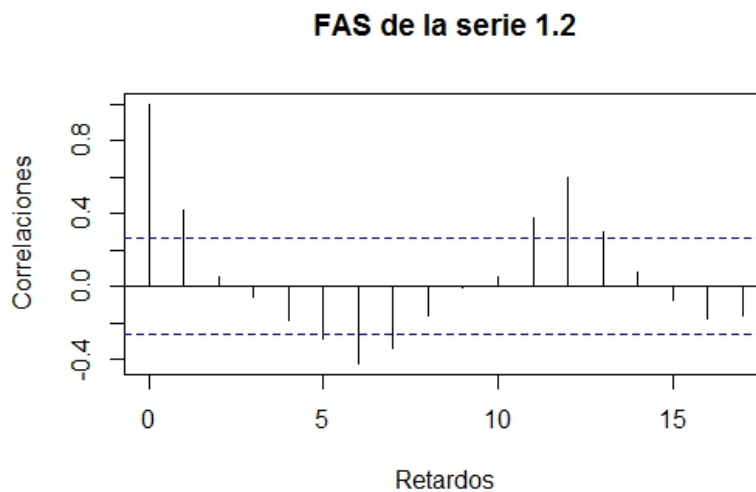


Figura 2.1: Gráfica de la FAS muestral de la producción de cerveza mensual en Australia medida en megalitros de enero de 1991 a agosto de 1995

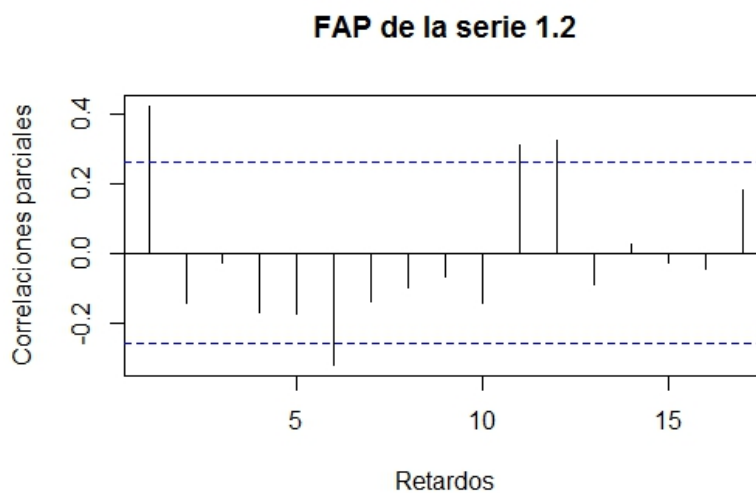


Figura 2.2: Gráfica de la FAP muestral de la producción de cerveza mensual en Australia medida en megalitros desde enero de 1991 hasta agosto de 1995

2.2. Proceso de ruido blanco

Aunque los valores sucesivos de una serie de tiempo son dependientes se hace la suposición de que está generada por una serie de choques aleatorios independientes, es decir, por una realización de un proceso estocástico

especial llamado **proceso de ruido blanco**.

Definición 2.8. Sea $\{X_t\}$ un proceso estocástico. $\{X_t\}$ es un **proceso de ruido blanco** si las variables aleatorias que lo conforman tienen media cero, tienen la misma varianza y son incorreladas.

De ahora en adelante se denotará a un proceso de ruido blanco por $\{Z_t\}$. Una característica de estos procesos es que el conocimiento de su pasado no aporta alguna información sobre su futuro. Algunos autores imponen la condición extra de que las variables aleatorias que los conforman deben tener distribución Normal y como resultado las variables no sólo están incorreladas sino que son independientes.

Para un proceso de ruido blanco estacionario $\gamma_k = 0$ para $k \neq 0$, $\rho_0 = \sigma^2$ donde σ^2 es la varianza de las variables aleatorias que lo conforman. Por lo tanto $\rho_k = 0$ para $k > 0$. Entonces se espera que la función $\bar{\rho}_k$ de una realización de un proceso Z_t también se aproxime a cero. De hecho, para n grande, las funciones $\bar{\rho}_1, \dots, \bar{\rho}_k$ con $k > 0$ son variables aleatorias independientes con distribución $N(\mu = 0, \sigma^2 = 1/n)$. A continuación se da un ejemplo de una serie generada por un ruido blanco.

Ejemplo 2.2. La gráfica siguiente representa una realización de tamaño 100 de un proceso de ruido blanco generado por una distribución uniforme con parámetros $(-1,1)$. A los procesos de este tipo se les llama **procesos puramente aleatorios** ya que están formados por variables aleatorias independientes e idénticamente distribuidas. Además son procesos estacionarios. En la Figura 2.3 se observa que no existe tendencia o periodicidad en el comportamiento de los datos, comprobando así que el proceso es estacionario. Las Figuras 2.4 y 2.5 contienen el correlograma y la FAP del proceso, donde en ambas se observa que las correlaciones simples y parciales son no significativas desde los primeros retrasos.

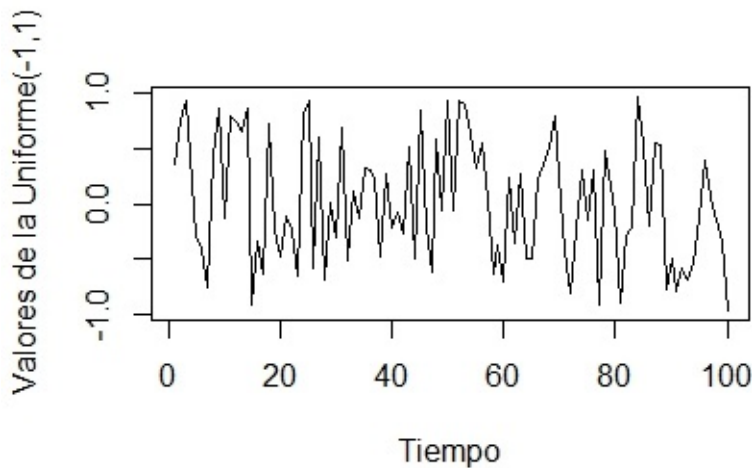


Figura 2.3: Gráfica de una realización de un proceso de ruido blanco generado por una distribución $U(-1,1)$

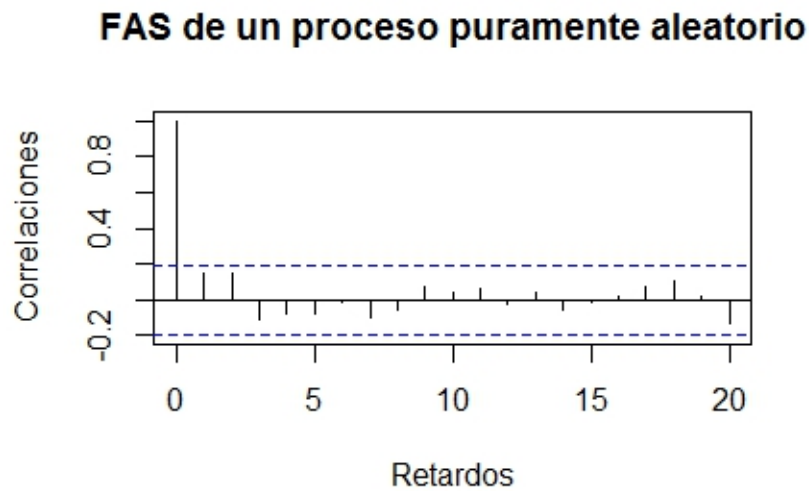


Figura 2.4: Gráfica de la FAS muestral de una realización de un proceso de ruido blanco generado por una distribución $U(-1,1)$

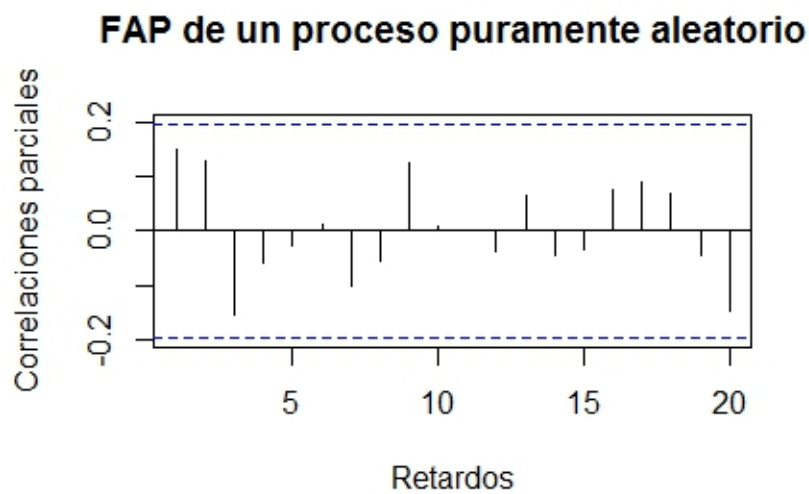


Figura 2.5: Gráfica de la FAP muestral de una realización de un proceso de ruido blanco generado por una distribución $U(-1,1)$

2.3. Procesos lineales

La clase de los procesos lineales, la cual incluye a los procesos autorregresivos (procesos AR), a los de medias móviles (procesos MA) y a los mixtos (procesos ARMA), provee un marco general para el estudio de los procesos estacionarios. Así, los procesos lineales son ampliamente usados en el análisis de las series de tiempo

(Brockwell y Davis, 2006, pág. 83), (Chatfield, 2016, pág. 41, 45 y 47).

Definición 2.9. Sea $\{X_t\}$ un proceso estocástico. $\{X_t\}$ es un **proceso lineal** si puede representarse como

$$X_t = \sum_{i=-\infty}^{\infty} \Psi_i Z_{t-i} \quad (2.6)$$

para toda $t = 0, \pm 1, \dots$, donde Ψ_i es una sucesión de números reales tal que $\sum_{i=-\infty}^{\infty} |\Psi_i|$ converge y $\{Z_t\}$ es un proceso de ruido blanco.

Existe una manera más compacta de expresar la Ecuación 2.6 mediante el operador de retraso.

Definición 2.10. Sea $\{X_t\}$ un proceso estocástico. El **operador de retraso** B^k transforma el proceso original $\{X_1, \dots, X_n\}$ en el proceso $\{X_{1-k}, \dots, X_{n-k}\}$, es decir, transforma un proceso con n variables aleatorias en uno con $n - k$ variables. B^k funciona de la siguiente manera:

- $B^0 = I$ donde I es la transformación identidad, es decir, $B^0 X_t = X_t$ para $t = 1, \dots, n$,
- $B X_t = X_{t-1}$ para $t = 1, \dots, n$ y
- $B^k X_t = B(B^{k-1} X_t) = X_{t-k}$ para $t = 1, \dots, n$.

Usando el operador de retraso y haciendo $\Psi(B) = \sum_{i=-\infty}^{\infty} \Psi_i B^i$ tenemos que la Fórmula (2.6) se puede reescribir como $X_t = \Psi(B)Z_t$.

Proposición 2.1. Sea $\{X_t\}$ un proceso estacionario con media 0 y función de covarianza γ_X . Si $\{\Psi_i\}$ es una sucesión de valores reales tal que $\sum_{i=-\infty}^{\infty} |\Psi_i|$ converge entonces el proceso

$$Y_t = \sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$$

es estacionario con media cero y función de covarianza

$$\gamma_Y(k) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Psi_i \Psi_j \gamma_X(k - i + j)$$

para toda $k \geq 0$.

Proposición 2.2. Si en la proposición anterior el proceso $\{X_t\}$ es ruido blanco entonces el proceso

$$Y_t = \sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$$

es un proceso lineal con función de covarianza

$$\gamma_Y(k) = \sigma^2 \sum_{i=-\infty}^{\infty} \Psi_i \Psi_{i+k}$$

para toda $k \geq 0$ donde σ^2 es la varianza de X_t .

Las demostraciones de las dos proposiciones anteriores se pueden consultar en (Brockwell y Davis, 2006, pág. 52).

Capítulo 3

Modelos ARIMA

3.1. Introducción

La clase más popular de modelos lineales para series de tiempo consiste en los modelos autorregresivos de medias móviles (ARMA) que incluyen los modelos autorregresivos (AR) y los de medias móviles (MA). Los modelos ARMA se utilizan para describir estructuras lineales dinámicas, para representar relaciones lineales entre los retrasos de las variables y para pronosticar las series en estudio. Una clase de modelos particularmente útil contiene a los modelos autorregresivos integrados de medias móviles (ARIMA) que incluye a los modelos ARIMA estacionarios.

3.2. Modelos de medias móviles

Existen sucesos que influyen en ciertos fenómenos produciendo efectos inmediatos que no duran mucho pero que pueden seguir afectando en varios periodos posteriores. Resultan muy apropiados los *procesos de medias móviles* para describir los fenómenos mencionados, por ejemplo para describir los índices económicos los cuales son altamente influenciados por las decisiones gubernamentales desde que se ponen en práctica e incluso hasta en periodos posteriores.

Definición 3.1. Sea $\{X_t\}$ un proceso estocástico. $\{X_t\}$ es un **proceso de medias móviles de orden q** , denotado por $MA(q)$, si puede representarse mediante la ecuación

$$X_t = Z_t - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q} \quad (3.1)$$

para toda $t = 0, \pm 1, \dots$ donde $\{Z_t\}$ es un proceso de ruido blanco y $\{\theta_i\}_{i=1}^q$ son números reales.

La Fórmula (3.1) se puede reescribir en forma más compacta mediante el operador de retraso B como

$$X_t = \theta_q(B)Z_t$$

donde $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$. Por tanto el proceso $MA(q)$ es un proceso lineal donde $\psi_0 = 1$, $\psi_j = -\theta_j$, $j = 1, \dots, q$ y $\psi_j = 0$ en otro caso, y en consecuencia este proceso es estacionario para cualquier valor de θ_j .

Sea $\{X_t\}$ un proceso $MA(q)$. La función de autocovarianza de $\{X_t\}$ es

$$\gamma_k = \begin{cases} \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2) & \text{si } k = 0 \\ \sigma^2(-\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{i+k}) & \text{si } k = 1, 2, \dots, q \\ 0 & \text{si } k > q \end{cases}$$

donde σ^2 es la varianza del proceso de ruido blanco. La función de autocorrelación de $\{X_t\}$ es

$$\rho_k = \begin{cases} 1 & \text{si } k = 0 \\ \frac{(-\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{i+k})}{\sum_{i=1}^q \theta_i^2} & \text{si } k = 1, 2, \dots, q \\ 0 & \text{si } k > q \end{cases}$$

Observe que la función de autocorrelación de los procesos $MA(q)$ es nula para todos los retrasos mayores a q .

Como se verá más adelante la función de autocorrelación juega un papel importante en la identificación de un modelo que explique una serie de observaciones dada. Es deseable que exista una biyección entre una función de autocorrelación y el modelo $MA(q)$ del que proviene, sin embargo esto no siempre ocurre. Por ejemplo la función

$$\rho_k = \begin{cases} \frac{-\theta}{1+\theta^2} & \text{si } k = 1 \\ 0 & \text{si } k > 1 \end{cases}$$

podría pertenecer a cualquiera de los siguientes procesos $MA(1)$:

$$X_t = Z_t - \theta Z_{t-1} \quad \text{o} \quad \widehat{X}_t = \widehat{Z}_t - \frac{1}{\theta} \widehat{Z}_{t-1}.$$

Además, como sólo se pueden observar las series X_t y \widehat{X}_t y no los procesos de ruido blanco Z_t y \widehat{Z}_t , es imposible distinguir los procesos $MA(1)$ anteriores por lo menos a primera vista. Reescribiendo el primer modelo como

$$X_t = Z_t - \theta X_{t-1} - \theta^2 X_{t-2} - \dots = Z_t - \sum_{j=1}^{\infty} \theta^j X_{t-j} \quad (3.2)$$

se observa que cuando $|\theta| < 1$ el efecto de los retrasos X_{t-k} en X_t tiende a cero cuando k aumenta, mientras que si $|\theta| > 1$ el efecto de los retrasos también aumenta cuando k aumenta, es decir, un suceso ocurrido en el pasado tendría más influencia en el presente que uno ocurrido recientemente. Además si $|\theta| < 1$ la Expresión (3.2) se puede escribir como una suma infinita que converge:

$$Z_t = X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \dots = \sum_{j=0}^{\infty} \theta^j X_{t-j}.$$

De aquí surge el concepto de invertibilidad.

Definición 3.2. Sea $\{X_t\}$ un proceso estocástico. $\{X_t\}$ es *invertible* si se puede representar como

$$Z_t = \pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

para toda $t = 0, \pm 1, \dots$ donde $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, $\pi_0 = 1$ y $\sum_{j=0}^{\infty} |\pi_j| < \infty$.

Observe que $\pi(B)$ es el operador inverso de $\theta_q(B)$ por lo que es común usar la notación $\theta_q(B)^{-1}$ en lugar de $\pi(B)$. Entonces los coeficientes π_j en la expresión anterior se pueden obtener a partir de que

$$\pi(B)\theta_q(B) = 1.$$

Proposición 3.1. Sea $\{X_t\}$ un proceso $MA(q)$. $X_t = \theta_q(B)Z_t$ es *invertible* si y sólo si las raíces de la ecuación

$$\theta_q(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0$$

son en módulo mayores que la unidad, es decir que el valor absoluto de las raíces de la ecuación anterior son mayores a uno.

La demostración de la proposición anterior puede encontrarse en (Brockwell y Davis, 2013, pág. 85). Por otro lado la función de autocorrelación parcial nunca se corta y tiende a cero cuando k tiende a infinito. Un ejemplo de esto es el siguiente. Considere el proceso $MA(1)$ descrito por $X_t = Z_t - \theta Z_{t-1}$ con $|\theta| < 1$. Mediante la Ecuación (2.3) se deduce que

$$\alpha_k = \phi_{k,k} = -\frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}}$$

con $k \geq 1$. Observe que la FAP tiende a cero geométricamente cuando k tiende a infinito.

Ejemplo 3.1. A continuación se muestran la gráfica, la FAS y la FAP de una realización del proceso $MA(1)$

$$X_t = Z_t - 0.7Z_{t-1} \quad (3.3)$$

donde $\{Z_t\}$ es un proceso de ruido blanco con $\sigma^2 = 1$.

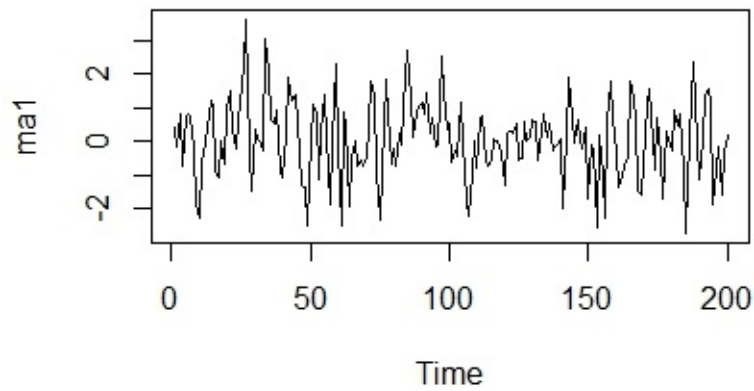


Figura 3.1: Gráfica de una realización del proceso $MA(1)$ $X_t = Z_t - 0.7Z_{t-1}$

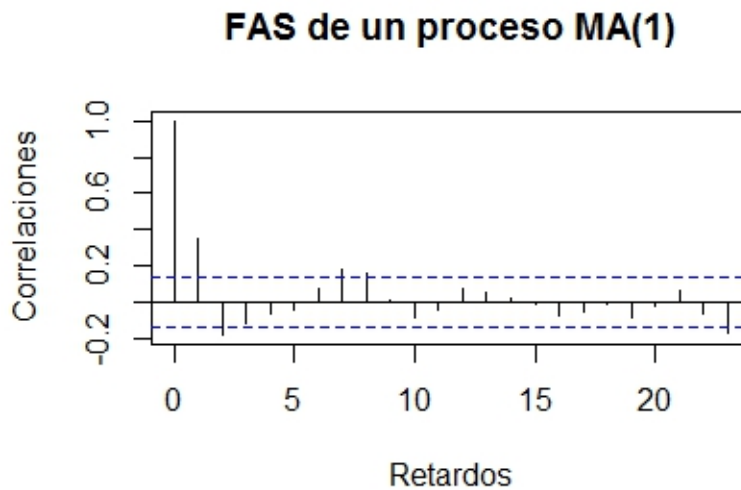


Figura 3.2: Gráfica de la FAS muestral de una realización del proceso $MA(1)$ $X_t = Z_t - 0.7Z_{t-1}$

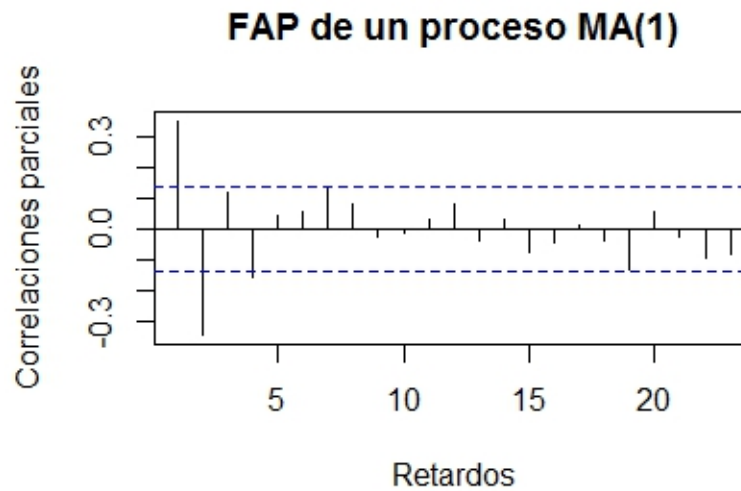


Figura 3.3: Gráfica de la FAP muestral de una realización del proceso MA(1) $X_t = Z_t - 0.7Z_{t-1}$

La Figura 3.1 representa 200 observaciones simuladas del proceso con el Modelo (3.3). No se observa tendencia o periodicidad alguna. La Figura 3.2 representa su correlograma. En ella se observa que $\bar{\rho}_1$ sobresale de las bandas de confianza al 95 % para los coeficientes de autocorrelación de un ruido blanco, lo cual conduce a rechazar la hipótesis de que los datos provienen de un proceso de ruido blanco. Además el hecho de que sea el valor de $\bar{\rho}_1$ el que sobresalga de las bandas está relacionado con el orden del proceso *MA*. La Figura 3.3 representa la FAP de la serie, la cual tampoco muestra comportamientos periódicos.

Ejemplo 3.2. A continuación se muestran la gráfica, la FAS y la FAP de una realización del proceso $MA(2)$

$$X_t = Z_t - 0.1Z_{t-1} - 0.7Z_{t-2} \quad (3.4)$$

donde $\{Z_t\}$ es un proceso de ruido blanco con $\sigma^2 = 1$.

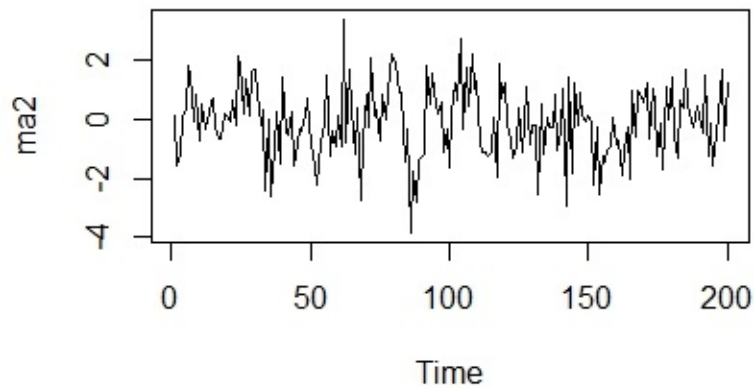


Figura 3.4: Gráfica de una realización del proceso $MA(2)$ $X_t = Z_t - 0.1Z_{t-1} - 0.7Z_{t-2}$

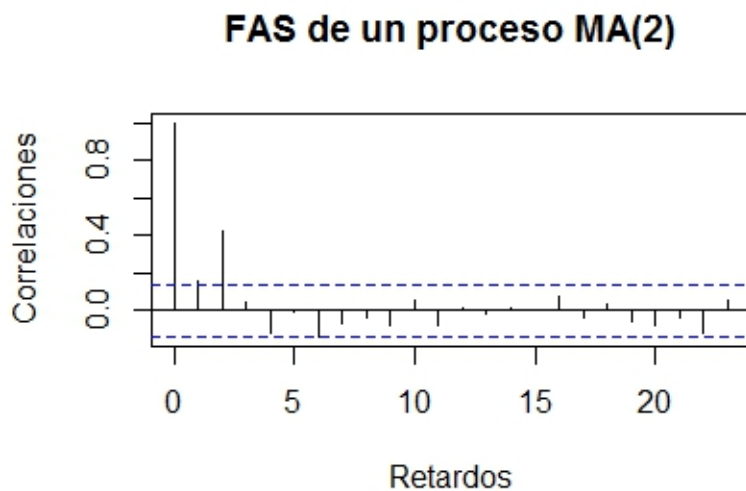


Figura 3.5: Gráfica de la FAS de una realización del proceso $MA(2)$ $X_t = Z_t - 0.1Z_{t-1} - 0.7Z_{t-2}$

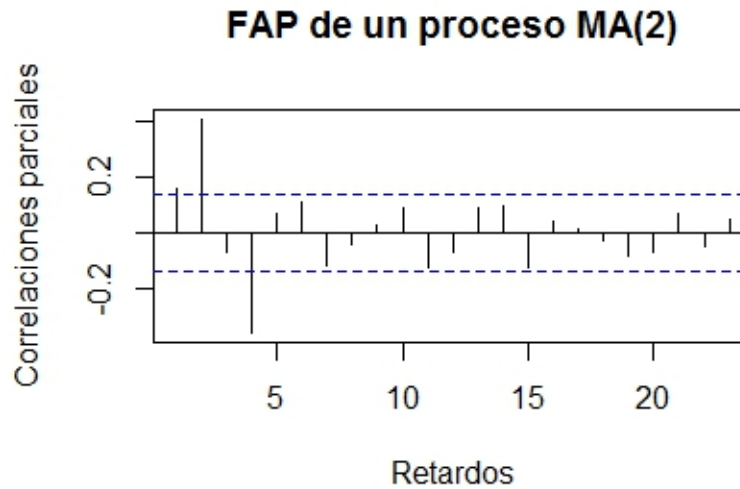


Figura 3.6: Gráfica de la FAP de una realización del proceso MA(2) $X_t = Z_t - 0.1Z_{t-1} - 0.7Z_{t-2}$

La Figura 3.4 representa 200 observaciones simuladas del proceso (3.4). Igual que en el ejemplo anterior no se observa tendencia o periodicidad alguna. La Figura 3.5 representa su correlograma. En este caso es $\bar{\rho}_2$ la correlación que sobresale de las bandas de confianza, lo cual indica que la serie no proviene de un ruido blanco y que el posible orden del modelo *MA* del que proviene es dos. Observe que la Figura 3.6, que representa la FAP de la serie, tampoco exhibe tendencia ni periodicidad.

3.3. Modelos autorregresivos

Un proceso estocástico que presenta dependencia lineal entre sus variables se llama *proceso autorregresivo*. Este modelo se utiliza cuando se supone que el valor actual de la serie depende de los valores pasados junto con un error aleatorio.

Definición 3.3. Sea $\{X_t\}$ un proceso estocástico. $\{X_t\}$ es un **proceso autorregresivo de orden p** , denotado por **AR(p)**, si puede caracterizarse mediante la ecuación

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t$$

para toda $t = 0, \pm 1, \dots$ donde $\{Z_t\}$ es un proceso de ruido blanco y $\{\phi_i\}_{i=1}^p$ son números reales.

Al igual que en los modelos MA, la expresión anterior se puede escribir de forma más compacta mediante el operador de retraso como

$$Z_t = \phi_p(B)X_t.$$

Observe que en virtud de la definición (3.2) con $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$. Los procesos AR son procesos invertibles.

Definición 3.4. Sea $\{X_t\}$ un proceso estocástico. $\{X_t\}$ es un proceso **causal** si puede expresarse como un **proceso MA**(∞), es decir si

$$X_t = \psi(B)Z_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

donde $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, $\psi_0 = 1$ y $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Observe que los procesos autorregresivos causales son procesos lineales para los que $\psi_i \neq 0$ con $i \geq 0$, y en consecuencia son procesos estacionarios.

Proposición 3.2. Sea $\{X_t\}$ un proceso $AR(p)$ definido por $Z_t = \phi_p(B)X_t$. $\{X_t\}$ es un proceso estacionario (causal) si y sólo si las soluciones de la ecuación

$$\phi_p(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0$$

son en módulo mayores que la unidad.

Definición 3.5. Sea $\{X_t\}$ un proceso estocástico. $\{X_t\}$ es un **proceso AR**(∞) si se puede representar como

$$Z_t = \sum_{i=0}^{\infty} \pi_i X_{t-i}$$

para todo $t = 0, \pm 1, \dots$ donde $\sum_{i=0}^{\infty} |\pi_i| < \infty$, $\pi_0 = 1$ y $\{Z_t\}$ es un proceso de ruido blanco.

Un proceso $AR(p)$ estacionario puede representarse como un proceso $MA(\infty)$ y un proceso $MA(q)$ invertible puede representarse como un $AR(\infty)$ con coeficientes que decrecen geométricamente. Por lo tanto la FAP de un proceso $MA(q)$ nunca se corta y tiende a cero cuando k aumenta. Además su forma depende de una mezcla de funciones exponenciales y senosoidales dependiendo de las raíces de la ecuación $\theta_q(z) = 0$.

3.3.1. Proceso AR(1)

Sea $\{X_t\}$ un proceso $AR(1)$ descrito por $Z_t = X_t - \phi X_{t-1}$ donde $\{Z_t\}$ es un proceso de ruido blanco con varianza σ^2 . Observe que el proceso anterior puede ser expresado como un $MA(\infty)$ siempre que $|\phi| < 1$, es decir puede expresarse como $X_t = \sum_{i=0}^{\infty} \phi^i Z_{t-i}$. Aplicando la proposición 2.2 a esta expresión se tiene que su función de autocovarianza es

$$\gamma_k = \sigma^2 \sum_{i=0}^{\infty} \phi^i \phi^{i+k} = \frac{\sigma^2 \phi^k}{1 - \phi^2}$$

con $k \geq 0$. Por lo tanto su varianza es

$$\gamma_0 = \frac{\sigma^2}{1 - \phi^2}$$

y su función de autocorrelación es

$$\rho_k = \phi^k.$$

Observe que a diferencia de los procesos *MA*, la función de autocorrelación del *AR(1)* no se corta en ningún retraso sino que siempre es no nula y decrece en forma geométrica debido a que $|\phi| < 1$.

Ejemplo 3.3. A continuación se muestran la gráfica, la FAS y la FAP de una realización del proceso *AR(1)*

$$Z_t = X_t - 0.2X_{t-1} \quad (3.5)$$

donde $\{Z_t\}$ es un proceso de ruido blanco con $\sigma^2 = 1$.

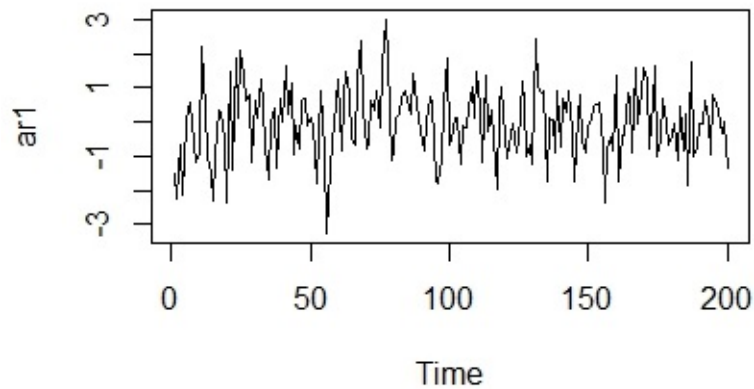


Figura 3.7: Gráfica de una realización del proceso *AR(1)* $Z_t = X_t - 0.2X_{t-1}$

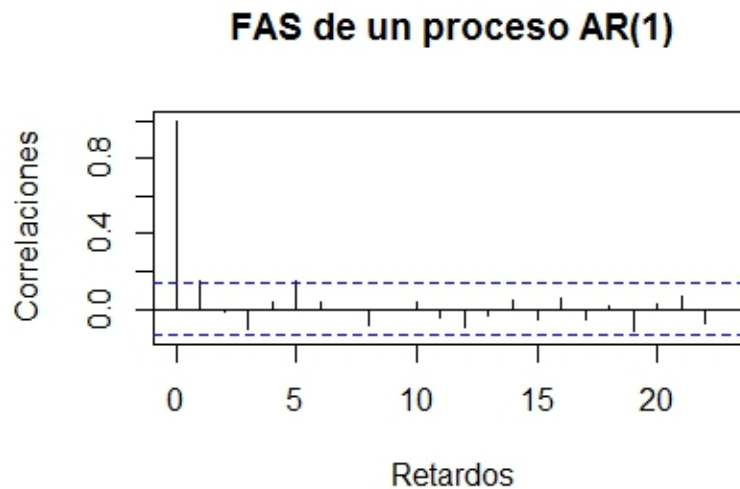


Figura 3.8: Gráfica de la FAS muestral de una realización del proceso *AR(1)* $Z_t = X_t - 0.2X_{t-1}$

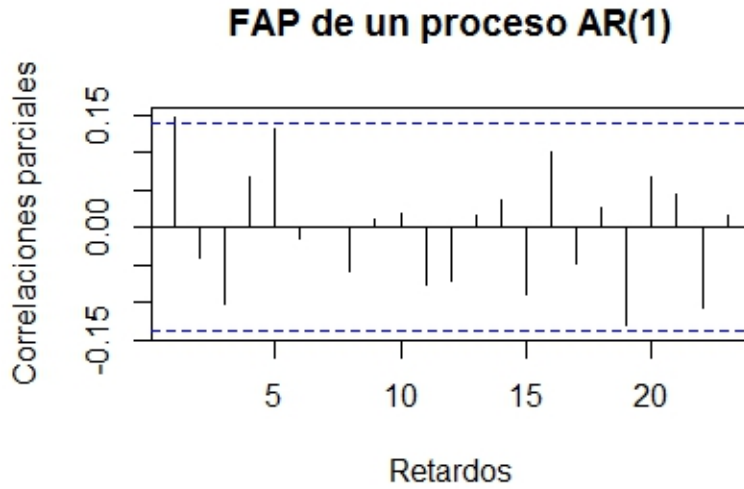


Figura 3.9: Gráfica de la FAP muestral de una realización del proceso AR(1) $Z_t = X_t - 0.2X_{t-1}$

La Figura 3.7 representa 200 observaciones simuladas del proceso (3.5). La Figura 3.8 representa su correlograma. Observe que a diferencia de los modelos *MA*, ningún valor de la FAS de los modelos *AR(1)* sobresale de las bandas de confianza y decrecen en forma geométrica. Observe que la Figura 3.9, que representa la FAP de la serie, tampoco exhibe tendencia ni periodicidad.

3.3.2. Proceso AR(2)

Sea $\{X_t\}$ un proceso estacionario *AR(2)* descrito por

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t \quad (3.6)$$

con $\{Z_t\}$ un proceso de ruido blanco de varianza σ^2 . Multiplicando por X_{t-k} y aplicando esperanza a la igualdad (3.6) se obtiene que

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \quad k \geq 1 \quad (3.7)$$

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2. \quad (3.8)$$

Dividiendo entre γ_0 las expresiones (3.7) y (3.8) resulta que

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 1 \quad (3.9)$$

$$1 = \phi_1 \rho_1 + \phi_2 \rho_2 + \frac{\sigma^2}{\gamma_0}. \quad (3.10)$$

Es posible encontrar expresiones para γ_0 , γ_k y ρ_k que no dependan de sus valores anteriores. Haciendo $k = 1, 2$ en la Expresión (3.9) y recordando que $\rho_k = \rho_{-k}$ y $\rho_0 = 1$ se tienen las siguientes igualdades:

$$\rho_1 = \phi_1\rho_0 + \phi_2\rho_1$$

$$\rho_2 = \phi_1\rho_1 + \phi_2\rho_0$$

y despejando ρ_1 y ρ_2 de cada una se obtiene que

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

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

Sustituyendo los valores anteriores en la igualdad (3.10) se obtiene la *varianza* del proceso $AR(2)$:

$$\rho_0 = \frac{\sigma^2(1 - \phi_2)}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)}$$

donde las ϕ 's satisfacen que

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

condiciones que aseguran la estacionaridad.

Para obtener la función de autocorrelación explícita del proceso $AR(2)$ habrá que considerar la expresión (3.9).

La solución general de ésta es (González Velasco y del Puerto García, 2009, pág.120)

$$\rho_k = A_1 z_1^{-k} + A_2 z_2^{-k}, \quad k \geq 0$$

donde z_1 y z_2 son las raíces de la ecuación característica

$$\phi_2 z^2 + \phi_1 z - 1 = 0$$

y las A 's son constantes que se determinan mediante las condiciones iniciales

$$\rho_0 = 1 \quad \text{y} \quad \rho_1 = \frac{\phi_1}{1 - \phi_2}.$$

Respecto a las raíces z_1 y z_2 de la ecuación característica pueden ocurrir los siguientes casos (ver (Brockwell y Davis, 2006, pág. 91)):

- Que las dos sean reales y diferentes. En este caso

$$\rho_k = A_1 z_1^{-k} + A_2 z_2^{-k},$$

por lo que ρ_k tiende a cero cuando k aumenta y la forma de la función depende de la igualdad o diferencia de los signos de z_1 y z_2 .

- Que las dos sean reales e iguales. Entonces

$$\rho_k = z^{-k}(A_1 + kA_2)$$

donde $z = z_1 = z_2$. En este caso también ocurre que ρ_k tiende a cero cuando k tiende a infinito.

- Que las dos sean complejos conjugados, es decir $z_2 = \bar{z}_1$. Aquí

$$\rho_k = A|z_1|^{-k} \cos(k\theta + B)$$

donde θ es el argumento de z_1 y las constantes A y B se determinan mediante las condiciones iniciales.

Entonces ρ_k decrece en forma senosoidal.

Ejemplo 3.4. A continuación se muestran la gráfica, la FAS y la FAP de una realización del proceso $AR(2)$

$$Z_t = X_t - 0.3X_{t-1} + 0.8X_{t-2} \quad (3.11)$$

donde $\{Z_t\}$ es un proceso de ruido blanco con $\sigma^2 = 1$.

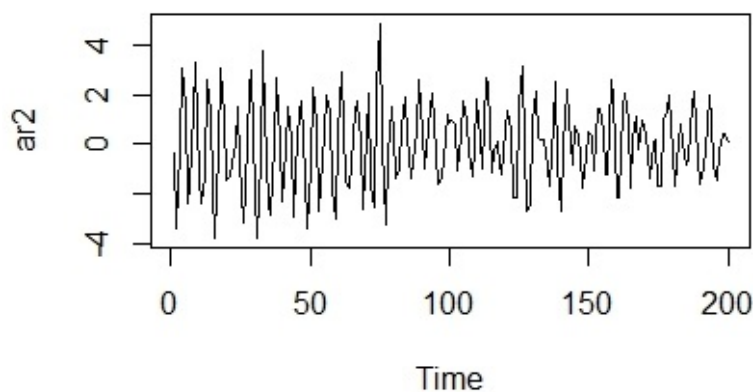


Figura 3.10: Gráfica de una realización del proceso $AR(2)$ $Z_t = X_t - 0.3X_{t-1} + 0.8X_{t-2}$

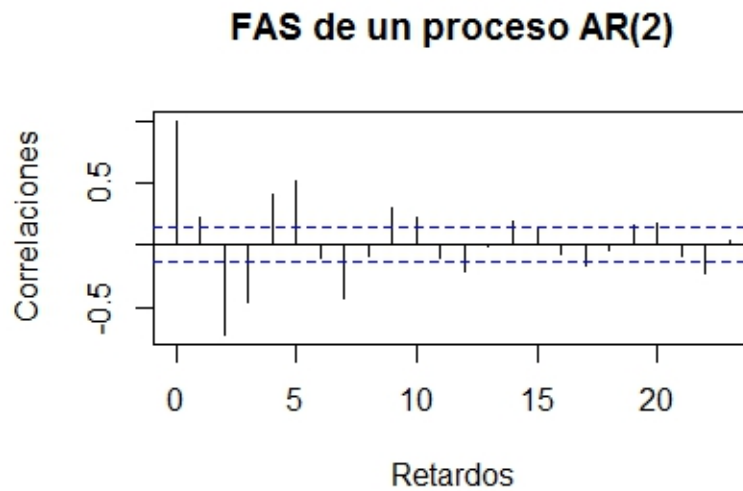


Figura 3.11: Gráfica de la FAS de una realización del proceso AR(2) $Z_t = X_t - 0.3X_{t-1} + 0.8X_{t-2}$

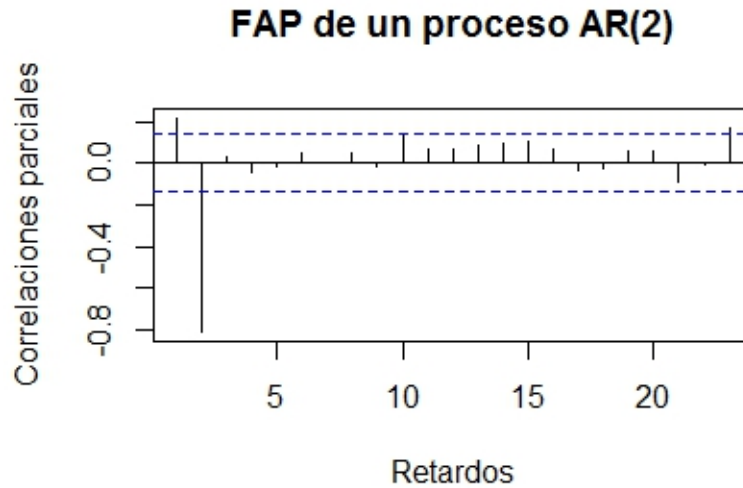


Figura 3.12: Gráfica de la FAP de una realización del proceso AR(2) $Z_t = X_t - 0.3X_{t-1} + 0.8X_{t-2}$

La Figura 3.10 representa 200 observaciones simuladas del proceso (3.11) y las Figuras 3.11 y 3.12 representan su función de correlación simple y parcial. Se observa que en la FAP el valor de la segunda correlación parcial es el que sobresale más de las bandas de confianza, lo que se relaciona con el orden del proceso AR que en este caso es dos.

3.3.3. Proceso AR(p)

Sea $\{X_t\}$ un proceso estacionario $AR(p)$ descrito por

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t \quad (3.12)$$

con $\{Z_t\}$ un proceso de ruido blanco con varianza σ^2 . Con el mismo procedimiento que para el proceso $AR(2)$ se puede obtener una expresión para γ_k a partir de sus valores en retrasos anteriores. Así

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} + 0, \quad k \geq 1. \quad (3.13)$$

Dividiendo entre γ_0 la expresión (3.13) se tiene

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \quad k \geq 1. \quad (3.14)$$

La solución de la Ecuación (3.14) es

$$\rho_k = \sum_{i=1}^p A_i z_i^{-k}, \quad k \geq 0$$

donde z_i , $i = 1, \dots, p$, son las soluciones de la ecuación característica y las constantes A_i , $i = 1, \dots, p$ se encuentran a partir de las condiciones iniciales. La estructura de ésta función de autocorrelación podría ser complicada debido a los términos con raíces reales que aportan funciones exponenciales y funciones senosoidales que provienen de las raíces complejas conjugadas.

Como se vio en la sección anterior la función de autocorrelación de un proceso $MA(q)$ se anula a partir del retraso q , por lo que ésta se usa generalmente para identificar su orden. No ocurre lo mismo con los procesos AR puesto que dicha función no se anula en ningún retraso ni presenta alguna otra característica relacionada con el orden del proceso, sin embargo la función de autocorrelación parcial sí aporta información acerca del orden del proceso AR .

Considere un proceso estocástico $AR(p)$ descrito por la Ecuación (3.12). Los primeros p valores de su función de autocorrelación determinan de manera única los parámetros del modelo mediante las llamadas *ecuaciones de Yule-Walker*:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \cdots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \cdots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \cdots + \phi_p. \end{aligned}$$

Tomando en cuenta el sistema anterior y que $k = p$, la solución del sistema (2.2) es

$$\phi_{p,j} = \phi_j$$

donde $j = 1, \dots, p$, por lo que

$$\alpha_p = \phi_p.$$

Si $k > p$, de la Ecuación (3.14) se deduce que

$$\phi_{k,j} = \phi_j$$

donde $j = 1, \dots, p$ y

$$\phi_{k,j} = 0$$

con $j = p + 1, \dots, k$ es solución del sistema (2.2) por lo que

$$\alpha_k = 0.$$

Por lo tanto la función de autocorrelación parcial de un proceso $AR(p)$ es nula a partir del retraso p siendo ésta la que generalmente se utiliza para identificar el orden de un proceso AR .

3.4. Modelos autorregresivos de medias móviles

Los *modelos autoregresivos de medias móviles*, mejor conocidos como modelos $ARMA$, son una combinación de los modelos anteriores. Debido a que los modelos MA requieren un número finito de parámetros sin imponer restricciones sobre sus valores y a que los modelos AR pueden representarse como modelos $MA(\infty)$, lo que impone condiciones sobre el decrecimiento de los coeficientes ψ_i , los modelos $ARMA$ pueden representar procesos mediante un número no muy grande de parámetros, donde los primeros q coeficientes son cualesquiera y los siguientes decrecen de acuerdo a alguna ley.

Definición 3.6. Sea $\{X_t\}$ un proceso estocástico. $\{X_t\}$ es un **proceso mixto autorregresivo de media móvil de orden (p,q)** denotado por $ARMA(p,q)$, si

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q} \quad (3.15)$$

para todo $t = 0, \pm 1, \dots$ donde $\{Z_t\}$ es un proceso de ruido blanco, $\{\phi_i\}_{i=1}^p$, $\{\theta_i\}_{i=1}^q$ son constantes reales y los polinomios $\phi_p(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ y $\theta_q(z) = 1 - \theta_1 z - \dots - \theta_q z^q$ no tienen factores comunes.

El modelo $ARMA(p, q)$ se puede escribir de manera compacta con el operador de retraso B como

$$\phi_p(B)X_t = \theta_q(B)Z_t.$$

Observe que la Definición 3.6 prohíbe que los polinomios $\phi_p(z)$ y $\theta_q(z)$ tengan raíces comunes para evitar la sobreparametrización. Suponga que $\lambda_1, \dots, \lambda_p$ son las raíces reales del polinomio $\phi_p(z) = 0$ donde $\lambda_i \neq \lambda_j$, $i, j = 1, \dots, p$. También suponga que $\delta_1, \dots, \delta_q$ son las raíces reales del polinomio $\theta_q(z) = 0$ con $\delta_i \neq \delta_j$, $i, j = 1, \dots, q$. Entonces la Ecuación (3.15) sería equivalente a la ecuación

$$(1 - \lambda_1 B) \cdots (1 - \lambda_p B) X_t = (1 - \delta_1 B) \cdots (1 - \delta_q B) Z_t$$

y si los polinomios tuvieran una raíz en común se cancelarían dos factores en la fórmula anterior quedando un modelo $ARMA(p-1, q-1)$.

Observe que haciendo $\theta_q(B) = 1$ en la Ecuación (3.15) se tiene un modelo autorregresivo de orden p mientras que si $\phi_p(B) = 1$ se tiene un modelo $MA(q)$. Las definiciones de invertibilidad y estacionaridad se extienden para procesos $ARMA$.

Definición 3.7. Sea $\{X_t\}$ un modelo $ARMA(p, q)$ descrito por $\phi_p(B)X_t = \theta_q(B)Z_t$. $\{X_t\}$ es **invertible** si se puede expresar como

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

para todo $t = 0, \pm 1, \dots$ donde $\pi_0 = 1$ y $\sum_{j=0}^{\infty} |\pi_j| < \infty$.

Definición 3.8. Sea $\{X_t\}$ un modelo $ARMA(p, q)$ descrito por $\phi_p(B)X_t = \theta_q(B)Z_t$. $\{X_t\}$ es **causal** o estacionario si se puede expresar como

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

para todo $t = 0, \pm 1, \dots$ donde $\psi_0 = 1$ y $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Proposición 3.3. Sea $\{X_t\}$ un proceso $ARMA(p, q)$ descrito por $\phi_p(B)X_t = \theta_q(B)Z_t$. $\{X_t\}$ es invertible si y sólo si las raíces del polinomio $\theta_q(z)$ son en módulo mayores que la unidad.

Proposición 3.4. Sea $\{X_t\}$ un proceso $ARMA(p, q)$ descrito por $\phi_p(B)X_t = \theta_q(B)Z_t$. $\{X_t\}$ es causal si y sólo si las raíces del polinomio $\phi_p(z)$ son en módulo mayores que la unidad.

La demostración de las proposiciones anteriores puede consultarse en (Brockwell y Davis, 2013, pág. 85 y 87).

La función de autocovarianza de un proceso $ARMA(p, q)$ es

$$\gamma_h = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$$

donde σ^2 es la varianza de $\{Z_t\}$ y su función de autocorrelación simple es

$$\rho_h = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+h}}{\sum_{j=0}^{\infty} \psi_j^2}.$$

Si $q - p < 0$ la función de autocovarianza consiste en una mezcla de funciones exponenciales y ondas senosoidales que tienden a cero según las raíces del polinomio $\phi_p(z)$. Ni la función de autocovarianza ni la de autocorrelación simple se cortan en algún retraso. La función de autocorrelación parcial de un $ARMA(p, q)$ nunca se anula y decrece de acuerdo a una mezcla de funciones exponenciales y senosoidales según las raíces de los polinomios autorregresivo y de media móvil. Por lo tanto la FAS y la FAP tienen estructuras complejas con un número infinito de elementos distintos de cero, por lo que es difícil determinar el orden del proceso mediante ellas.

3.5. Modelos autorregresivos integrados de medias móviles

Las secciones anteriores se han referido a procesos estacionarios, sin embargo muchas series temporales no lo son. En este caso la no estacionaridad se puede manifestar de formas distintas, por ejemplo cuando la serie no es estable en media, es decir cuando existe cierta tendencia y el nivel de la serie no es el mismo a lo largo del tiempo. También puede ocurrir que la serie no sea estable en varianza, es decir cuando la variabilidad de los datos crece o disminuye junto con el nivel de la serie. Otra manifestación de la no estacionaridad podría ocurrir en las autocorrelaciones cuando éstas se modifican con el tiempo. Es posible convertir una serie no estacionaria en estacionaria mediante ciertas transformaciones.

Por ejemplo considere un proceso $\{X_t\}$ correspondiente a un AR(1) donde $\phi_1 = 1$. Así

$$X_t = X_{t-1} + Z_t \quad (3.16)$$

para toda $t = 0, \pm 1, \dots$ con $\{Z_t\}$ un proceso de ruido blanco con varianza σ^2 . Suponiendo que el proceso se inicia en un pasado infinitamente lejano es posible reescribirlo como

$$X_t = \sum_{j=0}^{\infty} Z_{t-j}$$

de donde la varianza de $\{X_t\}$ sería infinita por lo que el proceso no sería estacionario. Incluso si el proceso iniciara en $t = 0$ seguiría siendo no estacionario ya que

$$X_t = Z_t + Z_{t-1} + \dots + Z_1$$

por lo que $Var[X_t] = t\sigma^2 = Cov[X_t, X_{t+k}]$, $\rho_{t,t+k} = \frac{t}{\sqrt{t(t+k)}}$ y en consecuencia si t tendiera a infinito también lo harían la varianza y la covarianza. No así la FAS que tendería a uno. De este modo la covarianza y las correlaciones no sólo dependen de la distancia entre las variables sino también de su posición. Es posible convertir el proceso (3.16) en estacionario mediante el *operador diferencia*.

Definición 3.9. Sea $\{X_t\}$ un proceso estocástico. El **operador diferencia de orden 1**, denotado por ∇ , transforma el proceso original en un nuevo proceso Y_t resultado de la resta entre X_t y X_{t-1} , es decir

$$Y_t = \nabla X_t = X_t - X_{t-1}.$$

El **operador diferencia de orden d** denotado por ∇^d funciona de la siguiente forma:

- $\nabla^2 X_t = \nabla(\nabla X_t) = X_t - 2X_{t-1} + X_{t-2}$
- $\nabla^d = \nabla(\nabla^{d-1})$.

Existe una relación entre el operador de retraso B^k y el operador ∇^d . Observe que

$$\nabla X_t = X_t \smile X_{t-1} = B^0 X_t - B X_t = (I - B) X_t$$

para $t = 1, \dots, n$. Aplicando el operador diferencia nuevamente se tiene

$$\begin{aligned} \nabla^2 X_t &= \nabla(\nabla X_t) \\ &= \nabla(X_t \smile X_{t-1}) \\ &= (X_t - X_{t-1}) - (X_{t-1} \smile X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2} \\ &= B^0 X_t - 2B X_t + B^2 X_t \\ &= (I - 2B + B^2) X_t \\ &= (I - B)^2 X_t \end{aligned}$$

para $t = 1, \dots, n$. Por lo tanto aplicando d veces el operador ∇ se tiene la expresión general

$$\nabla^d X_t = (I - B)^d X_t \tag{3.17}$$

para $t = 1, \dots, n$.

Regresando al ejemplo, si se aplica el operador diferencia de orden uno al proceso (3.16) se obtiene el proceso

$$Y_t = \nabla X_t = Z_t. \tag{3.18}$$

Entonces el nuevo proceso obtenido mediante la primera diferencia del proceso original es un proceso de ruido

blanco y por lo tanto es estacionario. Es posible regresar al proceso $\{X_t\}$ a partir de $\{Y_t\}$ como sigue:

$$\begin{aligned} X_t &= Y_t + X_{t-1} \\ &= Y_t + Y_{t-1} + X_{t-2} \\ &= \vdots \\ &= Y_t + Y_{t-1} + Y_{t-2} + Y_{t-3} + \dots \end{aligned}$$

Ya que el proceso original $\{X_t\}$ se obtiene sumando el proceso $\{Y_t\}$ se dice que éste pertenece a la clase de *procesos integrados* conformada por todos los procesos que se pueden transformar en estacionarios mediante el operador diferencia.

Definición 3.10. Sea $\{X_t\}$ un proceso estocástico. $\{X_t\}$ es un *proceso autorregresivo integrado de media móvil de orden (p,d,q)* llamado también **ARIMA (p,d,q)** si

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = (1 - \theta_1 B - \dots - \theta_q B^q) Z_t$$

donde $Y_t = \nabla^d X_t$ y $\{Z_t\}$ es un proceso de ruido blanco.

Así un proceso $\{X_t\}$ es un *ARIMA (p,d,q)* si el proceso $Y_t = \nabla^d X_t$ es un proceso *ARMA (p,q)* estacionario e invertible. Como en los modelos anteriores la expresión anterior se escribe en forma compacta mediante

$$\phi_p(B)(1 - B)^d X_t = \theta_q(B) Z_t \quad (3.19)$$

donde $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ y $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$. Observe que el proceso (3.18) es un *ARIMA $(0,1,0)$* .

La función de autocorrelación simple de un proceso *ARIMA* decrece lentamente al aumentar el retraso k para t suficientemente grande. Por otro lado, aunque el operador diferencia podría transformar un proceso no estacionario en estacionario existen muchos otros procesos no estacionarios que requieren alguna otra transformación. Generalmente la varianza de un proceso no estacionario cambia junto con su nivel, es decir que su varianza es una función de su media que depende del tiempo. Teniendo una serie con tal comportamiento es necesario aplicar otras transformaciones, como las de Box-Cox, para estabilizar la varianza y luego aplicar el operador diferencia. El proceso obtenido al aplicar esas transformaciones con parámetro λ es

$$X_k^{(\lambda)} = \begin{cases} \frac{(X_t^\lambda - 1)}{\lambda} & \text{si } \lambda \neq 0 \\ \log X_t & \text{si } \lambda = 0 \end{cases}$$

para toda $t = 0, \pm 1, \dots$. Una de sus ventajas es que permiten obtener distribuciones aproximadas a la Normal.

3.6. Otros modelos

Como ya se había mencionado existen series temporales con comportamientos estacionales, es decir, comportamientos que se repiten después de un periodo regular de tiempo. El periodo de tiempo más pequeño en que se observan las repeticiones se llama periodo estacional y se denotará por s . Así, en un proceso estacionario puro hay dependencia lineal sólo entre las observaciones separadas por múltiplos de s . Los modelos correspondientes a éstos procesos son los llamados **modelos autorregresivos de media móvil estacionales puros** denotados por $\text{ARMA}(\mathbf{P}, \mathbf{Q})_s$. Si $\{X_t\}$ es un proceso $\text{ARMA}(P, Q)_s$ entonces

$$X_t - \Phi_1 X_{t-s} - \dots - \Phi_P X_{t-Ps} = Z_t - \Theta_1 Z_{t-s} - \dots - \Theta_Q Z_{t-Qs}$$

para toda $t = 0, \pm 1, \dots$ con $\{Z_t\}$ un proceso de ruido blanco, $\{\Phi_i\}_{i=1}^P$ y $\{\Theta_i\}_{i=1}^Q$ constantes reales. Las funciones de autocorrelación simple y parcial de procesos estacionarios puros tienen las mismas características en los múltiplos del periodo estacional que las de los procesos no estacionales. Ambas funciones son no nulas en cualquier retraso y decrecen de acuerdo a una mezcla de funciones exponenciales y senosoidales.

Además de las series puramente estacionales existen otras que además de presentar dependencia lineal entre observaciones separadas en múltiplos del periodo estacional s también presentan dependencia lineal con observaciones adyacentes. En estos casos convienen los *modelos multiplicativos* que son una combinación de modelos estacionales y no estacionales. $\{X_t\}$ es un **proceso estacional multiplicativo mixto autorregresivo de media móvil** denotado por $\text{ARMA}(\mathbf{p}, \mathbf{q}) \times \text{ARMA}(\mathbf{P}, \mathbf{Q})_s$ si

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps})X_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs})Z_t$$

para toda $t = 0, \pm 1, \dots$ con $\{Z_t\}$ un proceso de ruido blanco, $\{\phi_i\}_{i=1}^p$, $\{\Phi_i\}_{i=1}^P$, $\{\theta_i\}_{i=1}^q$ y $\{\Theta_i\}_{i=1}^Q$ constantes reales. En este modelo la dependencia regular asociada a los intervalos de medida de la serie se describe mediante $\phi_p(B)$ y $\theta_q(B)$ mientras que la dependencia estacional asociada a observaciones separadas s unidades de tiempo se describe mediante $\Phi_P(B^s)$ y $\Theta_Q(B^s)$. Las funciones de autocorrelación simple y parcial de estos procesos son una mezcla de sus componentes ordinarios y estacionales.

Observe que la estacionalidad en una serie puede provocar que ésta no sea estacionaria. Un ejemplo es la serie conformada por medidas de temperatura tomadas mensualmente durante diez años, en donde podría ser que las temperaturas de los mismos meses sean muy parecidas, es decir, las observaciones de los eneros sean casi

las mismas, y así para los otros meses. En este caso $X_t = \text{temperatura media mensual}$ y

$$X_t = S_t^{(s)} + N_t \quad (3.20)$$

donde N_t es un proceso estacionario y $S_t^{(s)}$ toma un valor distinto en cada periodo de la estación de forma que

$$S_t^{(s)} = S_{t-k}^{(s)} \quad (3.21)$$

para $k = \pm 1, \pm 2, \dots$. Note que $\{X_t\}$ no es estacionario en media. También podría ocurrir que la estacionalidad no fuera constante, es decir que

$$S_t^{(s)} = S_{t-ks}^{(s)} + V_t^{(s)} \quad (3.22)$$

con $k = \pm 1, \pm 2, \dots$ y $V_t^{(s)}$ un proceso estacionario. Para series con estos comportamientos son adecuados los modelos *SARIMA*. $\{X_t\}$ es un **proceso estacional multiplicativo autorregresivo integrado de medias móviles** llamado $\text{ARIMA}(\mathbf{p}, \mathbf{d}, \mathbf{q}) \times \text{ARIMA}(\mathbf{P}, \mathbf{D}, \mathbf{Q})_s$ o $\text{SARIMA}(\mathbf{p}, \mathbf{d}, \mathbf{q}) \times (\mathbf{P}, \mathbf{D}, \mathbf{Q})_s$ si el proceso diferenciado

$$Y_t = (1 - B)^d (1 - B^s)^D X_t$$

es un modelo estacional multiplicativo $\text{ARMA}(p, q) \times \text{ARMA}(P, Q)_s$ causal e invertible en la forma

$$\begin{aligned} (1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}) Y_t = \\ (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}) Z_t \end{aligned}$$

para toda $t = 0, \pm 1, \dots$ con $\{Z_t\}$ un proceso de ruido blanco, $\{\phi_i\}_{i=1}^p$, $\{\Phi_i\}_{i=1}^P$, $\{\theta_i\}_{i=1}^q$ y $\{\Theta_i\}_{i=1}^Q$ constantes reales.

Capítulo 4

Metodología de Box & Jenkins

Una metodología estadística para llevar a cabo el ajuste y pronóstico de un modelo *ARIMA* es la *metodología de Box & Jenkins*, la cual consiste en cuatro etapas en las que se busca un modelo parsimonioso, es decir, un modelo que tiene el menor número posible de parámetros. La primera etapa de la metodología es la *identificación* del modelo, es decir la identificación de la estructura no estacionaria, si es que existe, y luego la identificación de la estructura *ARMA* estacionaria, mediante la observación de las funciones de autocorrelación simple y parcial muestrales. La segunda etapa es la *estimación* de los parámetros del modelo elegido. Sin embargo, debido a la gran capacidad computacional que existe en la actualidad es posible estimar los parámetros de no sólo uno sino de varios modelos. La tercera etapa consiste en la *validación* del o los modelos, en la que se revisa que éstos cumplan los requisitos teóricos. La última etapa es la *predicción* de valores futuros de la serie.

4.1. Identificación

El primer paso de la identificación del modelo es el análisis de la estacionaridad. Como ya se mencionó en la sección anterior, si la serie de tiempo no es estacionaria es posible convertirla en estacionaria mediante una transformación.

4.1.1. Identificación de la estructura no estacionaria

El primer paso en el análisis de una serie de tiempo es dibujar su gráfica ya que en ella se puede observar si existe tendencia, estacionalidad, observaciones atípicas, varianza no constante o algún otro comportamiento que indique que la serie no es estacionaria. Si se observa alguna característica anterior, generalmente se aplica una transformación, pero si no se observa con claridad alguna conducta o si los puntos se alinean alrededor de una

línea paralela al eje x entonces no es necesario aplicar alguna transformación a los datos.

Las transformaciones más usadas son para estabilizar la varianza o la media. Una serie con varianza no constante usualmente requiere una transformación de Box Cox, en particular una transformación logarítmica (ver final de la Sección 3.5 del Capítulo 1). En las gráficas de estas series, la dispersión de los datos no se mantiene alrededor del nivel de cada momento y es común que la dispersión aumente cuando el nivel aumenta.

Otra gráfica que refleja el grado de variabilidad es la representación de la desviación típica contra la media de subconjuntos de los elementos de la serie. Es deseable que dichos subconjuntos sean lo más homogéneos posible, así que si la serie no es estacional se forman grupos de cinco o seis datos, mientras que para series estacionales se incluye un ciclo completo en cada grupo. En el caso de que se tuvieran m grupos, si las m parejas de medias - desviaciones típicas siguieran la relación $f(x) = kx^\alpha$, una transformación de Box Cox donde $\lambda = 1 - \alpha$ llevaría a una serie con variabilidad independiente al nivel. En este caso α se puede estimar mediante la recta de regresión $\log s_i = \log k + \alpha \log \bar{x}_i$ donde s_i es la desviación típica y \bar{x}_i es la media muestral del grupo i . Es posible que la serie no sea sensible al valor de λ , de forma que la gráfica de las parejas dispersión-media de la serie transformada sea muy parecida a la de la serie original. En este caso tampoco es necesario aplicar alguna transformación.

Por otro lado, para estabilizar una serie respecto a su media generalmente se aplican diferenciaciones ordinarias o estacionales. La gráfica de la serie y los correlogramas reflejan la existencia de la tendencia. Se requiere una diferenciación ordinaria cuando en la gráfica de la serie se observa una tendencia muy marcada. En ese caso los valores de la FAS muestral decrecerán lenta y linealmente, mientras que el primer valor de la FAP muestral será muy cercano a uno. Aplicando una diferencia se eliminarán tendencias de tipo lineal y aplicando una diferencia de orden 2 se eliminarán las de tipo cuadrático. Usualmente sólo se aplican una o dos diferenciaciones.

Si la serie es estacional es común que la FAS muestral tenga valores grandes que decrecen lentamente en los retrasos que son múltiplos del periodo estacional. En este caso es común que aplicando una sola diferenciación estacional desaparezca dicho comportamiento. También puede ocurrir que sean necesarias ambas diferenciaciones, la ordinaria y la estacional, donde no importa el orden en que se apliquen ya que se obtiene la misma serie. Por lo tanto es recomendable, cuando se tienen series estacionales, considerar las gráficas y correlogramas de las series ∇X_t , $\nabla_s X_t$ y $\nabla \nabla_s X_t$ para verificar la estacionaridad de cada una. Observe que al aplicar alguna diferenciación a la serie ésta reduce su tamaño. Así, si se aplicó una diferenciación ordinaria de orden d

y una diferenciación estacional de orden D con periodo s a una serie con n datos entonces la nueva serie tendrá $n - d - sD$ observaciones.

4.1.2. Identificación del modelo ARMA

Una vez que la serie en estudio es estacionaria, es decir que es factible que sea generada por un proceso estacionario, se puede realizar la identificación del modelo *ARMA*, la cual consiste en la selección del orden de las partes de medias móviles ordinaria y estacional y de las partes autorregresivas ordinaria y estacional. Así, se deben encontrar los órdenes q , Q , p y P del modelo basándose en los correlogramas. Conviene tomar en cuenta que no se pretende identificar de inmediato el mejor modelo, sino más bien reducir el conjunto de todos los posibles modelos *ARMA* que puedan representar las principales características de la serie.

Para identificar la parte ordinaria del modelo es necesario observar los primeros valores de las funciones de autocorrelación simple y parcial, y para la parte estacional observar los valores múltiples de la estacionalidad. Conviene buscar modelos simples que reflejen los rasgos más obvios de los correlogramas como los valores claramente significativos, es decir, los que caen fuera de los límites de referencia, o como el decrecimiento geométrico o sinusoidal. También es recomendable iniciar la identificación con modelos *AR* o *MA* de orden bajo y no con modelos mixtos. Finalmente se requiere que exista concordancia entre la parte ordinaria y la estacional. Esto puede verificarse a través de las interacciones alrededor de los retrasos estacionales, principalmente en la FAS. Es común que en la práctica las series puedan ajustarse a modelos *ARMA* con órdenes p y q menores o iguales a tres y P y Q menores o iguales a dos.

4.2. Estimación de los parámetros

Dada la serie de tiempo $\{X_t\}$ suponga que en la etapa de identificación se encontró que ésta es una muestra de el modelo *ARMA*(\hat{p}, \hat{q}) estacionario e invertible con \hat{p} y \hat{q} conocidos. Entonces la serie puede ser descrita mediante la fórmula

$$Z_t = \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} + (X_t - \mu) - \phi_1 (X_{t-1} - \mu) - \dots - \phi_p (X_{t-p} - \mu).$$

para $t = 0, \pm 1, \dots$. La estimación de los parámetros consiste en calcular los valores de $\{\phi_i\}_{i=1}^p$, $\{\theta_i\}_{i=1}^q$, μ y la varianza del proceso de ruido blanco σ^2 . Algunos métodos para calcular los parámetros son *el método de momentos*, *el método de máxima verosimilitud*, *estimación por mínimos cuadrados no condicionales*, *estimación por mínimos cuadrados condicionales* y *estimación no lineal*.

4.2.1. Método de momentos

Consiste en igualar los momentos teóricos con los momentos muestrales y resolver la ecuación o el sistema de ecuaciones obtenido. El estimador de la media del proceso que se obtiene por este método es la media muestral. En el caso de los modelos *MA* y *ARMA* los estimadores obtenidos así no se comportan tan bien como los de los modelos *AR*, sin embargo pueden utilizarse como valores iniciales en un procedimiento de estimación más elaborado. Respecto a los modelos *AR*(p) se tiene que

$$\hat{\phi}_p = \hat{\mathbf{R}}_p^{-1} \hat{\rho}_p \quad \text{y} \quad \hat{\sigma}^2 = \hat{\gamma}_0(1 - \hat{\rho}_p^t \hat{\mathbf{R}}_p^{-1} \hat{\rho}_p)$$

donde $\hat{\phi}_p = (\hat{\phi}_1, \dots, \hat{\phi}_p)^t$, $\hat{\sigma}^2$ son los estimadores de ϕ_p y de la varianza, $\hat{\mathbf{R}}_p = (\hat{\rho}_{i-j})_{i,j=1}^p$ y $\hat{\rho}_p = (\hat{\rho}_1, \dots, \hat{\rho}_p)^t$. Para una serie grande los estimadores anteriores se distribuyen aproximadamente Normal y $\hat{\sigma}^2$ está muy cercano al verdadero valor del parámetro (ver (Brockwell y Davis, 2013, pág. 241)).

4.2.2. Método de máxima verosimilitud

Sea $\{X_t\}$ un proceso *ARMA*(p, q) tal que el ruido blanco que lo conforma tiene variables aleatorias independientes, todas distribuidas *Normal*($0, \sigma^2$). El método de máxima verosimilitud consiste en maximizar la función de verosimilitud, es decir la función de densidad conjunta de $\mathbf{X} = (X_1, \dots, X_n)$ vista como función de los parámetros que en este caso son $(\phi_p, \theta_q, \sigma^2) = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)$. La función de verosimilitud de un proceso *ARMA*(p, q) es de la forma

$$L((\phi_p, \theta_q, \sigma^2) | \mathbf{X}) = g_1(\phi_p, \theta_q, \sigma^2) \exp \left[-\frac{S(\phi_p, \theta_q)}{2\sigma^2} \right] \quad (4.1)$$

donde g_1 es una función que sólo depende de los parámetros $(\phi_p, \theta_q, \sigma^2)$ y

$$S(\phi_p, \theta_q) = \sum_{t=1-p-q}^n E^2[U_t | \mathbf{X}, \phi_p, \theta_q, \sigma^2] \quad \text{con} \quad U_t = \begin{cases} Z_t & \text{si } t = 1, \dots, n \\ g_2(\mathbf{Z}_*, \mathbf{X}_*) & \text{si } t \leq 0 \end{cases}$$

donde g_2 es una función lineal de los valores iniciales no observables $\mathbf{Z}_* = (Z_{1-q}, \dots, Z_{-1}, Z_0)$ y $\mathbf{X}_* = (X_{1-p}, \dots, X_{-1}, X_0)$ que son necesarios para los cálculos de los valores Z_1, \dots, Z_n en la ecuación que define el modelo *ARMA*. Las funciones g_1 y g_2 dependen del modelo que se considere. La principal complicación al maximizar la función (4.1) es la presencia de la función g_1 y el cálculo de $E[\mathbf{U}_* | \mathbf{X}]$ donde $\mathbf{U}_* = (U_{1-p-q}, \dots, U_{-1}, U_0)$.

4.2.3. Estimación por mínimos cuadrados no condicionales

Ignorando la parte correspondiente a g_1 de la Función (4.1), observe que maximizar $\exp \left[-\frac{S(\phi_p, \theta_q)}{2\sigma^2} \right]$ es equivalente a minimizar $S(\phi_p, \theta_q)$. A este procedimiento se le conoce como estimación por mínimos cuadrados

no condicionales. Aún así este método presenta dificultades ya que se requiere calcular los valores esperados iniciales, además de que la expresión que se debe minimizar puede ser complicada.

4.2.4. Estimación por mínimos cuadrados condicionales

Si en lugar de minimizar la función $S(\phi_p, \theta_q)$ se minimiza la expresión

$$S_c(\phi_p, \theta_q) = \sum_{t=p+1}^n Z_t^2 \quad (4.2)$$

es más sencillo calcular los estimadores. Este método se llama estimación por mínimos cuadrados condicionales y la función (4.2) se llama *suma de cuadrados condicionales*, la cual se obtiene de la suma de cuadrados no condicionales donde los valores iniciales $Z_p, Z_{p-1}, \dots, Z_{p+1-q}$ en los cálculos de los correspondientes Z_t se igualan a cero.

4.2.5. Estimación no lineal

Si el modelo sobre el que se quiere hacer la estimación contiene términos de medias móviles entonces la suma de cuadrados condicional y también la no condicional son funciones no cuadráticas de los parámetros. Por este motivo son útiles los algoritmos de estimación no lineal, ya sea para minimizar $S(\phi_p, \theta_q)$ y $S_c(\phi_p, \theta_q)$ o para maximizar $L((\phi_p, \theta_q, \sigma^2))$. Uno de los algoritmos no lineales más usados es el *algoritmo de Gauss-Newton*.

4.3. Validación del modelo

La validación del modelo consta de varios pasos: el análisis de los residuales, la reformulación del modelo y la aplicación de un método de selección de modelos.

4.3.1. Análisis de los residuales

Ya que se estimaron los parámetros es necesario evaluar si se cumplen las hipótesis teóricas del modelo. La hipótesis primaria es que $\{Z_t\}$ sea un proceso de ruido blanco gaussiano. Suponga que se ajustó un modelo $ARMA(p, q)$ a los datos. Se llaman *residuales* a los valores obtenidos de la fórmula

$$\hat{Z}_t = X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_p X_{t-p} + \hat{\theta}_1 \hat{Z}_{t-1} + \dots + \hat{\theta}_q \hat{Z}_{t-q}$$

para $t = 1, \dots, n$ donde $\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q$ son las estimaciones de los parámetros. Para calcularlos existen dos métodos: el método condicional que toma los valores iniciales $Z_{1-q}, \dots, Z_{-1}, Z_0$ y $X_{1-p}, \dots, X_{-1}, X_0$ igualados a su valor esperado o el método exacto que calcula los residuales mediante el método *backcasting*.

4.3.2. Método backcasting o backforecasting

Suponga que la serie $\{X_n\} = \{X_1, \dots, X_n\}$ proviene del modelo

$$\phi_p(B)X_t = \theta_q(B)Z_t. \quad (4.3)$$

La estructura probabilística de la serie no solo puede ser explicada por el Modelo (4.3), sino también mediante el modelo

$$\phi_p(F)X_t = \theta_q(F)W_t \quad (4.4)$$

donde F es el *operador de adelanto* tal que $F^j X_t = X_{t+j}$, $j \geq 1$ y $\{W_t\}$ es un proceso de ruido blanco con la misma varianza que $\{Z_t\}$. De hecho la función de autocorrelación simple del Modelo (4.4) es la misma que la del Modelo (4.3). Se llama backcasting al cálculo de $E[Z_t|X_1, \dots, X_n]$ con $t = 1 - q, \dots, -1, 0$ y $E[X_t|X_1, \dots, X_n]$ con $t = 1 - p, \dots, -1, 0$ mediante el Modelo (4.4).

Regresando a los residuales, note que éstos son aproximaciones de los verdaderos valores de Z_t no observables del ruido blanco. Por lo tanto en la validación del modelo se verifica que los residuales tengan un comportamiento muy parecido al de un ruido blanco, es decir, que tengan media cero, varianza constante y que no exista correlación en ningún retraso. Además se revisa que su distribución sea aproximadamente Normal ya que en ese caso se tendría la independencia con lo que se garantizaría que toda la información de los datos está reflejada en ellos.

Usualmente lo primero que se revisa es si existe correlación entre los residuales. Su función de autocorrelación simple muestral se define como

$$\hat{\rho}_{\hat{Z}}(k) = \frac{\sum_{t=1}^{n-k} (\hat{Z}_t - \bar{Z})(\hat{Z}_{t+k} - \bar{Z})}{\sum_{t=1}^n (\hat{Z}_t - \bar{Z})^2}$$

donde n es el número de residuales disponibles y \bar{Z} es su media muestral. Si todos los valores $\hat{\rho}_{\hat{Z}}(k)$ se encuentran dentro de las bandas $\pm 1.96n^{-1/2}$ en la FAS y la FAP entonces se considera que no existe correlación. Pero si uno de cada 20 valores estimados de $\hat{\rho}_{\hat{Z}}(k)$ está fuera de las bandas o cercano a sus límites es indicativo de que sí existe correlación entre los residuales y por lo tanto de que el modelo es inadecuado.

Además de observar el comportamiento de cada autocorrelación estimada mediante el correlograma se puede aplicar un test *Portmanteau* el cual considera todos los valores de $\hat{\rho}_{\hat{Z}}(k)$ como un grupo. El *contraste de Box - Ljung - Pierce* es un contraste global para la hipótesis de que *los primeros h valores son cero*. El estadístico de

Box - Ljung - Pierce es

$$Q(h) = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_{\hat{Z}}^2(k)}{n-k}.$$

Bajo la hipótesis nula de que los residuales son un ruido blanco, éste se distribuye asintóticamente como una χ_{h-r}^2 , donde r es el número de parámetros estimados del modelo. Si los modelos no son estacionales entonces $r = p + q$. Si los modelos son estacionales entonces $r = p + q + P + Q$. Se concluye que la hipótesis se rechaza al nivel α si el valor $Q(h)$ es mayor al cuantil $1 - \alpha$ de la distribución χ_{h-r}^2 . El estadístico de Box -Ljung - Pierce no es el único test Portmanteau utilizado en el análisis de series de tiempo. Arranz (2005) hace una revisión de otros test de éste tipo. Además McLeod y Li (1983) encontraron que las autocorrelaciones del cuadrado de los residuos son útiles para detectar dependencia no lineal en los residuos de un modelo ARMA ajustado.

Para verificar que los residuales tienen media cero se calcula la media muestral

$$\bar{Z} = \sum_{t=1}^n \frac{\hat{Z}_t}{n}$$

donde n es el número de residuales y la varianza muestral

$$\hat{S}_{\hat{Z}}^2 = \sum_{t=1}^n \frac{(\hat{Z}_t - \bar{Z})^2}{(n-r)}.$$

Finalmente si

$$\frac{|\bar{Z}|}{\hat{S}_{\hat{Z}}/\sqrt{n}} \tag{4.5}$$

es significativamente grande respecto a la distribución $N(0, 1)$ se concluye que $E[Z_t]$ es diferente de cero. Esta prueba debe aplicarse después de haber comprobado que los residuales son incorrelados para asegurar que $\hat{S}_{\hat{Z}}^2$ es un buen estimador de la varianza.

La estabilidad de la varianza se puede comprobar estudiando la gráfica de los residuales respecto al tiempo y la normalidad de los residuales puede verificarse mediante un histograma o un $Q - Q$ plot o aplicando algún test como por ejemplo el de Shapiro - Wilk, el de Kolmogorov - Smirnov - Lillieford o el de Jarque - Bera. Note que el contraste de normalidad solo tiene sentido si ya se ha verificado que los residuales no son correlados.

4.3.3. Reformulación del modelo

Si se encuentra evidencia de que existe correlación en los residuales del modelo será necesario reajustarlo de la siguiente manera. Suponga que se identificó y estimó el siguiente modelo $ARMA(p, q)$:

$$\phi_p(B)X_t = \theta_q(B)\hat{Z}_t,$$

pero que al observar la FAS y FAP de los residuales \widehat{Z}_t se concluye que no son un ruido blanco y que más bien siguen el modelo $ARMA(r, s)$:

$$\psi_r(B)\widehat{Z}_t = \gamma_s(B)A_t$$

donde $\{A_t\}$ es un proceso de ruido blanco. Sustituyendo el segundo modelo en el primero, ya que los operadores son conmutativos, resulta el nuevo modelo

$$\psi_r(B)\phi_p(B)\widehat{Z}_t = \theta_q(B)\gamma_s(B)A_t.$$

Así que para la serie original se estimará un modelo $ARMA(p + r, q + s)$ de donde se obtendrán nuevos residuales en los que se repetirán las etapas de reformulación, estimación y verificación hasta que se obtenga un modelo adecuado.

Una manera de verificar que un modelo podría ser el adecuado es ajustar uno o varios modelos con órdenes mayores al elegido y revisar que los parámetros aumentados no sean significativos. Por ejemplo, suponga que el modelo $ARIMA(p, d, q)$ ha parecido el más adecuado y ya se ha ajustado. Ahora conviene estimar los modelos $ARIMA(p+1, d, q)$ y $ARIMA(p, d, q+1)$ y observar si los parámetros adicionales son significativos. Observe que conviene ampliar el orden de las partes AR y MA por separado para evitar que sus efectos se compensen.

Criterios de selección de modelos

Existen criterios de selección que se aplican para elegir de entre varios modelos el que mejor representa a los datos. Estos criterios se basan en resúmenes estadísticos de los residuales del modelo ajustado. Los criterios más usados son el llamado *Akaike Information Criterion* (AIC) y el *Bayesian Information Criterion* (BIC). El estadístico AIC es

$$AIC = \frac{-2 \ln L(\widehat{\beta}, \widehat{\sigma}^2) + 2r}{n}$$

y el estadístico BIC es

$$BIC = \frac{-2 \ln L(\widehat{\beta}, \widehat{\sigma}^2) + r \ln n}{n},$$

donde en ambos L es la función de verosimilitud, n es el tamaño de la muestra con el que se estimó el modelo, r es el número de parámetros estimados y $\widehat{\beta}$, $\widehat{\sigma}^2$ son los parámetros estimados. En estos criterios el primer término mide la precisión del modelo y el segundo penaliza el aumento de sus parámetros. Entonces se elige el modelo con el menor AIC o BIC según la elección del criterio, aunque generalmente los dos criterios conducirán al mismo modelo. Cuando se comparan los modelos es importante asegurarse de que n sea el mismo para todos ellos, por ejemplo no es posible comparar el BIC de un modelo $ARMA(p, q)$ con el de un modelo $ARIMA(p, 2, q)$.

Capítulo 5

Ejemplos de la metodología de Box & Jenkins

5.1. Introducción

El Nasdaq 100 es un índice bursátil de Estados Unidos que refleja la evolución de las 100 empresas con mayor capitalización bursátil que cotizan en el mercado electrónico del Nasdaq. Se incluyen dentro de este sector empresas de telecomunicaciones, de ventas al por mayor y al por menor, empresas de hardware y software y de biotecnología, siempre y cuando formen parte del Nasdaq Stock Market. Pueden formar parte de este índice empresas americanas e internacionales.

La principal característica del Nasdaq 100, a diferencia de otros índices bursátiles, es que no incluye valores financieros como las compañías de inversión. La ponderación o peso que tiene cada valor en este índice es realizada según su capitalización en el mercado. Todos los valores que forman parte del índice tienen una capitalización mínima de mercado de 500 millones de dólares. Entre las empresas que lo conforman se encuentran Microsoft Corporation, Cisco, Amazon, Apple Computer, Biomet, Dell, Ebay, Patterson Companies, Starbucks y Yahoo.

QQQ es el símbolo del índice Nasdaq 100, el cual se compra y vende como acciones llamadas ETF (Exchange Traded Funds) y es uno de los que tienen el mayor volumen de compra y venta. Como se puede comprar y vender igual que cualquier acción común y corriente, el inversionista que quiera ganar dinero con una subida o bajada de éste índice de la bolsa de valores, puede comprar o vender el ETF correspondiente. Esto ocurriría sin

que el inversionista tuviera que invertir en las acciones de alguna compañía específica. Cuando se hace daytrading, es decir comprar y vender el mismo día, el inversionista quiere asegurarse que una acción tiene suficiente volumen o liquidez durante el día para poder comprar y vender esa acción fácilmente. No hay muchos índices negociables que tengan bastante volumen, pero el índice Nasdaq 100 sí lo tiene, lo cual hace posible su compra y venta como si fueran acciones normales de alto volumen.

Por otro lado, Microsoft Corporation es una empresa multinacional de origen estadounidense, fundada el 4 de abril de 1975 por Bill Gates y Paul Allen. Dedicada al sector del software y el hardware, desarrolla, fabrica, licencia y produce software y equipos electrónicos, siendo sus productos más usados el sistema operativo Microsoft Windows y la suite Microsoft Office, los cuales tienen una importante posición entre las computadoras personales. Tiene 93 000 empleados en 102 países diferentes e ingresos de 72,360 millones de dólares durante el año 2012.

Microsoft afianzó su posición en otros mercados como el de sistemas operativos y suites de oficina con recursos como la red de televisión por cable MSNBC, el portal de Internet MSN y la enciclopedia multimedia Microsoft Encarta, producto y servicio cancelado por la empresa a principios de 2009 debido a la competencia de Wikipedia. La compañía también comercializa hardware como el ratón de Microsoft y productos de entretenimiento casero como Xbox, Xbox 360, Xbox One, Zune y MSN TV. Microsoft ha dado soporte a sus usuarios a través de Usenet en grupos de noticias y en Internet.

Cuando la compañía lanzó el 13 de marzo de 1986 su Oferta Pública de Venta, el precio de las acciones fue de 21 dólares por acción. Al cierre del primer día, las acciones valían 28 dólares. El cierre inicial y subsiguiente aumento en los años siguientes convirtió en millonarios a varios de sus empleados que poseían acciones de la empresa. El precio de las acciones llegó a su máximo en 1999, con un valor aproximado de 119 dólares por acción. La empresa sin embargo dejó de ofrecer dividendos hasta el año 2003. A partir de ese año, el valor de las acciones ha estado disminuyendo hasta situarse en 25.90 dólares (noviembre de 2010).

El índice Nasdaq 100 está fuertemente influenciado por los gigantes de la tecnología Intel, Cisco Systems, Microsoft Corporation, Google, Apple Inc. y Research in Motion. Ya que dichas compañías corresponden a un gran porcentaje del total de acciones que maneja el índice, cualquier suceso o actividad ocurrida con sus acciones ejerce gran influencia en el índice y por lo tanto en sus propias acciones. Debido a esto es común que los inversionistas deseen observar el comportamiento individual de las acciones de éstas grandes compañías, a

pesar de la tendencia general del índice o del mercado. Las subidas y bajadas de las grandes empresas ocurren a pesar de la tendencia del mercado debido a sus ciclos de producción, a sus ingresos o a los movimientos realizados por otros inversionistas.

5.2. Serie QQQ

Considere la serie compuesta por los precios de las acciones del Índice Nasdaq 100, a la que llamaremos serie QQQ. Esta serie contiene 2,520 datos tomados diariamente del 13 de diciembre del 2004 al 15 de diciembre del 2014.

5.2.1. Identificación de la estructura no estacionaria

La Figura 5.1 representa la variación del precio de las acciones respecto al tiempo. Es posible notar una tendencia creciente a partir del año 2009 y que no hay un comportamiento periódico marcado. Además la variación permanece constante respecto a la tendencia. Este comportamiento se corrobora en la gráfica de la dispersión contra la media.

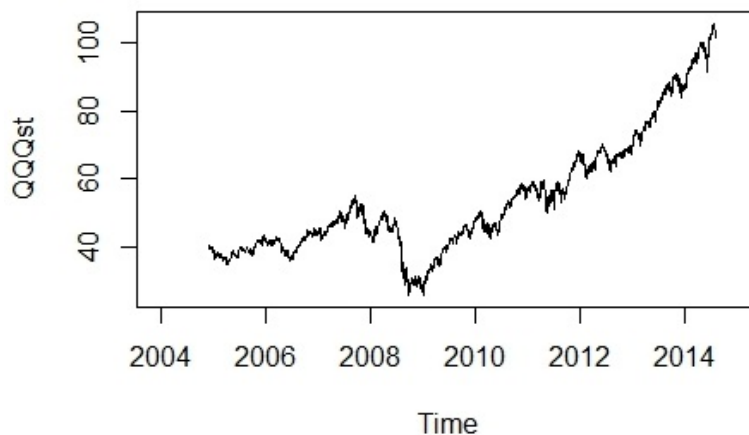


Figura 5.1: Gráfica de los precios de las acciones del Índice Nasdaq 100 tomados diariamente del 13/12/2004 al 15/12/2014

La Figura 5.2 es la gráfica de dispersión contra la media de la serie QQQ. Para dibujarla se dividió la serie en grupos con diez elementos cada uno y a cada grupo se le calculó su media y su desviación estándar. En la gráfica el eje x representa la media de cada grupo y el eje y su desviación. Es posible observar que no existe una dependencia lineal entre la desviación y la media, es decir la variación de la serie no aumenta cuando el nivel sí

lo hace. Por lo tanto se considera que no es necesario aplicar alguna transformación logarítmica. Para analizar la tendencia se observan las gráficas de la FAS y la FAP muestrales.

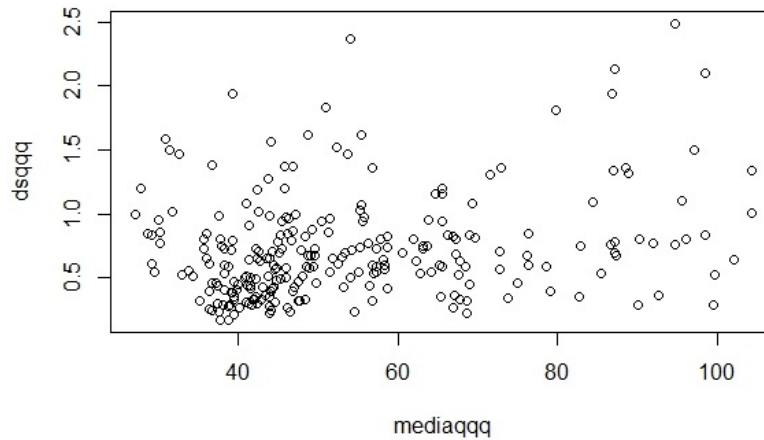


Figura 5.2: Gráfica de la dispersión contra la media de los precios de las acciones del Índice Nasdaq 100 tomados diariamente del 13/12/2004 al 15/12/2014

La Figura 5.3 corresponde al correlograma de la serie QQQ, la cual exhibe un decrecimiento lineal lento. La Figura 5.4 corresponde a la FAP muestral, en la que la autocorrelación parcial del primer retraso aparece fuera de las bandas de confianza. Dichos comportamientos son propios de una serie no estacionaria, por lo que es necesario aplicar una diferenciación ordinaria para volverla estacionaria.

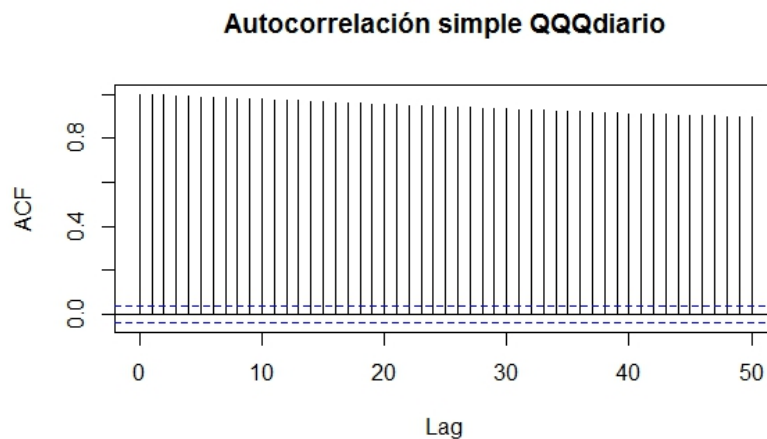


Figura 5.3: Gráfica de la FAS muestral de los precios de las acciones del Índice Nasdaq 100 tomados diariamente del 13/12/2004 al 15/12/2014

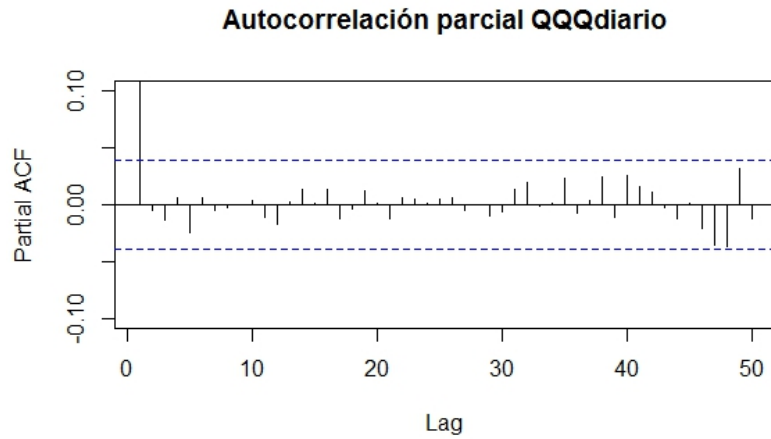


Figura 5.4: Gráfica de la FAP muestral de los precios de las acciones del Índice Nasdaq 100 tomados diariamente del 13/12/2004 al 15/12/2014

La Figura 5.5 muestra la gráfica de la primera diferencia ordinaria de la serie QQQ respecto al tiempo. En consecuencia ésta serie tiene 2,519 datos. En ella desapareció la tendencia creciente observada a partir del año 2009 en la Figura 5.1. En las Figuras 5.6 y 5.7 se observan la FAS y la FAP muestrales de la serie diferenciada, a la que a partir de ahora se llamará $Y_t = \nabla X_t$ donde X_t es la serie original. Estas gráficas exhiben un decrecimiento rápido y sus primeras autocorrelaciones son no significativas por lo que parecen corresponder a una estructura estacionaria. Por lo tanto ya se puede suponer que Y_t procede de un proceso estacionario. El siguiente paso es seleccionar el orden de las partes autorregresivas y de medias móviles.

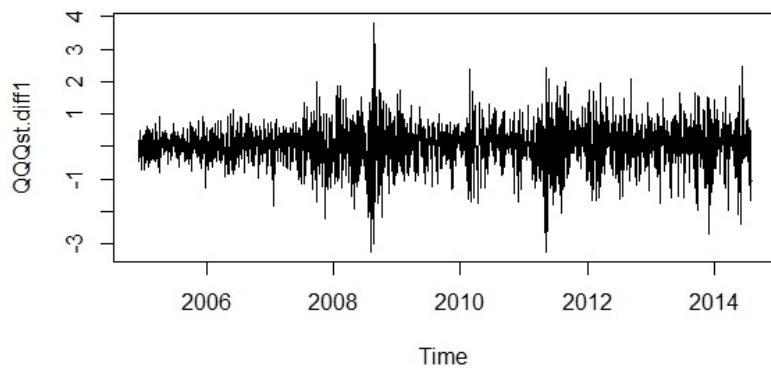


Figura 5.5: Gráfica de la primera diferencia de los precios de las acciones del Índice Nasdaq 100

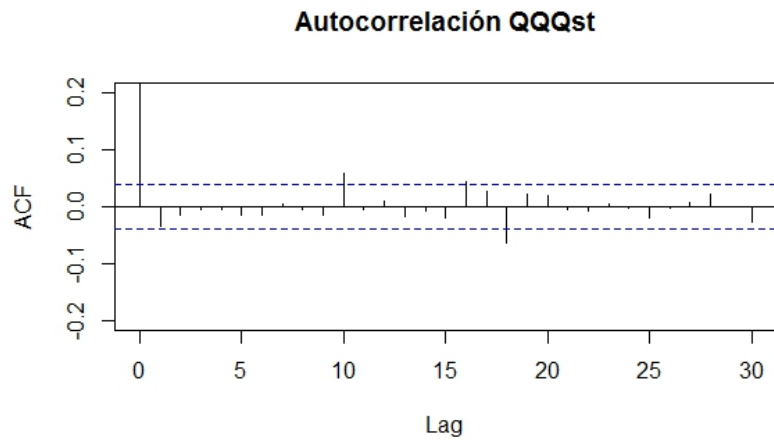


Figura 5.6: Gráfica de la FAS muestral de la primera diferencia de los precios de las acciones del Índice Nasdaq 100

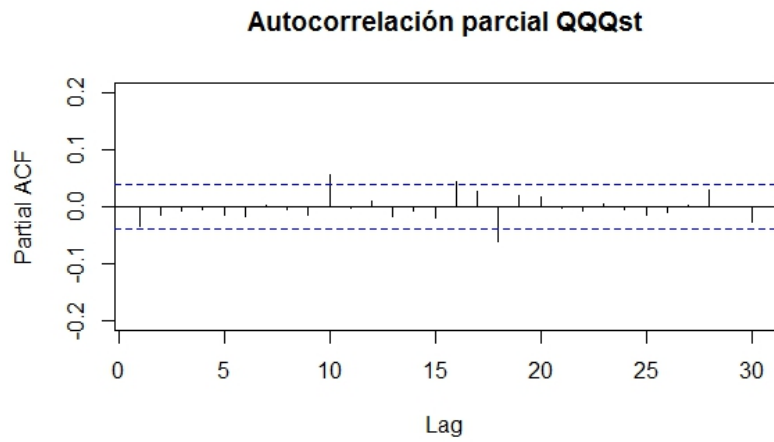


Figura 5.7: Gráfica de la FAP muestral de la primera diferencia de los precios de las acciones del Índice Nasdaq 100

5.2.2. Identificación del modelo ARMA

Tanto en la FAS como en la FAP muestrales de las Figuras 5.6 y 5.7 correspondientes a la serie Y_t , de los primeros cinco retrasos sólo el primero está en el límite de las bandas de confianza, lo cuál es indicativo de que los órdenes de las partes autorregresivas y de medias móviles podrían ser uno. Esto se corrobora aplicando la función `auto.arima` del paquete `forecast` del software estadístico R (Hyndman y Khandakar (2008)), la cual indica que el modelo más apropiado es el $ARMA(0, 1)$ e incluso indica que el nivel de la serie no es cero, por lo que el modelo tiene un intercepto distinto de cero. La función `auto.arima` proporciona estimaciones máximo verosímiles exactas utilizando el filtro de Kalman para calcular la función de verosimilitud exacta por medio de la representación del modelo en el espacio de estados (Maindonald y Braun, 2010, pág. 288). La Tabla 5.1 contiene las estimaciones para el modelo $ARMA(0, 1)$ en la serie Y_t . Las columnas ES MA1 y ES IN se refieren a los errores estándar de las estimaciones del parámetro MA1 y del intercepto IN respectivamente.

Modelo estimado	ES MA1	ES IN	$\hat{\sigma}^2$
$Y_t = (1 + 0.0344B)Z_t + 0.0243$	0.0202	0.0122	0.4008

Tabla 5.1: Modelo estimado $ARMA(0,1)$ para la serie $Y_t = \nabla X_t$ donde $X_t = QQQ$. La primera columna contiene el modelo estimado, la segunda y tercera columnas contienen los errores estándar de la estimación del parámetro MA1 y del intercepto respectivamente. La cuarta columna contiene la estimación de la varianza del modelo.

Se aplicó la prueba de Wald para verificar que los dos parámetros son significativos. En el caso de MA1 se tiene que $(MA1/ESMA1)^2 = 2.9001$ y que el valor-p es 0.0886. Dado que la hipótesis nula es que el parámetro es igual a cero, si $\alpha = 0.1$ entonces $p < \alpha$. Por lo tanto existe evidencia de que el parámetro MA1 es distinto de cero. En el caso del intercepto se tiene que $(IN/ESIN)^2 = 3.9673$ y que el valor-p es 0.0464. Ya que $p < \alpha$ se concluye que el intercepto es significativo. Con $\alpha = 0.1$ se tiene un riesgo del 10% de concluir erróneamente que los parámetros son significativos cuando en realidad no lo son.

5.2.3. Validación del modelo

A continuación se muestran las FAS y FAP de los residuales del modelo. Las gráficas se realizaron con las funciones `acf` y `pacf` del software estadístico R mediante las Fórmulas (2.4) y (2.5), ver Venables y Ripley (2002).

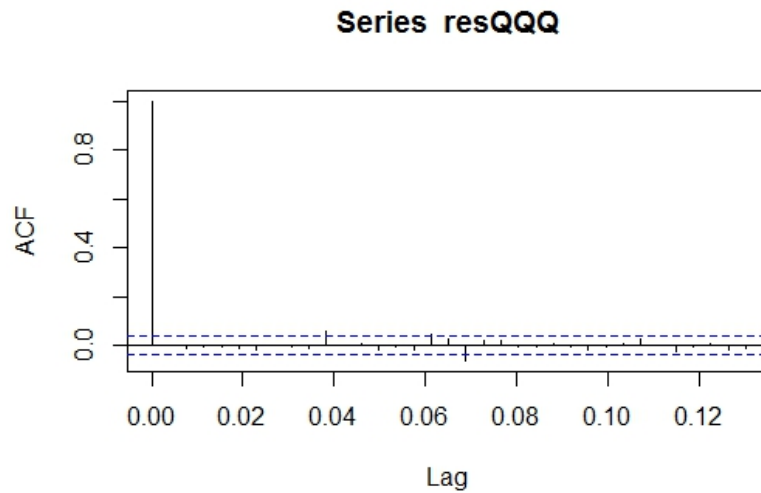


Figura 5.8: Gráfica de la FAS muestral de los residuales del modelo ARMA(0,1) para Y_t

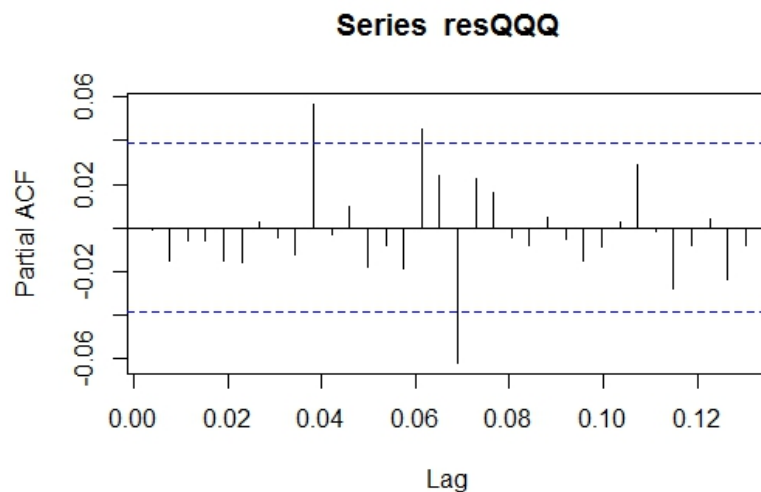


Figura 5.9: Gráfica de la FAP muestral de los residuales del modelo ARMA(0,1)

En las Figuras 5.8 y 5.9 no se observa evidencia clara de que exista dependencia (serial) entre los residuales. El hecho de que algunos retrasos tengan valores fuera de las bandas (10, 20 en la FAS y 10, 16 y 18 en la FAP) se debe a la aleatoriedad de las estimaciones. Aplicando el test de Box - Ljung - Pierce se tiene que $Q(30) = 36.4363$ donde Q es la estadística de la prueba distribuida χ^2 con 29 grados de libertad en este caso. El valor-p del contraste es 0.1612 y la hipótesis nula es que los residuales son (serialmente) independientes, es decir que sus correlaciones son cero y si existe alguna correlación cuyo valor es distinto de cero es debida a

la aleatoriedad. Con $\alpha = 0.05$ se tiene que $p > \alpha$ y en consecuencia se concluye que no existe evidencia de dependencia entre los residuales.

Para verificar que los residuales tienen media cero se calculó el valor de la estadística (4.5) donde σ se reemplazó por S ya que sólo se conoce la desviación muestral. El valor obtenido fue 0.0011. Sean $H_0 : \mu = 0$ donde μ es la media de los residuales, $H_1 : \mu \neq 0$ y $\alpha = 0.05$. El valor crítico en la tabla de la distribución $N(1, 0)$ es ± 1.96 . Ya que el valor de la estadística (4.5) no está dentro del intervalo $(-1.96, 1.96)$, no se rechaza la hipótesis nula por lo que no existe evidencia de que la media de los datos es distinta de cero.

La Figura 5.10 es la gráfica de los residuales respecto al tiempo. En ella se percibe que la variabilidad es aproximadamente constante excepto por los saltos que aparecen a finales del 2008 y a mediados del 2011. El primero es consecuencia de la baja que sufrió el precio de las acciones en dicho año y que también se observa en la Figura 5.1.

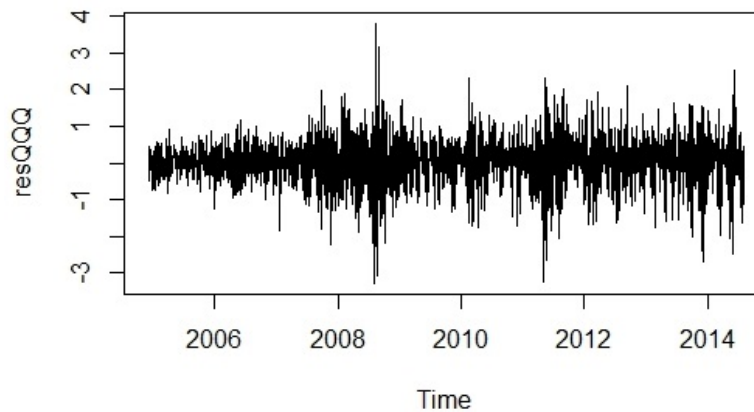


Figura 5.10: Gráfica de los residuales del modelo ARMA(0,1)

Para verificar la normalidad de los residuales se aplicó el test de Jarque Bera obteniendo $JB = 794.5296$. Como el valor-p del contraste resulta menor a $2.2e-16$ y la hipótesis nula es que los residuales son una muestra de una población que se distribuye Normal, con $\alpha = 0.05$ se tiene que $p < \alpha$ por lo que hay evidencia de la falsedad de la hipótesis nula y por lo tanto de que los residuales no tienen una distribución normal estándar. La Figura 5.11 muestra el histograma de los residuales el cual tiene forma de campana, aunque el sesgo vale -0.2934 . Por ser ésta medida negativa la cola izquierda es ligeramente más larga que la derecha. Además su curtosis vale 5.6880, la cual indica que la cima de la distribución es más afilada que la de la distribución normal estándar.

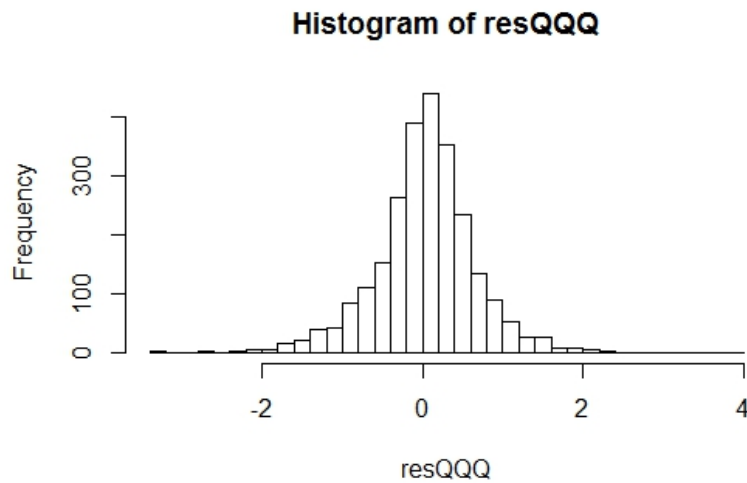


Figura 5.11: Histograma de los residuales del modelo ARMA(0,1)

En resumen se puede considerar que los residuales del modelo ARMA(0,1) son incorrelados (serialmente independientes), tienen media cero y tienen varianza constante. Entonces se cumple la hipótesis de que Z_t proviene de un proceso de ruido blanco y por lo tanto el modelo elegido es adecuado. Por último, una vez que ya se tiene un modelo que describa una serie de tiempo dada es posible hacer predicciones. Sin embargo ese tema no se tratará en esta tesis.

5.3. Serie MSFT

Se llamará MSFT a la serie compuesta por 2,520 precios de las acciones de Microsoft Corporation tomados diariamente del 13 de diciembre del 2004 al 15 de diciembre del 2014.

5.3.1. Identificación de la estructura no estacionaria

La Figura 5.12 representa la variación del precio de las acciones respecto al tiempo. Es evidente la tendencia creciente a partir del año 2013 pero no parece existir algún comportamiento periódico. Su variación no aumenta con la tendencia. Esto se verifica en la gráfica de dispersión de la desviación estándar contra la media en la Figura 5.13, la cual se obtuvo de la misma forma que la Figura 5.2 en el Ejemplo 5.2. No se aprecia la existencia de dependencia lineal entre la media y la desviación estándar por lo que no se requiere aplicar alguna

transformación logarítmica a la serie MSFT. El análisis de la tendencia se realizará mediante las FAP y FAS muestrales en las Figuras 5.14 y 5.15 respectivamente.

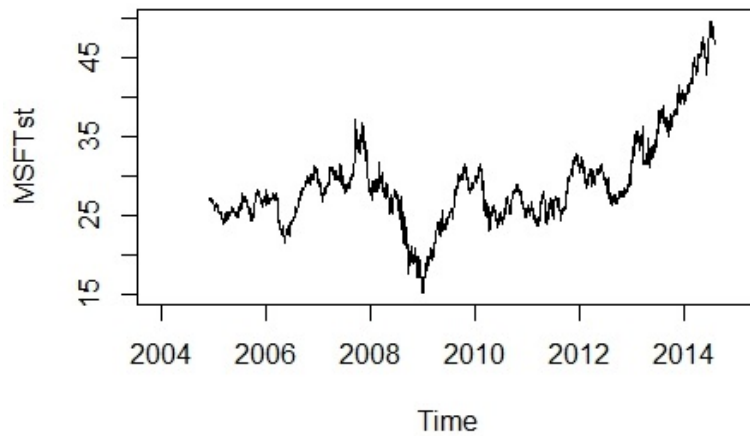


Figura 5.12: Gráfica de los precios de las acciones de Microsoft Corporation tomados diariamente del 13/12/2004 al 15/12/2014

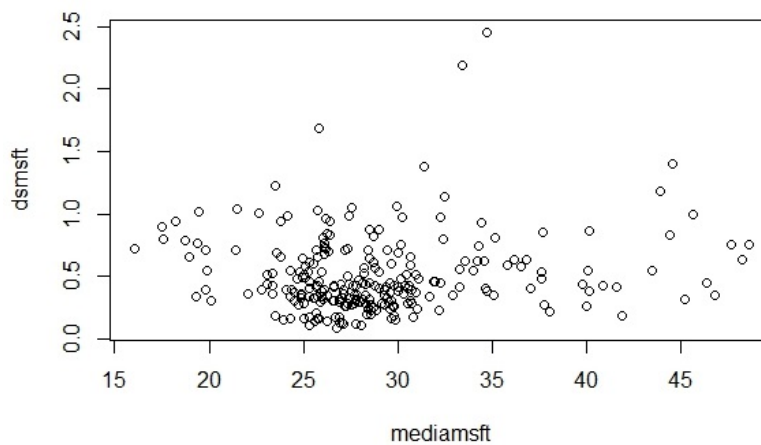


Figura 5.13: Gráfica de dispersión contra la media de los precios de las acciones de Microsoft Corporation tomados diariamente del 13/12/2004 al 15/12/2014

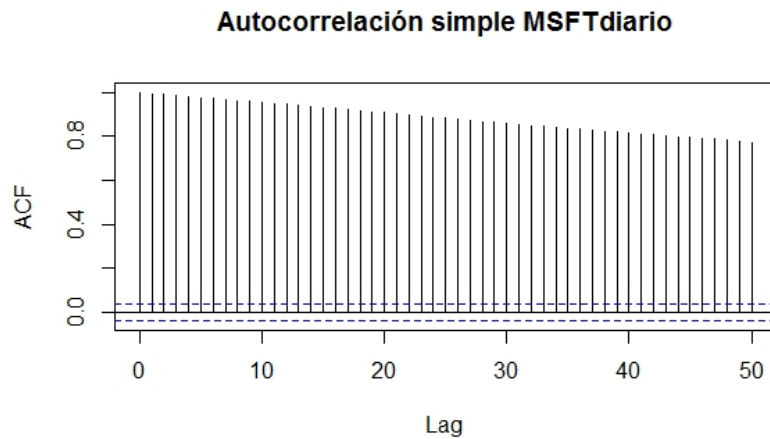


Figura 5.14: Gráfica de la FAS muestral de los precios de las acciones de Microsoft Corporation tomados diariamente del 13/12/2004 al 15/12/2014

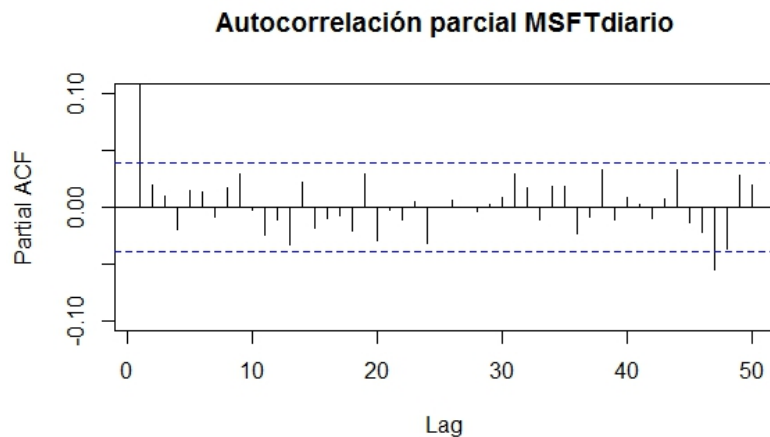


Figura 5.15: Gráfica de la FAP muestral de los precios de las acciones de Microsoft Corporation tomados diariamente del 13/12/2004 al 15/12/2014

La Figura 5.14 exhibe un decrecimiento lineal lento y en la Figura 5.15 se aprecia que la autocorrelación parcial del primer retraso es significativa, de donde se concluye que la serie MSFT no es estacionaria y que es necesario aplicarle una diferenciación ordinaria. La Figura 5.16 muestra la gráfica de la primera diferencia ordinaria de la serie MSFT respecto al tiempo, en la que ya no se percibe la tendencia creciente observada a partir del año 2013 en la Figura 5.12. En las Figuras 5.17 y 5.18 se ubican las FAS y FAP muestrales de la serie diferenciada. En ambas gráficas la mayoría de las autocorrelaciones simples y parciales de los retrasos están dentro de las bandas de confianza por lo que parecen corresponder a una estructura estacionaria.

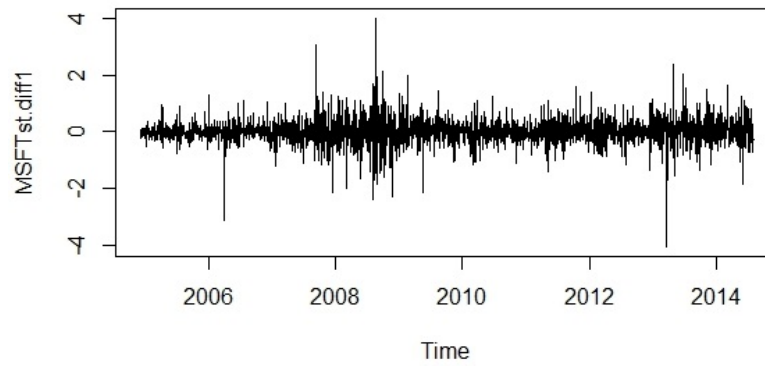


Figura 5.16: Gráfica de la primera diferencia de los precios de las acciones de Microsoft Corporation

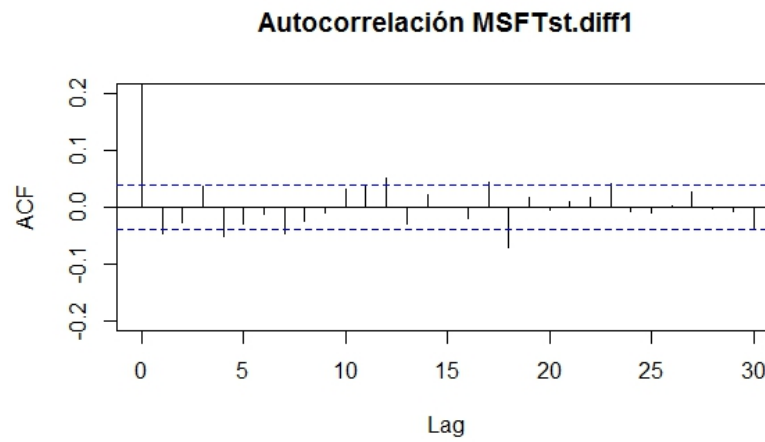


Figura 5.17: Gráfica de la FAS muestral de la primera diferencia de los precios de las acciones de Microsoft Corporation

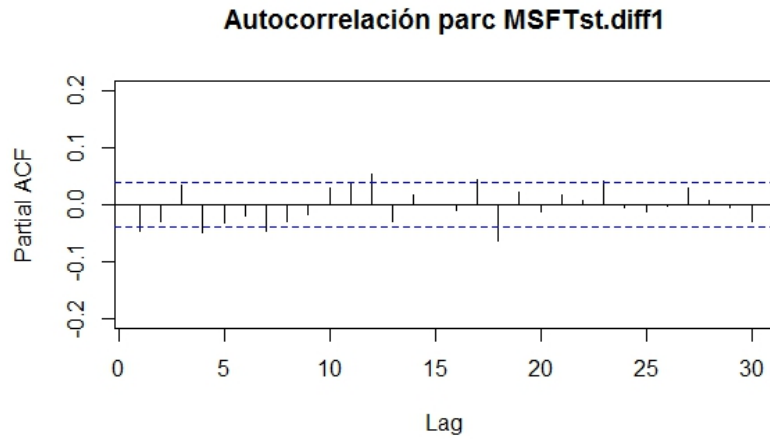


Figura 5.18: Gráfica de la FAP muestral de la primera diferencia de los precios de las acciones de Microsoft Corporation

5.3.2. Identificación del modelo ARMA

El hecho de que el primer retraso en la FAS y en la FAP muestrales sea significativo es indicativo de que un modelo ARMA(1,1) podría ser adecuado. Por otro lado la función auto.arima indica que el modelo ARMA(2,2) ajusta bien a la serie diferenciada. La Tabla 5.2 contiene las estimaciones de los parámetros de los modelos ARMA(1,1) y ARMA(2,2) en la serie $Y_t = \nabla X_t$ donde X_t es la serie MSFT.

Modelo	Modelo estimado	$\hat{\sigma}^2$
ARMA(1,1)	$(1 + 0.78326B)Y_t = (1 - 0.7424B)Z_t$ (0.0928) (0.1002)	0.2083
ARMA(2,2)	$(1 + 1.0421B + 0.7646B^2)Y_t = (1 - 1.0086B - 0.7147B^2)Z_t$ (0.1720) (0.1909) (0.1877) (0.2068)	0.2078

Tabla 5.2: Modelos estimados ARMA(1,1) y ARMA(2,2) para la serie $Y_t = \nabla X_t$ donde $X_t = MSFT$. La segunda columna contiene los modelos estimados donde las cantidades entre paréntesis que aparecen debajo de las estimaciones de los parámetros son sus errores estándar y la tercera columna contiene la estimación de la varianza de cada modelo.

Para ver si algún parámetro del modelo ARMA(2,2) es no significativo se utilizó la prueba de Wald. La hipótesis nula de cada prueba es que el valor del parámetro es cero contra la hipótesis alternativa de que es distinto de cero. En este caso todos los valores p son muy pequeños e incluso menores por mucho a $\alpha = 0.001$, por lo que existe evidencia muy fuerte de que todos los parámetros son significativos. Esto conduce a rechazar el modelo ARMA(1,1) ya que en él no aparecen (o son nulos) los términos de segundo orden del modelo ARMA(2,2).

5.3.3. Validación del modelo

A continuación se presenta el análisis de los residuales del modelo ARMA(2,2). En las Figuras 5.19 y 5.20 se encuentran las gráficas de la FAS y la FAP muestrales. En ninguna se aprecia evidencia clara de que exista dependencia (serial) entre los residuales. El hecho de que aparezcan algunos retrasos con valores relativamente altos, es decir sobre o fuera de las bandas, se debe a la aleatoriedad de las observaciones.

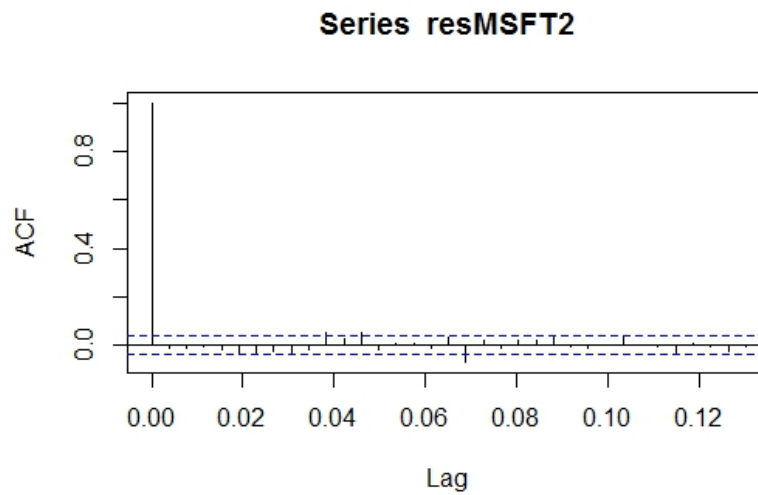


Figura 5.19: Gráfica de la FAS muestral de los residuales del modelo ARMA(2,2) para Y_t

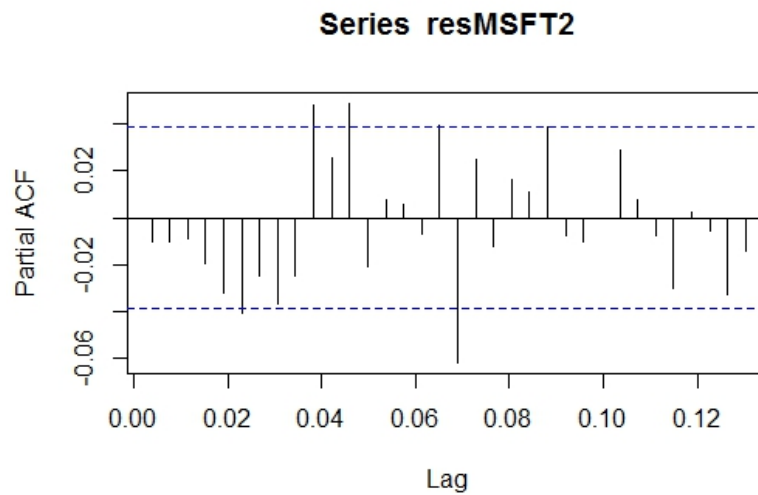


Figura 5.20: Gráfica de la FAP muestral de los residuales del modelo ARMA(2,2) para Y_t

Aplicando el test de Box - Ljung - Pierce se tiene que $Q(30) = 60.2798$ donde Q se distribuye χ^2 con 26 grados de libertad en este caso. El valor-p del contraste es 0.1830 y la hipótesis nula es que los residuales son (serialmente) independientes. Con $\alpha = 0.05$ se tiene que $p > \alpha$ y en consecuencia se concluye que no existe evidencia de dependencia entre los residuales.

Para verificar que los residuales del modelo ARMA(2,2) tienen media cero también se calculó el valor de la estadística (4.5) pero con S obteniendo 0.8776. Sean $H_0 : \mu = 0$ donde μ es la media de los residuales, $H_1 : \mu \neq 0$ y $\alpha = 0.05$. El valor crítico en la tabla de la distribución $N(0, 1)$ en este caso es ± 1.96 . Ya que el valor de la estadística (4.5) no está dentro del intervalo $(-1.96, 1.96)$, no se rechaza la hipótesis nula por lo que no existe evidencia de que la media de los datos es distinta de cero.

La variabilidad de los residuales se analizó mediante su gráfica respecto al tiempo. En la Figura 5.21 se aprecia que la variabilidad es aproximadamente constante excepto por los cuatro saltos que aparecen a principios del 2006, a finales del 2007, a finales del 2008 y a principios del 2013. Estos saltos son consecuencia de las altas y bajas que sufrió el precio de las acciones en esos años y que también se observan en la Figura 5.12.

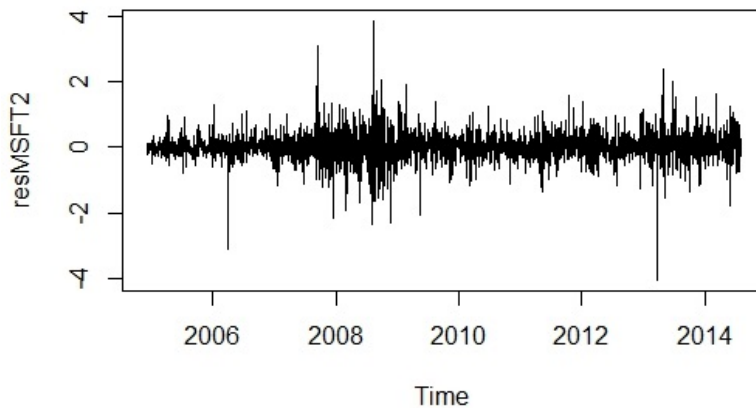


Figura 5.21: Gráfica de los residuales del modelo ARMA(2,2)

Para verificar que los residuales se distribuyen como una normal estándar se aplicó el test de Jarque Bera obteniendo $JB = 7153.651$. Como el valor-p del contraste resulta menor a $2.2e-16$, con $\alpha = 0.05$ se tiene que $p < \alpha$ por lo que hay evidencia de la falsedad de la hipótesis nula y por lo tanto de que los residuales no tienen distribución normal estándar. La Figura 5.22 es el histograma de los residuales. El sesgo es -0.1138 , es decir la distribución de los datos no es simétrica, y tiene la cola izquierda ligeramente más larga que la derecha. La curtosis es 11.2509, por lo que la cima de la distribución es más afilada que la de la distribución normal estándar.

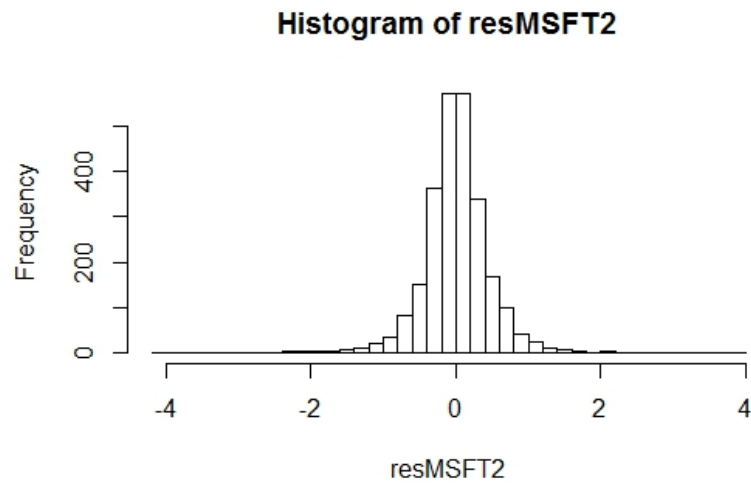


Figura 5.22: Histograma de los residuales del modelo ARMA(2,2)

De lo anterior se deduce que los residuales del modelo ARMA(2,2) provienen de un proceso de ruido blanco y por lo tanto el modelo elegido es adecuado. Por último, una vez que ya se tiene un modelo que describa una serie de tiempo dada es posible hacer predicciones. Sin embargo ese tema no se tratará en esta tesis.

Capítulo 6

Series de tiempo multivariadas

6.1. Introducción

Una **serie de tiempo multivariada** consiste en dos o más series univariadas concatenadas de tal manera que cada elemento de la serie es un vector con dos o más entradas. Es decir, $\mathbf{X}_t = (X_{1t}, X_{2t}, \dots, X_{kt})^T$ es una serie multivariada donde $X_{1t}, X_{2t}, \dots, X_{kt}$ son las k series univariadas que la conforman y a las que se les llama *componentes*. En las series multivariadas no solo hay dependencia serial dentro de cada componente sino también interdependencia entre sus distintos componentes, es decir entre X_{it} y X_{jt} , $i \neq j$. Un ejemplo de una serie multivariada de tres componentes es aquella conformada por los precios de las acciones de las compañías IBM, Microsoft Corporation y Wal Mart. En este caso X_{1t} es el precio de las acciones de IBM en el tiempo t , X_{2t} es el precio de las acciones de Microsoft Corporation y X_{3t} es el precio de las acciones de Wal Mart.

Muchos de los modelos discutidos en el Capítulo 4 se pueden generalizar directamente al caso multivariado. Sin embargo, hay situaciones en las que la generalización requiere modelos y métodos nuevos para manejar las relaciones entre los componentes de una serie multivariada.

6.1.1. Series de tiempo multivariadas estacionarias

Sea $\mathbf{X}_t = (X_{1t}, X_{2t}, \dots, X_{kt})^T$ una serie de tiempo k -dimensional.

- El vector de medias de \mathbf{X}_t es

$$\boldsymbol{\mu} = E(\mathbf{X}_t)$$

donde la esperanza se toma elemento por elemento en la distribución conjunta de \mathbf{X}_t . Así la media $\boldsymbol{\mu}$ es

un vector de dimensión k donde sus entradas son las esperanzas no condicionales de los componentes de \mathbf{X}_t . Si se quiere hacer referencia a sus elementos entonces se escribirá $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)^T$.

- La matriz de covarianzas de \mathbf{X}_t es

$$\boldsymbol{\Gamma}_0 = E[(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})^T].$$

donde $\boldsymbol{\Gamma}_0$ es de dimensión $k \times k$, el i -ésimo elemento de la diagonal es la varianza de X_{it} y el (i, j) -ésimo elemento es la covarianza entre X_{it} y X_{jt} . Se escribe $\boldsymbol{\Gamma}_0 = [\Gamma_{ij}(0)]$ cuando se hace referencia a sus elementos.

Definición 6.1. Sea \mathbf{X}_t una serie de tiempo k -dimensional. \mathbf{X}_t es **estacionaria de primer orden o estacionaria en media** si el vector de medias $\boldsymbol{\mu}$ es constante respecto al tiempo.

Definición 6.2. Sea \mathbf{X}_t una serie de tiempo k -dimensional. \mathbf{X}_t es **estacionaria de segundo orden o débilmente estacionaria** si

i) es estacionaria de primer orden y

ii) la matriz de covarianzas $\boldsymbol{\Gamma}_0$ es constante respecto al tiempo.

6.2. Matrices de correlaciones y covarianzas cruzadas

Definición 6.3. Sean $\mathbf{X}_t = (X_{1t}, X_{2t}, \dots, X_{kt})$ una serie de tiempo k -dimensional y

$\mathbf{D} = \text{diag}\{\sqrt{\Gamma_{11}(0)}, \dots, \sqrt{\Gamma_{kk}(0)}\}$ una matriz diagonal de dimensión $k \times k$ que consiste en las desviaciones estándar de X_{it} con $i = 1, \dots, k$. La **matriz de correlaciones cruzadas** de \mathbf{X}_t se define como

$$\boldsymbol{\rho} \equiv [\rho_{ij}(0)] = \mathbf{D}^{-1}\boldsymbol{\Gamma}_0\mathbf{D}^{-1}. \quad (6.1)$$

De la Ecuación (6.1) se deduce que

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\text{Cov}(x_{it}, x_{jt})}{\sqrt{\text{Var}(x_{it})}\sqrt{\text{Var}(x_{jt})}},$$

es decir, la entrada (i, j) de la matriz $\boldsymbol{\rho}$ es el coeficiente de correlación entre X_{it} y X_{jt} . Se puede verificar que $\rho_{ij}(0) = \rho_{ji}(0)$, $-1 \leq \rho_{ij}(0) \leq 1$ y $\rho_{ii}(0) = 1$ con $i, j = 1, \dots, k$. Por lo tanto $\boldsymbol{\rho}$ es una matriz simétrica con 1's en la diagonal.

Las matrices de correlaciones cruzadas miden qué tan fuerte es la dependencia lineal entre los componentes de una serie multivariada. La **matriz de covarianzas cruzadas en el retraso ℓ** de \mathbf{X}_t se define como

$$\boldsymbol{\Gamma}_\ell \equiv [\Gamma_{ij}(\ell)] = E[(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_{t-\ell} - \boldsymbol{\mu})^T].$$

Por lo tanto, el (i, j) -ésimo elemento de Γ_ℓ es la covarianza entre X_{it} y $X_{j,t-\ell}$. Para series débilmente estacionarias, la matriz de covarianzas cruzadas Γ_ℓ es una función de ℓ y no del tiempo t .

Definición 6.4. Sea X_t una serie de tiempo k -dimensional. X_t es un **ruido blanco** si es una serie estacionaria de segundo orden cuyo vector de medias es el vector $\mathbf{0}$ y la matriz de covarianzas es

$$\Gamma_\ell = \begin{cases} \Gamma_{ij}(\ell) & \text{si } \ell \neq 0 \\ \mathbf{0} & \text{e.o.c.} \end{cases}$$

La **matriz de correlaciones cruzadas en el retraso** ℓ de X_t se define como

$$\rho_\ell \equiv [\rho_{ij}(\ell)] = \mathbf{D}^{-1} \Gamma_\ell \mathbf{D}^{-1}. \quad (6.2)$$

De la Ecuación (6.2) se deduce que

$$\rho_{ij}(\ell) = \frac{\Gamma_{ij}(\ell)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\text{Cov}(x_{it}, x_{j,t-\ell})}{\sqrt{\text{Var}(x_{it})\sqrt{\text{Var}(x_{jt})}}}, \quad (6.3)$$

es decir, la entrada (i, j) de la matriz ρ_ℓ es el coeficiente de correlación entre X_{it} y $X_{j(t-\ell)}$. Si $\ell > 0$, este coeficiente mide la dependencia lineal de X_{it} sobre $X_{j(t-\ell)}$, la cual ocurre antes del tiempo t . Entonces si $\rho_{ij}(\ell) \neq 0$ con $\ell > 0$ se dice que *la serie X_{jt} guía a la serie X_{it} en el retraso ℓ* . De la misma forma, $\rho_{ji}(\ell)$ mide la dependencia lineal entre X_{jt} y $X_{i(t-\ell)}$ y se dice que *la serie X_{it} guía a la serie X_{jt} en el retraso ℓ* si $\rho_{ji}(\ell) \neq 0$ con $\ell > 0$. Además de la Ecuación (6.3) se deduce que los elementos de la diagonal $\rho_{ii}(\ell)$ son los coeficientes de autocorrelación de X_{it} en el retraso ℓ .

De lo anterior se obtienen algunas propiedades importantes de las correlaciones cruzadas cuando $\ell > 0$. La primera es que, en general, $\rho_{ij}(\ell) \neq \rho_{ji}(\ell)$ para $i \neq j$ ya que los coeficientes de correlación miden diferentes relaciones lineales entre X_{it} y X_{jt} . Por lo tanto, generalmente Γ_ℓ y ρ_ℓ son no simétricas. La segunda es que, de acuerdo con la propiedad $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ y suponiendo estacionaridad débil,

$$\begin{aligned} \text{Cov}(X_{it}, X_{j(t-\ell)}) &= \text{Cov}(X_{j(t-\ell)}, X_{it}) \\ &= \text{Cov}(X_{jt}, X_{i(t+\ell)}) \\ &= \text{Cov}(X_{jt}, X_{i(t-(-\ell))}) \end{aligned}$$

por lo que $\Gamma_{ij}(\ell) = \Gamma_{ji}(-\ell)$ de donde $\Gamma_\ell = (\Gamma_{-\ell})^T$ y por lo tanto $\rho_\ell = (\rho_{-\ell})^T$.

6.2.1. Dependencia lineal

Las matrices de correlaciones cruzadas $\{\rho_\ell | \ell = 0, 1, \dots\}$ de una serie de tiempo multivariada débilmente estacionaria contienen la siguiente información:

1. Los elementos de la diagonal $\{\rho_{ii}(\ell) | \ell = 0, 1, \dots\}$ son los valores que toma la función de autocorrelación de X_{ii} .
2. Los elementos fuera de la diagonal $\rho_{ij}(0)$ miden la relación lineal actual entre X_{it} y X_{jt} .
3. Para $\ell > 0$, los elementos fuera de la diagonal $\rho_{ij}(\ell)$ miden la dependencia lineal de X_{it} respecto de los valores pasados $X_{j(t-\ell)}$.
4. Si $\rho_{ij}(\ell) = 0$ para toda $\ell > 0$ entonces X_{it} no depende linealmente de algún valor pasado $X_{j(t-\ell)}$ de la serie X_{jt} .

En general, la relación lineal entre dos series de tiempo univariadas $\{X_{it}\}$ y $\{X_{jt}\}$ se puede resumir de la siguiente manera:

- Si $\rho_{ij}(\ell) = \rho_{ji}(\ell) = 0$ para todo $\ell \geq 0$ entonces no existe relación lineal entre X_{it} y X_{jt} . Si $\rho_{ij}(\ell) = \rho_{ji}(\ell) = 0$ se cumple solo para $\ell > 0$ entonces no existe relación guía en algún retraso ℓ y se dice que las dos series están *desacopladas*.
- Si $\rho_{ij}(\ell) \neq 0$ para $\ell = 0$ entonces X_{it} y X_{jt} están correlacionados.
- Si $\rho_{ij}(\ell) = 0$ para todo $\ell > 0$ y $\rho_{ij}(\ell') \neq 0$ para algún $\ell' > 0$ entonces X_{it} no depende de algún valor anterior de X_{jt} pero X_{jt} sí depende de algún valor anterior de X_{it} . En este caso se dice que hay una *relación unidireccional*.
- Si $\rho_{ij}(\ell) \neq 0$ y $\rho_{ij}(\ell') \neq 0$ para los valores $\ell > 0$ y $\ell' > 0$ entonces se dice que existe una *relación de retroalimentación* entre X_{it} y X_{jt} .

Es posible obtener más información sobre la relación entre dos o más series de tiempo construyendo un modelo multivariado ya que éste considera simultáneamente las correlaciones cruzadas y seriales entre las series en estudio, ver (Makridakis, Wheelwright, y Hyndman, 1998, pág. 423).

6.2.2. Matriz de correlaciones cruzadas muestrales

Sea $\mathbf{X}_t = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T\}$ una serie de tiempo multivariada observada.

- El vector de medias muestrales de \mathbf{X}_t es

$$\bar{\mathbf{X}} = \frac{1}{T} \sum_{t=1}^T \mathbf{X}_t.$$

- La matriz de covarianzas cruzadas muestrales en el retraso ℓ de \mathbf{X}_t es

$$\widehat{\Gamma}_\ell = \frac{1}{T} \sum_{t=\ell+1}^T (\mathbf{X}_t - \overline{\mathbf{X}})(\mathbf{X}_t - \overline{\mathbf{X}})^T, \quad \ell \geq 0.$$

- La matriz de correlaciones cruzadas muestrales en el retraso ℓ de \mathbf{X}_t es

$$\widehat{\rho}_\ell = \widehat{\mathbf{D}}^{-1} \widehat{\Gamma}_\ell \widehat{\mathbf{D}}^{-1}$$

donde $\widehat{\mathbf{D}}$ es la matriz diagonal de dimensión $k \times k$ de las desviaciones estándar muestrales de los componentes de la serie \mathbf{X}_t .

6.3. Modelos vectoriales autorregresivos

6.3.1. Proceso VAR(1)

Definición 6.5. Sea $\{\mathbf{X}_t\}$ un proceso estocástico multivariado con variables aleatorias de dimensión k . $\{\mathbf{X}_t\}$ es un **proceso vectorial autorregresivo de orden 1**, denotado por **VAR(1)**, si se puede representar mediante la ecuación

$$\mathbf{X}_t = \phi_0 + \Phi \mathbf{X}_{t-1} + \mathbf{Z}_t \quad (6.4)$$

donde ϕ_0 es un vector de dimensión k , Φ es una matriz de dimensión $k \times k$ y $\{\mathbf{Z}_t\}$ es una sucesión de vectores aleatorios de dimensión k no correlacionados serialmente con media cero y matriz de covarianzas Σ .

En las aplicaciones se requiere que la matriz de covarianzas Σ sea definida positiva, (Lutkepohl, 2005, pág. 30). De no ser así la dimensión de \mathbf{X}_t podría reducirse. Otra suposición que se hace con frecuencia es que el proceso $\{\mathbf{Z}_t\}$ se distribuye como una Normal multivariada.

Ejemplo 6.1. Considere la serie bivariada $\mathbf{X}_t = \{X_{1t}, X_{2t}\}$. El modelo VAR(1) consiste en las siguientes dos ecuaciones:

$$\begin{aligned} X_{1t} &= \phi_{10} + \Phi_{11}X_{1,t-1} + \Phi_{12}X_{2,t-1} + Z_{1t} \\ X_{2t} &= \phi_{20} + \Phi_{21}X_{1,t-1} + \Phi_{22}X_{2,t-1} + Z_{2t} \end{aligned}$$

donde Φ_{ij} es el (i, j) -ésimo elemento de Φ y ϕ_{i0} es el i -ésimo elemento de ϕ_0 . En la primera ecuación, Φ_{12} denota la dependencia lineal de X_{1t} sobre $X_{2,t-1}$ en la presencia de $X_{1,t-1}$. Por lo tanto Φ_{12} es el efecto condicional de $X_{2,t-1}$ sobre X_{1t} dado $X_{1,t-1}$. Si $\Phi_{12} = 0$ entonces X_{1t} no depende de $X_{2,t-1}$ y el modelo indica que X_{1t} sólo depende de su valor anterior. De manera análoga pasa cuando $\Phi_{22} = 0$ en la segunda ecuación. Considerando las dos ecuaciones simultáneamente, si $\Phi_{12} = 0$ y $\Phi_{21} \neq 0$, existe una relación

unidireccional de X_{1t} sobre X_{2t} . Si $\Phi_{12} = \Phi_{21} = 0$, entonces X_{1t} y X_{2t} están desacopladas. Si $\Phi_{12} \neq 0$ y $\Phi_{21} \neq 0$, entonces las dos series se retroalimentan. En resumen,

- Si $\Phi_{12} = 0$ y $\Phi_{21} \neq 0$, existe una relación unidireccional de X_{1t} sobre X_{2t} . La relación unidireccional opuesta se tiene si $\Phi_{21} = 0$ pero $\Phi_{12} \neq 0$.
- Si $\Phi_{12} = \Phi_{21} = 0$, entonces X_{1t} y X_{2t} están desacopladas.
- Si $\Phi_{12} \neq 0$ y $\Phi_{21} \neq 0$, entonces las dos series se retroalimentan.

Como ya se mencionó, los coeficientes de la matriz Φ de la Ecuación 6.4 miden la dinámica de dependencia de \mathbf{X}_t . Además la relación actual entre X_{1t} y X_{2t} se refleja en los elementos fuera de la diagonal de la matriz de covarianzas Σ de \mathbf{Z}_t . Si $\sigma_{12} = 0$ entonces no existe alguna relación lineal actual entre los dos componentes de la serie. En Econometría al modelo VAR(1) de la Ecuación 6.4 se le llama *modelo en forma reducida* debido a que no se muestran explícitamente los elementos de la matriz Σ , pero es posible pasar a una expresión explícita a través de una transformación lineal, ver (Tsay, 2005, pág. 400).

En general, para una serie multivariada $\mathbf{X}_t = \{X_{1t}, X_{2t}, \dots, X_{kt}\}$, ésta forma explícita se llama la *ecuación estructural* de X_{kt} , la cual es

$$X_{kt} + \sum_{i=1}^{k-1} w_{ki} X_{it} = \phi_{k0}^* + \sum_{i=1}^k \Sigma_{ki}^* X_{i(t-1)} + b_{kt},$$

donde ϕ_{k0}^* es el k -ésimo elemento de ϕ_0^* , Φ_{ki}^* es el (k, i) -ésimo elemento de Φ^* y $\mathbf{b}_t = (b_{1t}, \dots, b_{kt})^T = \mathbf{L}^{-1} \mathbf{Z}_t$ donde \mathbf{L} es una matriz triangular inferior con 1's en la diagonal. Lo que asegura que la matriz \mathbf{L} exista es el hecho de que la matriz Σ es definida positiva y mediante la descomposición de Cholesky se puede encontrar la matriz \mathbf{L} y una matriz diagonal \mathbf{G} tal que $\Sigma = \mathbf{LGL}^T$.

Suponga que el modelo VAR(1) en la Ecuación 6.4 es débilmente estacionario. Aplicando la esperanza al modelo y tomando en cuenta que $E(\mathbf{Z}_t) = \mathbf{0}$ se obtiene

$$E(\mathbf{X}_t) = \phi_0 + \Phi E(\mathbf{X}_{t-1}).$$

Como $E(\mathbf{X}_t)$ es invariante en el tiempo,

$$\boldsymbol{\mu} \equiv E(\mathbf{X}_t) \equiv (\mathbf{I} - \Phi)^{-1} \phi_0$$

siempre que la matriz $\mathbf{I} - \Phi$ sea no singular, donde \mathbf{I} es la matriz identidad de $k \times k$. Si $\phi_0 = (\mathbf{I} - \Phi)\boldsymbol{\mu}$, el modelo VAR(1) en la Ecuación 6.4 se puede escribir como

$$(\mathbf{X}_t - \boldsymbol{\mu}) = \Phi(\mathbf{X}_{t-1} - \boldsymbol{\mu}) + \mathbf{Z}_t.$$

Sea $\widetilde{\mathbf{X}}_t = \mathbf{X}_t - \boldsymbol{\mu}$ la serie de tiempo con media corregida. Entonces el modelo VAR(1) es

$$\widetilde{\mathbf{X}}_t = \boldsymbol{\Phi} \widetilde{\mathbf{X}}_{t-1} + \mathbf{Z}_t. \quad (6.5)$$

Después de varias sustituciones la Ecuación (6.5) puede reescribirse como

$$\widetilde{\mathbf{X}}_t = \mathbf{Z}_t + \boldsymbol{\Phi} \mathbf{Z}_{t-1} + \boldsymbol{\Phi}^2 \mathbf{Z}_{t-2} + \boldsymbol{\Phi}^3 \mathbf{Z}_{t-3} + \dots. \quad (6.6)$$

A partir de la Ecuación (6.6) pueden deducirse las siguientes propiedades del modelo VAR(1).

- Ya que \mathbf{Z}_t no es correlacionada serialmente, se sigue que $\text{Cov}(\mathbf{Z}_t, \mathbf{X}_{t-1}) = 0$. De hecho, \mathbf{Z}_t no está correlacionada con $\mathbf{X}_{t-\ell}$ para ningún $\ell > 0$.
- Multiplicando por \mathbf{Z}_t^T el lado derecho de la Ecuación (6.6), tomando la esperanza y usando el hecho de que el proceso \mathbf{Z}_t no es correlacionado serialmente se obtiene que $\text{Cov}(\mathbf{X}_t, \mathbf{Z}_t) = \boldsymbol{\Sigma}$.
- Para un modelo VAR(1), \mathbf{X}_t depende de \mathbf{Z}_{t-j} . Para que tal dependencia sea significativa, $\boldsymbol{\Phi}^j$ debe converger a cero cuando $j \rightarrow \infty$. Esto significa que el módulo del eigenvalor de $\boldsymbol{\Phi}$ debe ser menor que 1. De otro modo $\boldsymbol{\Phi}^j$ divergirá o convergirá a una matriz distinta de cero cuando $j \rightarrow \infty$. De hecho, el requisito de que el módulo de todos los eigenvalores de $\boldsymbol{\Phi}$ sean menores a 1 es una condición necesaria y suficiente para la estacionaridad débil de \mathbf{X}_t siempre que la matriz de covarianzas de \mathbf{Z}_t exista.
- Es posible verificar que

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi} \boldsymbol{\Gamma}_{\ell-1}, \quad \ell > 0,$$

donde $\boldsymbol{\Gamma}_j$ es el j -ésimo retraso de la matriz de covarianzas cruzadas de \mathbf{X}_t .

6.3.2. Proceso VAR(p)

Definición 6.6. Sea $\{\mathbf{X}_t\}$ un proceso estocástico multivariado donde cada una de sus variables son de dimensión k . $\{\mathbf{X}_t\}$ es un **proceso vectorial autorregresivo de orden p** , denotado por **VAR(p)**, si puede representarse mediante la ecuación

$$\mathbf{X}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \mathbf{X}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{X}_{t-p} + \mathbf{Z}_t \quad (6.7)$$

donde $p > 0$, $\boldsymbol{\phi}_0$ es un vector de dimensión k , $\boldsymbol{\Phi}_i$ son matrices de dimensión $k \times k$ y $\{\mathbf{Z}_t\}$ es una sucesión de vectores aleatorios de dimensión k no correlacionados serialmente con media cero y matriz de covarianza $\boldsymbol{\Sigma}$.

Usando el operador de retraso B la Ecuación (6.7) se puede escribir como

$$(\mathbf{I} - \boldsymbol{\Phi}_1 B - \dots - \boldsymbol{\Phi}_p B^p) \mathbf{X}_t = \boldsymbol{\phi}_0 + \mathbf{Z}_t \quad (6.8)$$

donde \mathbf{I} es la matriz identidad de tamaño $k \times k$.

De la ecuación

$$\widetilde{\mathbf{X}}_t = \Phi_1 \widetilde{\mathbf{X}}_{t-1} + \dots + \Phi_p \widetilde{\mathbf{X}}_{t-p} + \mathbf{Z}_t$$

que es similar a la Ecuación (6.5) del modelo VAR(1) y con las mismas técnicas se pueden obtener las siguientes propiedades:

- $\text{Cov}(\mathbf{X}_t, \mathbf{Z}_t) = \Sigma$,
- $\text{Cov}(\mathbf{X}_{t-\ell}, \mathbf{Z}_t) = 0$ para $\ell > 0$,
- $\Gamma_\ell = \Phi_1 \Gamma_{\ell-1} + \dots + \Phi_p \Gamma_{\ell-p}$ para $\ell > 0$.

La condición necesaria y suficiente para la estacionaridad débil de un VAR(p) es que los eigenvalores de Φ^* sean menores a uno en módulo, donde

$$\Phi^* = \begin{pmatrix} 0 & \mathbf{I} & 0 & \dots & 0 \\ 0 & 0 & \mathbf{I} & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{I} \\ \Phi_p & \Phi_{p-1} & \Phi_{p-2} & \dots & \Phi_1 \end{pmatrix}$$

es una matriz de dimensión $kp \times kp$, donde $\mathbf{0}$ e \mathbf{I} son las matrices nula e identidad de dimensiones $k \times k$.

Por otro lado observando las entradas de la matriz Φ_ℓ del modelo VAR(p), si el (i, j) -ésimo elemento de ésta matriz es cero para toda ℓ entonces X_{it} no depende de los valores pasados de X_{jt} . La estructura de las entradas de la matriz Φ_ℓ brinda información de la relación retraso - guía entre los componentes de \mathbf{X}_t .

6.3.3. Ajuste de un modelo VAR

La metodología estadística para llevar a cabo el ajuste de un modelo VAR es muy parecida a la metodología de Box & Jenkins. Ésta consiste en la especificación del orden del modelo VAR con ayuda de la FAP generalizada al caso multivariado, la estimación de los parámetros y la validación. Para un modelo VAR específico, la estimación de los parámetros se puede llevar a cabo mediante mínimos cuadrados o máxima verosimilitud. Los dos métodos son asintóticamente equivalentes. Una vez ajustado un modelo se debe verificar cuidadosamente. El estadístico $Q_k(m)$ puede aplicarse a los residuales de las series para verificar la suposición de que no hay correlaciones seriales o cruzadas en los residuales. Para un modelo VAR(p) ya ajustado el estadístico $Q_k(m)$

de los residuales se distribuye asintóticamente χ^2 con $k^2m - g$ grados de libertad donde g es el número de parámetros estimados en las matrices que son los coeficientes AR en el modelo.

Una forma de especificar el orden p del modelo VAR es contrastando $H_0 : \Phi_\ell = 0$ contra $H_a : \Phi_\ell \neq 0$ secuencialmente para $\ell = 1, 2, \dots$. Por ejemplo, para saber si el modelo $VAR(i)$ es mejor que el modelo $VAR(i-1)$ se contrasta $H_0 : \Phi_i = 0$ con $H_a : \Phi_i \neq 0$ con la estadística

$$M(i) = - \left(T - k - i - \frac{3}{2} \right) \ln \left(\frac{|\widehat{\Sigma}_i|}{|\widehat{\Sigma}_{i-1}|} \right)$$

donde $\widehat{\Sigma}_i$ está definida como en la Ecuación (6.9) y $|A|$ es el determinante de la matriz A . El estadístico $M(i)$ se distribuye asintóticamente χ^2 con k^2 grados de libertad, ver (Reinsel, 1993, pág. 79).

Otra forma de seleccionar el orden p es mediante el Akaike Information Criterion (AIC) o el Bayesian Information Criterion (BIC). Suponga que Z_t se distribuye Normal multivariada y que la estimación de los parámetros se realizó con el método de máxima verosimilitud. Los estimadores de ϕ_0 y Φ_j son los mismos tanto para el método de mínimos cuadrados como para máxima verosimilitud. El estimador máximo verosímil para Σ es

$$\widehat{\Sigma}_i = \frac{1}{T} \sum_{t=i+1}^T \widehat{Z}_t (\widehat{Z}_t)^T.$$

El AIC para un $VAR(i)$ bajo la suposición de normalidad es

$$AIC(i) = \ln(|\widehat{\Sigma}_i|) + \frac{2k^2i}{T}.$$

Para una serie multivariada dada se puede seleccionar el orden $VAR(p)$ tal que

$$AIC(p) = \min_{0 \leq i \leq p_0} AIC(i)$$

donde p_0 es un entero positivo preespecificado. Otro criterio de información disponible es el BIC para el modelo $VAR(i)$:

$$BIC(i) = \ln(|\widehat{\Sigma}_i|) + \frac{k^2i \ln(T)}{T}.$$

Suponga que la estimación de los parámetros se puede llevar a cabo mediante el método de mínimos cuadrados. Sean $\widehat{\Phi}_j$ y $\widehat{\phi}_0$ los estimadores por mínimos cuadrados de Φ_j y ϕ_0 respectivamente. La matriz de covarianzas de los residuales

$$\widehat{Z}_t = \mathbf{X}_t - \widehat{\phi}_0 - \widehat{\phi}_1 \mathbf{X}_{t-1} - \dots - \widehat{\phi}_i \mathbf{X}_{t-i}$$

es

$$\widehat{\Sigma}_i = \frac{1}{T - 2i - 1} \sum_{t=i+1}^T \widehat{Z}_t (\widehat{Z}_t)^T \quad (6.9)$$

donde $i \geq 0$.

Es posible generalizar el estadístico $Q(m)$ mencionado en la Sección 4.3.2 del Capítulo 4 al caso multivariado. En este caso el estadístico es un test conjunto para verificar que las primeras m matrices de correlaciones cruzadas de \mathbf{X}_t son nulas.

Dada una serie multivariada \mathbf{X}_t , la hipótesis nula del test estadístico es $H_0 = \rho_1 = \dots = \rho_m = 0$ y la hipótesis alternativa es $H_1 = \rho_i \neq 0$ para algún $i = 1, \dots, m$. Por lo tanto el estadístico verifica que no hay autocorrelaciones o correlaciones cruzadas en la serie de tiempo multivariada. El estadístico se calcula mediante la fórmula

$$Q_k(m) = T^2 \sum_{\ell=1}^m \frac{1}{T-\ell} \text{tr}(\widehat{\Gamma}_\ell^T \widehat{\Gamma}_0^{-1} \widehat{\Gamma}_\ell \widehat{\Gamma}_0^{-1}) \quad (6.10)$$

donde T es el tamaño de la muestra, k es la dimensión de \mathbf{X}_t y $\text{tr}(A)$ es la traza de la matriz A . Bajo la hipótesis nula $Q_k(m)$ sigue una distribución χ^2 con $k^2 m$ grados de libertad.

Si se rechaza la hipótesis nula en el test anterior conviene ajustar un modelo multivariado a la serie para analizar las relaciones retraso - guía entre los componentes de la serie multivariada en estudio. En las secciones siguientes se introducen los modelos VARIMA que corresponden a los modelos ARIMA para series multivariadas.

6.4. Modelos vectoriales de medias móviles

6.4.1. Proceso VMA(q)

Definición 6.7. Sea $\{\mathbf{X}_t\}$ un proceso estocástico multivariado con variables aleatorias de dimensión k . $\{\mathbf{X}_t\}$ es un **proceso vectorial de medias móviles de orden q** , denotado por **VMA(q)**, si se puede representar mediante la ecuación

$$\mathbf{X}_t = \boldsymbol{\theta}_0 + \mathbf{Z}_t - \boldsymbol{\Theta}_1 \mathbf{Z}_{t-1} - \dots - \boldsymbol{\Theta}_q \mathbf{Z}_{t-q} \quad (6.11)$$

donde $\boldsymbol{\theta}_0$ es un vector k dimensional, $\boldsymbol{\Theta}_i$ son matrices de tamaño $k \times k$ y $\{\mathbf{Z}_t\}$ es una sucesión de vectores aleatorios de dimensión k no correlacionados serialmente con media cero y matriz de covarianzas $\boldsymbol{\Sigma}$.

Con el operador de retraso B el modelo anterior es

$$\mathbf{X}_t = \boldsymbol{\theta}_0 + \boldsymbol{\Theta}(B) \mathbf{Z}_t \quad (6.12)$$

donde $\boldsymbol{\Theta}(B) = I - \boldsymbol{\Theta}_1 B - \dots - \boldsymbol{\Theta}_q B^q$. Similar al caso univariado, los procesos VMA(q) son débilmente estacionarios siempre que exista la matriz de covarianzas $\boldsymbol{\Sigma}$ de \mathbf{Z}_t . Tomando la esperanza en la Ecuación (6.11)

se obtiene que $\boldsymbol{\mu} = E(\mathbf{X}_t) = \boldsymbol{\theta}_0$.

Algunas propiedades de los modelos VMA(q) son las siguientes:

- $Cov(\mathbf{X}_t, \mathbf{Z}_t) = \boldsymbol{\Sigma}$
- $\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma} + \boldsymbol{\Theta}_1 \boldsymbol{\Sigma} \boldsymbol{\Theta}_1^T + \dots + \boldsymbol{\Theta}_q \boldsymbol{\Sigma} \boldsymbol{\Theta}_q^T$
- $\boldsymbol{\Gamma}_\ell = 0$ si $\ell > q$
- $\boldsymbol{\Gamma}_\ell = \sum_{j=\ell}^q \boldsymbol{\Theta}_j \boldsymbol{\Sigma} \boldsymbol{\Theta}_{j-\ell}$ si $1 \leq \ell \leq q$ donde $\boldsymbol{\Theta}_0 = -\mathbf{I}$.

Ejemplo 6.2. Considere el modelo bivariado MA(1)

$$\mathbf{X}_t = \boldsymbol{\theta}_0 + \mathbf{Z}_t - \boldsymbol{\Theta}_1 \mathbf{Z}_{t-1} = \boldsymbol{\mu} + \mathbf{Z}_t - \boldsymbol{\Theta}_1 \mathbf{Z}_{t-1} \quad (6.13)$$

que, visto en forma vectorial es

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} - \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix}. \quad (6.14)$$

El modelo indica que los valores actuales de las series X_{it} solo dependen de los valores actuales y pasados del ruido aleatorio Z_{it} . En la Ecuación (6.14) el parámetro Θ_{12} indica la dependencia lineal de X_{1t} sobre $Z_{2,t-1}$ en la presencia de $Z_{1,t-1}$. Si $\Theta_{12} = 0$ entonces X_{1t} no depende de los valores pasados de Z_{2t} . Igualmente, si $\Theta_{21} = 0$, entonces X_{2t} no depende de los valores pasados de Z_{1t} . Los elementos fuera de la diagonal de $\boldsymbol{\Theta}_1$ muestran la dependencia dinámica entre los componentes de la serie. En este ejemplo VMA(1) la clasificación de las relaciones entre X_{1t} y X_{2t} es la siguiente:

- Si $\Theta_{12} = \Theta_{21} = 0$ entonces los componentes X_{1t} y X_{2t} están desacoplados.
- Si $\Theta_{12} = 0$ pero $\Theta_{21} \neq 0$ entonces existe una relación dinámica unidireccional de X_{1t} sobre X_{2t} . La relación unidireccional opuesta se tiene si $\Theta_{21} = 0$ pero $\Theta_{12} \neq 0$.
- Si $\Theta_{12} \neq 0$ y $\Theta_{21} \neq 0$ entonces hay una relación de retroalimentación entre X_{1t} y X_{2t} .

La correlación actual entre X_{it} es la misma que la que hay entre Z_{it} . La clasificación anterior también puede generalizarse al modelo VMA(q).

6.4.2. Ajuste de un modelo VMA

A diferencia de los modelos VAR, la estimación de los modelos VMA es más complicada (Fuller, 1995, pág. 421). Se puede obtener una aproximación de la función de verosimilitud mediante dos métodos. Uno es el método de la verosimilitud condicional que asume que $\mathbf{Z}_t = \mathbf{0}$ para $t \leq 0$. El otro es el método de la verosimilitud

exacta que considera a Z_t , con $t \leq 0$, como parámetro adicional al modelo. La verosimilitud exacta requiere cálculos más complicados, pero ofrece estimadores de los parámetros más exactos, especialmente cuando algunos eigenvalores de Θ se acercan a 1 en módulo.

El ajuste de un modelo VMA requiere tres pasos: la especificación del orden q mediante las matrices de correlaciones cruzadas, de tal forma que $\rho_\ell = 0$ para $\ell > q$, la estimación del modelo especificado usando el método de la verosimilitud condicional o exacta, aunque es preferible el segundo cuando el tamaño de la serie no es grande, y finalmente la verificación del ajuste, por ejemplo aplicando el estadístico $Q_k(m)$ a la serie de los residuales.

6.5. Modelos vectoriales autorregresivos y de medias móviles

6.5.1. Proceso VARMA(p,q)

Definición 6.8. Sea $\{X_t\}$ un proceso estocástico multivariado con variables aleatorias de dimensión k . $\{X_t\}$ es un **proceso vectorial autorregresivo y de medias móviles de orden (p,q)**, denotado por **VARMA(p,q)**, si se puede representar mediante la ecuación

$$X_t - \Phi_1 X_{t-1} - \dots - \Phi_p X_{t-p} = Z_t - \Theta_1 Z_{t-1} - \dots - \Theta_q Z_{t-q} \quad (6.15)$$

donde Φ_i y Θ_i son matrices de tamaño $k \times k$ y $\{Z_t\}$ es una sucesión de vectores aleatorios de dimensión k no correlacionados serialmente con media cero y matriz de covarianzas Σ .

Los modelos VARMA son una generalización al caso multivariado de los modelos ARMA. Un aspecto desfavorable de éstos que no se tiene en los modelos univariados es el problema de identificabilidad. A diferencia de los ARMA, los modelos VARMA no están definidos de forma única. Por ejemplo el modelo VMA(1)

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Z_{1(t-1)} \\ Z_{2(t-1)} \end{bmatrix} \quad (6.16)$$

es idéntico al modelo VAR(1)

$$\begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} = \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \end{bmatrix}. \quad (6.17)$$

Analizando los modelos univariados que conforman los modelos anteriores se observa que para el VMA(1)

$$\begin{aligned} X_{1t} &= Z_{1t} - 2Z_{2(t-1)} \\ X_{2t} &= Z_{2t}. \end{aligned}$$

Para el VAR(1) las ecuaciones de los modelos univariados son

$$\begin{aligned} X_{1t} + 2X_{2(t-1)} &= Z_{1t} \\ X_{2t} &= Z_{2t}. \end{aligned}$$

Sustituyendo X_{2t} en X_{1t} del modelo VMA(1) se tiene que X_{1t} es el mismo que en el modelo VAR(1) y por lo tanto los modelos de las Ecuaciones (6.16) y (6.17) son idénticos.

Otro problema de identificabilidad que es más peligroso es el siguiente. Considere el modelo VARMA(1,1)

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \end{bmatrix} = \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Z_{1(t-1)} \\ Z_{2(t-1)} \end{bmatrix}. \quad (6.18)$$

Este modelo es idéntico al VARMA(1,1) siguiente

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2 + \eta \\ 0 & \varphi \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \end{bmatrix} = \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & \eta \\ 0 & \varphi \end{bmatrix} \begin{bmatrix} Z_{1(t-1)} \\ Z_{2(t-1)} \end{bmatrix} \quad (6.19)$$

para cualquier η y φ distintos de cero. En este caso la equivalencia se dá ya que $X_{2t} = Z_{2t}$ en ambos modelos. Los efectos de los parámetros η y φ en el sistema se contrarrestan entre las partes AR y MA del segundo modelo. Esta situación es seria ya que con ciertos valores de η y φ la función de verosimilitud del modelo vectorial ARMA(1,1) anterior no está definida de manera única, resultando en una situación similar a la multicolinealidad en un análisis de regresión. Los dos ejemplos anteriores muestran los posibles problemas de identificabilidad que pueden ocurrir en los modelos VARMA. Por lo tanto la construcción de éstos modelos requiere ciertos cuidados.

Los modelos VARMA(p,q) pueden escribirse en forma compacta con el operador B :

$$\Phi(B)\mathbf{X}_t = \phi_0 + \Theta(B)\mathbf{Z}_t$$

donde $\Phi(B) = I - \Phi_1 B - \dots - \Phi_p B^p$ y $\Theta(B) = I - \Theta_1 B - \dots - \Theta_q B^q$ son polinomios matriciales de dimensión $k \times k$. Aquí se asume que los dos polinomios matriciales no tienen factores en común, de otra forma el modelo podría simplificarse. La condición necesaria y suficiente de estacionaridad débil para \mathbf{X}_t es la misma que para el modelo VAR(p) con polinomio matricial $\Phi(B)$. Para $\nu > 0$, los elementos (i, j) de las matrices Φ_ν y Θ_ν indican la dependencia lineal de X_{it} sobre $X_{j(t-\nu)}$ y $Z_{j(t-\nu)}$ respectivamente. Si el elemento (i, j) es cero para todos los coeficientes matriciales en las partes AR y MA entonces X_{it} no depende de los valores retrasados de X_{jt} . La proposición contraria no se cumple en un modelo VARMA. En otras palabras, los coeficientes distintos de cero en la posición (i, j) de las matrices AR y MA pueden existir incluso cuando X_{it} no

dependa de algún retraso de X_{jt} .

Para ilustrar lo anterior considere el modelo

$$\begin{bmatrix} \Phi_{11}(B) & \Phi_{12}(B) \\ \Phi_{21}(B) & \Phi_{22}(B) \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \Theta_{11}(B) & \Theta_{12}(B) \\ \Theta_{21}(B) & \Theta_{22}(B) \end{bmatrix} \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix}. \quad (6.20)$$

En este caso las condiciones necesarias y suficientes para la existencia de la relación dinámica unidireccional entre X_{1t} y X_{2t} son

$$\Phi_{22}(B)\Theta_{12}(B) - \Phi_{12}(B)\Theta_{22}(B) = 0 \quad (6.21)$$

y además

$$\Phi_{11}(B)\Theta_{21}(B) - \Phi_{21}(B)\Theta_{11}(B) \neq 0. \quad (6.22)$$

La condición en la Ecuación (6.21) muestra que X_{1t} no depende de valores pasados de Z_{2t} ni de X_{2t} y la condición en la Ecuación (6.22) muestra que X_{2t} sí depende de valores pasados de Z_{1t} . De las Ecuaciones (6.21) y (6.22), $\Theta_{12}(B) = \Phi_{12}(B) = 0$ es una condición suficiente pero no necesaria para que exista una relación unidireccional entre X_{1t} y X_{2t} , ver (Tsay, 2005, pág. 423).

La estimación de un modelo VARMA puede llevarse a cabo por el método condicional o exacto de verosimilitud. La estadística $Q_k(m)$ también se puede aplicar a la serie de los residuales de un modelo ajustado, donde su distribución asintótica es χ^2 con $k^2m - g$ grados de libertad donde g es el número de parámetros estimados en los coeficientes matriciales AR y MA.

6.5.2. Modelos marginales de los componentes

Dado un modelo vectorial para \mathbf{X}_t , los modelos univariados implicados respectivos a los componentes X_{it} son los modelos marginales. Para un modelo k -dimensional VARMA(p,q) los modelos marginales son modelos ARMA[kp,(k-1)p + q]. Este resultado se puede obtener en dos pasos. Primero, los modelos marginales de un modelo VMA(q) son modelos MA(q) ya que si \mathbf{X}_t es un VMA(q) su matriz de correlación cruzada se anula después del retraso q y el ACF de X_{it} también se anula después del retraso q . Entonces X_{it} es un proceso MA y su modelo univariado es de la forma

$$X_{it} = \theta_{i0} + \sum_{j=1}^q \theta_{ij} b_{i,t-j}$$

donde b_{it} es una sucesión de variables no correladas con media cero y varianza σ_{ib}^2 . Los parámetros θ_{ij} y σ_{ib} son funciones de los parámetros del modelo VMA de \mathbf{X}_t .

El segundo paso para obtener el resultado es diagonalizar el polinomio matricial AR del modelo VARMA(p,q).

Por ejemplo considere el modelo bivariado AR(1)

$$\begin{bmatrix} 1 - \Phi_{11}B & -\Phi_{12}B \\ -\Phi_{21}B & 1 - \Phi_{22}B \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix}. \quad (6.23)$$

Multiplicando la Ecuación (6.23) por la matriz

$$\begin{bmatrix} 1 - \Phi_{22}B & \Phi_{12}B \\ \Phi_{21}B & 1 - \Phi_{11}B \end{bmatrix}$$

se obtiene

$$[(1 - \phi_{11}B)(1 - \phi_{22}B) - \phi_{12}\phi_{22}B^2] \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} 1 - \phi_{22}B & -\phi_{12}B \\ -\phi_{21}B & 1 - \phi_{11}B \end{bmatrix} \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix}. \quad (6.24)$$

Del lado derecho de la Ecuación (6.24) se observa que los polinomios univariados AR de X_{it} son de orden 2.

En contraste, el lado izquierdo de la ecuación está en la forma VMA(1). Usando el resultado de los modelos

VMA del paso uno se tiene que el modelo univariado para X_{it} es un modelo ARMA(2,1). La técnica se puede

generalizar al caso k -dimensional VAR(1), obteniendo como resultado que los modelos marginales son modelos

ARMA(k,k-1). Para un modelo k dimensional VAR(p), los modelos marginales son ARMA[kp,(k-1)p]. Para los

modelos VARMA, el orden [kp,(k-1)p+q] es el máximo orden posible respecto a los modelos marginales. Sin

embargo, en la práctica el orden marginal de X_{it} puede ser mucho más bajo.

6.6. Ajuste de la serie bivariada (QQQ,MSFT)

Como ya se mencionó en la Sección 5.1, Microsoft Corporation es una empresa multinacional de origen es-

tadounidense fundada en 1975 por Bill Gates y Paul Allen. Dedicada al sector del software y el hardware,

desarrolla, fabrica, licencia y produce software y equipos electrónicos, siendo sus productos más usados el sis-

tema operativo Microsoft Windows y la suite Microsoft Office, los cuales tienen una importante posición entre

las computadoras personales. Cuenta con 93 000 empleados en 102 países diferentes y con ingresos de 72 360

millones de dólares durante el año 2012.

Por otro lado el Nasdaq 100 es un índice bursátil de Estados Unidos que refleja la evolución de las 100 em-

presas con mayor capitalización bursátil que cotizan en el mercado electrónico del Nasdaq. Dichas empresas se

dedican a las telecomunicaciones, a ventas al mayoreo y menudeo, al hardware y software o biotecnología y

forman parte del Nasdaq Stock Market. QQQ es el símbolo del índice Nasdaq 100, el cual se compra y vende

como acciones, y es uno de los que tienen el mayor volumen de compra y venta. Como se puede comprar y

vender igual que cualquier acción común y corriente, el inversionista que quiera ganar dinero con las subidas o bajadas de éste índice puede comprar o vender sus acciones. Esto puede pasar sin que el inversionista invierta en las acciones de alguna compañía específica. Cuando se hace daytrading, es decir comprar y vender el mismo día, el inversionista quiere asegurarse que una acción tiene suficiente liquidez durante el día para poder comprar y vender esa acción fácilmente. No hay muchos índices negociables que tengan bastante volumen, pero el índice Nasdaq 100 sí lo tiene, lo cual hace posible su compra y venta como si fueran acciones normales de alto volumen.

El índice Nasdaq 100 está fuertemente influenciado por Intel, Cisco Systems, Microsoft Corporation, Google, Apple Inc., Oracle y Research in Motion (ver Lerman (2001), Israel (2007) y Turner (2005)). Ya que dichas compañías corresponden a un gran porcentaje del total de acciones que maneja el índice, cualquier suceso o actividad ocurrida con sus acciones ejerce gran influencia en el índice y por lo tanto en sus propias acciones. Debido a esto es común que los inversionistas deseen observar el comportamiento individual y conjunto de las acciones de éstas grandes compañías con las del índice. Las subidas y bajadas de dichas empresas ocurren según sus ciclos de producción, sus ingresos o los movimientos realizados por otros inversionistas, a pesar de la tendencia general del mercado.

En el Capítulo 5 se analizaron las series QQQ y MSFT referentes a los precios de las acciones del Índice Nasdaq 100 y de Microsoft Corporation respectivamente, ambos tomados diariamente desde el 13 de diciembre del 2004 al 15 de diciembre del 2014. Se encontró que las series no son estacionarias pero que las series compuestas por sus primeras diferencias sí lo son y que los modelos ARMA más adecuados para explicarlas son el ARMA(0,1) para la serie X_t donde $X_t = \nabla QQQ$ y ARMA(2,2) para la serie Y_t donde $Y_t = \nabla MSFT$. Considere la serie de tiempo bivariada (X_t, Y_t) . Como análisis inicial se realizó la Figura 6.1 que contiene las gráficas de dispersión entre a) las series X_t y Y_t , b) las series Y_t y el primer retraso de X_t , c) las series X_t y el primer retraso de Y_t y finalmente d) los primeros retrasos de X_t y Y_t . En las cuatro gráficas se aprecia que los puntos forman una línea inclinada a la derecha, lo cuál indica que existe una dependencia lineal fuerte entre las series, así como entre las series y sus respectivos retrasos, e incluso entre sus retrasos. Esto se confirma observando la matriz de autocorrelaciones de la serie bivariada:

$$\hat{\rho}_0 = \begin{pmatrix} 1 & 0.667 \\ 0.667 & 1 \end{pmatrix}.$$

Ya que las entradas (2,1) y (1,2) de la matriz son cercanas a uno la correlación entre las series X_t y Y_t es relativamente alta como se puede observar en la gráfica a) de la Figura 6.1.

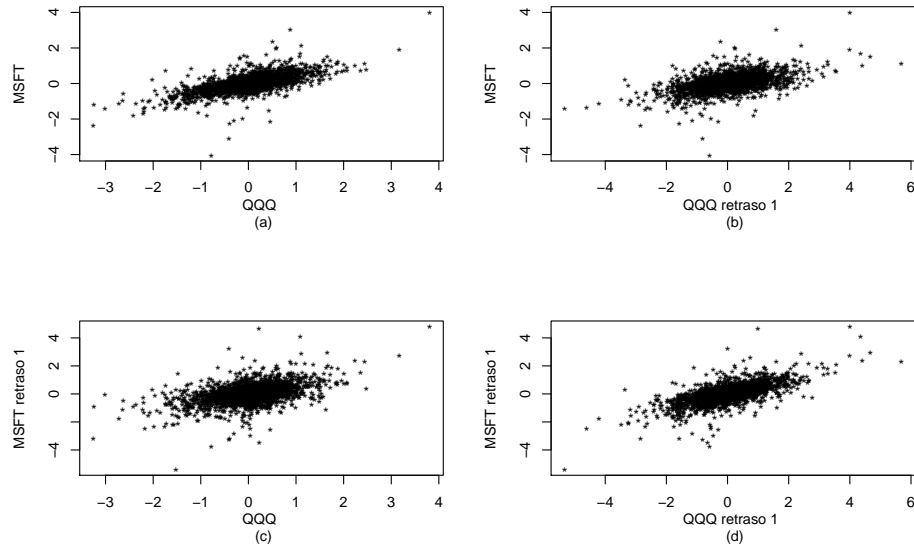


Figura 6.1: Gráficas de dispersión entre a) las series X_t y Y_t , b) Y_t y el primer retraso de X_t , c) X_t y el primer retraso de Y_t , d) los primeros retrasos de X_t y Y_t , donde $X_t = \nabla QQQ$ y $Y_t = \nabla MSFT$.

La parte superior de la Tabla 6.1 contiene las matrices de correlaciones cruzadas de la serie (X_t, Y_t) hasta el retraso 5. Por tratarse de una serie bivariada dichas matrices son de dimensión 2×2 . Ya que la experiencia empírica indica que es difícil absorber simultáneamente muchas matrices de correlación cruzada, especialmente cuando la dimensión de la serie multivariada es mayor a 3, la parte inferior de la tabla contiene las mismas matrices pero en la notación simplificada de donde cada matriz de correlación cruzada se conforma por tres símbolos (+, - y .) con el siguiente significado:

- El símbolo + significa que el correspondiente coeficiente de correlación es mayor o igual a $1/\sqrt{T}$.
- El símbolo - significa que el correspondiente coeficiente de correlación es menor o igual a $-1/\sqrt{T}$.
- El símbolo . significa que el correspondiente coeficiente de correlación está entre $-1/\sqrt{T}$ y $1/\sqrt{T}$.

donde $1/\sqrt{T}$ es el valor crítico asintótico del 5% de la correlación muestral sobre la suposición de que X_t y Y_t son ruidos blancos.

En la Tabla 6.1 se observa lo siguiente. Ya que sólo la autocorrelación en la entrada (1,1) de la matriz correspondiente al retraso 1 es significativa se concluye que la serie X_t no depende de sus valores pasados, pero que Y_t sí depende de sus valores pasados ya que todas las entradas (2,2) de las matrices son significativas. Además la serie X_t depende de valores pasados de la serie Y_t . Esto se deduce de que las entradas (1,2) de las matrices en los retrasos 1, 3 y 4 son significativas. Finalmente, ya que las entradas (2,1) de las matrices en los retrasos 1, 2 y 3 son significativas se deduce que Y_t también depende de los valores pasados de la serie X_t .

a) Matrices de correlación cruzada										
	Retraso 1		Retraso 2		Retraso 3		Retraso 4		Retraso 5	
	difQQQ	difMSFT	difQQQ	difMSFT	difQQQ	difMSFT	difQQQ	difMSFT	difQQQ	difMSFT
difQQQ	-0.0333	-0.0277	-0.0147	-0.0126	-0.00555	0.0199	0.00511	-0.0432	-0.01446	-0.0197
difMSFT	-0.0664	-0.0456	-0.0279	-0.0262	0.02031	0.0364	-0.0144	-0.0517	-0.00213	-0.0289
b) Notación simplificada										
difQQQ	-	-	.	.	.	+	.	-	.	.
difMSFT	-	-	-	-	+	+	.	-	.	-

Tabla 6.1: a) Matrices de correlaciones cruzadas hasta el retraso 5 de la serie bivariada (X_t, Y_t) . b) Matrices de correlaciones cruzadas hasta el retraso 5 de la serie bivariada (X_t, Y_t) en notación simplificada.

La Figura 6.2 contiene las autocorrelaciones y las correlaciones cruzadas de las dos series. La gráfica en el a) contiene el ACF de la serie X_t y la del b) muestra la dependencia de X_t respecto al retraso de la serie Y_t . La gráfica en el c) contiene el ACF de la serie Y_t y la del d) muestra la dependencia de Y_t respecto al retraso de la serie X_t . De las gráficas b) y c) se deduce que las correlaciones cruzadas entre X_t y Y_t son estadísticamente significativas.

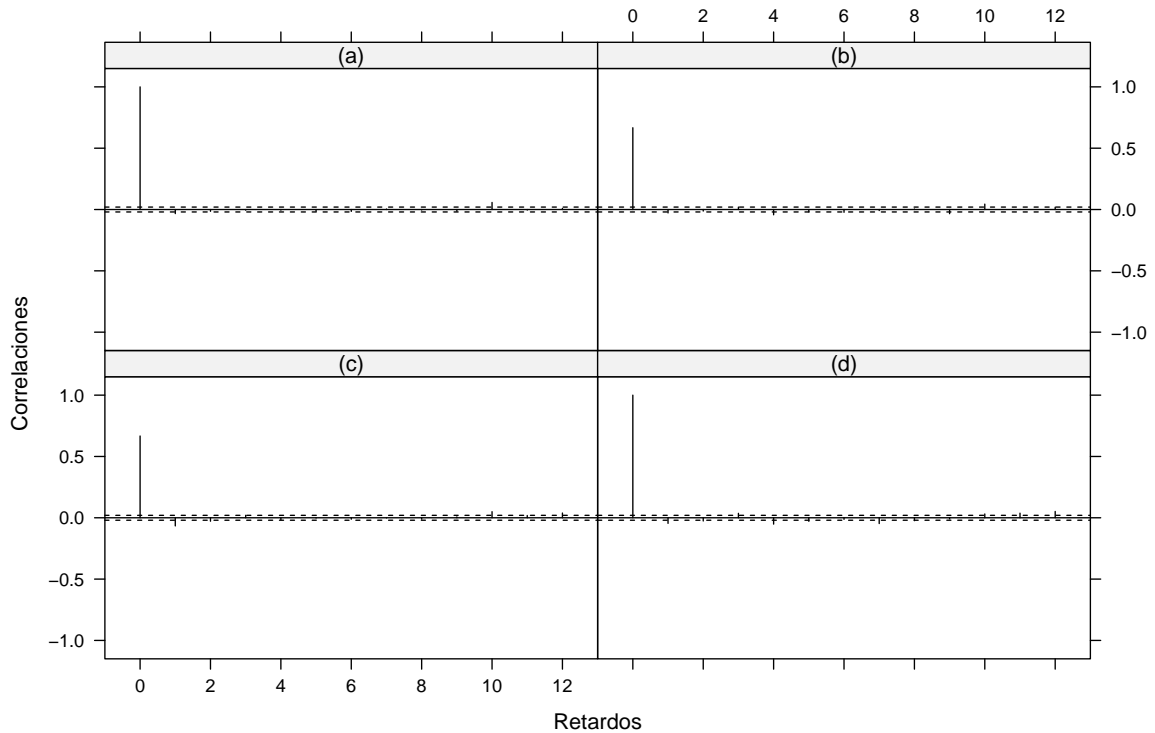


Figura 6.2: Funciones de autocorrelación y de correlaciones cruzadas de las series X_t y Y_t . a) ACF de la serie X_t , b) correlaciones cruzadas entre X_t y la serie Y_t retrasada, c) correlaciones cruzadas entre Y_t y la serie X_t retrasada y d) ACF de la serie Y_t .

Aplicando a la serie bivariada (X_t, Y_t) la estadística $Q_k(m)$ descrita en la Fórmula (6.10) del test Portmanteau se obtiene la Tabla 6.2. Observe que en este caso $k = 2$ ya que es una serie bivariada y m es el retraso hasta el que se toma en cuenta para calcular Q . La hipótesis nula del test es $H_0 : \rho_1 = \dots = \rho_m = \mathbf{0}$ y la alternativa es $H_a : \rho_i \neq \mathbf{0}$ para algún $i \in \{1, \dots, m\}$.

m	$Q_2(m)$	Grados de libertad	valor p
5	34.43226	20	2.3345e-02
10	66.95955	40	4.7808e-03
15	93.72282	60	3.4961e-03
20	136.86800	80	7.8541e-05
25	157.95994	100	1.9691e-04
30	180.50912	120	2.9653e-04

Tabla 6.2: Valores de la estadística $Q_2(m)$, grados de libertad y valores p del test Portmanteau aplicado a la serie bivariada (X_t, Y_t) con m tomando valores de 5,10,15,20,25 y 30.

Por lo tanto, ya que los valores p son menores al nivel de confianza $\alpha = 0.05$ se rechaza la hipótesis nula por lo que se confirma la existencia de dependencia serial en la serie bivariada. Ya que existe una relación de dependencia y retroalimentación entre las series X_t y Y_t , es necesario construir un modelo multivariado para capturar las relaciones existentes entre dichas series.

La elección del modelo VAR se realizará con la función VARselect del software estadístico R, la cuál enlista los valores de los test AIC, HQ y SC obtenidos de 10 modelos, ver Pfaff (2008). En la Tabla 6.3 se encuentran estos valores.

	VAR(1)	VAR(2)	VAR(3)	VAR(4)	VAR(5)	VAR(6)	VAR(7)	VAR(8)	VAR(9)	VAR(10)
AIC	-3.06	-3.058	-3.057	-3.057	-3.056	-3.054	-3.056	-3.054	-3.053	-3.054
HQ	-3.057	-3.051	-3.047	-3.044	-3.039	-3.034	-3.032	-3.027	-3.023	-3.02
SC	-3.051	-3.039	-3.029	-3.02	-3.01	-2.998	-2.991	-2.98	-2.97	-2.961

Tabla 6.3: Valores de AIC, HQ y SC aplicados a los modelos VAR(1) hasta VAR(10).

Ya que los tres valores de los estadísticos del modelo VAR(1) son los menores de entre los estadísticos de los diez modelos, ese se elige. Con la función VAR se estiman los parámetros del modelo, ver Pfaff (2008). El modelo ajustado se encuentra en la Ecuación 6.25.

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} -0.0247 & -0.0146 \\ (0.0267) & (0.037) \\ -0.0462 & -0.0026 \\ (0.0192) & (0.0267) \end{bmatrix} \begin{bmatrix} X_{(t-1)} \\ Y_{(t-1)} \end{bmatrix} + \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} \quad (6.25)$$

La validación del ajuste se realizó aplicando el test Portmanteau de la Fórmula (6.10) a los residuales para revisar si existe dependencia serial entre ellos. El valor-p del contraste es 0.1483. Con $\alpha = 0.05$ se tiene que $p > \alpha$ por lo que se puede concluir que no existe evidencia de dependencia entre los residuales. Para verificar la normalidad se aplicó el test de Jarque Bera. En este caso el valor-p del contraste es menor a $2.2e^{-16}$. Si $\alpha = 0.05$ p es menor que α por lo que existe evidencia de que los residuales no se distribuyen como una Normal estándar. Por último, una vez que ya se tiene un modelo que describa una serie de tiempo dada es posible hacer predicciones. Sin embargo ese tema no se tratará en esta tesis.

Capítulo 7

Modelo MVAR

7.1. Introducción

Es de gran interés para el ámbito financiero y científico modelar el comportamiento variante en el tiempo de los mercados financieros ya que las ganancias y pérdidas de los inversionistas dependen de las alzas y bajas de los precios de las acciones de las compañías. El modelo ARIMA de Box & Jenkins ha sido utilizado ampliamente para pronosticar el comportamiento de una serie de tiempo y en particular de una acción porque ofrece una buena aproximación. Sin embargo, a pesar del respaldo teórico del modelo ARIMA, los analistas continúan buscando nuevas maneras de obtener mejores pronósticos, y es que un supuesto común al modelar series es el de que las distribuciones marginales o condicionales son unimodales pero en la realidad muchas de ellas exhiben multimodalidad. Es por esta razón que se han construido modelos resultado de combinar los modelos ARIMA con los modelos de mezclas que caracterizan la multimodalidad. Ejemplos de esto son el modelo de mezclas con componentes gaussianos (GMTD) Le y cols. (1996), el modelo de mezclas con componentes autorregresivos (MAR) Wong y Li (2000) y el modelo de mezclas con componentes vectoriales gaussianos y autorregresivos (GMVAR) Kalliovirta, Meitz, y Saikkonen (2014).

El objetivo de este capítulo es implementar un modelo multivariado de series de tiempo para el estudio de las acciones incluidas por ejemplo en un portafolio de inversiones. El modelo MVAR (Mixture Vector Autoregressive Model) propuesto por P. W. Fong y cols. (2007) consiste en una mezcla de componentes gaussianos vectoriales autorregresivos. Un ejemplo de su aplicación se encuentra en T. P. Fong y Wong (2008) donde es utilizado para modelar variables macroeconómicas en diferentes situaciones de mercado ya que es capaz de capturar la dinámica de su evolución en el tiempo.

El siguiente capítulo está estructurado de la siguiente manera. La primera sección contiene definiciones y aspectos básicos de los modelos de mezclas. La segunda sección se refiere al modelo MAR que es la versión del modelo MVAR para series de tiempo univariadas. Finalmente la tercera sección se refiere al modelo MVAR.

7.2. Modelos de mezclas

Sea X_1, \dots, X_n una muestra aleatoria donde X_i es un vector aleatorio de dimensión p con función de densidad $f(x_i)$. Se dice que $f(x_i)$ es una *mezcla finita* si es de la forma

$$f(x_i) = \alpha_1 f_1(x_i) + \dots + \alpha_k f_k(x_i) \quad (7.1)$$

donde α_i con $i = 1, \dots, k$, son valores positivos que suman uno y $f_i(x_i)$ son funciones de densidad. A los parámetros α 's se les llama *pesos de la mezcla* y a las funciones $f_i(\cdot)$ se les llama *densidades de los componentes de la mezcla*. No se requiere que las densidades de los componentes pertenezcan a la misma familia paramétrica pero en la mayoría de las aplicaciones así se hace, ver Titterington, Smith, y Makov (1985).

La interpretación del modelo de mezclas es la siguiente. Sea Z_j una variable aleatoria categórica que toma valores de 1 hasta k con probabilidades α_1 hasta α_k respectivamente, es decir

$$P[Z_j = i] = \alpha_i$$

donde $i = 1, \dots, k$ y que

$$P[X_j | Z_j = i] = f_i(x_j)$$

con $i = 1, \dots, k$. Entonces la función de densidad marginal de X_j está dada por $f(x_j)$, es decir, la variable Z_j es la etiqueta del componente del vector X_j . Observe que es más conveniente trabajar con una variable Z_j de dimensión k en lugar de trabajar con una variable categórica. Sea Z_{ij} el i -ésimo elemento de Z_j . Entonces Z_{ij} vale uno si el componente i es el componente de origen de X_j en la mezcla, y vale cero en el caso contrario. Por lo tanto Z_j se distribuye Multinomial con k categorías y probabilidades $\alpha_1, \dots, \alpha_k$, (McLachlan y Peel, 2004, pág. 7).

7.2.1. Aplicaciones de los modelos de mezclas

Se ha observado que los modelos de mezclas se usan de dos maneras. Una de ellas es cuando existe conocimiento o se tiene una creencia de que la unidad experimental de dónde se obtuvo la muestra X pertenece a una de entre

k categorías existentes. En este caso el investigador no observa directamente la fuente de X . La función $f_j(\cdot)$ en la Ecuación (7.1) resume la función de densidad de X dado que la observación realmente proviene de la categoría j y α_j es la probabilidad de que la observación provenga de dicha fuente. La otra forma en la que se aplica un modelo de mezclas es cuando se utiliza como una herramienta matemática para poder realizar un análisis más flexible. Por ejemplo, una mezcla de dos componentes con distribución Normal donde uno de ellos tiene una varianza grande puede usarse para aproximar una distribución con colas pesadas.

7.2.2. Número de componentes

En varios contextos existe incertidumbre respecto al número de componentes y esto genera problemas estadísticos relacionados con el análisis de clusters, donde es necesario hacer suposiciones sobre la estructura paramétrica del modelo, como por ejemplo la suposición de que las densidades son normales, ya sean univariadas o multivariadas. Dadas las suposiciones paramétricas acerca de los componentes el problema se convierte en uno de ajuste de curvas. En algunas aplicaciones es conveniente encontrar la mezcla con menos componentes que proporcione un ajuste satisfactorio. En particular, en principio deseamos saber si es adecuado considerar dos componentes o solo uno. Dado un modelo de mezclas en concreto, es decir, con suposiciones acerca de las formas paramétricas de las funciones de densidad de los componentes, una mezcla con pocos componentes puede considerarse como la hipótesis nula en el marco del modelo original.

7.2.3. Estimación de los parámetros

Suponiendo que se conoce el número de componentes en la mezcla, incluso sólo como paso provisional, se tendrán que hacer inferencias referentes a los parámetros desconocidos del modelo de mezclas. En algunos casos es posible consultar estudios de las distribuciones individuales de los componentes. Por ejemplo, las distribuciones del tamaño de los cristales de varios minerales pueden establecerse individualmente en el laboratorio como parte de un análisis preliminar al estudio de un depósito de ceniza compuesto de una mezcla de minerales (Titterington y cols., 1985, pág. 8). Con ésto ya se conocerían las densidades de los componentes y sólo faltaría estimar los pesos de la mezcla. En otras aplicaciones podrían desconocerse los parámetros de las densidades de los componentes y conocerse los pesos de la mezcla. Por ejemplo, en estudios de poblaciones de peces es posible conocer la distribución del sexo en el banco de peces pero no es práctico revisar el sexo de cada pez, así que no se tiene un conocimiento detallado de las densidades de los componentes. Haciendo una suposición respecto a la distribución de cada componente el ejemplo es una situación en la que se conocen los pesos de la mezcla pero se desconocen los parámetros de los componentes. Además puede ocurrir que se desconozcan ambos, los parámetros de los componentes y los pesos. Probablemente se da ésta situación en la mayoría de las

aplicaciones. En cualquiera de los casos, ya que se han definido las distribuciones de los componentes, se puede utilizar alguno de los métodos de estimación de parámetros: método de momentos, de máxima verosimilitud, mínimos cuadrados o métodos Bayesianos.

7.3. Modelo MAR

Sea Y_t una serie de tiempo univariada. El modelo $MAR(K; p_1, \dots, p_K)$ propuesto en Wong y Li (2000) se define como

$$F(y_t | y_1, y_2, \dots, y_{t-1}) = \sum_{k=1}^K \alpha_k \Phi\left(\frac{y_t - \theta_{k0} - \theta_{k1}y_{t-1} - \theta_{k2}y_{t-2} - \dots - \theta_{kp_k}y_{t-p_k}}{\sigma_k}\right) \quad (7.2)$$

donde p_k es el orden AR para el k -ésimo componente, $\Phi(\cdot)$ es la función de distribución acumulada gaussiana estándar, $\theta_{k0}, \theta_{k1}, \theta_{k2}, \dots, \theta_{kp_k}$ son los coeficientes AR del k -ésimo componente y las α 's son los pesos de los componentes de la mezcla en donde $0 \leq \alpha_i \leq 1$ y $\sum \alpha_i = 1$.

Como se puede observar el modelo MAR es una mezcla donde sus componentes son modelos AR gaussianos. Ya que la esperanza condicional de los componentes depende de los valores pasados, la forma de las distribuciones condicionales de la serie cambia con el tiempo. Por ejemplo, las distribuciones condicionales podrían cambiar de unimodales a multimodales. Ya que la distribución condicional podría ser multimodal, la esperanza condicional de la serie

$$E(y_t | y_1, y_2, \dots, y_{t-1}) = \sum_{k=1}^K \alpha_k (\theta_{k0} + \theta_{k1}y_{t-1} + \dots + \theta_{kp_k}y_{t-p_k}) = \sum_{k=1}^K \alpha_k \mu_{kt}$$

no podría ser el mejor predictor de los valores futuros. El mérito del modelo MAR radica en su habilidad para describir la distribución condicional de la serie. Otra característica importante del modelo es que logra describir la varianza condicional que también cambia con el tiempo, ya que depende de las esperanzas condicionales de los componentes. Dicha varianza condicional es

$$Var(y_t | y_1, y_2, \dots, y_{t-1}) = \sum_{k=1}^K \alpha_k \sigma_k^2 + \sum_{k=1}^K \alpha_k \mu_{kt}^2 - \left(\sum_{k=1}^K \alpha_k \mu_{kt}\right)^2.$$

La resta en el lado derecho de la expresión anterior es positivo y es cero solo cuando las medias μ_{it} de todos los componentes son iguales. La varianza condicional es grande cuando estas medias difieren mucho. En este caso la distribución condicional podría ser multimodal. La varianza condicional es más pequeña cuando las μ_{it} son todas iguales.

La validación del modelo se puede llevar a cabo con los residuales normalizados

$$e_t = \frac{y_t - E[y_t | y_1, y_2, \dots, y_{t-1}]}{\sqrt{\text{Var}(y_t | y_1, y_2, \dots, y_{t-1})}}. \quad (7.3)$$

Las autocorrelaciones de éstos y de sus cuadrados se pueden utilizar para comprobar si los modelos ajustados son adecuados, y su gráfico de normalidad podría usarse para verificar que su distribución condicional es la Normal. Sin embargo, hay un inconveniente para este enfoque. Como ya se mencionó, la esperanza condicional de la serie puede no ser el mejor predictor cuando las distribuciones condicionales son multimodales. Por lo tanto, los residuos definidos en el sentido usual, es decir como en la Fórmula (7.3), pueden no ser apropiados para los modelos MAR. Sin embargo éstos podrían dar una indicación aproximada de la adecuación del modelo.

7.4. Modelo MVAR

7.4.1. Supuestos del modelo

El modelo más general para estudiar una serie de tiempo $\{Y_1, \dots, Y_T\}$ es la función de distribución conjunta

$$F(Y_1, \dots, Y_T; \theta) \quad (7.4)$$

donde θ es el vector de parámetros que determinan de manera única a la función de distribución $F(\cdot)$. La distribución de probabilidad $F(\cdot)$ modela el comportamiento estocástico de la serie. El análisis empírico de la serie consiste en estimar los parámetros en θ y realizar inferencias estadísticas acerca del comportamiento de Y_t dado su pasado. El Modelo (7.4) es muy general pero provee un esbozo general con respecto a lo que debería considerar un modelo para series de tiempo. Una manera de facilitar el estudio o el cálculo de la función de distribución conjunta de la serie es escribirla mediante la forma

$$F(Y_1, \dots, Y_T) = F(Y_1)F(Y_2|Y_1) \cdots F(Y_T|Y_{T-1}, \dots, Y_1) \quad (7.5)$$

donde θ se omitió por simplicidad. La Expresión (7.5) resalta la necesidad de conocer las distribuciones condicionales $F(Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_1)$ para t igual a $1, 2, \dots, T$. Por ejemplo, en finanzas, diferentes especificaciones distribucionales dan lugar a teorías distintas. Así, una versión de la hipótesis de caminata aleatoria es que la distribución condicional es igual a la distribución marginal. En este caso las variables de la serie son temporalmente independientes y por lo tanto no predecibles. En finanzas generalmente se considera que las variables aleatorias que conforman las series de tiempo en estudio son continuas, especialmente para los rendimientos de las acciones.

Por otro lado, una serie de tiempo es lineal si es una realización de un proceso lineal (ver Definición 2.9 en la Sección 2.3), es decir si se puede representar mediante la forma

$$X_t = \sum_{i=0}^{\infty} \Psi_i Z_{t-i} \quad (7.6)$$

para toda $t = 0, \pm 1, \dots$, donde Ψ_i es una sucesión de números reales tal que $\sum_{i=-\infty}^{\infty} |\Psi_i|$ converge y $\{Z_t\}$ es un proceso de ruido blanco. Al proceso $\{Z_t\}$ se le conoce como innovación o shock. En el análisis de series de tiempo lineales es común la suposición de que el proceso innovación tiene distribución gaussiana y por lo tanto su distribución es unimodal. Bajo dicha suposición ambas distribuciones, la marginal y la condicional, también son gaussianas. Si solo se tiene la suposición de que la distribución del proceso innovación es unimodal entonces la distribución marginal de la serie y/o la condicional también es unimodal. En el caso de que la distribución condicional sea gaussiana entonces la estacionaridad débil implica la estacionaridad fuerte haciendo equivalentes los dos conceptos. Sin embargo en la vida real muchas series muestran características que violan la suposición de normalidad y por lo tanto la de unimodalidad. Por ejemplo, se ha demostrado que la serie conformada por el número de línices canadienses atrapados en el río Mackenzie en el periodo de 1821 a 1934 tiene distribución marginal bimodal, ver Tong (1990). En el ámbito financiero, cuando el mercado de acciones llega a ser volátil en ciertos periodos se espera que sus precios aumenten o disminuyan bruscamente, por lo que en este caso una distribución condicional bimodal para los precios de las acciones parece conveniente. Por lo tanto un modelo no lineal es un candidato natural para describir el comportamiento de la serie.

Algunos modelos no lineales para series de tiempo univariadas capaces de modelar la multimodalidad de la distribución marginal o condicional son los siguientes. El modelo SETAR (zeroth-order self-exciting threshold autorregresivo) tiene distribuciones marginales que son una mezcla de distribuciones gaussianas y por lo tanto puede exhibir multimodalidad, aunque esto es difícil de probar matemáticamente, ver Tong (1990). Otro modelo no lineal de interés es el MATS (modelo multipredictor autorregresivo para series de tiempo) que identifica explícitamente las modas y antimodas en la distribución condicional. En este caso el modelo ajustado a los datos de los línices canadienses resultó con muchos parámetros no significativos, de manera que las distribuciones condicionales resultaron unimodales, ver Martin (1992). El modelo GMTD (gaussian mixture transition distribution model) captura características tales como periodos en los que la serie permanece constante, o periodos con mucha volatilidad o outliers (valores que sobresalen de los demás). Aunque dicho modelo es capaz de capturar algunas características no lineales falla en modelar series de tiempo con ciclos, ver Le y cols. (1996).

Wong y Li (2000) generalizan el modelo GMTD con el modelo MAR (modelo de mezclas con componentes autorregresivos) para series univariadas y P. W. Fong y cols. (2007) generalizan este último modelo para series

multivariadas proponiendo el modelo MVAR. Ambos modelos consisten en una mezcla de K componentes gaussianos con las siguientes propiedades:

- No es necesario que todos los componentes de la mezcla sean estacionarios para que la mezcla sea estacionaria.
- Las distribuciones condicionales $F(Y_t|Y_{t-1}, \dots, Y_1)$ cambian a lo largo del tiempo bajo este modelo y podrían ser multimodales, al igual que las distribuciones marginales.
- Como consecuencia de la forma cambiante de las distribuciones condicionales, estos modelos pueden capturar la heteroscedasticidad condicional, la cual es común en series de tiempo financieras.

A continuación se enlistan los supuestos de los modelos MAR y MVAR.

- Las series de tiempo que describen los modelos son series univariadas y multivariadas en tiempo discreto conformadas por variables aleatorias continuas. Además son series débilmente estacionarias (Definición 2.5 en la Sección 2.1), o que ya han sido transformadas para lograr su estacionaridad.
- Los comportamientos de las series de tiempo que pueden explicarse mediante estos modelos son los estallidos, outliers, periodos constantes o con alta volatilidad y ciclos. También se incluyen las series con distribuciones marginales o condicionales multimodales.
- En el caso de series de tiempo multivariadas se requiere que exista alguna relación de dependencia o retroalimentación entre los componentes de las series.
- Ningún parámetro del modelo cambia con el tiempo.

7.4.2. Definición

Sea Y_t una serie de tiempo vectorial n -dimensional. El $MVAR(n, K; p_1, \dots, p_K)$ se define como

$$F(y_t|y_1, y_2, \dots, y_{t-1}) = \sum_{k=1}^K \alpha_k \Phi(\Omega_k^{-\frac{1}{2}}(Y_t - \Theta_{k0} - \Theta_{k1}Y_{t-1} - \Theta_{k2}Y_{t-2} - \dots - \Theta_{kp_k}Y_{t-p_k})) \quad (7.7)$$

donde p_k es el orden VAR para el k -ésimo componente, $\Phi(\cdot)$ es la función de distribución multivariada acumulada gaussiana estándar, Θ_{k0} es un vector de coeficientes n -dimensional, $\Theta_{k1}, \Theta_{k2}, \dots, \Theta_{kp_k}$ son matrices de coeficientes de tamaño $n \times n$ para el k -componente, Ω_k es la matriz de varianzas y covarianzas para el k -ésimo componente y las α 's son los pesos de los componentes de la mezcla en donde $0 \leq \alpha_i \leq 1$ y $\sum \alpha_i = 1$.

Entonces el modelo MVAR de la Ecuación (7.7) es una mezcla con K componentes, donde las α 's, que son sus pesos, son parámetros desconocidos. Las distribuciones de los componentes de la mezcla son Normales estándar multivariadas aplicadas al modelo VAR

$$\Omega_k^{-\frac{1}{2}} (Y_t - \Theta_{k0} - \Theta_{k1}Y_{t-1} - \Theta_{k2}Y_{t-2} - \dots - \Theta_{kp_k}Y_{t-p_k}).$$

Dada una serie de tiempo vectorial Y_t n -dimensional no se conocen el número K de componentes de la mezcla ni tampoco los órdenes de los modelos VAR en cada componente. En la Sección 7.4.5 se explica cómo se eligen éstos números y en la Sección 7.4.4 se explica el método de máxima verosimilitud mediante el cual se calcularon los pesos de la mezcla y los parámetros de los modelos VAR.

7.4.3. Condiciones de estacionaridad

La estacionaridad de primer orden se establece en el Teorema 7.2 y el Teorema 7.1 es un caso especial del Teorema 7.2 para p igual a uno. Los Teoremas 7.3 y 7.4 establecen las condiciones de estacionaridad de segundo orden para los modelos $MVAR(n, K; 1, \dots, 1)$ y $MVAR(n, K; 2, \dots, 2)$ respectivamente. Estos resultados implican que no es necesario que todos los componentes de la mezcla sean estacionarios para que el modelo sea estacionario. Para órdenes VAR mayores a dos la condición requiere retrasos mayores en la autocovarianza, lo cual es complicado aún en el caso univariado. El Teorema 7.5 establece la condición de estacionaridad de segundo orden para el $MVAR(n, K; p, \dots, p)$. Las demostraciones de éstos teoremas se pueden revisar en P. W. Fong y cols. (2007).

Teorema 7.1. *El modelo $MVAR(n, K; 1, \dots, 1)$ es estacionario en media si y sólo si todos los eigenvalores de la matriz $\sum_{k=1}^K \alpha_k \Theta_{k1}$ tienen módulo menor a uno.*

Teorema 7.2. *El modelo $MVAR(n, K; p, \dots, p)$ es estacionario respecto a su media si y sólo si todas las raíces de la ecuación $|I_n - \lambda \Theta_1^\alpha - \lambda^2 \Theta_2^\alpha - \dots - \lambda^p \Theta_p^\alpha| = 0$ tienen módulo menor a uno, donde $\Theta_i^\alpha = (\alpha_1 \Theta_{1i} + \dots + \alpha_K \Theta_{Ki})$ para $i = 0, 1, \dots, p$.*

Teorema 7.3. *Supongamos que el proceso Y_t se ajusta al modelo $MVAR(n, K; 1, \dots, 1)$ y que además satisface la condición de estacionaridad de primer orden. El proceso es estacionario de segundo orden si y sólo si todos los eigenvalores de la matriz $\sum_{k=1}^K \alpha_k (\Theta_{k1} \otimes \Theta_{k1})$ tienen módulo menor a uno donde \otimes denota el producto de Kronecker.*

Teorema 7.4. *Supongamos que el proceso Y_t se ajusta al modelo $MVAR(n, K; 2, \dots, 2)$ y que además satisface la condición de estacionaridad de primer orden. El proceso es estacionario de segundo orden si y sólo si $|I_n - \lambda \beta_1 - \lambda^2 \beta_2| = 0$ tiene raíces menores a uno en módulo donde $\beta_1 = \sum_{k=1}^K \alpha_k M_k$, $\beta_2 = \sum_{k=1}^K \alpha_k (\Theta_{k2} \otimes$*

Θ_{k2}), $M_k = (\Theta_{k1} \otimes \Theta_{k1}) + (\Theta_{k2} \otimes \Theta_{k1})N_1 + (\Theta_{k1} \otimes \Theta_{k2})N_2$, $N_1 = (I - \Theta_2^\alpha \otimes \Theta_2^\alpha)^{-1}((I \otimes \Theta_1^\alpha) + (\Theta_1^\alpha \otimes \Theta_2^\alpha))$
y $N_2 = (I - \Theta_2^\alpha \otimes \Theta_2^\alpha)^{-1}((\Theta_1^\alpha \otimes I) + (\Theta_2^\alpha \otimes \Theta_1^\alpha))$.

Teorema 7.5. *Supongamos que el proceso Y_t se ajusta al modelo $MVAR(n, K; p, \dots, p)$ y que además satisface la condición de estacionaridad de primer orden. El proceso es estacionario de segundo orden si y sólo si $(1 - C_1 z^1 - \dots - C_p z^p) = 0$ tiene raíces menores a uno en módulo donde*

$$\begin{aligned} C_u &= \sum_{k=1}^K \alpha_k (\theta_{ku}^2 - \sum_{v=1}^{p-1} \sum_{|i-j=v|} \theta_{ki} \theta_{kj} b_{vu} \beta_{u,0}) \\ C_p &= \sum_{k=1}^K \alpha_k \theta_{kp}^2, \\ \beta_{l,0} &= \sum_{k=1}^K \alpha_k \theta_{kl}, \\ \beta_{l,u} &= \sum_{k=1}^K \alpha_k \sum_{|i-l=u|} \theta_{ki}, \\ \beta_{l,l} &= \left(\sum_{k=1}^K \alpha_k \sum_{|i-l=l|} \theta_{ki} \right) - 1 \end{aligned}$$

para $u, l = 1, \dots, p-1$ y B, B^{-1} son matrices de $(p-1) \times (p-1)$ tales que $B = (\beta_{i,j})_{i,j=1}^{p-1}$ y $B^{-1} = (b_{i,j})_{i,j=1}^{p-1}$.

7.4.4. Estimación de los parámetros

La estimación de los parámetros en esta tesis se llevó a cabo mediante el método de máxima verosimilitud que consiste en encontrar el máximo de la función de log verosimilitud

$$\ln(\mathcal{L}) = \sum_{i=1}^t \ln(f(y_i | \theta))$$

donde f es la función de densidad que se obtiene derivando la función de distribución (7.7). Si $y_t = (x_1, \dots, x_n)$ entonces

$$\begin{aligned}
 f(y_t | y_1, y_2, \dots, y_{t-1}) &= \frac{\partial^n}{\partial x_n \dots \partial x_{n-1}} F(y_t | y_1, y_2, \dots, y_{t-1}) \\
 &= \alpha_1 (2\pi)^{-1} |\Omega_1|^{-\frac{1}{2}} \\
 &\quad \exp\left[-\frac{1}{2} (Y_t - \Theta_{10} - \Theta_{11}Y_{t-1} - \dots - \Theta_{1p_1}Y_{t-p_1})^T \Omega_1^{-1} \right. \\
 &\quad \left. (Y_t - \Theta_{10} - \Theta_{11}Y_{t-1} - \dots - \Theta_{1p_1}Y_{t-p_1})\right] + \\
 &\quad \vdots \\
 &\quad \alpha_K (2\pi)^{-1} |\Omega_K|^{-\frac{1}{2}} \\
 &\quad \exp\left[-\frac{1}{2} (Y_t - \Theta_{K0} - \Theta_{K1}Y_{t-1} - \dots - \Theta_{Kp_K}Y_{t-p_K})^T \Omega_K^{-1} \right. \\
 &\quad \left. (Y_t - \Theta_{K0} - \Theta_{K1}Y_{t-1} - \dots - \Theta_{Kp_K}Y_{t-p_K})\right].
 \end{aligned}$$

La maximización de la función $\ln(\mathcal{L})$ se realizó mediante la función `nlm` del software estadístico R que usa un algoritmo de tipo Newton.

7.4.5. Elección del modelo

El BIC (Bayesian Information Criterion) se define como

$$BIC = -2 \ln \mathcal{L} + r \ln(s)$$

donde $\ln(\mathcal{L})$ es la función de log verosimilitud, r es el número de parámetros que se estiman en el modelo y s es el tamaño de la muestra. Para el modelo $MVAR(n, k; p_1, \dots, p_k)$ tenemos que

$$BIC = -2 \ln \mathcal{L} + \ln(T - p_{max}) \left\{ n^2 \sum_{i=1}^K p_i + K \left(\frac{n(1+n)}{2} + n + 1 \right) - 1 \right\} \quad (7.8)$$

donde p_{max} es el máximo permitible entre los ordenes AR de los componentes de la mezcla, n es el número de entradas de los vectores que conforman la serie de tiempo en estudio, T es el tamaño de la serie de tiempo Y_t y K es el número de componentes en la mezcla. Observe que $s = T - p_{max}$ ya que las series de tiempo $Y_t, Y_{t-1}, \dots, Y_{t-p_k}$ utilizadas en el modelo son el resultado de retrasar una serie de tiempo p_{max} momentos.

Además

$$r = n^2 \sum_{i=1}^K p_i + K \left(\frac{n(1+n)}{2} + n + 1 \right) - 1$$

ya que $n^2 \sum_{i=1}^K p_i$ es el total de entradas que tienen las matrices Θ_{ip_i} en toda la mezcla puesto que n^2 es el número de entradas en cada una de estas matrices y $\sum_{i=1}^K p_i$ es el total de matrices Θ_{ip_i} en cada componente.

Además $K \left(\frac{n(1+n)}{2} + n + 1 \right)$ es el total de parámetros contenidos en los vectores n -dimensionales Θ_{i0} , en las matrices Ω_i y los escalares α_i en toda la mezcla, esto es, las n entradas de un vector Θ_{i0} más las $\frac{n(1+n)}{2}$ entradas en el triángulo inferior de la matriz Ω_i más uno del escalar α y ésta suma multiplicada por K que es el total de componentes en la mezcla ya que en cada uno hay un vector Θ_{i0} , una matriz Ω_i y un escalar α . Para finalizar, a este último total se le resta uno ya que aunque aparecen K α 's en el modelo, como su suma es uno, en realidad sólo se estiman $K - 1$ y la última se obtiene a partir de las demás.

Al principio del proceso de elección del modelo no se conocen ni el número de componentes ni el orden AR de cada uno de ellos, por lo que primero se fijó un orden AR alto para cada componente variando el número de componentes y luego se eligió el modelo con el BIC más pequeño. Ya teniendo el número de componentes adecuado se encontró el orden AR para cada uno de ellos combinando los órdenes de 1 hasta el orden AR más alto elegido en el primer paso. Finalmente se calculó el BIC a cada uno de estos modelos y se eligió el que fuera menor.

7.4.6. Validación del modelo

Como se mencionó en la Sección 7.3, no es conveniente realizar la validación de los modelos MAR, ni de los modelos MVAR, mediante los residuales normalizados calculados con la Fórmula (7.3). Esto se debe a que la esperanza condicional no es un buen predictor cuando las distribuciones condicionales son multimodales. Por ejemplo, en el caso de tratarse de una distribución bimodal el modelo MVAR tendría dos componentes pero la variable observada y_t pertenece en realidad a sólo uno de los componentes, y la media condicional, que es un promedio ponderado de las medias de los dos componentes distintos, podría ubicarse en una región de baja probabilidad. Por lo tanto, la media condicional podría ser engañosa en la predicción.

Por lo tanto, para analizar el rendimiento del modelo MVAR más claramente, se sugiere que el cálculo de los residuos del modelo MVAR tenga en cuenta los valores esperados condicionales finales de las variables $Z_{t,i}$:

$$\tau_{t,k} = \frac{\alpha_k |\Omega_k|^{-\frac{1}{2}} \exp(\xi)}{\sum_{k=1}^K \alpha_k |\Omega_k|^{-\frac{1}{2}} \exp(\xi)} \quad (7.9)$$

donde $\xi = -\frac{1}{2}(Y_t - \Theta_{10} - \Theta_{11}Y_{t-1} - \dots - \Theta_{kp_k}Y_{t-p_k})^T \Omega_k^{-1} (Y_t - \Theta_{10} - \Theta_{11}Y_{t-1} - \dots - \Theta_{kp_k}Y_{t-p_k})$, $k = 1, \dots, K$, y si Y_1, \dots, Y_T son T vectores de dimensión n generados a partir del modelo $MVAR(n, K; p_1, \dots, p_K)$,

$$Z_{t,i} = \begin{cases} 1 & \text{si } Y_t \text{ viene del componente } i, 1 \leq i \leq K, \\ 0 & \text{en otro caso.} \end{cases} \quad (7.10)$$

Se considera que la variable observada en el tiempo t va al componente k si $\tau_{t,k}$ es el mayor de entre $\tau_{t,1}, \dots, \tau_{t,K}$. En el caso de dos componentes se considera que el vector observado Y_t va al primer componente si $\tau_{t,1}$ es mayor

que $\tau_{t,2}$. Después de agrupar las observaciones en componentes, el valor esperado de la variable en el tiempo t se obtiene mediante

$$E[y_t|y_1, y_2, \dots, y_{t-1}, \max(\tau_{t,i}, i = 1, \dots, K) = \tau_{t,k}] = \Theta_{k0} + \Theta_{k1}Y_{t-1} + \dots + \Theta_{kp_k}Y_{t-p_k}. \quad (7.11)$$

El residual de dimensión n $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})$ dado las variables $Z_{t,i}$ se calcula de la siguiente manera:

$$\varepsilon_t = Y_t - E[Y_t|y_1, y_2, \dots, y_{t-1}, \max(\tau_{t,i}, i = 1, \dots, K) = \tau_{t,k}]. \quad (7.12)$$

Para $t = 1, \dots, T$, los residuales $e_t = (e_{1t}, \dots, e_{nt})$ para cada componente se estandarizan por

$$e_{it} = \frac{\varepsilon_{it}}{(\text{Var}[y_{it}|y_1, y_2, \dots, y_{t-1}, \max(\tau_{t,i}, i = 1, \dots, K) = \tau_{t,k}])^{\frac{1}{2}}} \quad (7.13)$$

y luego se agrupan para formar un conjunto final de residuales.

Finalmente, la verificación del modelo se realiza mediante los Test Portmanteau univariados Q^1 y Q^2 aplicados a los residuales estandarizados y a sus cuadrados respectivamente. Dichas estadísticas son:

$$Q^1(m) = T(T+2) \sum_{i=1}^m (T-i)^{-1} \rho_i^2(e_{it}), \quad (7.14)$$

$$Q^2(m) = T(T+2) \sum_{i=1}^m (T-i)^{-1} \rho_i^2(e_{it}^2) \quad (7.15)$$

donde T es el tamaño de la serie de residuales y $\rho_i(e_{it})$ son las primeras i autocorrelaciones de la serie de los residuales e_{it} . Se usa la distribución $\chi^2(m)$ como referencia, donde m es el número de autocorrelaciones usadas en Q^1 y Q^2 , ver Box y cols. (1994).

Para evaluar el modelo mediante las correlaciones cruzadas se utilizan los siguientes Test Portmanteau. Para $n = 2$,

$$Q^3(m) = T(T+2) \sum_{i=1}^m (T-i)^{-1} \rho(i)^2, \quad (7.16)$$

$$Q^4(m) = T(T+2) \sum_{i=1}^m (T-i)^{-1} \rho(-i)^2 \quad (7.17)$$

donde $\rho(i)$ es la correlación entre $e_{1,t}$ y $e_{2,t+i}$, y $\rho(-i)$ es la correlación entre $e_{1,t}$ y $e_{2,t-i}$. En este caso también se usa la distribución $\chi^2(m)$.

Capítulo 8

Ajuste del MVAR a una serie bivariada

8.1. Introducción

En la actualidad muchas compañías dividen el conjunto de sus bienes, es decir su capital social, en partes iguales para venderlos total o parcialmente y así conseguir financiamiento, recaudar dinero para reinvertirlo o iniciar nuevos proyectos. Una acción es un título emitido por una empresa que representa el valor de una de esas partes iguales en las que fue dividido su capital social.

Las acciones otorgan a sus propietarios, llamados accionistas, derechos económicos ya que los hace beneficiarios de una parte de las ganancias de la compañía llamada dividendo, y derechos políticos ya que cumpliendo ciertos requisitos pueden tener voz y voto en las juntas que establecen sus políticas. La compra-venta de las acciones se lleva a cabo en un mercado financiero llamado Bolsa de Valores donde su precio varía continuamente de acuerdo a la oferta y la demanda. Así, además de los dividendos, el accionista también puede ganar dinero si consigue vender una acción, por ejemplo un viernes en cien pesos pero que compró el lunes en cincuenta pesos. Sin embargo, así como puede ganar si el precio de las acciones sube, puede perder si baja o si la empresa tiene pérdidas. Es por esta razón que para adquirir una acción se requiere tener confianza en el futuro de la compañía y si muchas personas creen que seguirá teniendo ganancias y se interesan por comprar sus acciones entonces su precio subirá.

Tomando en cuenta lo anterior se puede afirmar que modelar el comportamiento variante en el tiempo de los mercados financieros es de gran interés, tanto en el ámbito científico como en el financiero. El modelo ARI-MA de Box & Jenkins ha sido utilizado ampliamente para pronosticar el comportamiento de una acción ya

que ofrece una buena aproximación. Sin embargo, a pesar del respaldo teórico del modelo ARIMA, los analistas continúan buscando nuevas maneras de obtener mejores pronósticos. Ejemplos de ellas son el modelo multipredictor autorregresivo para series de tiempo (MATS) de Martin Martin (1992) que describe los saltos multimodales en series de tiempo, el modelo de mezclas Gaussiano (GMTD) de Le et. al Le y cols. (1996) para modelar estiramientos, estallidos y valores atípicos en series de tiempo, el modelo MAR de Wong & Le Wong y Li (2000) que es una generalización del modelo GMTD cuya distribución predictiva puede ser también multimodal, el modelo MVAR de Fong et. al P. W. Fong y cols. (2007) que consiste en una mezcla de componentes vectoriales Gaussianos autorregresivos, entre otros.

Como ya se mencionó en la Sección 5.1, Microsoft Corporation es una empresa multinacional de origen estadounidense fundada en 1975 dedicada al sector del software y el hardware. Además desarrolla, fabrica, licencia y produce software y equipos electrónicos, siendo sus productos más usados el sistema operativo Microsoft Windows y la suite Microsoft Office.

Por otro lado el Nasdaq 100 es un índice bursátil de Estados Unidos que refleja la evolución de las 100 empresas dedicadas a las telecomunicaciones, a las ventas por mayoreo y menudeo, al hardware y software o a la biotecnología. Dichas empresas cuentan con la mayor capitalización bursátil y cotizan en el mercado electrónico del Nasdaq.

No hay muchos índices negociables que tengan bastante volumen, pero el índice Nasdaq 100 sí lo tiene, lo cual hace posible su compra y venta como si fueran acciones normales de alto volumen. Así, el símbolo QQQ se compra y vende como acciones, siendo uno de los que tienen el mayor volumen de compra y venta.

Un gran porcentaje del índice Nasdaq 100 corresponde a empresas como Intel, Cisco Systems, Microsoft Corporation, Google, Apple Inc., Oracle y Research in Motion (ver Lerman (2001), Israel (2007) y Turner (2005)), por lo que cualquier suceso o actividad ocurrida con sus acciones ejerce gran influencia en él y por lo tanto en sus propias acciones. Debido a esto es común que los inversionistas deseen observar el comportamiento individual y conjunto de las acciones de éstas grandes compañías con las del índice.

La cartera analizada se compone de los precios de las acciones del Índice Nasdaq 100 (serie QQQ) y de Microsoft Corporation (serie MSFT). Cada serie se compone de 2,520 datos tomados diariamente del 13 de diciembre del 2004 al 15 de diciembre del 2014. Las Figuras 5.1 y 5.12 del Capítulo 5 muestran las gráficas de las

series de tiempo correspondientes a los precios de ambas acciones. En ellas no se aprecia periodicidad pero sí cierta tendencia a la alza a medida que el tiempo aumenta, lo cual indica que es necesario aplicar a ambas una diferenciación y así lograr que el nivel medio se estabilice.

8.2. Elección del modelo MVAR

El primer paso encaminado al ajuste del modelo MVAR consiste en el ajuste de un modelo ARMA a cada una de las series ya que esto proporciona alguna idea del mínimo orden AR adecuado para el MVAR. En el Capítulo 5 se afirma que el modelo más adecuado para describir la serie *QQQ* es el $ARIMA(0,1,1)$ y para la serie *MSFT* es el $ARIMA(2,1,2)$. El segundo paso es concatenar las series *QQQ* y *MSFT* para formar una serie bivariada $Y_t = (QQQ, MSFT)$ y graficar su distribución de frecuencias. De esta gráfica se deducirá el número de componentes en la mezcla del MVAR. La Figura 8.1 contiene el histograma de la serie Y_t . En él sobresalen tres barras, por lo que se consideraron a lo más tres componentes en la mezcla.

Por lo tanto, de acuerdo con los ajustes ARIMA, se tomará al 1 como el orden mínimo y al 3 como el orden máximo de los posibles órdenes AR del modelo MVAR. Además la distribución de frecuencias sugiere a lo más 3 componentes en la mezcla. Finalmente, para elegir el número de componentes adecuado se fijó el 3 como máximo orden AR en cada componente y se ajustaron tres modelos: el primero con un solo componente y los dos subsecuentes con un componente adicional a la vez. Se les calculó el BIC y se observó que el más pequeño es 5773.165 correspondiente a la mezcla con tres componentes. Para encontrar el orden AR adecuado en cada componente fue necesario ajustar cada una de las 27 combinaciones posibles de órdenes del 1 al 3 en los tres componentes. En este caso el BIC más pequeño es 5691.451 correspondiente al modelo con orden AR 1 en todos los componentes. Por lo tanto la evidencia sugiere que el mejor modelo es el $MVAR(2,3;1,1,1)$. En la Tabla 8.1 se encuentran los valores obtenidos al calcular el BIC a cada modelo. La parte de arriba de esta Tabla se refiere a la elección del número de componentes y la segunda a los órdenes AR. En la última columna aparece el número de parámetros a estimar en cada modelo. En este conteo se incluyen las entradas de los parámetros que son vectores y matrices.

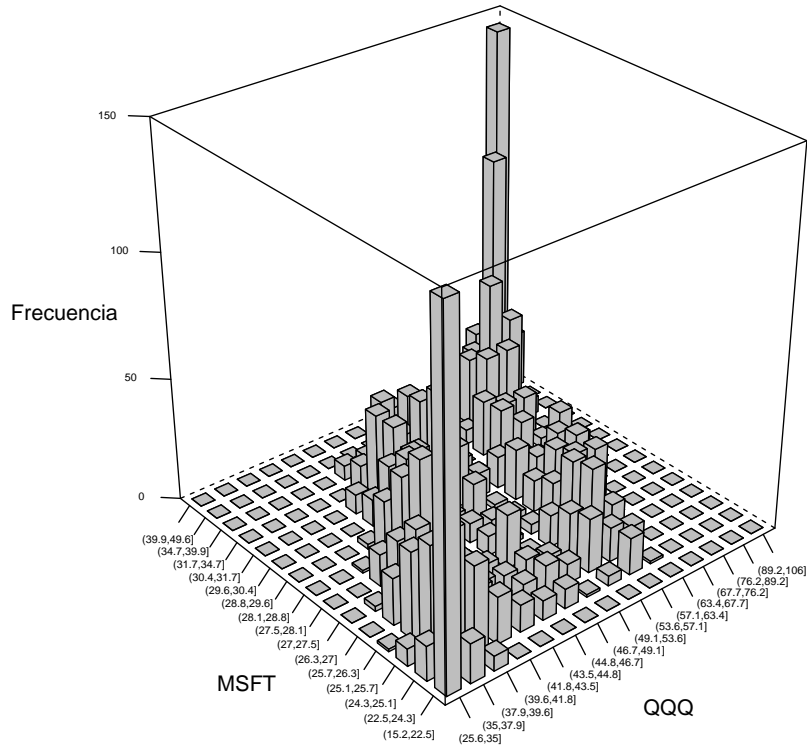


Figura 8.1: Histograma en 3D de la serie bivariada (QQQ,MSFT). En la gráfica los ejes x y y representan los valores de los precios de las acciones de las series QQQ y MSFT respectivamente, mientras que el eje z es la frecuencia en la que aparecen las parejas de la serie bivariada (QQQ,MSFT).

8.3. Eliminación de parámetros no significativos

El modelo elegido es el MVAR(2,3;1,1,1):

$$\begin{aligned}
 F(y_t | y_1, y_2, \dots, y_{t-1}) = & \alpha_1 \Phi[\Omega_1^{-\frac{1}{2}}(Y_t - \Theta_{10} - \Theta_{11}Y_{t-1})] + \\
 & \alpha_2 \Phi[\Omega_2^{-\frac{1}{2}}(Y_t - \Theta_{20} - \Theta_{21}Y_{t-1})] + \\
 & \alpha_3 \Phi[\Omega_3^{-\frac{1}{2}}(Y_t - \Theta_{30} - \Theta_{31}Y_{t-1})].
 \end{aligned} \tag{8.1}$$

No. de componentes	log verosimilitud	BIC	Orden AR	No. de parámetros por estimar
1	-3271.255	6675.614	3	17
2	-2797.293	5868.624	3,3	35
3	-2679.097	5773.165	3,3,3	53
3	-2732.173	5691.451	1,1,1	29
3	-2730.419	5719.242	1,1,2	33
3	-2711.671	5713.039	1,1,3	37
3	-2729.507	5717.418	1,2,1	33
3	-2728.483	5746.692	1,2,2	37
3	-2708.801	5738.616	1,2,3	41
3	-2727.542	5744.781	1,3,1	37
3	-2726.646	5774.307	1,3,2	41
3	-2694.8	5741.934	1,3,3	45
3	-2725.454	5709.313	2,1,1	33
3	-2722.557	5734.839	2,1,2	37
3	-2703.179	5727.372	2,1,3	41
3	-2715.05	5719.826	2,2,1	37
3	-2704.652	5730.351	2,2,2	41
3	-2702.86	5758.053	2,2,3	45
3	-2714.632	5750.279	2,3,1	41
3	-2703.636	5759.605	2,3,2	45
3	-2684.839	5753.33	2,3,3	49
3	-2722.714	5735.124	3,1,1	37
3	-2719.611	5760.237	3,1,2	41
3	-2700.267	5752.867	3,1,3	45
3	-2711.043	5743.1	3,2,1	41
3	-2700.993	5754.319	3,2,2	45
3	-2697.849	5779.35	3,2,3	49
3	-2707.401	5767.136	3,3,1	45
3	-2697.658	5778.968	3,3,2	49
3	-2679.097	5773.165	3,3,3	53

Tabla 8.1: Elección del número de componentes y órdenes AR del modelo MVAR para la serie bivariada (QQQ,MSFT). La primera parte de la tabla contiene los datos de los tres primeros modelos de los cuales se eligió el número de componentes de la mezcla del modelo MVAR. La segunda parte de la tabla contiene los 27 modelos que se ajustaron de entre los cuales se eligieron los órdenes AR de cada componente de la mezcla. La última columna de la tabla indica el número de parámetros que se estimaron contando las entradas de los vectores $\Theta_{i,0}$ y de las matrices $\Theta_{i,1}$, $\Theta_{i,2}$, $\Theta_{i,3}$, Ω_i donde $i = \{1, 2, 3\}$.

Observe que en él hay 12 parámetros: los tres valores univariados de las α 's, los tres vectores bivariados Θ_{10} , Θ_{20} y Θ_{30} , las tres matrices de tamaño 2×2 Θ_{11} , Θ_{21} y Θ_{31} y las tres matrices simétricas de tamaño 2×2 Ω_1 , Ω_2 y Ω_3 . Ya que las α 's suman uno, estimando sólo dos la otra se puede obtener a partir de ellas. Además como las matrices Ω son matrices de varianzas y covarianzas entonces son simétricas por lo que en lugar de estimar sus cuatro entradas solo se estiman tres. De acuerdo a lo anterior se puede decir que el modelo $MVAR(2,3;1,1,1)$ tiene 29 parámetros desconocidos, en donde ya se incluyen las entradas de los vectores y de las matrices. En la Tabla 8.2 se encuentran las estimaciones de sus parámetros.

Primer componente	Segundo componente	Tercer componente
$\alpha_1 = 0.4422$	$\alpha_2 = 0.5133$	$\alpha_3 = 0.0443$
$\Theta_{10} = \begin{pmatrix} 0.1338 \\ (0.0812) \\ 0.1210 \\ (0.0803) \end{pmatrix}$	$\Theta_{20} = \begin{pmatrix} 0.0884 \\ (0.1040) \\ 0.1069 \\ (0.1023) \end{pmatrix}$	$\Theta_{30} = \begin{pmatrix} -0.0095 \\ (0.7267) \\ 0.0428 \\ (0.7191) \end{pmatrix}$
$\Theta_{11} = \begin{pmatrix} 1.0037 & -0.0006 \\ (0.0019) & (0.0011) \\ -0.0087 & 0.9966 \\ (0.0051) & (0.0040) \end{pmatrix}$	$\Theta_{21} = \begin{pmatrix} 1.0045 & 0.0055 \\ (0.0028) & (0.0019) \\ -0.0121 & 0.9867 \\ (0.0070) & (0.0056) \end{pmatrix}$	$\Theta_{31} = \begin{pmatrix} 0.9898 & 0.0076 \\ (0.0167) & (0.0154) \\ 0.0128 & 0.9857 \\ (0.0477) & (0.0461) \end{pmatrix}$
$\Omega_1 = \begin{pmatrix} 0.1070 & 0.0413 \\ 0.0413 & 0.0422 \end{pmatrix}$	$\Omega_2 = \begin{pmatrix} 0.5545 & 0.2649 \\ 0.2649 & 0.2286 \end{pmatrix}$	$\Omega_3 = \begin{pmatrix} 1.3572 & 0.8010 \\ 0.8010 & 1.5300 \end{pmatrix}$

Tabla 8.2: Estimaciones de los parámetros del modelo $MVAR(2,3;1,1,1)$. La primera columna de la tabla contiene las estimaciones de los parámetros del primer componente o sumando en la mezcla de la Ecuación (8.1): el escalar α_1 , el vector bivariado $\Theta_{1,0}$ y las matrices de tamaño 2×2 $\Theta_{1,1}$ y Ω_1 . Los números entre paréntesis debajo de las estimaciones son sus correspondientes errores estándar. La segunda y tercera columnas contienen las estimaciones de los parámetros del segundo y tercer componente respectivamente.

El primer paso para obtener un modelo con todos sus parámetros significativos fue aplicarles la prueba de Wald. Con $\alpha = 0.05$ se encontraron cinco parámetros no significativos: las entradas (2,1) y (1,2) de la matriz $\Theta_{1,1}$, la entrada (2,1) de la matriz $\Theta_{2,1}$ y las entradas (2,1) y (1,2) de la matriz $\Theta_{3,1}$, con valores p 0.0911, 0.5306, 0.0857, 0.7872 y 0.6192 respectivamente. Por lo tanto no se rechaza la hipótesis nula que esos parámetros son cero. El segundo paso fue elegir el mejor modelo de entre cinco modelos nuevos en los que en cada uno se hizo cero un parámetro de los cinco parámetros antes mencionados. Se encontró que el menor BIC vale 5683.732 y corresponde al modelo que tiene la entrada (2,1) del parámetro $\Theta_{3,1}$ igual a cero, por lo que se considera que éste es hasta ahora el mejor modelo. A continuación se repitió el primer paso ahora con éste nuevo modelo. Se

encontraron cuatro parámetros no significativos: las entradas (2,1) y (1,2) de la matriz $\Theta_{1,1}$, la entrada (2,1) de la matriz $\Theta_{2,1}$ y la entrada (1,2) de la matriz $\Theta_{3,1}$. El resultado del segundo paso fue que el mejor modelo de los cuatro nuevos modelos fue el modelo con entradas (2,1) del parámetro $\Theta_{3,1}$ y (2,1) del parámetro $\Theta_{2,1}$ iguales a cero. Después de llevar a cabo este procedimiento dos veces más se obtuvo un modelo lo más reducido posible. La Tabla 8.3 contiene los cálculos del BIC de todos los modelos descritos anteriormente y la Tabla 8.4 muestra los parámetros estimados del modelo restringido final. Un procedimiento similar para reducir un modelo MVAR se puede encontrar en Wong (2011).

En resumen, el modelo elegido de entre todos los modelos considerados en la Tabla 8.1 es el MVAR(2,3;1,1,1) con la siguiente función de densidad:

$$\begin{aligned}
 F(y_t | y_1, y_2, \dots, y_{t-1}) = & \\
 & 0.4422 \Phi \left[\begin{pmatrix} 0.1070 & 0.0413 \\ 0.0413 & 0.0422 \end{pmatrix}^{-\frac{1}{2}} \left(Y_t - \begin{pmatrix} 0.1338 \\ 0.1210 \end{pmatrix} - \begin{pmatrix} 1.0037 & -0.0006 \\ -0.0087 & 0.9966 \end{pmatrix} Y_{t-1} \right) \right] + \\
 & 0.5133 \Phi \left[\begin{pmatrix} 0.5545 & 0.2649 \\ 0.2649 & 0.2286 \end{pmatrix}^{-\frac{1}{2}} \left(Y_t - \begin{pmatrix} 0.0884 \\ 0.1069 \end{pmatrix} - \begin{pmatrix} 1.0045 & 0.0055 \\ -0.0121 & 0.9867 \end{pmatrix} Y_{t-1} \right) \right] + \\
 & 0.0443 \Phi \left[\begin{pmatrix} 1.3572 & 0.8010 \\ 0.8010 & 1.5300 \end{pmatrix}^{-\frac{1}{2}} \left(Y_t - \begin{pmatrix} -0.0095 \\ 0.0428 \end{pmatrix} - \begin{pmatrix} 0.9898 & 0.0076 \\ 0.0128 & 0.9857 \end{pmatrix} Y_{t-1} \right) \right].
 \end{aligned} \tag{8.2}$$

y el modelo MVAR(2,3;1,1,1) restringido tiene la función de densidad siguiente:

$$\begin{aligned}
 F(y_t | y_1, y_2, \dots, y_{t-1}) = & \\
 & 0.4485 \Phi \left[\begin{pmatrix} 0.1091 & 0.0420 \\ 0.0420 & 0.0432 \end{pmatrix}^{-\frac{1}{2}} \left(Y_t - \begin{pmatrix} 0.1807 \\ 0.1664 \end{pmatrix} - \begin{pmatrix} 1.0055 & 0 \\ -0.0133 & 0.9937 \end{pmatrix} Y_{t-1} \right) \right] + \\
 & 0.5067 \Phi \left[\begin{pmatrix} 0.5610 & 0.2707 \\ 0.2707 & 0.2307 \end{pmatrix}^{-\frac{1}{2}} \left(Y_t - \begin{pmatrix} -0.0213 \\ -0.0002 \end{pmatrix} - \begin{pmatrix} 1 & 0.0034 \\ 0 & 0.9943 \end{pmatrix} Y_{t-1} \right) \right] + \\
 & 0.0447 \Phi \left[\begin{pmatrix} 1.3780 & 0.8060 \\ 0.8060 & 1.5460 \end{pmatrix}^{-\frac{1}{2}} \left(Y_t - \begin{pmatrix} 0.0201 \\ 0.0533 \end{pmatrix} - \begin{pmatrix} 0.9966 & 0 \\ 0 & 0.9996 \end{pmatrix} Y_{t-1} \right) \right].
 \end{aligned} \tag{8.3}$$

Observe que en este modelo el primer componente contiene varianzas 0.1091 y 0.0432, además de que su peso en la mezcla es de 0.4485. El hecho de que el valor de la entrada (1,2) de la matriz $\Theta_{1,1}$ es cero indica que

Parámetros no significativos	Componentes	Logvero	BIC	No. parámetros por estimar	No. componentes
Ninguno	—	-2732.173	5691.451	29	3
Entrada (2,1) de $\Theta_{2,1}$	2	-2733.828	5686.93	28	3
Entrada (2,1) de $\Theta_{1,1}$	1	-2733.072	5685.419	28	3
Entrada (1,2) de $\Theta_{1,1}$	1	-2732.418	5684.11	28	3
Entrada (1,2) de $\Theta_{3,1}$	3	-2732.386	5684.046	28	3
Entrada (2,1) de $\Theta_{3,1}$	3	-2732.229	5683.732	28	3
Entradas (2,1) de $\Theta_{3,1}$ y (2,1) de $\Theta_{1,1}$	3 y 1	-2733.746	5678.934	27	3
Entradas (2,1) de $\Theta_{3,1}$ y (1,2) de $\Theta_{1,1}$	3 y 1	-2733.126	5677.695	27	3
Entradas (2,1) de $\Theta_{3,1}$ y (2,1) de $\Theta_{2,1}$	3 y 2	-2732.461	5676.364	27	3
Entradas (2,1) y (1,2) de $\Theta_{3,1}$	3	-2732.475	5676.393	27	3
Entradas (2,1) de $\Theta_{3,1}$, (2,1) de $\Theta_{2,1}$ y (1,2) de $\Theta_{1,1}$	3, 2 y 1	-2733.79	5671.191	26	3
Entradas (2,1) de $\Theta_{3,1}$, (2,1) de $\Theta_{2,1}$ y (1,2) de $\Theta_{3,1}$	3 y 2	-2732.454	5668.521	26	3
Entradas (2,1) de $\Theta_{3,1}$, (1,2) de $\Theta_{1,1}$, (1,2) de $\Theta_{3,1}$ y (2,1) de $\Theta_{2,1}$	3, 1 y 2	-2734.296	5664.372	25	3

Tabla 8.3: Valores del BIC de los modelos MVAR(2,3;1,1,1). El primer modelo al que se hace referencia en la tabla es el modelo que se eligió de entre los modelos de la Tabla 8.1. En seguida se muestran los valores del BIC de los cinco modelos que resultan de anular los parámetros no significativos indicados en la primera columna. La segunda columna se refiere al número de componente en la mezcla donde se anularon los parámetros no significativos. La quinta columna de la tabla indica el número de parámetros que se estimaron contando las entradas de los vectores $\Theta_{i,0}$ y de las matrices $\Theta_{i,1}$, $\Theta_{i,2}$, $\Theta_{i,3}$, Ω_i donde $i = \{1, 2, 3\}$.

Primer componente	Segundo componente	Tercer componente
$\alpha_1 = 0.4485$	$\alpha_2 = 0.5067$	$\alpha_3 = 0.0447$
$\Theta_{10} = \begin{pmatrix} 0.1807 \\ (0.0669) \\ 0.1664 \\ (0.0663) \end{pmatrix}$	$\Theta_{20} = \begin{pmatrix} -0.0213 \\ (0.0770) \\ -0.0002 \\ (0.0758) \end{pmatrix}$	$\Theta_{30} = \begin{pmatrix} 0.0201 \\ (0.4324) \\ 0.0533 \\ (0.4260) \end{pmatrix}$
$\Theta_{11} = \begin{pmatrix} 1.0055 & 0 \\ (0.0015) & (---) \\ -0.0133 & 0.9937 \\ (0.0038) & (0.0023) \end{pmatrix}$	$\Theta_{21} = \begin{pmatrix} 1 & 0.0034 \\ (0.0012) & (0.0015) \\ 0 & 0.9943 \\ (---) & (0.0030) \end{pmatrix}$	$\Theta_{31} = \begin{pmatrix} 0.9966 & 0 \\ (0.0072) & (---) \\ 0 & 0.9996 \\ (---) & (0.0135) \end{pmatrix}$
$\Omega_1 = \begin{pmatrix} 0.1091 & 0.0420 \\ 0.0420 & 0.0432 \end{pmatrix}$	$\Omega_2 = \begin{pmatrix} 0.5610 & 0.2707 \\ 0.2707 & 0.2307 \end{pmatrix}$	$\Omega_3 = \begin{pmatrix} 1.3780 & 0.8060 \\ 0.8060 & 1.5460 \end{pmatrix}$

Tabla 8.4: Estimaciones de los parámetros del modelo MVAR(2,3;1,1,1) sin parámetros no significativos. La primera columna de la tabla contiene las estimaciones de los parámetros del primer componente o sumando en la mezcla de la Ecuación (8.1): el escalar α_1 , el vector bivariado $\Theta_{1,0}$ y las matrices de tamaño 2×2 $\Theta_{1,1}$ y Ω_1 . Los números entre paréntesis debajo de las estimaciones son sus correspondientes errores estándar. La segunda y tercera columnas contienen las estimaciones de los parámetros del segundo y tercer componente respectivamente.

los precios de las acciones de Microsoft dependen de los precios anteriores de las acciones del Índice Nasdaq pero no al revés. Es decir, los cambios de los precios en las acciones del Índice Nasdaq influyen en los precios de las acciones de Microsoft. El segundo componente tiene un peso de mezcla de 0.5067 y varianzas 0.5610 y 0.2307. La entrada (2,1) de la matriz $\Theta_{2,1}$ vale cero de donde se infiere que los precios de las acciones del Índice Nasdaq dependen de los precios pasados de las acciones de Microsoft. En el tercer componente se tiene un peso de mezcla 0.0447 y varianzas 1.3780 y 1.5460. Las entradas (1,2) y (2,1) de la matriz $\Theta_{3,1}$ valen cero, lo cual indica que están desacopladas.

8.4. Validación del modelo

Observe que la serie bivariada compuesta por las series de tiempo QQQ y MSFT cumple con los supuestos del modelo MVAR:

- Ya que las series QQQ y MSFT se componen de precios de acciones se cumple que las variables aleatorias que las conforman son continuas, observadas en tiempo discreto. Además son series que fueron transformadas para lograr su estacionaridad (Subsecciones 5.2.1 y 5.3.1 del Capítulo 5).
- Para verificar que el modelo MVAR(2,3;1,1,1) es estacionario de primer orden se calcularon los eigen-

valores de la matriz $\sum_{k=1}^3 \alpha_k \Theta_{k1}$ de acuerdo al Teorema 7.1. Los valores obtenidos para el modelo completo son 0.943 y 0.9336. Para el modelo restringido son 0.9265 y 0.9576. Ya que ambos pares de eigenvalores son menores a uno ambos modelos son estacionarios en media.

- De acuerdo al Teorema 7.3, si el modelo $MVAR(2,3;1,1,1)$ es estacionario en media, para que sea estacionario de segundo orden los eigenvalores de la matriz $\sum_{k=1}^3 \alpha_k (\Theta_{k1} \otimes \Theta_{k1})$ deben ser menores a uno. Los eigenvalores de esta matriz calculada con los parámetros del modelo completo son 1.001, 0.9943, 0.9928 y 0.9876, y calculada con los parámetros del modelo restringido son 1.0013, 0.9964, 0.9946 y 0.9916. Ya que no todos los eigenvalores son menores a uno existe evidencia de que ningún modelo es estacionario de segundo orden.
- En la Figura 8.1 se encuentra la gráfica de la distribución de frecuencias de la serie bivariada (QQQ,MSFT), donde se aprecia un comportamiento multimodal, por lo que es probable que las distribuciones marginales o condicionales de la serie bivariada también sean multimodales.
- Un gran porcentaje del índice Nasdaq 100 corresponde a Microsoft, por lo que cualquier suceso o actividad ocurrida con sus acciones ejerce gran influencia en él y por lo tanto en sus propias acciones. Esto se refleja en el análisis de las matrices de autocorrelaciones cruzadas de la serie (QQQ,MSFT) en la Subsección 6.6 del Capítulo 6.
- Los parámetros del modelo se calcularon de tal manera que no cambian respecto al tiempo.

Por otro lado, para realizar la validación del modelo completo y del restringido se calcularon los residuales estandarizados mediante la Fórmula (7.13). Se calcularon las estadísticas Q^1 y Q^2 con cada una de las dos series que componen la serie de residuales y con las primeras 5, 10, 15, 20 y 25 autocorrelaciones. Esto se hizo utilizando las Fórmulas (7.14) y (7.15) de la Sección 7.4.6. Además se calcularon las estadísticas Q^3 y Q^4 con las primeras 5, 10, 15, 20 y 25 correlaciones cruzadas como en las Fórmulas (7.16) y (7.17).

En la Tabla 8.5 se aprecian los valores de las estadísticas Q^1 y Q^2 con sus respectivos valores p y la Tabla 8.6 contiene los valores de las estadísticas Q^3 y Q^4 a los residuales del modelo $MVAR(2,3;1,1,1)$ completo. Con un nivel de significación del 0.05 se concluye que no existe evidencia de dependencia entre los residuales ya que todos los valores p son mayores a 0.05.

Por otro lado, en la Tabla 8.7 se encuentran los valores de las estadísticas Q^1 y Q^2 con sus respectivos valores p referentes a los residuales del modelo restringido. Sea $\alpha = 0.05$. Ya que los valores p de Q^1 calculados con las primeras 15, 20 y 25 autocorrelaciones son menores que α existe evidencia de dependencia entre los residuales.

Estadística	QQQ	valor p QQQ	MSFT	valor p MSFT
$Q^1(5)$	4.6456	0.4606	4.7057	0.4528
$Q^1(10)$	6.6059	0.7620	11.7251	0.3038
$Q^1(15)$	15.1137	0.4432	21.294	0.1276
$Q^1(20)$	27.9554	0.1104	31.8338	0.3451
$Q^1(25)$	33.203	0.126	35.4137	0.0809
$Q^2(5)$	6.1001	0.2965	6.3083	0.2773
$Q^2(10)$	7.6299	0.6649	10.7402	0.3781
$Q^2(15)$	19.1978	0.2049	20.9991	0.1368
$Q^2(20)$	27.7275	0.116	29.2084	0.08373
$Q^2(25)$	29.5854	0.2401	31.0951	0.1858

Tabla 8.5: Valores de las estadísticas Q^1 y Q^2 con cada una de las dos series que componen la serie de residuales del modelo MVAR(2,3;1,1,1) completo con las primeras 5, 10, 15, 20 y 25 autocorrelaciones. La columna QQQ corresponde a la primera entrada de los residuales y la columna MSFT a la segunda entrada. Los valores p se calcularon en base a la distribución $\chi^2(m)$ donde m es el número de autocorrelaciones usadas en el cálculo de Q^1 y Q^2 .

Estadística	(QQQ,MSFT)	valor p
$Q^3(5)$	6.7668	0.2385
$Q^3(10)$	10.8186	0.3718
$Q^3(15)$	21.3824	0.125
$Q^3(20)$	32.4272	0.1389
$Q^3(25)$	37.0376	0.0572
$Q^4(5)$	3.1755	0.6729
$Q^4(10)$	7.4846	0.679
$Q^4(15)$	15.1256	0.4424
$Q^4(20)$	28.0673	0.1078
$Q^4(25)$	31.7772	0.1645

Tabla 8.6: Valores de las estadísticas Q^3 y Q^4 a la serie bivariada de residuales del modelo MVAR(2,3;1,1,1) completo con las primeras 5, 10, 15, 20 y 25 correlaciones cruzadas. Los valores p se calcularon en base a la distribución $\chi^2(m)$ donde m es el número de correlaciones cruzadas usadas en el cálculo de Q^3 y Q^4 .

Sin embargo, analizando la estadística Q^2 no se obtiene la misma conclusión ya que en este caso todos los valores p resultan mayores que α . La Tabla 8.8 contiene los valores de las estadísticas Q^3 y Q^4 a los residuales del modelo $MVAR(2,3;1,1,1)$ restringido. Los p valores referentes a las dos estadísticas calculados con las primeras 15, 20 y 25 correlaciones cruzadas son menores que α por lo que hay nueva evidencia de dependencia entre los residuales.

Estadística	QQQ	valor p QQQ	MSFT	valor p MSFT
$Q^1(5)$	2.2334	0.8159	3.2468	0.6619
$Q^1(10)$	8.5189	0.5782	11.8294	0.2966
$Q^1(15)$	25.4416	0.0443	27.9537	0.0218
$Q^1(20)$	41.6598	0.003	39.2365	0.0062
$Q^1(25)$	46.676	0.0053	44.9878	0.0083
$Q^2(5)$	1.6483	0.8953	2.5277	0.7723
$Q^2(10)$	8.7007	0.5607	12.2535	0.2684
$Q^2(15)$	18.7324	0.2261	22.8344	0.0877
$Q^2(20)$	30.8693	0.0569	31.0509	0.0545
$Q^2(25)$	34.6662	0.0944	33.4525	0.1201

Tabla 8.7: Valores de las estadísticas Q^1 y Q^2 con cada una de las dos series que componen la serie de residuales del modelo $MVAR(2,3;1,1,1)$ restringido con las primeras 5, 10, 15, 20 y 25 autocorrelaciones. La columna QQQ corresponde a la primera entrada de los residuales y la columna MSFT a la segunda entrada. Los valores p se calcularon en base a la distribución $\chi^2(m)$ donde m es el número de autocorrelaciones usadas en el cálculo de Q^1 y Q^2 .

Ya que no se encontró evidencia de dependencia entre los residuales del modelo $MVAR(2,3;1,1,1)$ completo se concluye que éste es el modelo que describe mejor a la serie en estudio. Por último, una vez que ya se tiene un modelo que describa una serie de tiempo dada es posible hacer predicciones. La discusión respecto a este tema se encuentra en el Capítulo 9.

Estadística	(QQQ,MSFT)	valor p
$Q^3(5)$	2.1198	0.8323
$Q^3(10)$	8.9124	0.5404
$Q^3(15)$	24.768	0.0531
$Q^3(20)$	36.97	0.0117
$Q^3(25)$	43.6286	0.0119
$Q^4(5)$	3.3459	0.6468
$Q^4(10)$	11.6868	0.3065
$Q^4(15)$	29.345	0.0145
$Q^4(20)$	44.8316	0.0011
$Q^4(25)$	49.105	0.0027

Tabla 8.8: Valores de las estadísticas Q^3 y Q^4 a la serie bivariada de residuales del modelo MVAR(2,3;1,1,1) restringido con las primeras 5, 10, 15, 20 y 25 correlaciones cruzadas. Los valores p se calcularon en base a la distribución $\chi^2(m)$ donde m es el número de correlaciones cruzadas usadas en el cálculo de Q^3 y Q^4 .

Capítulo 9

Conclusiones

Se puede concluir que los objetivos de ésta tesis se cumplieron ya que en los primeros cuatro capítulos se hizo una revisión de la teoría de series de tiempo incluyendo algunos conceptos básicos y ejemplos de la estructura matemática sobre la cual descansa el concepto de serie de tiempo, es decir de los procesos estocásticos. En el Capítulo 5 se desarrollaron dos ejemplos de ajustes de los modelos ARIMA a series de tiempo univariadas y en el Capítulo 6 se revisaron conceptos básicos sobre las series de tiempo multivariadas. El segundo objetivo, la implementación del modelo MVAR, se llevó a cabo en los dos últimos capítulos. En el Capítulo 7 se desarrollaron detalles del modelo relacionados con las condiciones de estacionaridad, la estimación de sus parámetros y la elección de el número de componentes y de los órdenes AR. En el Capítulo 8 se realizó el ajuste a la serie bivariada formada por las series de los ejemplos del Capítulo 5, utilizando la información que ya se tenía de los órdenes AR de cada serie.

Los modelos ARIMA propuestos para modelar las series de tiempo QQQ y $MSFT$ univariadas fueron satisfactorios y se considera que describen adecuadamente su comportamiento general. Además las estimaciones de los parámetros fueron estadísticamente significativas. Lo mismo puede decirse del modelo MVAR para modelar la serie bivariada ($QQQ, MSFT$).

Por otro lado se deduce que el modelo MVAR es de gran utilidad para describir el comportamiento de series de tiempo multivariadas que exhiben características no lineales como ciclos, outliers, periodos de valores constantes o de alta volatilidad. En particular para las series que forman parte de una cartera de inversiones, aún cuando sus distribuciones marginales o condicionales sean multimodales. Además, aunque sus parámetros son escalares, vectores y matrices es posible calcularlos mediante el método de máxima verosimilitud y la maximización

se puede llevar a cabo de forma numérica con la función `nlm` del software estadístico R.

Por ser una mezcla de K componentes AR gaussianos, y ya que la esperanza condicional de los componentes depende de los valores pasados de la serie, la forma de las distribuciones condicionales cambia respecto al tiempo, por ejemplo de unimodal a multimodal. Ya que la distribución condicional de la serie puede ser multimodal, la esperanza condicional de la serie podría no ser el mejor predictor. El mérito de los modelos MVAR radica en su habilidad para describir la distribución condicional de la serie, y por lo tanto, de las distribuciones predictivas de m pasos

$$F(y_{t+m}|y_t, y_{t-1}, \dots, y_1).$$

Wong y Li (2000) calculan la distribución predictiva de dos pasos para el modelo MAR mediante el método naïve, de forma exacta y por el método de Monte Carlo.

En la aproximación naïve se utiliza el pronóstico de un paso $\hat{y}_{t+1} = E(y_{t+1}|y_t, y_{t-1}, \dots, y_1)$ como si fuera el verdadero valor de y_{t+1} , de manera que la distribución condicional es

$$F(y_{t+2}|y_t, y_{t-1}, \dots, y_1) = F(y_{t+2}|y_t, y_{t-1}, \dots, y_1, y_{t+1} = \hat{y}_{t+1}).$$

Con este método se ignora cualquier información que la forma de la distribución predictiva de un paso pueda aportar. Para el modelo MAR, la pérdida de información puede conducir a errores sobretodo cuando

$F(y_{t+1}|y_t, y_{t-1}, \dots, y_1)$ es multimodal. Una mejor aproximación es calcular la distribución predictiva exacta de dos pasos mediante

$$F(y_{t+2}|y_t, y_{t-1}, \dots, y_1) = \int F(y_{t+2}|y_t, y_{t-1}, \dots, y_1, y_{t+1}) dF(y_{t+1}|y_t, y_{t-1}, \dots, y_1).$$

La integral podría ser intratable analíticamente. En ese caso la aproximación Monte Carlo podría ser la más adecuada. Este método aplicado a la evaluación de la distribución predictiva de dos pasos es

$$F(y_{t+2}|y_t, y_{t-1}, \dots, y_1) = \frac{1}{N} \sum_{j=1}^N F(y_{t+2}|y_t, y_{t-1}, \dots, y_1, y_{t+1}^j)$$

donde $\{y_{t+1}^j\}$ se muestrea a partir de $F(y_{t+1}|y_t, y_{t-1}, \dots, y_1)$.

En trabajos posteriores a ésta tesis podría analizarse el pronóstico de series multivariadas mediante el modelo MVAR. Respecto al número de componentes del modelo puede llevarse a cabo una investigación ya que existen métodos para estimar el número de componentes en un modelo de mezclas para datos independientes. Por ejemplo la aproximación bayesiana en Richardson y Green (1997) podría extenderse a datos dependientes y

por lo tanto a series de tiempo. Por otro lado, la distribución condicional gaussiana en los modelos MVAR podría sustituirse por otras distribuciones. Por ejemplo, si se reemplaza por la exponencial, gama o Weibull, estos modelos podrían representar series de tiempo positivas. Finalmente, modelos de co-integración también pueden ser investigados en modelos de mezclas para ver si dos o más series de tiempo son co-integradas en ciertos periodos pero independientes en otros periodos.

Anexos

Apéndice A

Función de log verosimilitud del modelo

MVAR(2,3;1,1,1) en R

```
1 logmvar <- function(theta , datos){
2   h1 <- theta [1]
3   h2 <- theta [2]
4   h3 <- theta [3]
5   h4 <- theta [4]
6   h5 <- theta [5]
7   h6 <- theta [6]
8   tao7 <- theta [7]
9   tao8 <- theta [8]
10  h7 <- exp(tao7)
11  h8 <- exp(tao8)
12  delta9 <- theta [9]
13  h9 <- tanh(delta9)
14  gamma.0 <- theta [14]
15
16  h10 <- theta [10]
17  h11 <- theta [11]
18  h16 <- theta [16]
19  h17 <- theta [17]
```

```
20 h18 <- theta[18]
21 h19 <- theta[19]
22 tao20 <- theta[20]
23 tao21 <- theta[21]
24 h20 <- exp(tao20)
25 h21 <- exp(tao21)
26 delta22 <- theta[22]
27 h22 <- tanh(delta22)
28 gamma.1 <- theta[15]
29
30 h12 <- theta[12]
31 h13 <- theta[13]
32 tao23 <- theta[23]
33 tao24 <- theta[24]
34 h23 <- exp(tao23)
35 h24 <- exp(tao24)
36 delta25 <- theta[25]
37 h25 <- tanh(delta25)
38 h26 <- theta[26]
39 h27 <- theta[27]
40 h28 <- theta[28]
41 h29 <- theta[29]
42
43 gammas <- c(gamma.0, gamma.1, 0)
44 alphas <- exp(gammas)/sum(exp(gammas))
45 alpha.0 <- alphas[1]
46 alpha.1 <- alphas[2]
47 alpha.2 <- alphas[3]
48
49 Theta.11 <- matrix(c(h1, h2, h3, h4), nrow=2, ncol=2, byrow=FALSE)
50 Theta.10 <- c(h5, h6)
51 Omega.1 <- matrix(c(h7, h9, h9, h8), nrow=2, ncol=2, byrow=FALSE)
```

```

52
53   Theta.20 <- c(h10,h11)
54   Theta.21 <- matrix(c(h16,h18,h19,h17), nrow=2, ncol=2, byrow=FALSE)
55   Omega.2 <- matrix(c(h20,h22,h22,h21), nrow=2, ncol=2, byrow=FALSE)
56
57   Theta.30 <- c(h12,h13)
58   Theta.31 <- matrix(c(h26,h28,h29,h27), nrow=2, ncol=2, byrow=FALSE)
59   Omega.3 <- matrix(c(h23,h25,h25,h24), nrow=2, ncol=2, byrow=FALSE)
60
61   y.t0 <- window(PARQMmst, start=c(2004,244), end=c(2004,2762))
62   y.t1 <- window(PARQMmst, start=c(2004,243), end=c(2004,2761))
63
64   det.Ome.1 <- det(Omega.1)^{-1/2}
65   inv.Ome.1 <- solve(Omega.1)
66   e1t <- y.t0 - Theta.10 - y.t1 %*% Theta.11
67
68   det.Ome.2 <- det(Omega.2)^{-1/2}
69   inv.Ome.2 <- solve(Omega.2)
70   e2t <- y.t0 - Theta.20 - y.t1 %*% Theta.21
71
72   det.Ome.3 <- det(Omega.3)^{-1/2}
73   inv.Ome.3 <- solve(Omega.3)
74   e3t <- y.t0 - Theta.30 - y.t1 %*% Theta.31
75
76   f.y <- matrix(1:dim(y.t0)[1], ncol = 1)
77   for (i in 1:dim(y.t0)[1]){
78     f.y[i] <- (alpha.0*det.Ome.1*(2*pi)^{-1})*
79     exp( ( ( e1t[i,] %*% inv.Ome.1 ) %*% t(t(e1t[i,])) ) * -0.5) +
80     (alpha.1*det.Ome.2*(2*pi)^{-1})*
81     exp( ( ( e2t[i,] %*% inv.Ome.2 ) %*% t(t(e2t[i,])) ) * -0.5) +
82     (alpha.2*det.Ome.3*(2*pi)^{-1})*
83     exp( ( ( e3t[i,] %*% inv.Ome.3 ) %*% t(t(e3t[i,])) ) * -0.5)

```

```
84     }  
85  
86     loglik <- sum(log(f.y))  
87     -1*loglik  
88   }  
89  
90   min <- nlm(logmvar , paropred , hessian=TRUE, datos=PARQMmst)
```

Referencias

- Arranz, M. A. (2005). Portmanteau Test Statistics in Time Series. *Tol - Project*, www.tol-project.org.
- Box, G. E. P., Jenkins, G. M., y Reinsel, G. C. (1994). *Time series analysis* (3.^a ed.). Englewood Cliffs, NJ [u.a.]: Prentice-Hall. (Literaturverz. S. 556 - 567)
- Brockwell, P. J., y Davis, R. A. (2006). *Introduction to time series and forecasting*. Springer Science & Business Media.
- Brockwell, P. J., y Davis, R. A. (2013). *Time series: theory and methods*. Springer Science & Business Media.
- Broda, S. A., y Paoletta, M. S. (2009). CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation. *Journal of financial econometrics*, 4, 412-436.
- Chatfield, C. (2016). *The Analysis of Time Series: An Introduction*. Chapman and Hall, CRC Press.
- Fong, P. W., Li, W. K., Yau, C. W., y Wong, C. S. (2007). On a mixture vector autoregressive model. *Canadian Journal of Statistics*, 35(1), 135–150.
- Fong, T. P., y Wong, C. (2008). Stress Testing Banks Credit Risk Using Mixture Vector Autoregressive Models. *Hong Kong Monetary Authority*.
- Fuller, W. A. (1995). *Introduction to statistical time series* (2.^a ed.). Wiley Series in Probability and Statistics.
- González Velasco, M., y del Puerto García, I. M. (2009). *Series temporales*. Universidad de Extremadura, Servicio de Publicaciones.
- Guerrero Guzmán, V. M. (2003). *Análisis estadístico de series de tiempo económicas*. Editorial Thomson.
- Hyndman, R. J., y Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3), 1–22.
- Israel, I. J. (2007). Patterns of relative strength. *Lulu.com*.
- Kalliovirta, L., Meitz, M., y Saikkonen, P. (2014). Gaussian mixture vector autoregression. *Discussion Papers, Helsinki Center of Economic Research*.
- Le, N. D., Martin, R. D., y Raftery, A. E. (1996). Modeling Flat Stretches, Bursts Outliers in Time Series Using Mixture Transition Distribution Models. *Journal of the American Statistical Association*, 91(436),

1504–1515.

- Lerman, D. (2001). *Exchange traded funds and e-mini stock index futures* (1.^a ed.). John Wiley and Sons.
- Lutkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag GmbH.
- Maindonald, J., y Braun, W. J. (2010). *Data Analysis and Graphics using R. An Example Based Approach*. Cambridge University Press. Cambridge University Press, New York.
- Makridakis, S. G., Wheelwright, S. C., y Hyndman, R. J. (1998). *Forecasting: methods and applications* (3.^a ed.). Wiley.
- Martin, V. L. (1992). Threshold time series models as multimodal distribution jump processes. *Journal of Time Series Analysis*, 13(1), 79–94.
- McLachlan, G., y Peel, D. (2004). *Finite Mixture Models*. John Wiley & Sons.
- McLeod, A. I., y Li, W. K. (1983). Diagnostic Model Checking ARMA Time Series Models Using Squared-Residual Autocorrelations. *Journal of Time Series Analysis*, Vol. 4, No. 4.
- Pfaff, B. (2008). VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software*, 27(4), 1-32.
- Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*. Springer.
- Richardson, S., y Green, P. J. (1997). On bayesian analysis of mixtures with an unknown number of components (with discussion). *J. R. Statist. Soc.*, B(59), 731 - 792.
- Shumway, R. H., y Stoffer, D. S. (2010). *Time series analysis and its applications: with r examples*. Springer Science & Business Media.
- Titterton, M., Smith, A. F., y Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley.
- Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford University Press.
- Tsay, R. S. (2005). *Analysis of Financial Time Series* (Vol. 543). John Wiley & Sons.
- Turner, T. (2005). *Short-term trading in the new stock market* (Reprint ed.; Griffin, Ed.). St. Martin's Press.
- Venables, W. N., y Ripley, B. D. (2002). *Modern applied statistics with s-plus* (4.^a ed.). Springer-Verlag.
- Wong, C. S. (2011). Constrained Estimation of Mixture Vector Autoregressive Model. *19th International Congress on Modelling and Simulation, Perth, Australia, 12 - 16 December 2011*.
- Wong, C. S., y Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1), 95–115.