



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
ESPECIALIDAD EN ESTADÍSTICA APLICADA

UN ESTUDIO INTRODUCTORIO AL ANÁLISIS DE SERIES DE TIEMPO CON APLICACIONES
AL SECTOR FINANCIERO

TESINA

PARA OPTAR POR EL GRADO DE:
ESPECIALISTA EN ESTADÍSTICA APLICADA

PRESENTA:
FRANCISCO JOSUÉ MARTÍNEZ CERVANTES

DIRECTOR DE LA TESINA:
M. EN C. ANTONIO SORIANO FLORES

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

MÉXICO, D. F. , SEPTIEMBRE 2015



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

Quiero darle el más sincero agradecimiento a mis padres y a mis hermanas por darme todo tu su apoyo y hacer que todo esto sea posible.

A mi Josselin por darme la fuerza y los ánimos para seguir adelante por difícil que parezca.

Al M. en C. Antonio Soriano por todo el apoyo y conocimiento que me brindó en la elaboración de este trabajo, trabajo del cual me siento muy orgulloso.

A mis sinodales Dr. Alberto Contreras, M. en C. Patricia Romero, M. en C. Leticia Gracia-Medrano y al Act. y Esp. Antonio Aranda por sus valiosos comentarios que me fueron de gran utilidad para enriquecer este trabajo.

Finalmente, quiero agradecer también a la vida y a la UNAM por darme la maravillosa experiencia de estar aquí y poder superarme día a día.

“Por mi raza hablará el espíritu”

“La ciencia es una esfera finita que crece en el espacio infinito; cada nueva expansión le hace comprender una zona mayor de lo desconocido, pero lo desconocido es inagotable.”

J. L. Borges

Índice general

Introducción	5
1. Conceptos básicos	6
1.1. Definición, clasificación y objetivo	6
1.2. Análisis descriptivo de las series de tiempo	8
1.2.1. Tipos de variación	8
1.2.2. Descomposición aditiva y multiplicativa	9
1.3. Breve análisis de los componentes de las series de tiempo	10
1.3.1. Tendencia	10
1.3.2. Estacionalidad	12
1.4. Autocorrelación	12
1.4.1. Correlograma	13
2. Los procesos estocásticos estacionarios y sus propiedades	14
2.1. Procesos estocásticos	14
2.1.1. Series de tiempo estacionarias	14
2.2. Estimación de los momentos muestrales	15
2.2.1. Estimación de la media	15
2.2.2. Estimación de la varianza	16
2.2.3. Estimación de la autocovarianza y la autocorrelación	17
2.3. Propiedades espectrales de los modelos estacionarios	17
2.4. Teorema de Wold	19
3. Algunos modelos estocásticos	20
3.1. El modelo básico: ruido blanco	20
3.2. Modelos lineales de series de tiempo	21
3.2.1. Procesos lineales generales	21
3.2.2. Procesos de medias móviles	22
3.2.3. Procesos autorregresivos	23
3.2.4. ARMA	23
3.3. Modelos heterocedásticos condicionales	24
3.3.1. Procesos ARCH	24
3.3.2. Procesos GARCH	25
3.4. Modelos no estacionarios	26
3.4.1. Caminata aleatoria	26

3.4.2. ARIMA	27
4. Modelación ARIMA	28
4.1. Metodología Box-Jenkins	28
4.2. Identificación del modelo	29
4.2.1. Instrumentos de identificación del modelo	30
4.2.2. Identificación de modelos no estacionarios	34
4.3. Estimación de los parámetros del modelo	36
4.4. Diagnóstico del modelo	36
4.4.1. Análisis de los coeficientes	36
4.4.2. Análisis de los residuos	37
4.5. Predicción	42
4.5.1. Inconvenientes de la metodología ARIMA de Box y Jenkins	44
4.5.2. Método Stepwise	44
5. Análisis del Índice de Precios y Cotizaciones en México	45
5.1. Análisis descriptivo	45
5.1.1. Breve descripción de la información	45
5.1.2. Análisis de la estacionariedad	46
5.2. Metodología ARIMA	48
5.2.1. Identificación	48
5.2.2. Estimación	50
5.2.3. Validación del modelo	51
5.3. Predicción	55
Conclusiones	57
Bibliografía	59
A. Métodos para corregir la no estacionariedad	60
B. Ajuste del modelo en R	62

Introducción

Una *serie de tiempo* es una secuencia de datos relacionada con un momento en el tiempo, cuyas características intrínsecas y sumamente fundamentales son que dentro del conjunto de datos el orden es importante y que no se sostiene el supuesto de independencia, puesto que las observaciones adyacentes de la serie son dependientes o están correlacionadas entre ellas a lo largo del tiempo. Debido a estas características, no es posible aplicar los métodos estadísticos tradicionales, cuyos supuestos se basan en la independencia de las observaciones que componen la muestra, pues esto no sólo resultaría en inconvenientes al momento de ajustar los respectivos modelos, sino que también se generarían modelos poco eficientes. Por lo tanto, el conjunto de técnicas de estudio de series de datos ordenados cuya estructura de correlación entre los datos resulta inherente se denomina como *Análisis de Series de Tiempo*.

Ejemplos de series de tiempo abundan en muchos campos de conocimiento como la economía, la geografía, la ingeniería o las ciencias naturales, entre otros, considerando datos como la presión atmosférica, la temperatura, los precios de materias primas, los tipos de cambio y un sinnúmero de información que puede ser modelada a través de estas técnicas, lo cual les da una gran importancia dentro de la ciencia estadística. Es por ello, que el presente trabajo tiene por objeto dar una breve pero completa introducción a las técnicas de *Análisis de Series de Tiempo*, desde sus conceptos básicos hasta los modelos más comunes y los métodos mayormente utilizados con la finalidad de adentrarse en el estudio de este tipo de técnicas y métodos con un gran valor teórico y, principalmente, aplicativo.

De esta forma, en el primer capítulo se presentan los conceptos básicos que incluyen la clasificación de acuerdo al objetivo que se pretenda con la serie, un primer acercamiento descriptivo sobre cómo se componen las *series de tiempo* y un elemento fundamental dentro de este tipo de análisis como lo es la *autocorrelación*. En el segundo capítulo se expone a detalle la importancia del concepto de estacionariedad en el *Análisis de Series de Tiempo*, así como las propiedades y el por qué lo ideal es buscar que las *series de tiempo* se comporten de esta manera. El capítulo 3 es una presentación de los modelos básicos a través de los cuales se articula toda la teoría de las *Series de Tiempo*, mientras que el capítulo 4 tiene como principal propósito exponer la *metodología ARIMA*, la cual es un proceso iterativo de construcción de modelos de *series de tiempo*, que por su popularidad y alcance, es nodal en el *Análisis de Series de Tiempo*. Finalmente, como este tipo de técnicas tienen un amplio uso aplicativo e incluso algunos de los métodos más importantes tienen una importante base empírica, en la última sección de este trabajo se presenta un ejercicio de *series de tiempo* aplicado al sector financiero, pues de acuerdo a Tsay [13, p. 1], las series de tiempo financieras poseen la característica especial de contener un elemento de incertidumbre y como resultado de ello, estos métodos estadísticos juegan un papel importante en el análisis de este tipo de series.

Capítulo 1

Conceptos básicos

1.1. Definición, clasificación y objetivo

De acuerdo a Box & Jenkins [2, p. 1] una serie de tiempo se define como un conjunto de observaciones de cierto fenómeno consideradas secuencialmente en el tiempo. Matemáticamente este conjunto de observaciones se pueden denotar como $\{x_{t_1}, x_{t_2}, \dots, x_{t_n}\} = \{x_{t_i} : i \in T\}$ donde $\{x_{t_i}\}$ es el valor de la variable x en el tiempo t_i y T es el conjunto de índices [14, Woodward, p. 3].

De acuerdo al objetivo a desarrollar, las series de tiempo pueden ser clasificadas en varias categorías según diferentes criterios [9, Kitagawa, p. 6-7]:

- Series de tiempo continuas y discretas: siendo continuas cuando las observaciones son recolectadas permanentemente en el tiempo, es decir, cuando el conjunto de índices es un conjunto no numerable (por ejemplo $T = \mathbb{R}$) y discretas cuando la información observada es recolectada en ciertos intervalos de tiempo, es decir, cuando el conjunto de índices T es numerable. A su vez, las series de tiempo discretas tienen otra subcategoría de acuerdo al intervalo de tiempo en el que son recolectadas: datos observados en intervalos igualmente espaciados (serie de tiempo equiespaciada) y datos observados en intervalos desigualmente espaciados. Un ejemplo de una serie de tiempo continua es el registro constante que se puede tener del precio de un activo en el mercado financiero desde que se originó, mientras que el ejemplo de una serie de tiempo discreta puede ser ese mismo precio pero considerando sólo una(s) temporada(s).
- Series de tiempo univariadas y multivariadas: refiriéndose a una serie de tiempo univariada cuando se trata solamente de un conjunto de datos a lo largo del tiempo. Por otro lado, cuando se trata de series de tiempo en las que se obtiene simultáneamente el registro de dos o más fenómenos, se habla de series de tiempo multivariadas. Ejemplo de una serie de tiempo univariada son las ganancias de una empresa registrada en cierto tiempo, mientras que una serie de tiempo multivariada sería el registro de las ganancias de varias empresas que se desenvuelven en cierto sector de la economía.
- Series de tiempo estacionarias y no estacionarias: en el análisis de series de tiempo, la variación irregular que presentan las series es expresada generalmente a través de modelos estocásticos y, en algunos casos, dicho fenómeno aleatorio es la realización de un modelo estocástico con una estructura invariante a lo largo del tiempo. Cuando es así, a la serie de tiempo se le conoce como estacionaria, es decir, $\mathbb{E}(x_t) = \mu$. Por el otro lado, si la estructura estocástica de la serie cambia con el tiempo, la serie de tiempo es no estacionaria, o sea, $\mathbb{E}(x_t) = \mu_t$. Como ejemplos, en el caso de una serie de tiempo estacionaria se tiene el precio de un título o una acción en el mercado financiero, en el cual las

cotizaciones varían de forma constante alrededor de un punto medio; mientras que en el caso de la serie de tiempo no estacionaria se tiene el Producto Interno Bruto de un país el cual posee una tendencia que modifica su estructura a lo largo del tiempo.

- Series de tiempo Gaussianas y no Gaussianas: considerándose Gaussianas cuando la distribución de las variables aleatorias de la serie de tiempo siguen una distribución normal ($x_t \sim N(\mu_t, \sigma_t)$), de otra manera se trata de una serie de tiempo no Gaussianas. Un ejemplo de esto es cuando la serie presenta de forma ocasional un patrón asimétrico por el cual la distribución marginal no puede ser considerada como Gaussianas.
- Series de tiempo lineales y no lineales: una serie de tiempo que se puede expresar como una realización (variable respuesta) de un modelo lineal es llamada serie de tiempo lineal. En contraste, de no ser así es una serie de tiempo no lineal. En el caso de las series de tiempo financieras existen muchos ejemplos para ambos casos: en la primer situación donde generalmente se observa precios de acciones o ventas de empresas que son realización de modelos lineales; mientras que el segundo caso pueden ser índices de volatilidad, como el nasdaq, que presentan correlación bivariada o son modelos bilineales.

Además debe considerarse que en muchas series de tiempo existen observaciones perdidas y valores atípicos, los cuales requieren de un tratamiento especial considerando que en las series de tiempo las problemáticas provienen de situaciones reales, por lo que hacer caso omiso de esta información puede modificar seriamente el resultado final. En ese sentido, los métodos robustos están diseñados para que el modelo sea insensible a los valores atípicos y a las observaciones perdidas [4, Chatfield, p. 6].

Para los términos que corresponden a esta investigación, la atención se enfocará en series de tiempo univariadas, Gaussianas, lineales y discretas con intervalos de tiempo igualmente distribuidos. Respecto a la estacionariedad, se abordará el caso de las dos categorías antes descritas, es decir, las series de tiempo estacionarias y las no estacionarias.

A partir de esta clasificación, y de acuerdo a la investigación, los principales objetivos de las series de tiempo pueden agruparse en tres principales fases [9, Kitagawa, p. 8]:

- Descripción: con la información que aporta una serie de tiempo observada se usan métodos descriptivos y gráficos que expresan o resumen eficazmente las características de la serie temporal. De esta forma, es posible capturar las características esenciales de las series de tiempo y obtener una pista sobre cómo se puede modelar dicha serie de tiempo, así como para detectar si existen datos atípicos.
- Modelación: en esta etapa capturamos la estructura estocástica de la serie de tiempo mediante la identificación de un modelo apropiado. Puesto que hay varios tipos de series de tiempo, es necesario seleccionar el tipo de modelo adecuado y estimar los parámetros correspondientes, dependiendo de las características y el objetivo de la serie de tiempo. Una representación esquemática de este proceso, en el caso de un modelo lineal, se muestra en la *figura 1.1*.

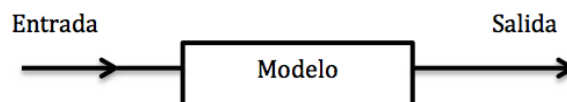


Figura 1.1: Ejemplo de modelación lineal

- **Predicción:** asumiendo que la serie de tiempo observada es una realización de un modelo, sobre la base de las correlaciones en el tiempo y entre las variables, se puede estimar el comportamiento futuro de la serie de tiempo mediante el uso de la información extraída de las observaciones presentes y pasadas.

1.2. Análisis descriptivo de las series de tiempo

Una primera aproximación hacia las series de tiempo es la idea de los componentes no observados. Como explica Chatfield [4, p. 11], podría pensarse que al igual que en la mayoría de las áreas de la estadística, en las series de tiempo el análisis comienza con el cálculo de la media y la desviación estándar para medir su ubicación y dispersión. Sin embargo, es diferente pues existen componentes sistemáticos no observados que pueden afectar el cálculo de dichos estadísticos. Incluso cuando la serie de tiempo no posee ninguno de los componentes sistemáticos, las estadísticas de resumen no tienen sus propiedades habituales como se verá más adelante. Por lo tanto, el análisis descriptivo de las series de tiempo se centra en las formas de entender los efectos típicos de las series de tiempo como la tendencia, la estacionalidad¹ y las correlaciones entre observaciones sucesivas.

1.2.1. Tipos de variación

Los métodos clásicos de análisis de series de tiempo se basan principalmente en la descomposición de su varianza en componentes sistemáticos que representan la tendencia, estacionalidad y otros cambios cíclicos. Cualquier otro tipo de variación restante en la serie de tiempo se atribuye a fluctuaciones irregulares. De acuerdo a Chatfield [4, p. 11], este enfoque no es siempre el mejor, sin embargo, resulta particularmente valioso cuando los principales componentes de variación son la tendencia y/o la estacionalidad. Así, las diferentes fuentes de variación se pueden agrupar en cuatro principales:

- (I) **Tendencia:** es un cambio suave a través del tiempo en el valor medio de los datos. Generalmente es el componente de largo plazo que constituye la base del crecimiento o declinación de una serie de tiempo.
- (II) **Variación estacional:** es una variación periódica de amplitud constante que se encuentra típicamente en los datos clasificados por trimestres, meses o semanas. La variación estacional se refiere a un patrón de cambio a través del tiempo.
- (III) **Otras variaciones cíclicas:** refiriéndose a la variación de la serie que tiene amplitud no constante. Este conjunto de fluctuaciones en forma de onda o ciclos, de más de un año de duración, es producido por cambios en las condiciones en las que se desenvuelve la serie.
- (IV) **Otras fluctuaciones irregulares:** una vez que se han removido de la serie de tiempo los componentes de tendencia y variaciones cíclicas, las fluctuaciones irregulares son el tipo de variación que queda. Éstas poseen la característica de que son causadas por situaciones no controlables o no predecibles y pueden ser o no aleatorias.

¹Nótese la diferencia entre estacionalidad y estacionariedad: la primera se refiere al componente cíclico que presenta la serie en periodos de tiempo menores a un año, mientras que la segunda se refiere a la característica de ciertos procesos estocásticos de mantenerse invariantes a lo largo de tiempo

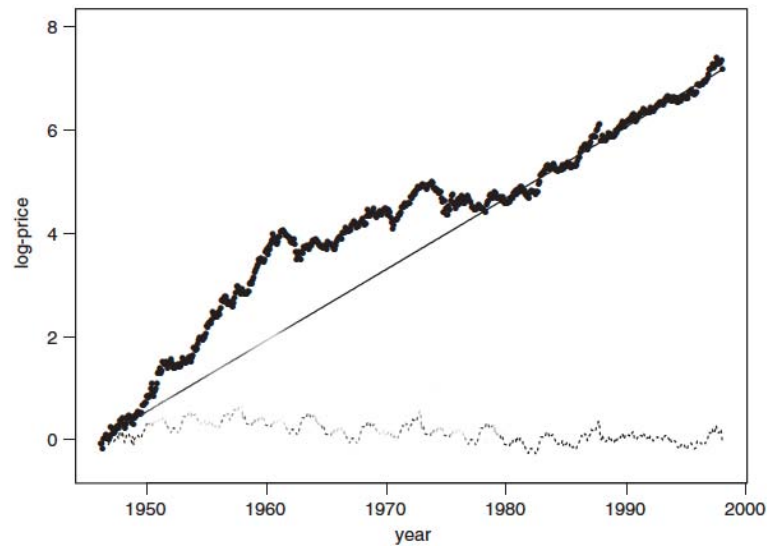


Figura 1.2: Ejemplo de una serie de tiempo con y sin tendencia: Log precios de los valores de 3M de 1946 a 1997

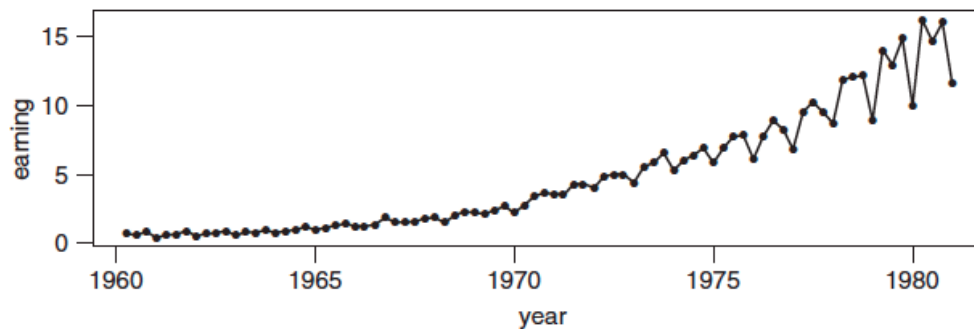


Figura 1.3: Ejemplo de una serie de tiempo estacional: rendimientos por acción de Johnson & Johnson de 1960 a 1980

1.2.2. Descomposición aditiva y multiplicativa

Una vez identificados los tipos de variación, las series de tiempo pueden tener un comportamiento que sugiera la posibilidad de representar los datos como una realización de un proceso que puede tener todos o alguno de los componentes. Entonces, el modelo de dicha serie se representará de alguna de las siguientes dos formas:

$$X_t = f(TR_t, SN_t, CL_t, IR_t) \begin{cases} TR_t \cdot SN_t \cdot CL_t \cdot IR_t, & \text{modelo multiplicativo} \\ TR_t + SN_t + CL_t + IR_t, & \text{modelo aditivo} \end{cases}$$

Donde:

- TR_t es la componente de tendencia,
- SN_t es el componente estacional de periodo,
- CL_t es el componente cíclico,
- IR_t es el componente aleatorio o de error.

Cabe resaltar que el modelo de descomposición multiplicativa es útil al modelar series de tiempo que manifiestan una

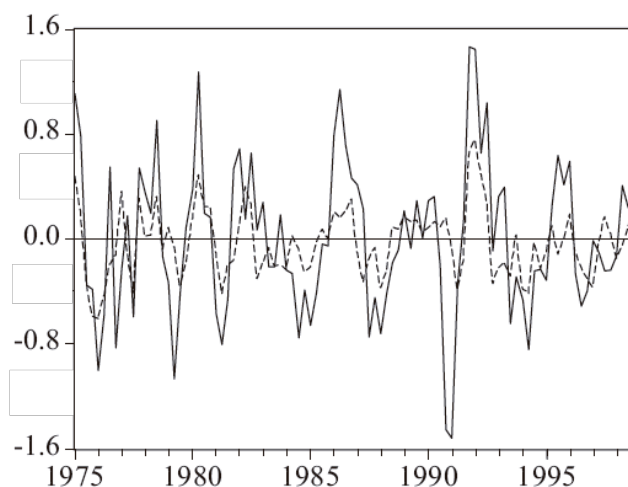


Figura 1.4: Ejemplo de una serie de tiempo con ciclo: tasa de inflación de Alemania de 1975 a 2000

variación estacional creciente o decreciente. Por su parte, el modelo de descomposición aditiva es útil al modelar series de tiempo con una variación estacional constante [1, Bowerman, p. 326-341].

1.3. Breve análisis de los componentes de las series de tiempo

1.3.1. Tendencia

En muchas de las series de tiempo un componente recurrente es la tendencia. El tipo de tendencia más sencillo es el habitual “tendencia lineal + ruido”, para el cual la observación en el tiempo t es la variable aleatoria x_t y expresando el término de tendencia (TR_t) como una función paramétrica de la observación, es decir, $TR_t = \alpha + \beta t$, tenemos que:

$$X_t = f(TR_t, IR_t) \begin{cases} TR_t \cdot IR_t = (\alpha + \beta t) \cdot \epsilon_t & \text{Para el modelo multiplicativo} \\ TR_t + IR_t = \alpha + \beta t + \epsilon_t & \text{Para el modelo aditivo} \end{cases}$$

Donde α, β son constantes y ϵ_t denota término de error aleatorio con media cero. De esta forma, TR_t puede ser identificada a través de diversos métodos.

Uno de los métodos más usados es la *estimación por mínimos cuadrados*, procedimiento por el cual es fácil calcular α y β minimizando $\sum(x_t - TR_t)^2$. Dentro de este marco, β puede ser descrito como la pendiente de la tendencia, es decir, el cambio en el nivel de TR_t por unidad de tiempo. Sin embargo, debe considerarse que la *estimación por mínimos cuadrados* suele tener ciertos inconvenientes como:

- (I) Que se debe asumir una tendencia fija para toda la duración del conjunto de datos, lo cual generalmente no es cierto, pues en realidad la tendencia puede estar cambiando constantemente en el tiempo. En todo caso, debería de utilizarse un método adaptativo que permitiese ajustar la tendencia calculada a los cambios.
- (II) Para que el método sea efectivo, se debe de tratar con una simple forma restringida de TR_t , como los modelos lineales por ejemplo.

Filtros lineales

Otro procedimiento para modelar la tendencia es el uso de filtros lineales o suavizamiento, con los cuales se convierte una serie x_t en otra serie y_t mediante una operación lineal:

$$y_t = Sm(x_t) = \sum_{r=-q}^{+s} a_r x_{t+r}$$

Y se puede representar la relación entre la salida y_t y la entrada x_t como:

$$x_t \rightarrow \boxed{\text{filtro}} \rightarrow Sm(x_t) = y_t$$

Donde el conjunto de pesos $\{a_r\}$ de los filtros usualmente se asume como simétrico y normalizado, es decir ($a_r = a_{-r}$ y $\sum a_r = 1$). Un ejemplo derivado de esta primer formulación es el filtro de medias móviles, dado por:

$$y_t = \frac{1}{2q+1} \sum_{r=-q}^q x_{t+r}$$

El tamaño de este filtro está determinado por el número q , que es el orden de la media móvil. Cuando $q = 1$ se tiene una media móvil simple de tres puntos. Sin embargo, los pesos no tienen que ser los mismos en cada punto. De esta forma, los filtros como las medias móviles son procedimientos que se aplican retroactivamente a una serie de tiempo, utilizando tanto observaciones pasadas como posteriores al momento del cálculo del suavizamiento, es decir, al momento de cálculo se pierden observaciones y entonces se tienen $t = n - q$ observaciones. Además, como esta técnica no requiere de un modelo predeterminado, no produce una fórmula para extrapolar y dar pronósticos. En ese sentido, el modelar líneas de tendencia resulta más ventajoso, aunque también existen otros métodos de suavizamiento que permiten que no haya datos faltantes ($t = n$) y se pueda proyectar un pronóstico utilizando los valores pasados y presentes de x_t . Una popular técnica para esto es el suavizamiento exponencial que considera un filtro de la forma:

$$Sm(x_t) = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j x_{t-j}$$

Donde α es una constante que va de $0 < \alpha < 1$ y los pesos $a_j = \alpha(1-\alpha)^j$ decrecen geométricamente con j . La experiencia sugiere que α ronda entre 0.1 y 0.3, pero como puede apreciarse, a partir del objeto que se persiga se elegirá el filtro más adecuado para encontrar una tendencia específica [3, Chan, p. 7].

Finalmente, cabe decir que los métodos precedentes ayudan a estimar la tendencia mediante un suavizador \hat{T}_t , pero en la mayoría de las aplicaciones resulta poco relevante estimarla. En su lugar, el interés puede estar en eliminar su efecto. Un método sencillo para lograrlo es diferenciando. Para ello, consideremos un operador retraso denotado por B que actúa sobre el tiempo en forma de $Bx_t = x_{t-1}$. Entonces, el operador diferencia se define como:

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

$$\nabla^h x_t = (1 - B)^h x_t, \quad h = 1, 2, \dots, n$$

Con lo cual, si se considera que $p < n$, la serie $x_t = TR_t + IR_t$ con $TR_t = \sum_{h=0}^p a_p t^p$, entonces $\nabla^p x_t = p!a_p + \nabla^p IR_t$ y TR_t es eliminado. Por lo tanto, la diferenciación es una especie de filtro de alta frecuencia, en el que la baja frecuencia, la tendencia TR_t en este caso, es eliminada y sólo se conserva la alta frecuencia, es decir, IR_t [3, Chan, p. 7-8]. En principio,

podemos eliminar cualquier tendencia polinómica diferenciando la serie, sin embargo este proceso tiene el inconveniente de que cada que se diferencia la serie, se pierden datos. En consecuencia, no es aconsejable diferenciar muchas veces la serie. En ese sentido, un criterio empírico bastante útil es comparar las varianzas estimadas de la serie en cada diferenciación, donde la varianza mínima corresponderá al nivel adecuado de diferenciación. Un ejemplo de esto es si se define a x_t como una serie de tiempo y a y_t , z_t y w_t como sus respectivas diferenciaciones, es decir, $y_t = \nabla x_t$, $z_t = \nabla^2 x_t$, y $w_t = \nabla^3 x_t$, el criterio para determinar cuándo se debe de dejar de diferenciar está dado por:

$$\text{Var}(\hat{x}_t) > \text{Var}(\hat{y}_t) > \text{Var}(\hat{z}_t) < \text{Var}(\hat{w}_t)$$

Donde z_t representa el nivel adecuado hasta el cual diferenciar [7]p. 23-25

1.3.2. Estacionalidad

Cuando el componente estacional SN_t está presente en la ecuación, las técnicas antes vistas deben ajustarse a éste. De acuerdo a las formulaciones anteriores, la ecuación con el componente estacional queda:

$$X_t = f(TR_t, SN_t, IR_t) \begin{cases} TR_t \cdot SN_t \cdot IR_t, & \text{Para el modelo multiplicativo} \\ TR_t \cdot SN_t + IR_t, & \text{Para el modelo mixto} \\ TR_t + SN_t + IR_t, & \text{Para el modelo aditivo} \end{cases}$$

Al igual que en la tendencia, dependiendo de la meta, se puede estimar la estacionalidad mediante algún tipo de suavizado estacional o eliminarlo de los datos mediante la respectiva operación. Supongamos que la parte estacional tiene un período d , tal que $S_{t+d} = S_t$, $\sum_{j=1}^d S_j = 0$.

A. Método de medias móviles: primero se estima la tendencia mediante el filtro de medias móviles en todo un ciclo completo de manera que el efecto de estacionalidad sea promediado. Dependiendo si es par o impar, se lleva a cabo uno de los dos siguientes pasos:

1. Si $d = 2q$, tenemos $\hat{TR}_t = \frac{1}{d}(\frac{1}{2}x_{t-q} + x_{t-q+1} \dots + x_{t+q-1} + \frac{1}{2}x_{t+q})$, para $t = q + 1, \dots, n - q$
2. Si $d = 2q + 1$, tenemos $\hat{TR}_t = \frac{1}{d} \sum_{r=-q}^q x_{t+r}$, para $t = q + 1, \dots, n - q$

Después de estimar TR_t , se filtra la serie y se estima la parte estacional de los residuos de $x_t - \hat{TR}_t$.

B. Diferenciación estacional. Por otro lado, se puede aplicar la diferenciación estacional para eliminar el efecto de este término. Considerando la d -enésima diferenciación de $x_t - x_{t-d}$, se elimina el efecto de SN_t [3, Chan, p. 8-9].

1.4. Autocorrelación

Un instrumento fundamental para analizar las propiedades de una serie de tiempo, en términos de la interrelación temporal de sus observaciones, son los coeficientes de autocorrelación de la muestra. Estos miden la correlación, si la hay, entre observaciones separadas en h períodos y proporcionan información descriptiva útil.

El coeficiente de correlación ($\rho_{x_t, x_{t+h}}$) entre dos variables x_t y x_{t+h} , mide el grado de asociación lineal entre ambas variables y se define como:

$$\rho_{x_t, x_{t+h}} = \frac{\text{COV}(x_t, x_{t+h})}{\sqrt{V(x_t)V(x_{t+h})}} \quad h = \pm 1, \pm 2, \dots, \pm n - 1$$

Este coeficiente toma valores $|\rho_{x_t, x_{t+h}}| \leq 1$. Si $\rho_{x_t, x_{t+h}} = 0$, no existe relación lineal entre x_t y x_{t+h} . Si $|\rho_{x_t, x_{t+h}}| = 1$ existe una relación lineal perfecta entre las variables. Entonces, dada una estructura de correlación, en las series de tiempo se define por función de correlación, la cual es estimada a partir de la serie de tiempo observada. Entonces, tenemos que:

$$\hat{\rho}_h = \frac{\sum_{t=1}^{n-h} (x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad h = 1, 2, \dots, n - 1$$

Esta ecuación es conocida como el coeficiente de correlación al rezago h y como se demostrará más adelante, este estimador es insesgado asintóticamente y es una herramienta muy útil en la construcción de modelos de series de tiempo.

1.4.1. Correlograma

Una ayuda útil en la interpretación de un conjunto de coeficientes de autocorrelación es un gráfico que se llama correlograma, el cual presenta los respectivos coeficientes de autocorrelación ρ_h contra el rezago h para $h = 0, 1, 2, 3, \dots, m$, donde m es por lo general mucho menor que n [4, Chatfield, p. 24].

La inspección visual de este gráfico puede resultar de gran ayuda, pues permite detectar si la serie de tiempo tiene tendencia, efectos estacionales o ciclos, o si bien, carece de ellos y es completamente aleatoria. Además el correlograma permite detectar la no presencia de correlación (cuando $\rho_h \approx 0$, $\hat{\rho}_h \sim N(0, 1/n)$), la correlación a corto plazo (cuando $\rho_h \rightarrow 0$ conforme las observaciones estén más separadas en el tiempo), la alternancia de la serie (mostrando observaciones “arriba-abajo-arriba” de una media global) y las observaciones atípicas [4, Chatfield, p. 20-24].

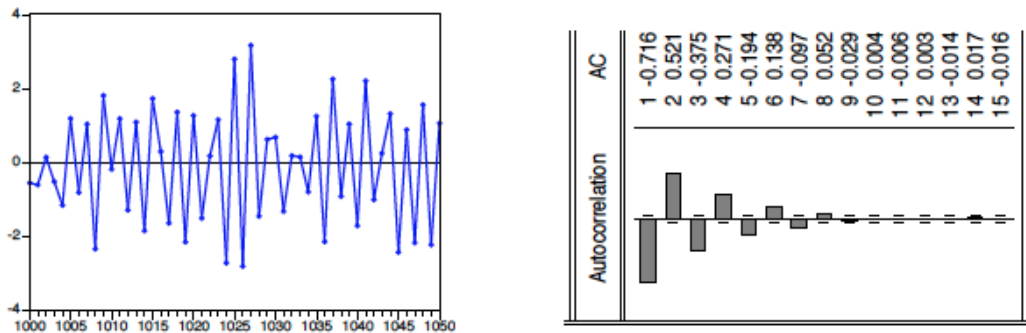


Figura 1.5: Ejemplo de un correlograma de una serie con alternancia y correlación de corto plazo

Capítulo 2

Los procesos estocásticos estacionarios y sus propiedades

2.1. Procesos estocásticos

Woodward [14, p. 3] define a un proceso estocástico como una sucesión de variables aleatorias que evolucionan en función de otra variable, es decir, tenemos $\{X(t); t \in T\}$ indexadas a un conjunto de índices T . Cuando esa otra variable es el tiempo, se hace referencia a ese proceso estocástico como una serie de tiempo.

Es importante recordar que una variable aleatoria, Υ , es una función definida en el espacio muestral Ω donde el rango son los números reales. Un valor observado de la variable aleatoria Υ es un número real $y = \Upsilon(\omega)$ para cada $\omega \in \Omega$. Entonces, una realización de la serie de tiempo $\{X(t); t \in T\}$ es el conjunto de resultados con valores reales $\{x(t, \omega); t \in T\}$ para cada valor fijo de $\omega \in \Omega$, es decir, una serie de tiempo de n datos es una muestra extraída (una realización) de un vector de n variables aleatorias $(x_{t_1}, x_{t_2}, \dots, x_{t_n})$ que forman parte de un proceso estocástico.

Si se considera que un proceso estocástico es un conjunto de variables aleatorias $\{X_t\}_{t \in T}$ indexadas a un conjunto de índices T sobre un espacio de probabilidad Ω y toma valores en \mathbb{R} , se tiene entonces que este proceso es la aplicación de dos argumentos, donde si se deja fijo a ω , tenemos la realización del proceso asociada a ω :

$$X : T \times \Omega \rightarrow \mathbb{R}$$

$$(t, \omega) \rightarrow X(t, \omega) = X_t(\omega)$$

2.1.1. Series de tiempo estacionarias

De acuerdo a Tsay [13, p. 25] una serie de tiempo $\{x_t\}$ es estrictamente estacionaria si su distribución conjunta $(x_{t_1}, \dots, x_{t_n})$ es idéntica a $(x_{t_1+h}, \dots, x_{t_n+h})$ para toda h , donde h es un entero positivo arbitrario y (t_1, \dots, t_n) es una colección de n enteros positivos arbitrarios. En otras palabras, un proceso es estrictamente estacionario cuando su distribución conjunta $(x_{t_1}, \dots, x_{t_n})$ es invariante a lo largo del tiempo. Además, se dice que el proceso estacionario es de primer orden o en media, si la función de medias es constante, es decir, que $\mathbb{E}(x_t) = \mu_t = k$, donde k es una constante y μ_t no depende de t .

Sin embargo, cabe destacar que este supuesto de invariabilidad a lo largo de tiempo es muy fuerte y aún más complicado de verificar empíricamente, por lo que se acepta una versión más débil de los procesos estocásticos estacionarios. Entonces, una serie de tiempo $\{x_t\}$ es débilmente estacionaria si tanto la media de x_t y la covarianza entre x_t y x_{t-h} son invariantes en el tiempo, donde h es un número entero arbitrario. Específicamente, $\{x_t\}$ es débilmente estacionario si (a) $\mathbb{E}(x_t) = \mu_t = k$, es decir, la función de medias es constante y (b) $Cov(x_t, x_{t-h}) = \gamma_h$, que sólo depende de h . De esta forma, un conjunto T de datos observados tiene estacionariedad débil si en el gráfico de tiempo los datos de este proceso fluctúan de forma constante en torno a un nivel fijo. Cabe agregar también, que a la estacionariedad débil también se le conoce como estacionariedad en covarianza.

Además, una condición implícita de las series de tiempo débilmente estacionarias es que se supone a los dos primeros momentos de x_t como finitos. Así, de acuerdo a las definiciones, si x_t es estrictamente estacionaria y sus dos primeros momentos son finitos, entonces x_t es también débilmente estacionaria.

2.2. Estimación de los momentos muestrales

Como se mencionó previamente, un objetivo principal de las series de tiempo es modelar la información de una serie de datos para obtener parámetros de dicha serie sobre la base de una realización. Por tanto, el interés es encontrar estimadores que nos brinden información sobre la media, la varianza y las funciones de covarianza y correlación. Respecto a la media y a la varianza, son frecuentemente usadas como estadísticas básicas para capturar las características de los fenómenos aleatorios, mientras que la función de autocovarianza es una herramienta para representar la relación entre los valores pasados y presentes de la serie de tiempo.

2.2.1. Estimación de la media

Para un proceso estocástico estacionario $\{X(t); t \in T\}$ y bajo la suposición de estacionariedad, la función de medias μ_t de una serie de tiempo se vuelve una constante y no depende del tiempo t , es decir, el proceso estocástico es estable en la media:

$$\mu_t = \mu = \mathbb{E}(x_t)$$

Ahora bien, dada una realización $\{x_t, t = 1, 2, \dots, n\}$ de una serie de tiempo estacionaria, el estimador natural de la media común μ es la media muestral:

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$$

Donde \bar{x} es un estimador insesgado de μ . Esto se puede demostrar fácilmente de la siguiente forma:

$$\mathbb{E}(\bar{x}) = \mathbb{E}\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(x_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu$$

2.2.2. Estimación de la varianza

Por su parte, la función de varianzas de un proceso estacionario $\{X(t); t \in T\}$ está dada por:

$$\sigma_t^2 = \text{Var}(x_t) = \mathbb{E}[(x_t - \mu)^2] = \sigma^2$$

Donde σ_t^2 es constante y finita para todo t . Entonces, si $\{x_t, t = 1, 2, \dots, n\}$ es una serie de tiempo discreta estacionaria en varianza, esta varianza constante poblacional σ^2 se puede estimar mediante la varianza muestral de tamaño n :

$$\text{Var}(x_t) = \hat{\sigma}^2 = \frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x})^2$$

Cabe recordar que en el caso de las *series de tiempo*, existen sucesiones de observaciones que podrían estar correlacionadas. Si la correlación es positiva, la $\text{Var}(x_t)$ tiende a subestimar a la varianza poblacional en el caso de series cortas (con n pequeño) debido a que observaciones sucesivas tienden a ser relativamente similares. Sin embargo, en la mayoría de los casos, esto no resulta ser un problema serio ya que disminuye el sesgo a medida que incrementa el tamaño n de la serie en función de qué tan correlacionados estén los datos, donde $\hat{\sigma}^2$ es un estimador asintóticamente insesgado de la varianza de la población [6, Cowpertwait, p. 32-33].

Ergodicidad

El supuesto de ergodicidad significa que los momentos muestrales, que son calculados sobre la base de una serie de tiempo con un número finito de observaciones, convergen (en algún sentido) cuando $n \rightarrow \infty$, en relación a los momentos correspondientes de la población. De manera más precisa, se dice que x_t es ergódico en la media si:

$$\lim_{n \rightarrow \infty} \mathbb{E}[(\bar{x} - \mu)^2] = 0$$

Y la varianza es ergódica si:

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[\left\{\frac{1}{n} \sum_{t=1}^n (x_t - \mu)^2 - \sigma^2\right\}^2\right] = 0$$

Estas condiciones se asumen y son llamadas como las “propiedades de consistencia” para variables aleatorias dependientes [8, Kirchgässner, p. 13].

2.2.3. Estimación de la autocovarianza y la autocorrelación

La función de autocovarianzas del proceso estacionario $\{X(t); t \in T\}$, suponiendo t fijo para todo μ_t (lo cual implica que $\gamma_{t_1, t_1+h} = \gamma_{t_2, t_2+h} = \gamma_h$), se define como:

$$\gamma_{t_1, t_2} = Cov(x_{t_1}, x_{t_2}) = \mathbb{E}[(x_{t_1} - \mu)(x_{t_2} - \mu)]$$

Donde en el caso de una serie de tiempo estacionaria la autocovarianza posee las propiedades de que: (a) γ_{t_1, t_2} depende solamente de $t_2 - t_1$, (b) $\gamma(0) = \sigma^2$ y (c) $|\gamma(h)| \leq \gamma(0)$.

Por su parte, la función de autocovarianza muestral $\hat{\gamma}_h$ que sirve para estimar γ_h es:

$$\hat{\gamma}_h = \gamma_{t, t+h} = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}), \quad h = 1, 2, \dots, n-1$$

En este caso, se utiliza un estimador insesgado como \bar{x} en lugar de μ , por lo tanto el estimador $\hat{\gamma}_h$ sólo es insesgado asintóticamente. Es importante considerar también que en el caso de esta función no es tan conveniente emplear un h tan grande, pues a medida que este número incrementa se tienen menos datos para estimar dicha autocovarianza.

Otro aspecto relevante de la función de autocovarianza es que depende mucho de las unidades (es decir, la escala) en las que se mide x_t , por lo que una alternativa es la función de autocorrelación (conocida como FAC), la cual se trata de una versión estandarizada de la covarianza y que en el caso del proceso $\{X(t); t \in T\}$ se describe por:

$$\rho_{t_1, t_2} = Cor(x_{t_1}, x_{t_2}) = \frac{\gamma_{t_1, t_2}}{\sqrt{\sigma_{t_1}^2, \sigma_{t_2}^2}}$$

Esta función tiene las propiedades de que $\rho_0 = 1$ y $|\rho_h| \leq 1$. Mientras que la función de correlación muestral se obtiene al estandarizar la autocovarianza:

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0} = \frac{\hat{\gamma}_h}{\hat{\sigma}^2}; \quad h = \pm 1, \pm 2, \dots, \pm n-1$$

La cual puede interpretarse como la correlación entre las observaciones de la misma variable en el tiempo x_t y el tiempo x_{t+h} [14, Woodward, p. 19]. Finalmente, al igual que el estimador de la autocovarianza, $\hat{\rho}_h$ es insesgado asintóticamente.

2.3. Propiedades espectrales de los modelos estacionarios

La mayoría de las series de tiempo presentan algunos patrones periódicos que pueden aproximarse por medio de la suma de ondas armónicas. Esto se hace mediante la descomposición de la serie en funciones trigonométricas en cada frecuencia. Para dicho fin, es necesario comprender algunos términos. El primero de ellos es la velocidad a la cual la serie oscila en términos del ciclo, donde un ciclo se define como un seno o coseno de onda definido durante un intervalo de tiempo de medida. La velocidad de la oscilación, *frecuencia*, se define como los ciclos por unidad de tiempo. Algunas veces, es más conveniente definir la frecuencia como radianes por unidad de tiempo λ [1 radian = $1/(2\pi)$ de un ciclo] y tenemos la relación:

$$\lambda = 2\pi f$$

Finalmente, el período T_0 de la oscilación de una serie se define como la longitud de tiempo requerido para un ciclo completo y se puede expresar como el recíproco de la frecuencia,

$$T_0 = 1/\lambda$$

Entonces, para un proceso estacionario se tiene una función de autocovarianza γ_h que decrece rápidamente mientras los rezagos se incrementan y además cumple la condición de que [9, Kitagawa, p. 31]:

$$\sum_{h=-\infty}^{\infty} |\gamma_h| < \infty$$

Entonces se puede definir una transformación de Fourier de γ_h (conocido como el teorema de representación espectral):

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_h e^{-ih\lambda} \geq 0, \quad \text{para todo } \lambda \in [-\pi, \pi]$$

La cual es llamada la función de densidad espectral de potencia o simplemente poder del espectro. Este poder del espectro representa una serie de tiempo en términos de funciones trigonométricas con varias frecuencias y expresa las características de la serie de tiempo con las magnitudes de estos componentes cíclicos. Esta función indica la contribución a la varianza de x_t por una frecuencia f . Además esta función satisface las siguientes propiedades:

1. f es inclusivo ($f(\lambda) = f(-\lambda)$), es decir, f es una función par.
2. $\gamma_h = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda = \int_{\pi}^{-\pi} \cos(h\lambda) f(\lambda) d\lambda$, la cual es una Ecuación de Euler en la que $i \int_{\pi}^{-\pi} \sen\lambda f(\lambda) d\lambda = 0$ por propiedades de la integral definida seno.

En la práctica tanto ρ_h como γ_h son desconocidas, por lo que evaluar la función de densidad espectral estimada \hat{f} puede ser difícil. Entonces, debe considerarse la siguiente función:

$$I(f) = \frac{1}{n} \left| \sum_{t=1}^n x_t e^{-itf_j} \right|$$

Donde $f_j = 2\pi j/n$, $j = 0, 1, \dots, [n/2]$. Esta ecuación se conoce como *periodograma* y es una estimación de la función de densidad espectral. A f_j se le conoce como las frecuencias de Fourier. Debe resaltarse que si el espectro del modelo ajustado es similar al periodograma basado en los datos, entonces el modelo capta la naturaleza del fenómeno, es decir, hay una relación directa entre el modelo y el espectro de potencia del proceso.

2.4. Teorema de Wold

Una importante postulación, principalmente para los modelos de series de tiempo lineales es el teorema de la descomposición de Wold, el cual hace hincapié en el papel central del proceso lineal general en el estudio de los procesos débilmente estacionarios (MA, AR y ARMA) que serán analizados más adelante [14, Woodward, p. 58].

Sea $\{x_t; t = 0, \pm 1, \pm 2, \dots\}$ un proceso débilmente estacionario con media cero, x_t puede ser expresado como la suma de dos procesos débilmente estacionarios con media cero, no correlacionados:

$$x_t = U_t + V_t$$

De tal forma que:

- (I) U_t es un proceso lineal general con un componente aleatorio (ruido blanco).
- (II) V_t es completamente determinado por una función lineal de sus valores pasados

El proceso de V_t es denominado generalmente como el componente “determinista” de x_t mientras U_t es el componente no determinista. Así, a partir de este Teorema se expresa a las series de tiempo débilmente estacionarias como una función lineal del choque de dos procesos no correlacionados, uno estocástico y otro determinista, entendiendo a la definición de “determinista” en el marco de los modelos lineales.

Capítulo 3

Algunos modelos estocásticos

3.1. El modelo básico: ruido blanco

El ejemplo más básico de un proceso estocástico estrictamente estacionario es un ruido blanco. Dicho proceso de tiempo discreto consiste en una secuencia de variables aleatorias independientes e idénticamente distribuidas, denotadas como a_1, a_2, \dots, a_t , las cuales se asume tienen media cero y varianza σ_a^2 . Por definición se deduce que este proceso posee media y varianza constantes y que hay independencia entre cada una de las variables a_t , por lo cual no están correlacionadas y la función de autocovarianza es simple:

$$\gamma_h = \text{cov}(a_t, a_{t+h}) = \begin{cases} \sigma_a^2 & h = 0 \\ 0 & h = \pm 1, \pm 2, \dots \end{cases}$$

Mientras que la función de autocorrelación está dada por:

$$\rho_h = \begin{cases} 1 & h = 0 \\ 0 & h = \pm 1, \pm 2, \dots \end{cases}$$

Como la media y función de autocovarianzas no dependen del tiempo, el proceso es estacionario de segundo orden, es decir, es débilmente estacionario. Esta situación hace que este proceso pueda ser referido como un proceso de ruido blanco en el sentido débil y que algunas referencias hacia él se centren en sus propiedades de segundo orden, pero de hecho, la hipótesis de independencia implica que el proceso es también estrictamente estacionario.

Este proceso de ruido blanco es fundamental pues además de que posee las propiedades básicas de los modelos estacionarios, éste juega un papel muy importante en la construcción de procesos con propiedades mucho más complicadas a través de operaciones de filtros lineales [2, Box & Jenkins, p. 29].

3.2. Modelos lineales de series de tiempo

3.2.1. Procesos lineales generales

Muchas de las series de tiempo se pueden expresar como la salida (*output*) de un filtro lineal con entrada (*input*) de un ruido blanco. Entonces, se define que un proceso estocástico $\{x_t\}_{t \in T}$ es un *proceso lineal general* si para todo $t \in \mathbb{Z}$ se puede representar como

$$x_t - \mu = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

Donde el primer término $x_t - \mu$ es la desviación del proceso de su media, $\{a_t\}_{t \in T}$ es un proceso de ruido blanco con media cero y varianza finita, que en la literatura suele conocerse también como *innovación* o *shock*, y finalmente, $\sum_{j=0}^{\infty} |\psi_j| < \infty$ es una sucesión de constantes reales absolutamente sumables.

Para las series de tiempo lineales de la ecuación anterior, la estructura dinámica de x_t está regida por ψ_j , los cuales son llamados *pesos* ψ de x_t . Si x_t es débilmente estacionaria, se puede obtener la media y la varianza fácilmente utilizando la independencia de $\{a_t\}$ como:

$$\mathbb{E}(x_t) = \mu, \quad \text{Var}(x_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty$$

Donde σ_a^2 es la varianza del proceso a_t . Como $\text{Var}(x_t) < \infty$, por teoría de la medida¹ se tiene que $\{\psi_j^2\}$ es convergente a la serie, esto es, $\psi_j^2 \rightarrow 0$ y $j \rightarrow \infty$. Esta última restricción implica que el valor presente depende de forma convergente del proceso aleatorio de ruido blanco [13, Tsay, p. 31], es decir, la influencia de la parte aleatoria a_{t-j} va desapareciendo de x_t conforme se incrementa j (se aleja en el pasado). A partir de ello, podemos determinar que la autocovarianza del rezago h es:

$$\begin{aligned} \gamma_h = \text{Cov}(x_t, x_{t+h}) &= \mathbb{E} \left[\left(\sum_{j=0}^{\infty} \psi_j a_{t-j} \right) \left(\sum_{k=0}^{\infty} \psi_k a_{t+h-k} \right) \right] = \mathbb{E} \left(\sum_{j,k=0}^{\infty} \psi_j \psi_k a_{t-j} a_{t+h-j} \right) \\ \gamma_h &= \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \mathbb{E}(a_{t+h-j}^2) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \end{aligned}$$

Consecuentemente, los pesos ψ están relacionados a las autocorrelaciones de x_t de la siguiente forma:

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+h}}{\sum_{j=1}^{\infty} \psi_j^2}, \quad h \geq 0,$$

Con lo cual, se destaca que la labor en estos *procesos lineales generalizados* es describir el patrón de los pesos ψ de x_t . Esto es de suma relevancia, pues como se verá más adelante, los procesos *AR*, *MA* y *ARMA* son casos especiales de este modelo.

¹Como esta teoría rebasa los límites de estudio de esta tesina, se asume como cierta. Ello implica que $\mathbb{E}(\prod_{i=0}^{\infty} x_i) = \prod_{i=0}^{\infty} \mathbb{E}(x_i)$

3.2.2. Procesos de medias móviles

Un proceso de promedios móviles de orden finito q , $MA(q)$, es una sucesión X_t cuyo modelo se compone como sigue:

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

Considerando que a_t es un proceso de ruido blanco con variables aleatorias independientes e idénticamente distribuidas, teniendo media cero y varianza σ_a^2 ; μ es la media del proceso MA y las $\{\theta_j\}$ son constantes reales, a las cuales se les conoce como los parámetros de promedios móviles e indican la intensidad de efecto del evento.

Un proceso de MA con media centrada, $\{x_t - \mu\}$, posee las siguientes características:

1. $\mathbb{E}(x_t) = 0$
2. $Var(x_t) = \sigma_a^2 \sum_{q=1}^q \theta_q^2$
3. $\gamma_h = Cov(x_t, x_{t+h}) \begin{cases} 0 & h > q - 1 \\ \sigma_a^2 \sum_{i=1}^{q+h} \theta_i \theta_{i-h} & h = 1, \dots, q - 1 \end{cases}$
4. $\rho_h = \begin{cases} \sum_{i=1}^{q+h} \theta_i \theta_{i-h} / \sum_{i=1}^{q+h} \theta_i^2 & h = 1, \dots, q - 1 \\ 0 & \text{otro caso} \end{cases}$

Es importante notar, que en el caso de la autocovarianza y la autocorrelación, la serie no depende de t , es decir, no depende del tiempo en el que la observación se encuentre sino más bien de qué tanto están separadas las observaciones, por lo que el modelo de orden q , $MA(q)$ está correlacionado únicamente con las últimas q observaciones y es un proceso estacionario de segundo orden, es decir, un proceso $MA(q)$ es siempre estacionario. Finalmente, otro aspecto importante en los procesos de medias móviles es la condición de invertibilidad, la cual supone que la correlación entre una variable y su pasado va reduciéndose a medida que se incrementa t , es decir, que el proceso es ergódico. Para el caso del proceso de media móvil más sencillo, el $MA(1)$, se obtiene esta condición a partir del polinomio de dicha media móvil:

$$x_t = a_t - \theta a_{t-1}$$

Este polinomio puede expresarse por medio de los operadores retraso y diferencia (*ver Apéndice A*). En el caso del primero, el operador retraso denotado por B , actúa sobre un proceso x_t de la siguiente forma:

$$Bx_t = x_{t-1} \quad B^2x_t = B(Bx_t) = Bx_{t-1} = x_{t-2} \quad \dots \quad B^jx_t = x_{t-j}$$

A partir de este operador retraso, entonces se define al operador diferencia:

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

Entonces, el polinomio de la media móvil está dado por:

$$\theta_1(B) = (1 - \theta B)$$

Se obtienen las raíces resolviendo la ecuación:

$$1 - \theta B = 0 \quad \text{despejando} \quad B = \frac{1}{\theta}$$

Entonces, la condición de invertibilidad para el proceso $MA(1)$ está dado por:

$$|B| = \left| \frac{1}{\theta} \right| > 1 \quad \text{o lo que es lo mismo} \quad |\theta| < 1$$

3.2.3. Procesos autorregresivos

Otra categoría de modelos que se utiliza comúnmente son los modelos autorregresivos (AR). Estos modelos tienen la particularidad de que se asemejan mucho al modelo de regresión lineal tradicional, pues cuando se reemplaza el predictor en el modelo clásico de regresión por los últimos rezagos de la serie de tiempo, se obtiene un modelo AR. Debido a esto, la mayoría de los resultados estadísticos obtenidos mediante la regresión clásica se pueden generalizar al caso AR, razón por la cual este modelo es uno de los más utilizados en el análisis de series de tiempo [3, Chan, p. 25].

Entonces, se dice que un proceso $\{x_t\}$ es un autoregresivo de orden p , denotado por $AR(p)$, si para un número entero $p \geq 1$ y ϕ_1, \dots, ϕ_p constantes reales (conocidas como coeficientes autoregresivos) se tiene que:

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

Donde $\{a_t\}$ es un proceso de ruido blanco con media cero, varianza σ_a^2 e i.i.d. Los procesos $AR(p)$ pueden ser o no estacionarios, dependiendo de los valores de sus coeficientes autorregresivos sobre los cuales se hacen pruebas de contraste llamadas *pruebas de raíz unitaria* de las que se hablará más adelante.

Finalmente si se traslada el proceso a la media, $\{x_t - \mu\}$, el modelo $AR(1)$ tiene las siguientes propiedades:

1. $\mathbb{E}(x_t) = 0$
2. $Cov(x_t, x_{t+h}) = \gamma_{(x_t, x_{t+h})} = \frac{\sigma_a^2 \phi^h}{(1-\phi^2)}$
3. $\gamma_0 = Var(x_t)$
4. $\rho_h = \frac{\gamma_h}{\gamma_0} = \phi^h$

Con lo anterior, se concluye que un proceso AR está determinado únicamente por su función de autocorrelación, la cual decrece de forma exponencial y alterna el signo si $\phi < 0$. Además, este proceso cumple la condición de estacionariedad cuando $|\phi| < 1$. En caso contrario, se denomina a este proceso como “caminata aleatoria” y donde $|\phi| = 1$.

3.2.4. ARMA

Los procesos autorregresivos de medias móviles, $ARMA(p,q)$ determinan x_t en función de su pasado hasta el retardo p , de la innovación contemporánea y el pasado de la innovación hasta el retardo q , es decir, son la combinación de un modelo $AR(p)$ y un $MA(q)$ con la intención de obtener una generalización de los procesos anteriores. La gran utilidad de los procesos ARMA es que la estructura compleja de una serie de tiempo estacionaria puede ser descrita completamente por éste.

Entonces, un modelo $ARMA(p,q)$ se define por:

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

Del misma forma que en los modelos AR y MA, el proceso se centra en $\mu = 0$ sin perder generalidad. Cuando $p \neq 0$ y $q \neq 0$ el modelo resultante es referido como modelo mixto.

Un proceso $ARMA(p, q)$ puede ser formalmente escrito en la forma lineal general, como una MA de orden infinito y en forma de operadores de retraso, siempre y cuando se cumplan las condiciones de estacionariedad e invertibilidad de este proceso, quedando de la siguiente forma:

$$x_t = \phi^{-1}(B)\theta(B)a_t = \psi(B)a_t$$

Y en forma de un AR de orden infinito:

$$\theta^{-1}(B)\phi(B)x_t = a_t$$

3.3. Modelos heterocedásticos condicionales

3.3.1. Procesos ARCH

En el caso de los modelos $ARMA$, un requisito indispensable era que el componente aleatorio fuese un proceso de ruido blanco, es decir, un proceso no correlacionado e idénticamente distribuido con media cero. Sin embargo, no es raro encontrar series en las que dicho ruido es más variable o volátil durante ciertos intervalos de tiempo respecto de otros. Esto se debe a que no existe una total independencia en las observaciones de la serie de tiempo x_t , y que por tanto, la información del pasado de x_t influye en el futuro. A partir de esta premisa, surgen los modelos $ARCH(p)$ introducidos por Engle en 1982², los cuales buscan relajar la hipótesis de normalidad permitiendo tener una serie no correlacionada pero formada por variables dependientes.

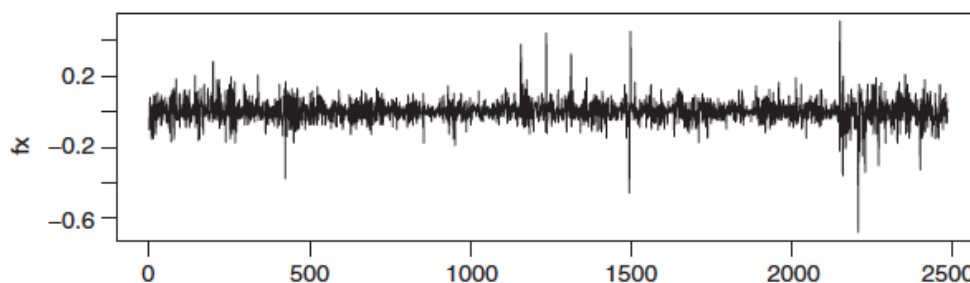


Figura 3.1: Ejemplo de una serie con dependencia en las observaciones: cambio porcentual en el tipo de cambio de marcos alemanes por dólares medidos en intervalos de 10 minutos en junio 1989

Shumway [12, p. 281] define el modelo $ARCH(p)$ como:

$$x_t = \sigma_t a_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i x_{t-i}^2$$

Donde a_t usualmente se asume como $a_t \sim N(0, 1)$, independiente e idénticamente distribuida (ruido blanco). Ahora, si $\mathcal{F}_{t-1} = \sigma(x_{t-1}, x_{t-2}, \dots, x_{t-n})$ representa el conjunto de sigmas generados por información pasada hasta el tiempo $t - 1$.

²Este tipo de modelos fue introducido por Robert F. Engle con el artículo *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations* en la Revista *Econometrica* número 50, páginas 987-1007.

Entonces:

$$\text{Var}(x_t^2 | \mathcal{F}_{t-1}) = \mathbb{E}(x_t^2 | \mathcal{F}_{t-1}) = \mathbb{E}(\sigma_t^2 a_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2 \mathbb{E}(a_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2$$

Esta identidad implica que la varianza condicional de la serie x_t evoluciona de acuerdo a valores previos de x_t^2 como un modelo $AR(p)$: de ahí el nombre de modelo autorregresivo condicional heteroscedástico o modelo de varianza condicional. De la misma forma, las condiciones necesarias para tener una ecuación bien definida son que:

$$\alpha_i \geq 0 \text{ para } i = 0, \dots, p \quad \text{y} \quad \alpha_1 + \dots + \alpha_p < 1$$

El ejemplo más sencillo de un proceso ARCH, es el ARCH(1) que se representa como:

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2$$

El $ARCH(1)$ cuenta con las siguientes propiedades [3, Chan, p. 102]:

1. $\mathbb{E}(x_t) = \mathbb{E}[x_t | \mathcal{F}_{t-1}] = \mu = 0$
2. $\text{Var}(x_t) = \frac{\alpha_0}{1-\alpha_1}$, si $0 \leq \alpha_1 < 1$ y $\text{Var}[x_t | \mathcal{F}_{t-1}] = \sigma_t^2$
3. $\text{Cov}(x_t, x_{t-h}) = 0$
4. $\{f(x_t)\}$ tiene colas más pesadas que una función de densidad normal $(0, \sigma^2)$

Lo cual implica que el ARCH(1) es un ruido blanco centrado en 0 con colas más pesadas, donde es más probable que aparezcan outliers en comparación con un ruido blanco Gaussiano. De esta forma, se puede decir:

$$x_t \sim ARCH(p) \Rightarrow x_t^2 \sim AR(p)$$

3.3.2. Procesos GARCH

Similar a la extensión de un modelo AR en un modelo $ARMA$, se puede extender la noción de un modelo ARCH a un modelo ARCH generalizado (GARCH). Específicamente, un modelo GARCH se puede expresar de la siguiente manera:

$$x_t = \sigma_t a_t \quad a_t \sim N(0, 1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 + \sum_{j=1}^q \alpha_j x_{t-j}^2$$

Donde $\alpha_0 > 0$ y $\alpha_j \geq 0$ con $j \in \{1, 2, \dots, q\}$ y $\beta_i \geq 0$ con $i \in \{1, 2, \dots, p\}$, que son las condiciones que los coeficientes de α y β deben cumplir para la varianza. De manera análoga con el modelo ARCH(p) y el $AR(p)$:

$$\{x_t\}_t \sim GARCH(p, q) \Rightarrow \{x_t^2\}_t \sim ARMA(p, q)$$

puede aproximarse a un proceso ARMA(1,1) [13, Tsay, p. 114]. Finalmente, el GARCH(p,q) comparte la propiedad de los modelos ARCH de tener colas más pesadas en comparación con la distribución normal.

3.4. Modelos no estacionarios

En la práctica es muy común encontrar series con diferentes tipos de componentes que producen que la serie de tiempo en cuestión sea no estacionaria. En términos generales, una serie de tiempo no estacionaria puede presentar un cambio sistemático en la media o en la varianza, produciendo que estos estadísticos no sean constantes y finitos. De la misma forma, también es posible que las autocovarianzas dependan del tiempo y que por tanto, muchos de los modelos desarrollados no puedan ser aplicados. Sin embargo, existen modelos que funcionan como alternativa para procesos que no cumplen con el supuesto de estacionariedad. En ese sentido, se abundará más en este apartado en aquellos modelos donde no se cumple el requisito de la estacionariedad cuando la fuente del problema es una tendencia estocástica (no confundir con la tendencia determinista explicada en el *Capítulo 1*).

Asimismo, también existen métodos para realizar transformaciones a la serie de tiempo que permitan a esta cumplir con el supuesto de estacionariedad y ajustarse a los modelos antes descritos (ver *Apéndice A*)

3.4.1. Caminata aleatoria

El modelo de caminata aleatoria (o paseo aleatorio) es un modelo $AR(1)$ con parámetro $\phi=1$, que usualmente se utiliza para describir el comportamiento del mercados de valores [14, Woodward, p. 192]. Sea x_t , con $t = 1, 2, \dots$ una caminata aleatoria denotada por:

$$x_t = x_{t-1} + a_t$$

Donde $x_0 = 0$ y a_t es un proceso ruido blanco i.i.d con media cero y varianza σ_a^2 . El nombre de “caminata aleatoria” viene del hecho de que para el tiempo $t - 1$, el movimiento del tiempo t está totalmente determinado por la cantidad aleatoria a_t , de tal suerte que a veces resulta conveniente escribir el modelo de la forma:

$$x_t = \sum_{h=1}^t a_h$$

Para $t = 1, 2, \dots$, donde se puede apreciar que $\mathbb{E}(x_t) = 0$ y la $Var(x_t) = t\sigma_a^2$, mediante lo cual se deduce que la caminata aleatoria es un proceso no estacionario, pues $Var(x_t)$ no es constante y depende del tiempo. Entonces, sus respectivas funciones de autocovarianza y de autocorrelación están dadas por:

$$Cov(x_{t_1}, x_{t_2}) = t_1\sigma_a^2 \quad \rho_x(t_1, t_2) = \sqrt{\frac{t_1}{t_2}}$$

Caminata aleatoria con deriva

Una extensión del modelo anterior es la caminata aleatoria con deriva, que está dada por:

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

Donde α es una constante diferente de cero llamada *drift* y al igual que en el modelo anterior $x_0 = 0$. Basándose en el modelo de caminata aleatoria, para el tiempo $t - 1$, el movimiento del tiempo t está determinado por α más la parte aleatoria a_t . Además la media es $\mathbb{E}(x_t) = \alpha t$ y, entonces, la media del proceso se incrementa o disminuye en el tiempo dependiendo de si α es positivo o negativo, respectivamente.

A menudo la caminata aleatoria con deriva es confundida con la tendencia determinista, pues ambos incluyen deriva y

un componente de ruido blanco, sin embargo, la diferencia fundamental es que en el caso de la caminata aleatoria el valor del tiempo t está correlacionado con el valor del periodo anterior x_{t-1} , mientras que en el caso de la tendencia determinista está correlacionada con la tendencia a lo largo de tiempo βt . Esto significa que la media del proceso no estacionario con tendencia determinista crece alrededor de una tendencia fija, la cual es constante e independiente del tiempo.

3.4.2. ARIMA

El modelo denominado Autorregresivo Integrado de Medias Móviles de orden (p,d,q) o ARIMA (p, d, q) es un modelo que considera la no estacionariedad de las series de tiempo, por lo tanto considera el remover las otras fuentes de variación mediante el proceso de diferenciación para obtener una serie estacionaria. Este modelo se describe de la siguiente forma:

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

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t + \dots + \theta_q a_{t-q}$$

$$\phi(B)w_t = \theta(B)a_t$$

Con lo cual se puede notar que un modelo ARIMA (p,d,q) es un modelo ARMA (p,q) aplicado a una nueva serie diferenciada d -veces.

Capítulo 4

Modelación ARIMA

La idea básica en los modelos de series de tiempo es poder construir una estructura que permita explicar el comportamiento de la serie y a partir de ello predecir valores futuros. Estos modelos, al igual que el resto de los modelos en la ciencia matemática, emplean ciertas constantes o parámetros cuyos valores deben ser estimados a partir de los datos, siendo sumamente importante que se emplee el menor número posible de parámetros para las representaciones adecuadas, pues esto además de significar un modelo más sencillo implica que pocos parámetros son capaces de explicar cantidades mayores de información. Esta característica se conoce como el *principio de parsimonia* y desarrolla un papel central en la construcción de modelos matemáticos.

En ese sentido, el objetivo dentro de esta etapa de modelación es la obtención de modelos adecuados pero parsimoniosos, puesto que los procedimientos de predicción pueden ser gravemente deficientes si estos modelos son insuficientes o no hacen un uso adecuado de los parámetros. Es por ello, que Box & Jenkins [2, p. 16] proponen un proceso de selección necesariamente iterativo, es decir, un proceso de evolución-adaptación, que considere los escenarios de ensayo y error, el cual se explicará a continuación.

4.1. Metodología Box-Jenkins

La construcción de los modelos *ARIMA* se lleva a cabo mediante un proceso en el que se pueden distinguir cuatro etapas como puede observarse en la *Figura 4.1*.

La primer etapa considera el proceso de identificación de los órdenes p, d, q apropiados para reproducir las características de la serie bajo estudio, tomando en cuenta que es posible identificar a más de un modelo como el posiblemente idóneo, por lo que una vez que se ha identificado la familia del modelo, a partir del principio de parsimonia se elige uno solo. Otra parte muy importante dentro de esta etapa es:

1. Buscar que la estructura sea estacionaria, por lo que si la serie es no estacionaria se filtran las partes de tendencia y/o estacionalidad, para obtener sólo la parte estacionaria si es que la hay.
2. Una vez obtenida la parte estacionaria se buscará el modelo *ARMA* que mejor ajuste los datos.

Posteriormente, en la etapa de estimación, una vez que se ha logrado un uso óptimo de los datos, se realiza inferencia sobre los parámetros condicionada a que el modelo investigado sea apropiado. En la siguiente etapa, la validación, se

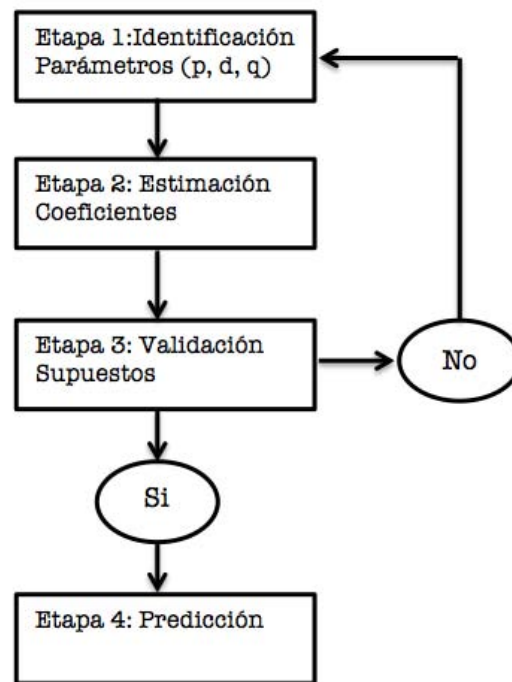


Figura 4.1: Diagrama del método iterativo de Box y Jenkins

realizan contrastes de diagnóstico para comprobar si el modelo se ajusta a los datos, o de no ser así, encontrar las discrepancias con la intención de mejorarlo. Finalmente, una vez que se han desarrollado correctamente las etapas anteriores, se realiza una validación cruzada para determinar si el modelo es adecuado o no. Esta técnica se utiliza para evaluar los resultados del modelo ARIMA con el objetivo de garantizar que el modelo es independiente y cumple con las características adecuadas para predecir valores. Una forma muy usual de emplear esta técnica en este tipo de modelos, es extraer las últimas observaciones y contrastar la capacidad predictiva del modelo con los datos reales extraídos.

A continuación se abunda más sobre cada una de estas etapas y los instrumentos que se utilizan en cada una de ellas con tal de encontrar el mejor modelo, recordando que esta metodología se basa completa y fundamentalmente en los principios de selección de *forma iterativa* y el principio de parametrización, *parsimonia*.

4.2. Identificación del modelo

El objetivo de esta etapa de la modelación, como se mencionó anteriormente, es seleccionar el modelo $ARIMA(p, d, q)$ apropiado para la serie, es decir, que reproduce de la mejor forma las características de la serie. Esta identificación del modelo se lleva a cabo en dos fases antes mencionadas, que son el análisis de la estacionariedad y la elección de los órdenes p y q .

4.2.1. Instrumentos de identificación del modelo

La función de autocorrelación y la función de autocorrelación parcial

Un instrumento sumamente fundamental en la etapa de identificación es la estructura de dependencia temporal del proceso estocástico recogida en la función de autocorrelación (FAC) y en la función de autocorrelación parcial (FACP), pues en estas funciones se recoge las características dinámicas del proceso estacionario.

En el caso de la FAC, los coeficientes de autocorrelación muestral de x_t son:

$$\hat{\rho}_h = \frac{\sum_{t=1}^{n-h} (x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad h = \pm 1, \pm 2, \dots, \pm n - 1$$

Para identificar los ordenes p y q , se compararán las funciones de autocorrelación muestrales con las FAC teóricas de los modelos ARMA cuyas características se conocen:

	FAC
MA(q)	Se anula para $j > q$
AR(p)	Decrecimiento rápido
ARMA(p, q)	Decrecimiento rápido

Si el correlograma muestral de la serie x_t presenta un corte a partir de un rezago finito j , se identifica con la FAC teórica de un MA(j). Pero si el correlograma muestral no presenta ningún corte sino que parece decrecer rápidamente siguiendo una estructura exponencial o de onda seno-coseno, la identificación no es tan clara, pues dicho comportamiento corresponde tanto para un modelo teórico AR o ARMA de cualquier orden. En ese sentido, una función complementaria que ayuda en la selección del modelo es la función de autocorrelación parcial, la cual justamente permite identificar de forma sencilla el orden p de un proceso AR.

El coeficiente de autocorrelación parcial de orden h , denotado por $\phi_{h,h}$, mide el grado de asociación lineal existente entre las variables x_t y x_{t+h} una vez ajustado el efecto lineal de todas las variables intermedias y aprovechando el hecho de que un proceso AR(p) tiene infinitas funciones de autocorrelación, se puede describir en términos de coeficientes de la siguiente regresión lineal [2, Box & Jenkins, p. 67-68]:

$$x_t = \alpha + \phi_{1,1}x_{t-1} + \phi_{1,2}x_{t-2} + \dots + \phi_{h,h}x_{t-h} + e_t$$

En cuanto a sus propiedades, la FACP posee propiedades equivalentes a la FAC como el hecho de que $\phi_0 = 1$ y $|\phi_{h,h}| \leq 1$. Además, la función de autocorrelación parcial se puede estimar a partir de los datos de la serie como una función de los coeficientes de autocorrelación simples estimados, $\hat{\rho}_h$.

Comparando la estructura de las funciones de autocorrelación simple y parcial estimadas con las características básicas de las funciones de autocorrelación teóricas de la tabla siguiente, se puede identificar el o los procesos que podrían haber generado la serie bajo estudio mediante un proceso iterativo como se muestra en la *Figura 4.2*:

	FAC	FACP
MA(q)	Se anula para $j > q$	Decrecimiento rápido
AR(p)	Decrecimiento rápido	Se anula para $j > p$
ARMA(p, q)	Decrecimiento rápido	Decrecimiento rápido

Finalmente, para decidir cuando $\hat{\rho}_h$ y $\hat{\phi}_{h,h}$ son estadísticamente distintos de cero, se utilizan las propiedades asintóticas de los estimadores y con un nivel de significancia de $\alpha = 0.05$ se tiene que:

A. Para un proceso AR(p):

$$h > p \quad \text{y} \quad \left| \hat{\phi}_{h,h} \right| \geq 1.96 \left(\frac{1}{\sqrt{n}} \right) \Rightarrow \hat{\phi}_{h,h} \text{ es estadísticamente diferente de cero.}$$

B. Para un proceso MA(q):

$$h > q \quad \text{y} \quad \left| \hat{\rho}_h \right| \geq 1.96 \left(\frac{1}{\sqrt{n}} \right) \Rightarrow \hat{\rho}_h \text{ es estadísticamente diferente de cero.}$$

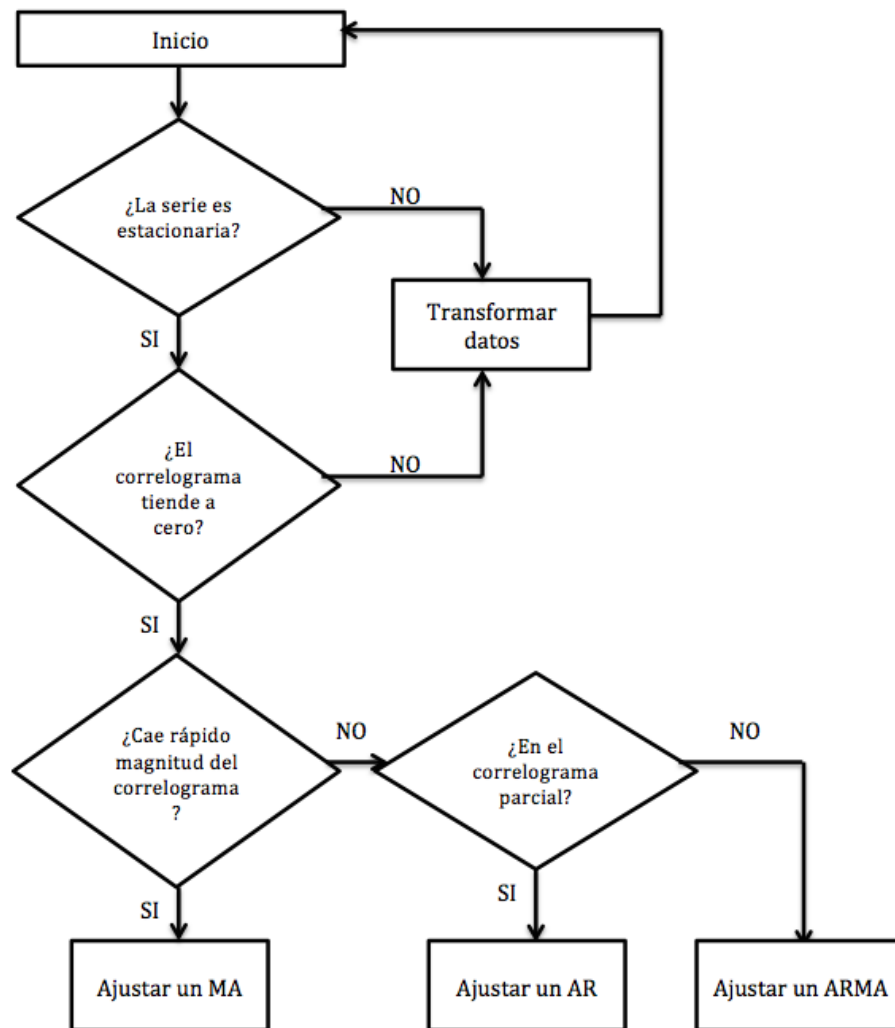


Figura 4.2: Proceso iterativo para identificar el modelo a ajustar

De esta forma, a partir de la FAC y FACP, se determina conjuntamente los parámetros del modelo tomando en cuenta las *regiones admisibles* de los parámetros. Dichas regiones, se obtienen como consecuencia de las condiciones de estacionariedad e invertibilidad (el efecto de los x_{t-h} sobre x_t debe tender a 0 cuando $h \rightarrow \infty$) de los respectivos modelos. Por ejemplo, de acuerdo a Guerrero [7, p. 120], un modelo $AR(1)$ es estacionario si $-1 < \phi < 1$, por lo tanto la región admisible se representa gráficamente en el intervalo siguiente:

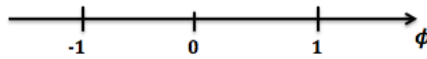


Figura 4.3: Región admisible para un modelo $AR(1)$

De igual manera, un modelo $AR(2)$ será estacionario si satisface las condiciones $\phi_2 - \phi_1 < 1$, $\phi_2 + \phi_1 = 1$ y $-1 < \phi_2 < 1$, de aquí que la región admisible venga a ser un triángulo donde también hay una curva definida por $\phi_1^2 + 4\phi_2 = 0$, la cual divide los valores de la ecuación característica $1 - \phi_1x - \phi_2x^2 = 0$ como se observa en la *Figura 4.4*:

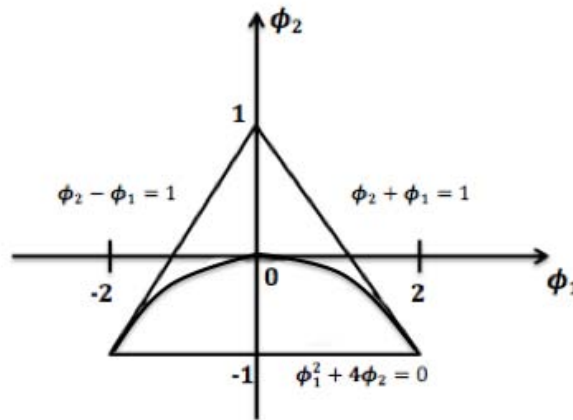


Figura 4.4: Región admisible para un modelo $AR(2)$

La función de autocorrelación parcial extendida

De acuerdo a Tsay [13, p. 59] ni la FAC ni la FACP contienen suficiente información como para determinar el orden de un modelo ARMA, razón por la cual él propone una nueva aproximación mediante el uso de la función de autocorrelación parcial extendida (FACPE). La idea básica es una generalización de la ecuación Yule-Walker que comprende dos pasos: primero estimar el componente $AR(p)$ para de forma subsecuente transformar la serie ARMA en un proceso $MA(q)$ puro, del cual se pueda determinar su respectivo orden a partir de la FAC.

La salida de una FACPE es una tabla de dos vías, donde las filas corresponden a un AR de orden p y las columnas un MA de orden q .

En la *figura 4.5* se muestra el ejemplo de un modelo $ARMA(1, 1)$, donde se puede ver que la característica clave de la tabla es que contiene un triángulo de O 's con el vértice superior izquierdo situado en el orden $(1, 1)$. Esta característica se utiliza para identificar el orden de un proceso $ARMA(p, q)$, donde se debe ubicar el respectivo triángulo de O 's con su vértice superior izquierdo en la posición (p, q) .

FAC	FACP	FACPE
Decrecimiento rápido	Decrecimiento rápido	un triángulo con vértice (p, q)

AR	MA							
	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	O	O	O	O	O	O	O
2	*	X	O	O	O	O	O	O
3	*	*	X	O	O	O	O	O
4	*	*	*	X	O	O	O	O
5	*	*	*	*	X	O	O	O

X significa no cero, O significa cero y * es otro número que podría ser cero o no cero, pero que no es relevante pues no juega papel alguno den la determinación del orden (1,1)

Figura 4.5: Tabla de una FACPE teórica para un $ARMA(1, 1)$

Akaike y otros criterios

Dentro de la etapa de identificación puede suceder que para una serie de tiempo observada existan varios modelos que la representen bien, por esta razón, existen otros criterios alternativos a la FAC y a la FACP que permiten establecer que modelo resulta mejor en términos de parsimonia y carácter explicativo. Uno de ellos es el criterio de información de Akaike (AIC). Este criterio consiste en penalizar la logverosimilitud (logaritmo natural de la función de máxima verosimilitud) del modelo, condicionado a las primeras m observaciones, más un factor que involucra la cantidad de parámetros estimados:

$$AIC = -2 \ln L(x_{m+1}, \dots, x_n | x_1, \dots, x_m) + \frac{2(p + q + 1)}{n}$$

Esta técnica toma como mejor modelo a aquel que presente el AIC más pequeño. Además, utilizando el AIC, la diferencia entre dos valores es significativa si es mayor a $\chi_{s,\alpha}^2$, donde s representa al número de parámetros más que un modelo tiene con respecto al otro, con lo cual se penaliza al modelo candidato a partir del número de parámetros con los que cuenta [13, Tsay p. 42], asumiendo que el modelo cumple con los supuestos necesarios en los residuales (*ver sección “Diagnóstico del modelo”*).

Un problema con el uso del AIC es que a medida que se incrementa el número de realizaciones, el AIC tiende a sobreestimar los órdenes del modelo, es decir, selecciona modelos con muchos parámetros [14, Woodward, p. 283]:. Es por ello, que muchos autores sugieren modificaciones al AIC para ajustar este criterio a dichos problemas. Las dos alternativas más populares son el AICC (Hurvich y Tsai, 1989) y el BIC (Criterio de información Schwarz, 1978) [14, Woodward, p. 284]:

$$AICC = \ln(\hat{\sigma}_a^2) + \frac{n + p + q + 1}{n - p - q - 3} \quad y \quad BIC = \ln(\hat{\sigma}_a^2) + (p + q + 1) \frac{\ln(n)}{n}$$

Estudios de simulaciones han mostrado que el BIC tiene un mejor comportamiento para muestras con un gran número realizaciones n , mientras que el AICC generalmente se tiene un mejor comportamiento con muestras n más pequeñas cuando la cantidad de parámetros es relativamente grande en ambos casos.

4.2.2. Identificación de modelos no estacionarios

Búsqueda de ruido blanco

En la búsqueda de un modelo parsimonioso, la investigación siempre debe comenzar con la decisión concerniente a si el modelo se puede o no se puede modelar adecuadamente usando el más simple de los modelos ARMA, el ruido blanco. Esto debido a que el análisis puede arrojar modelos innecesariamente complicados cuando desde un principio el proceso podía ser modelado fácilmente mediante un proceso de ruido blanco.

La forma de identificar si la serie puede explicarse mediante un proceso de ruido blanco, es examinando gráficamente las realizaciones y las autocorrelaciones de la muestra, como se muestra en la *Figura 4.6*:

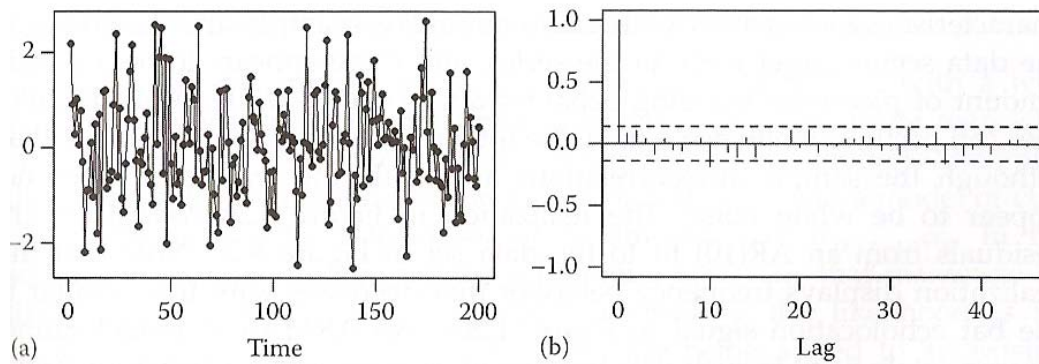


Figura 4.6: Realizaciones y correlograma de un proceso de ruido blanco

Como se observa, no hay ningún comportamiento cíclico o tendencia notable en los datos, y si bien las verdaderas autocorrelaciones para un proceso blanco tienen la propiedad de que $\rho_h = 0$ para cada $h \neq 0$, las autocorrelaciones de la muestra no son idénticas a cero. Esto se debe a que el estimador $\hat{\rho}_h$ es un estimador asintóticamente insesgado de ρ_h dado que $\mathbb{E}(\hat{\rho}_h) \approx 0$ para cada $h \neq 0$ y realización de tamaño n , lo cual implica que:

$$Cov(\hat{\rho}_h, \hat{\rho}_{h+r}) \approx 0, r \neq 0 \quad y \quad Var(\hat{\rho}_h) = \frac{1}{n}, h \neq 0$$

Por lo tanto, si una realización dada es a partir de un proceso de ruido blanco, entonces las autocorrelaciones de la muestra deben ser pequeñas y aproximadamente no correlacionadas a los diferentes rezagos. Para una h dada a un nivel de significancia $\alpha = 0.05$, se tiene:

$$|\hat{\rho}_h| > 1.96 \left(\frac{1}{\sqrt{n}} \right)$$

Estabilidad en media y varianza

En el tratamiento de las series de tiempo, se puede encontrar que una serie es no estacionaria tanto porque su media o su varianza no son constantes en el tiempo. Cuando una serie no es estacionaria en varianza, es decir, no se puede sostener el supuesto de que ha sido generada por un proceso con varianza constante en el tiempo, la solución es transformar la serie mediante algún método que estabilice la varianza (ver *Apéndice A*). Si suponemos que la varianza del proceso la podemos expresar como alguna función del nivel:

$$\text{Var}(x_t) = kf(\mu_t)$$

Siendo k una constante positiva y f una función conocida. El objetivo es conseguir alguna función que transforme la serie de forma que $h(x_t)$ tenga varianza constante.

Por su parte, en el caso de la media, cuando esta no es estacionaria generalmente se debe a que existe un componente de tendencia, que como se explicó en la sección primera de esta investigación, es el movimiento a largo plazo de la serie, una vez eliminados los ciclos y el término irregular, que hace que la serie sea creciente o decreciente. Retomando la ecuación de la serie con tendencia, se tiene que:

$$x_t = TR_t + IR_t$$

Donde se busca eliminar el componente de TR_t para que la serie sea estacionaria. Una forma de hacerlo, como se explicó en el primer apartado de esta investigación es transformándola tomando diferencias. Así, si la serie no es estacionaria en media, se toman d sucesivas diferencias de orden 1 sobre la serie hasta obtener una serie estacionaria:

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

Ahora el problema es identificar el número de diferencias exacto que se deben realizar para que la serie sea estacionaria, pues en dicho proceso pueden aparecer los siguientes problemas:

- Que la raíz autorregresiva este próxima a la unidad, lo cual haría que el modelo de la serie se convirtiese en una caminata aleatoria. Un ejemplo es, si se considera un modelo $AR(1)$:

$$x_t = 0.95x_{t-1} + a_t \rightarrow 1 - 0.95 B = 0 \rightarrow B = \frac{1}{0.95} = 1.05 > 1$$

El cual es un proceso estacionario, sin embargo, es muy cercano al proceso de caminata aleatoria $x_t = x_{t-1} + a_t$ porque la raíz unitaria está próxima a él. En la literatura de Series de Tiempo a las pruebas que hacen el contraste de $H_0 : \alpha = 1$ se les conoce como pruebas de raíz unitaria, entre ellas la más usada es la Dickey-Fuller [3, Chan, p. 96].

- Sobrediferenciación, lo cual pasa cuando a la serie se le realizan más diferencias de las necesarias. Cabe resaltar, que si se diferencia un proceso estacionario éste sigue siendo estacionario, sin embargo, es muy importante que el orden de diferenciación sea el adecuado.

4.3. Estimación de los parámetros del modelo

Una vez que se han identificado los órdenes del modelo que parece ha podido generar a la serie x_t , se procede a estimar los parámetros desconocidos $\phi = (\phi_1, \dots, \phi_p)$, $\theta = (\theta_1, \dots, \theta_q)$ y σ_a^2 , con lo cual se completa la identificación de un único modelo ARIMA(p, d, q).

Para la estimación de los parámetros se puede utilizar el método de Mínimos Cuadrados o Máxima Verosimilitud. En el caso del primero, se minimiza la suma de cuadrados de los residuos (valor ajustado menos valor observado):

$$\text{Min} \sum_{t=1}^n a_t^2$$

Sin embargo, este método puede producir estimadores sesgados. Por el otro lado, en el método de verosimilitud, se encuentran los parámetros de forma tal que se maximiza la verosimilitud de la serie observada, lo cual brinda que los estimadores sean asintóticamente insesgados y se distribuyan asintóticamente normal. Entonces, dado un modelo ARIMA, la función de verosimilitud para los parámetros ϕ , μ , θ y σ_a^2 está dado por

$$L = L(a_1, \dots, a_{t-1}, a_t | \mu, \phi, \theta, \sigma_a^2) = (2\pi\sigma_a^2)^{-T/2} \exp\left(\frac{-1}{2\sigma_a^2} \sum_{t=1}^T a_t^2\right)$$

Donde el proceso para estimar los parámetros consiste en derivar $\log(L)$ con respecto a ϕ , μ , θ y σ_a^2 , para posteriormente igualar a cero cada derivada y resolver el sistema de ecuaciones. El inconveniente de estimar los parámetros de esta forma, es que la función de verosimilitud debe ser exacta (es decir, que no esté condicionada a las primeras observaciones), lo cual se puede conseguir mediante una combinación de la función de verosimilitud condicionada y la función de densidad no condicionada para los valores iniciales [2, Box & Jenkins, p. 231-237].

En ese sentido, existen métodos que permiten estimar los parámetros mediante algoritmos de innovación o procedimientos de estimación recursivos, como el filtro de Kalman, que permiten obtener parámetros sin necesidad de una función de verosimilitud exacta [14, Woodward, p. 248-254].

4.4. Diagnóstico del modelo

Una vez que se ha identificado el modelo y se han estimado sus respectivos parámetros, en este diagnóstico, se evalúa la adecuación del modelo a los datos para determinar si el modelo es válido, o en caso de no serlo, descubrir de qué manera el modelo es inadecuado. Para ello debe comprobarse que las estimaciones de los coeficientes del modelo son significativas y cumplen las condiciones de estacionariedad, mientras que los residuos deben de tener un comportamiento aproximado al de un ruido blanco a_t , pues a_t es un proceso no observable.

4.4.1. Análisis de los coeficientes

Para comprobar que los coeficientes del modelo ARMA sean los adecuados, se deben de realizar los contrastes de significación individual para comprobar que el modelo propuesto no esté sobreidentificado. Entonces, considerando a los respectivos coeficientes:

$$\beta = (\delta, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \dots, \theta_q)$$

Se plantean los contrastes pertinentes:

$$H_0 : \delta = 0 \quad \text{frente a} \quad H_a : \delta \neq 0$$

$$H_0 : \phi_i = 0 \quad \text{frente a} \quad H_a : \phi_i \neq 0$$

$$H_0 : \theta_i = 0 \quad \text{frente a} \quad H_a : \theta_i \neq 0$$

Si los coeficientes fueron estimados por máxima verosimilitud, su distribución es asintótica $\hat{\beta}_i \sim N(\beta_i, V(\hat{\beta}_i))$, por lo que se contrasta la H_0 de no significancia individual de los parámetros mediante el estadístico t , que también sigue asintóticamente una distribución normal:

$$t = \frac{(\hat{\beta}_i - 0)}{\sqrt{V(\hat{\beta}_i)}} \sim N(0, 1)$$

Mientras que se rechaza la hipótesis nula para el coeficiente a un nivel de significancia α cuando:

$$\left| t = \frac{(\hat{\beta}_i - 0)}{\sqrt{V(\hat{\beta}_i)}} \right| > Z_{1-\alpha/2}$$

Además de esta prueba de significancia individual, se sugiere también analizar la matriz de covarianzas entre los parámetros estimados con el fin de detectar la posible presencia de una correlación excesivamente alta entre las estimaciones de los mismos lo que puede ser un indicativo de presencia de *factores comunes* en el modelo.

4.4.2. Análisis de los residuos

El análisis de residuos, es decir, los valores observados menos los valores ajustados, consiste en una serie de contrastes con el objetivo de determinar si los residuos replican el comportamiento de un ruido blanco, es decir, que se distribuyen normal (si los coeficientes fueron estimados por máxima verosimilitud), de forma independiente (los diferentes a_t no están correlacionados), con media cero y varianza constante.

Este análisis de los residuos es fundamental en la verificación del modelo, pues si por alguna razón en dicho análisis no se cumple alguno de los supuestos antes mencionados, existe una violación a los supuestos y el modelo no puede ser validado.

Supuesto de normalidad

Como se mencionó anteriormente, es de especial interés que los residuos se distribuyan normal con media cero y varianza constante, pues esto implica que son un ruido blanco y el modelo tiene buena capacidad explicativa. Para comprobar este supuesto fundamental sobre los residuos del modelo, existen diversas pruebas tanto gráficas como numéricas.

Por el lado gráfico el más utilizado es el histograma, el cual es un gráfico en el que se esperaría que los datos se concentraran en el centro y existiera simetría entre las colas de la distribución, es decir, que los datos presenten una forma acampanada. Sin embargo, un problema muy común dentro de este tipo de gráficos es el intervalo de amplitud, dado que a partir del tamaño de éste se puede cambiar la percepción sobre la distribución de los residuos. En ese sentido, otro gráfico más ilustrativo es el gráfico normal cuantil-cuantil (Q-Q norm en inglés) el cual permite observar cuan cerca está la distribución de los residuos de la distribución normal ideal de ruido blanco, a partir de las diferencias entre cada una de las distribuciones de probabilidad. Ambos tipos de gráficos pueden observarse en la *Figura 4.7*.

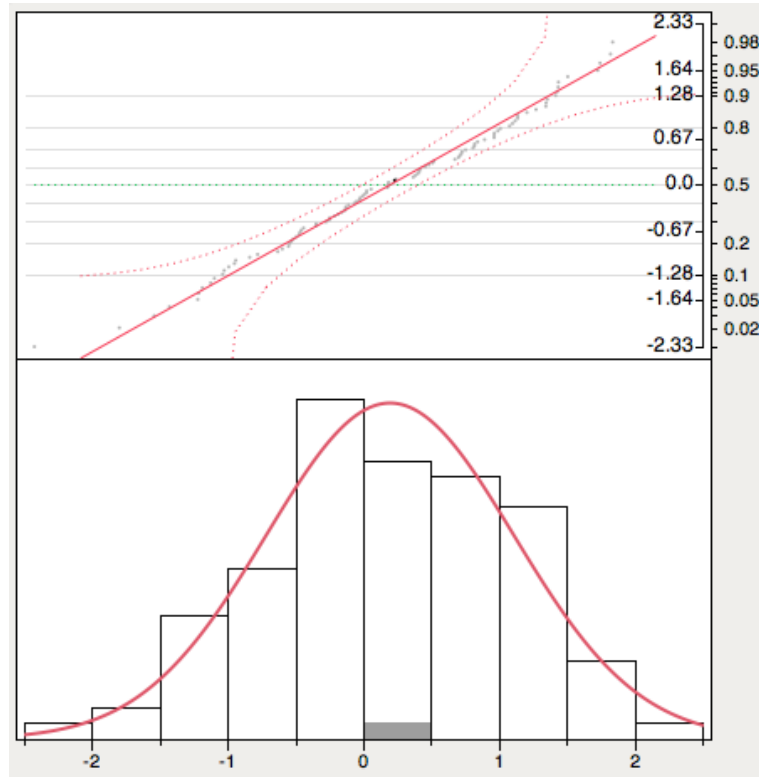


Figura 4.7: Ejemplo de un Histograma y un Q-Q norm de una simulación de la distribución normal

A pesar de que estas pruebas gráficas pueden ser útiles para identificar si los residuos del modelo se distribuyen normal, dada la importancia de este supuesto, se requieren de pruebas que lo respalden numéricamente para determinar de manera contundente la bondad del modelo.

Una prueba numérica comúnmente utilizada para probar que la sucesión ordenada $x_{(i)}$ es normal, es el Test Shapiro-Wilk, el cual se plantea de la siguiente forma:

$$w = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Con dicho estadístico se comprueba si la sucesión ordenada de datos se distribuye normal, comparando el resultado obtenido empíricamente (w) con el valor teórico (W) que corresponde a las tablas de normalidad de acuerdo a la siguiente prueba de contraste:

$$H_0 : a_t \sim N(0, \sigma_a^2) \quad \text{frente a} \quad H_0 : a_t \not\sim N(0, \sigma_a^2)$$

Entonces, se rechaza H_0 si $\mathbb{P}(W > w) \leq 0.05$. Esta prueba se considera potente para contraste de normalidad principalmente para muestras pequeñas ($n < 30$), sin embargo, un inconveniente sumamente importante de este estadístico es que suele ser sensible a la correlación de la muestra, por lo que para realizar esta prueba se requiere del supuesto de independencia de los errores (el cual a su vez requiere del supuesto de normalidad en algunas pruebas), puesto que en muchos casos la falta de independencia provoca falsos positivos, es decir, que acepte normalidad de los residuos aunque en realidad no la haya.

Una alternativa es una prueba no paramétrica Kolmogorov-Smirnov. Esta prueba permite contrastar la bondad de ajuste de distribuciones completamente especificadas y continuas, y es más sensible a los valores cercanos a la mediana que a los extremos de la distribución. Entonces, el estadístico de la Kolmogorov-Smirnov se compone por:

$$D = \max_{1 \leq i \leq n} \left(F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right)$$

Donde F es la distribución acumulada teórica de la distribución que se está contrastando y debe ser continua y estar completamente especificada. Por su parte, las hipótesis a probar, similar al Test Shapiro-Wilk es H_0 si los residuos siguen la específica función de distribución normal *vs* H_a los residuos no siguen la específica función de distribución. En la práctica esta prueba es poco útil pues requiere que la distribución esté completamente especificada y con todos sus parámetros como conocidos. Una adecuación útil y que convierte a esta prueba en una de amplio uso, es mediante la corrección de Lilliefors donde se asume que la distribución es normal. En este caso la prueba es exacta porque el estadístico de prueba y la familia de la distribución normal son invariantes bajo transformaciones de escala.

Una prueba con mayor potencia y cuya sensibilidad es proporcional incluso en las zonas extremas de la distribución, es la Anderson-Darling, la cual es una prueba no paramétrica (modificación de la Kolmogorov-Smirnov) que determina si una muestra ordenada $x_{(1)} < \dots < x_{(n)}$ proviene de una distribución específica, en este caso, una distribución normal. Entonces, el estadístico se compone de la siguiente forma:

$$A^2 = -n - S$$

Donde:

$$S = \sum_{k=1}^n \frac{2k-1}{n} [\ln F(x_{(k)}) + \ln(1 - F(x_{(n+1-k)}))]$$

Mientras que el contraste de la prueba, al igual que el Test Shapiro-Wilk y la Kolmogorov-Smirnov, es rechazar H_0 si $\mathbb{P}(A^2 > a^2) \leq 0.05$.

Supuesto de media cero y varianza constante

En el caso de la media, se puede realizar un análisis gráfico representando los residuos a lo largo del tiempo y observando si los valores oscilan alrededor de cero. Para reforzar este análisis se puede llevar a cabo el contraste de hipótesis de $H_0 : \mathbb{E}(a_t) = 0$ con el estadístico:

$$t = \sqrt{N} \frac{\bar{\hat{a}}}{\sqrt{\hat{V}(\hat{a})}} \sim N(0, 1)$$

Donde $\bar{\hat{a}}$ y $V(\hat{a})$ son la media y la varianza muestrales de los residuos. No se rechaza la hipótesis nula al nivel de significancia del 5% si $|t| \leq 1.96$.

Para el caso de la varianza constante, en el gráfico de los residuos la dispersión de los mismos deberá ser constante para concluir que la varianza de a_t permanece constante.

De forma alternativa a la inspección grafica, también existen pruebas estadísticas con las cuales se puede determinar si la varianza de los residuos es homogénea. Una de las pruebas más utilizadas es la prueba de Bartlett, en la cual se prueba si diferentes muestras poseen la misma varianza poblacional. Entonces, si hay k grupos de tamaño n_i con su respectiva varianza S_i^2 , el estadístico de prueba se define como:

$$U = \frac{1}{C} \left[(N - k) \ln(S^2) - \sum_{i=1}^k (n_i - 1) \ln(S_i^2) \right]$$

Donde:

$$C = 1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{N - k} \right) \quad \text{y} \quad N = \sum_{i=1}^k n_i$$

Mientras que S^2 es la estimación combinada de la varianza:

$$S^2 = \frac{1}{N - k} \sum_i (n_i - 1) S_i^2$$

Cuando la hipótesis nula es cierta, el estadístico tiene distribución aproximadamente χ^2 con $k - 1$ grados de libertad. La ventaja de esta prueba es que no requiere que los tamaños de los grupos sean iguales, sin embargo, también posee dos grandes desventajas: es sensible a la normalidad (es decir, se requiere que los residuos se distribuyan normal) [10, Montgomery p. 82] y la formación de los k grupos es generada artificialmente. Una opción menos sensible al supuesto de normalidad, pero que también comparte la disyuntiva de la formación de k grupos, es el test de Levene [10, Montgomery p. 83-84]. Este estadístico al ser más robusto, permite que haya menor probabilidad de rechazar una hipótesis verdadera de igualdad de varianzas solo porque las distribuciones de las poblaciones muestreadas no son normales. Este prueba tiene su estadístico de la siguiente forma:

$$W = \frac{(N - k) \sum_{i=1}^k n_i (\bar{Z}_{i.} - \bar{Z}_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} n_i (Z_{ij} - \bar{Z}_{i.})^2}$$

Donde Z_{ij} puede tener las siguientes definiciones:

1. $Z_{ij} = |X_{ij} - \bar{X}_{i.}|$, donde $\bar{X}_{i.}$ es la media del i -ésimo grupo.
2. $Z_{ij} = |X_{ij} - \tilde{X}_{i.}|$, donde $\tilde{X}_{i.}$ es la mediana del i -ésimo grupo.

El test de Levene rechaza la hipótesis de que las varianzas son iguales con un nivel de significancia α si $W > F_{\alpha, k-1, N-k}$ donde $F_{\alpha, k-1, N-k}$ es el valor crítico superior de la distribución F con $k-1$ grados de libertad en el numerador y $N-k$ grados de libertad en el denominador a un nivel de significancia α .

Como alternativa a estas pruebas donde se requiere formar grupos de manera artificial, está el Test Breusch-Pagan, el cual determina si la varianza estimada de los residuos del modelo depende de los valores de las variables independientes. Entonces, se tiene un modelo de la siguiente forma:

$$a_t = \beta_0 + \beta_1 t + u$$

Donde el estadístico Breusch-Pagan se construye mediante una ecuación auxiliar de la regresión, quedando de la forma:

$$\hat{u}^2 = \gamma_0 + \gamma_1 t + v$$

Por lo que a través del test- F se confirma si las variables independientes son significativas y entonces se rechaza H_0 de homoscedasticidad. Un inconveniente en la práctica de este test, es que justamente resulta difícil reconocer a las variables independientes causantes de la heteroscedasticidad.

Supuesto de no correlación

De la misma forma que en los supuestos anteriores, para comprobar este supuesto existen desde pruebas estadísticas hasta análisis gráficos. En el caso gráfico, si los residuos se comportan como un ruido blanco, el correlograma de la FAC y FACP muestrales deberían ser prácticamente nulos. En el caso de las pruebas estadísticas, esto se puede realizar ya sea mediante contrastes de significancia individual como conjunta:

- La primer prueba corresponde al análisis de significancia individual de los coeficientes de correlación

$$H_0 : \rho_h(a) = 0 \quad \text{frente a} \quad H_a : \rho_h(a) \neq 0$$

Entonces la regla de decisión es:

$$\left| \rho_h(a) \right| \geq \frac{\xi_{1-\alpha}}{\sqrt{n-p-d-q}} \Rightarrow \text{Rechazar } \rho_h = 0 \text{ con significancia de } \alpha = 0.05$$

Donde $\xi_{1-\alpha}$ es el cuantil $(1-\alpha)$ correspondiente a una distribución normal estándar.

- Para hacer una prueba conjunta sobre de los coeficientes de autocorrelación, donde la hipótesis nula es:

$$H_0 : \rho_1(a) = \rho_2(a) = \dots = \rho_T(a) = 0$$

Mientras que la hipótesis alternativa es que algún coeficiente ρ_h sea no nulo para $h = 1, \dots, T$. La prueba más conocida para este caso es el estadístico Ljung-Box, donde los grados de libertad se establecen en $n-p-d-q$, es decir, restando los parámetros que se ajustaron en el modelo. Se tiene entonces que:

$$Q = n(n+2) \sum_{h=1}^T \frac{\hat{\rho}_h^2}{n-h}$$

Se rechaza la hipótesis nula de ausencia de correlación, para un nivel de significancia $\alpha = 0.05$, si:

$$Q > \chi_{0.05}^2(n - p - d - q)$$

Si se presenta este caso, el modelo exhibiría correlación en los residuos por lo que se concluiría que el modelo no ha sido capaz de reproducir el patrón de comportamiento sistemático de la serie.

4.5. Predicción

En esta última etapa, el objetivo es obtener predicciones óptimas de x_t en algún momento en el futuro basándose en la información pasada disponible:

$$I_t = \{x_t, x_{t-1}, x_{t-2}, \dots\}$$

Entonces, recordando que el objetivo de esta investigación se centra en las series de tiempo univariadas, se puede denotar a cualquier pronóstico puntual univariado de x_{t+h} como $\hat{x}_t(h)$. Este pronóstico se deriva de una *función de pérdida*, la cual es posible definir si se considera a e como error en el pronóstico, entonces:

$$e = (\text{valores observados} - \text{valores ajustados})$$

Donde la función de pérdida, $L(e)$, especifica la *pérdida* asociada con el error en la predicción de tamaño e . Esta función tiene las propiedades de que (i) $L(0) = 0$ y (ii) $L(e)$ es función continua que incrementa con el valor absoluto de e . La importancia de esta función radica en una vez definida, un “buen pronóstico” puede definirse como pronóstico que minimiza el promedio de pérdida sobre la distribución de probabilidad de los errores del pronóstico.

A partir de la función de pérdida, se puede denominar al predictor óptimo como aquel que minimiza la función de pérdida o al Error Cuadrático Medio de Predicción (el cuadrado de la función de pérdida), es decir:

$$\mathbb{E}[x_{t+h} - \hat{x}_t(h)]^2 \leq \mathbb{E}[x_{t+h} - \hat{x}_t^*(h)]^2 \quad \forall \hat{x}_t^*(h)$$

Se puede demostrar que, bajo condiciones de regularidad muy débiles, el predictor por punto óptimo viene dado por la esperanza condicionada al conjunto de información:

$$x_t(h) = \mathbb{E}[x_{t+h}|I_t] = \mathbb{E}[x_{t+h}|x_t, x_{t-1}, x_{t-2}, \dots] = \mathbb{E}_t[x_{t+h}]$$

Donde se puede ver que el valor esperado de la distribución de $x_t(h)$ está condicionada a la información disponible. Por otro lado, nada garantiza que esta esperanza condicionada sea una función lineal del pasado de la serie, pero si el proceso sigue una distribución normal, se puede demostrar que la esperanza condicionada se puede expresar como una función lineal del conjunto de información, I_t

En cuanto a la predicción óptima por intervalo, ésta se construye a partir de la distribución de probabilidad de los errores del pronóstico, bajo el supuesto de que los errores se distribuyen como un ruido blanco:

$$e_t(h) = x_{t+h} - x_t(h) \sim N(0, V(e_t(h)))$$

De donde se puede obtener el estadístico el intervalo al nivel de confianza $(1 - \alpha) \%$:

$$\frac{x_{t+h} - x_t(h) - 0}{\sqrt{V(e_t(h))}} \sim N(0, 1)$$

Tanto la predicción puntual como la de intervalo en los modelos ARMA se pueden obtener de tres formas:

1. *Utilizando la ecuación del modelo directamente*

Asumiendo que la ecuación del modelo se conoce exactamente, entonces se obtiene $\hat{x}_t(h)$ directamente de la ecuación sustituyendo (i) los valores futuros de a por cero, (ii) los valores futuros de X por su esperanza condicional y (iii) los valores actuales y pasados de X y a por sus valores observados. Por ejemplo, considerando un modelo autoregresivo $AR(1)$, se tiene:

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

$$\hat{x}_t(1) = \phi x_t$$

$$\hat{x}_t(2) = \phi \hat{x}_t(1)$$

2. *Utilizando los pesos ψ*

Un modelo ARMA puede ser rescrito como un proceso de media móvil de orden infinito y los pesos ψ pueden ser utilizados para calcular pronósticos, pero son principalmente útiles para calcular las varianzas del error de pronóstico, pues:

$$x_{t+h} = a_{t+h} + \psi_1 a_{t+h-1} + \dots$$

Y los futuros a son desconocidos para el tiempo t , entonces $\hat{x}_t(h)$ es igual a $\sum_{j=0}^{\infty} \psi_{h+j,t-j}$. Por lo tanto el error de pronóstico del tiempo h es:

$$(a_{t+h} + \psi_1 a_{t+h-1} + \dots + \psi_{h-1} a_{t+1})$$

Mientras que la varianza del error de pronóstico para el tiempo h :

$$(1 + \psi_1^2 + \dots + \psi_{h-1}^2) \sigma_a^2$$

3. *Utilizando los pesos ϕ*

Un modelo ARMA puede ser rescrito como un proceso autoregresivo de orden infinito y los pesos ϕ pueden ser utilizados para calcular pronósticos, pues se tiene:

$$x_{t+h} = \phi_1 x_{t+h-1} + \dots + \phi_h x_t + \dots + a_{t+h}$$

Donde $\hat{x}_t(h)$ está dado por:

$$\hat{x}_t(h) = \phi_1 \hat{x}_t(h-1) + \phi_2 \hat{x}_t(h-2) + \dots + \phi_h x_t + \phi_{h+1} x_{t-1} + \dots$$

Estas predicciones pueden obtenerse recursivamente, reemplazando los valores futuros de x con valores pronosticados necesariamente.

En general, los métodos (1) o (3) son utilizados para calcular predicciones puntuales, mientras que el método (2) es utilizado para calcular las varianzas del error de pronóstico. En la práctica el modelo no puede ser conocido exactamente,

entonces se deben estimar los parámetros del modelo (y los respectivos ψ y ϕ si son necesarios) y se deben estimar también los valores pasados de Z , por los residuos observados o los errores de pronóstico futuros. [4, Chatfield, p. 84]

4.5.1. Inconvenientes de la metodología ARIMA de Box y Jenkins

Al igual que diversos modelos en la ciencia matemática, la metodología ARIMA de Box y Jenkins presenta diversos alcances que se han explicado a lo largo de esta investigación, sin embargo, también cuenta con algunas limitaciones de las cuales es importante hacer mención. La primera de ellas se refiere a que si bien en la práctica los paquetes estadísticos incluyen este tipo modelación, este procedimiento está diseñado dentro de un enfoque no-automatizado, es decir, el juicio subjetivo del investigador puede afectar seriamente el análisis, la elección y la construcción del modelo. Es por ello que los modelos ARIMA pueden ser muy versátiles, pero a la vez más complicados y requieren de una experiencia considerable para ser posible identificar los modelos adecuados, pues un investigador puede encontrar diversos modelos que expliquen bien a la serie de tiempo dada.

Otro inconveniente es que el método requiere de muchas observaciones para funcionar adecuadamente (De acuerdo a Chatfield [4, p. 84] al menos 50 observaciones para datos estacionales mensuales).

4.5.2. Método Stepwise

Un método alternativo para pronosticar información a partir del modelo es mediante la autoregresión Stepwise, propuesto por Granger y Newbold (1986) [5, Chatfield p. 75], este modelo se puede considerar como un subconjunto del procedimiento Box-Jenkins que posee la ventaja de ser completamente automático. Este método depende del hecho de que los modelos AR son más fáciles de ajustar en comparación con los modelos MA o ARMA aunque los modelos AR requieran de parámetros extras para hacer una buena representación de la información.

El procedimiento para este método es primero considerar las primeras diferencias de los datos para que exista estacionariedad en la media. Después se determina un rezago máximo posible r y se elige el mejor modelo AR con una sola variable rezagada entre el rezago 1 y p (dicha p se refiere al orden del proceso autorregresivo), entonces, se tiene:

$$w_t = \mu + \alpha_h^{(1)} w_{t-h} + a_t^{(1)} \quad 1 \leq h \leq p$$

Donde $w_t = x_t - x_{t-1}$ y $\alpha_h^{(1)}$ es el coeficiente de autoregresión en el rezago h cuando se ajusta una sola variable rezagada. Entonces, el mejor modelo autorregresivo se encuentra para cada uno de los rezagos. El proceso concluye cuando la reducción en la suma del cuadrado de los residuos en la j -ésima etapa es menor que la cantidad antes fijada. En ese sentido, Granger y Newbold proponen elegir $p = 13$ para información cuatrimestral y $p = 25$ para información mensual.

Capítulo 5

Análisis del Índice de Precios y Cotizaciones en México

A partir de la breve introducción a la teoría sobre el *Análisis de Series de Tiempo*, el objetivo de este capítulo es modelar una serie de tiempo financiera como el Índice de Precios y Cotizaciones en México, para poder explicar su comportamiento y realizar predicciones. La información presentada sobre este indicador fue obtenida del sitio *Yahoo! Finanzas*, siendo su periodicidad diaria y comprendiendo el periodo de octubre del 2014 a febrero del 2015.

En adelante, por practicidad, se denotará a este conjunto de información como IPC (o alguna de sus correspondientes transformaciones), entendiéndolo como un conjunto de variables x_t que se ajustarán a la *metodología Box y Jenkins*. Además se aclara que los paquetes estadísticos para realizar dicha labor fueron JMP en su versión 9.0.1 y R en su versión 3.2.1.

5.1. Análisis descriptivo

5.1.1. Breve descripción de la información

El Índice de Precios y Cotizaciones es el índice principal de la Bolsa Mexicana de Valores que expresa el volumen y rendimiento del mercado accionario en México a partir de una canasta de precios de acciones representativas cotizadas en la Bolsa. Esta canasta se compone por una muestra de acciones, balanceada y representativa del mercado accionario, que está integrada por emisoras de distintos sectores de la economía y la cual está sujeta a una revisión semestral pues las emisoras de las acciones deben cumplir ciertos criterios para ser tomadas en cuenta. Por lo tanto, el IPC es considerado como un índice ponderado del valor de capitalización de las acciones, lo cual quiere decir que el cambio en el precio de una acción integrante del índice influye en su evolución de acuerdo al peso relativo que la acción tiene en la muestra.

Este índice tiene su fecha base de cálculo el 30 de octubre de 1978 y desde entonces el tamaño de acciones que lo componen ha oscilado entre 35 y 50, siendo estas acciones clasificadas como de alta y media bursatilidad (acciones que se negocian mucho en el mercado tanto por volumen como por importe) y su principal objetivo es ser un indicador altamente representativo y confiable del Mercado Accionario Mexicano.

Para los términos que competen a esta investigación, se considera el valor de IPC al “cierre ajustado”, el cual proporciona el precio de cierre del día ajustado para todos los fraccionamientos y distribuciones de dividendos pertinentes, pues esta categoría permite realizar un mejor análisis de los rendimientos históricos considerando dicha cifra como el valor del IPC al día en cuestión. El periodo que abarca esta información es de octubre de 2014 a febrero de 2015, mientras que la periodicidad es diaria, por lo que tomando en cuenta que los mercados bursátiles sólo se labora días hábiles, se tiene una muestra compuesta por $n = 100$, a partir de la cual se buscará ajustar un modelo con el cual se pueda predecir las siguientes 3 observaciones correspondientes a la primeros días de marzo.

Finalmente, cabe decir que como el valor del IPC es un representativo periódico de la actividad bursátil en México es claro que se trata de una serie de tiempo, pues se relaciona con la actividad del día anterior y no con el valor de su fecha de inicio debido a que la muestra que lo compone se revisa con cierta frecuencia a fin de que las acciones que lo componen sean las de mayor peso y representatividad.

5.1.2. Análisis de la estacionariedad

Para poder realizar el ajuste a la *metodología Box y Jenkins* se requiere que el IPC sea una serie de tiempo estacionaria o que al menos tenga una transformación que permita que ésta sea un proceso estacionario. La forma de determinar si cumple este requisito es mediante la inspección visual de su comportamiento respecto del tiempo, con la cual buscamos *i)* un comportamiento periódico de la serie; *ii)* varianza homogénea y; *iii)* media constante. Asimismo, se transforma la serie a logaritmos de la forma $y_t = \ln(x_t)$ nombrandola LIPC, pues esta transformación (que es una transformación de Box-Cox muy utilizada en las series económicas, *ver Apéndice A*) permite que la escala de la serie sea más pequeña y esto pueda ayudar a reducir la heterogeneidad de la varianza. Entonces, se tiene la siguiente gráfica (*Figura 5.1*) con el IPC y el LIPC:

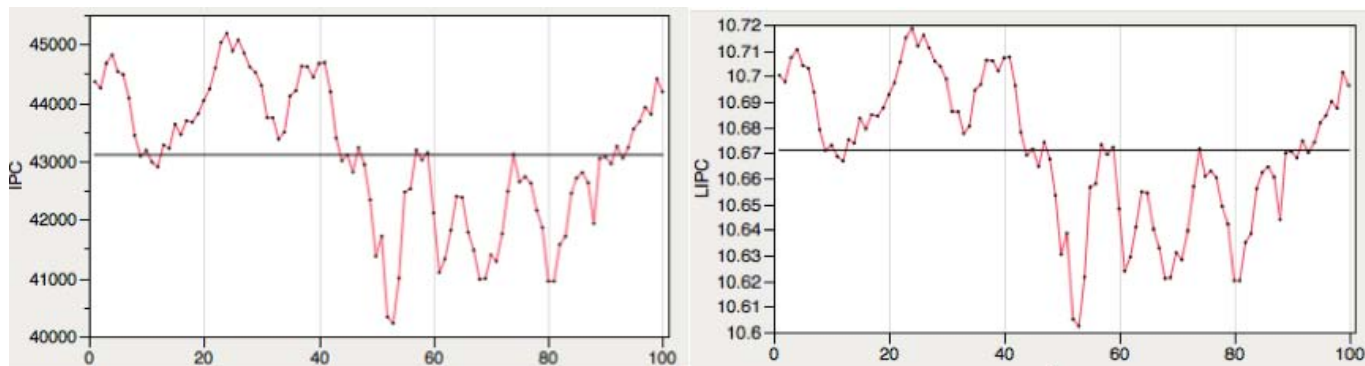


Figura 5.1: IPC y LIPC vs tiempo (t)

Como se observa tanto para el IPC como para LIPC parece no haber una alternada en ciertos lapsos de tiempo, mientras que en otros periodos de tiempo se observa una gran volatilidad, principalmente a partir del segundo mes (es decir, la observación 45). Dicho componente resulta acorde a la lógica financiera, pues la actividad bursátil suele tener variaciones importantes a lo largo del tiempo, mientras que la gran volatilidad del IPC se debe tanto a la caída en los precios del petróleo como a la incertidumbre que existe en los mercados debido a la política monetaria de la Reserva Federal de los Estados Unidos (Fed), pues el efecto de las políticas del mayor socio comercial de México tienen mucho efecto en este indicador.¹ En ese sentido, como el objetivo de esta investigación es modelar el IPC (o su transformación,

¹NOTIMEX, *Revista Punto de Vista*, 16 de diciembre del 2014, .

el LIPC) a través de la metodología Box-Jenkins, se busca que la serie cumpla con la propiedad de estacionariedad más que extraer el componente tendencial. Por lo tanto, para eliminar este componente (*ver Apéndice A* se transforma la serie a sus primeras diferencias de la forma $z_t = (1 - B)y_t = y_t - y_{t-a}$, nombrándose DLIPC y se observa gráficamente en la *Figura 5.2*:

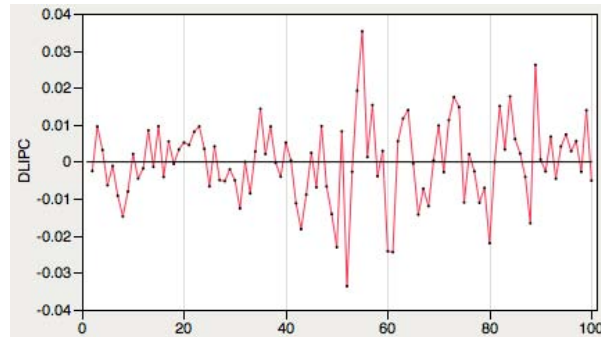


Figura 5.2: Primera diferencia de LIPC (DLIPC) vs tiempo (t)

Con lo cual se obtiene una serie estacionaria DLIPC (z_t , la cual tiene media constante $\bar{x} = -3.987477e - 05$ y con una varianza que parece ser homogénea oscilando entre entre 0.03 y -0.03 , cuyo valor es de $\hat{\sigma}^2 = 0.0001196149$).

Asimismo, también es importante identificar si esta nueva serie transformada (DLIPC) cuenta con algún efecto estacional que pueda influir en su modelación, por lo que se realiza el correspondiente periodograma, del cual se extrae el ciclo de la serie DLIPC y se observa en la *Figura 5.3*:

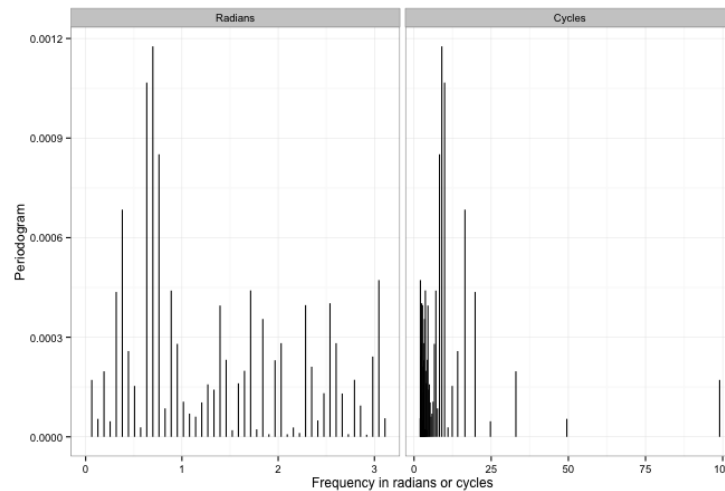


Figura 5.3: Periodograma del IPC

```
max.location j- which.max(per$per)
ciclo=per$c[max.location]
9
```

Con lo cual, como se muestra en el recuadro, se determina que existe un ciclo cada 9 días, por lo que considerando que este índice bursátil no cotiza los fines de semana, indica que hay un efecto estacional aproximadamente cada quincena.

5.2. Metodología ARIMA

5.2.1. Identificación

Una vez que se ha obtenido una serie estacionaria, en este caso en primeras diferencias $z_t = (1 - B)y_t$, es posible desarrollar el método iterativo de *Box y Jenkins* para encontrar un modelo que explique y describa el comportamiento del IPC. Para identificar el modelo, es necesario obtener la FAC y la FACP, las cuales se elaboran considerando 25 rezagos² y se tiene:

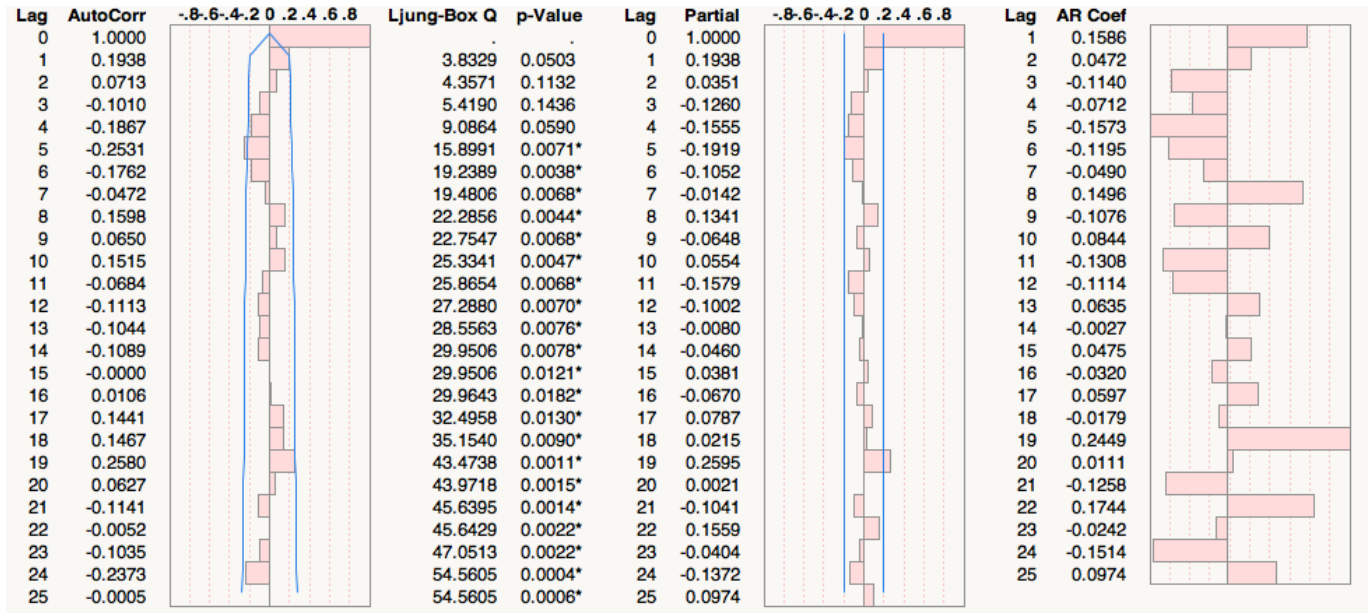


Figura 5.4: Correlograma con la FAC y FACP de DLIPC

La figura 5.4 muestra la función de autocorrelación estimada (FAC) tiene una estructura alternada, por lo que esa función sugiere ajustar una media móvil. Por su parte, el análisis de la función de autocorrelación parcial estimada (FACP) muestra, de acuerdo a las bandas y a los contrastes de significación individual, que hay un decrecimiento rápido, sin embargo, tanto en la FAC como FACP no se anula el efecto lo cual inclinaría la decisión hacia un modelo quizá de un orden mayor a 2.

Para ajustar el modelo ARIMA, también se considera la información proporcionada por el periodograma, el cual indicó que se presenta un ciclo en los datos cada 9 días, es decir, cada quincena. A partir de este criterio, se ajusta un modelo ARIMA estacional que considera un ciclo cada quincena $((1 - B)^9 y_t)$, con lo que se tiene:

Este modelo arroja un *Akaike* de -610.5614 y un $-2\log LH$ de -614.5614 . Sin embargo, como puede apreciarse en la Figura 5.5, este modelo presenta una correlación importante en los residuos de los rezagos 5 y 19. De esta forma, para determinar el modelo que mejor se ajusta al DLIPC, se realizan diversos modelos de forma iterativa, tomando como base el *Criterio de Akaike* y que los residuos se comporten como un ruido blanco, con lo cual se genera la siguiente tabla:

²En la práctica es común considerar que los rezagos no sean mayores al 25% de las observaciones. En este caso, se utilizan los rezagos que el programa da por *default* y cumple con dicho criterio.

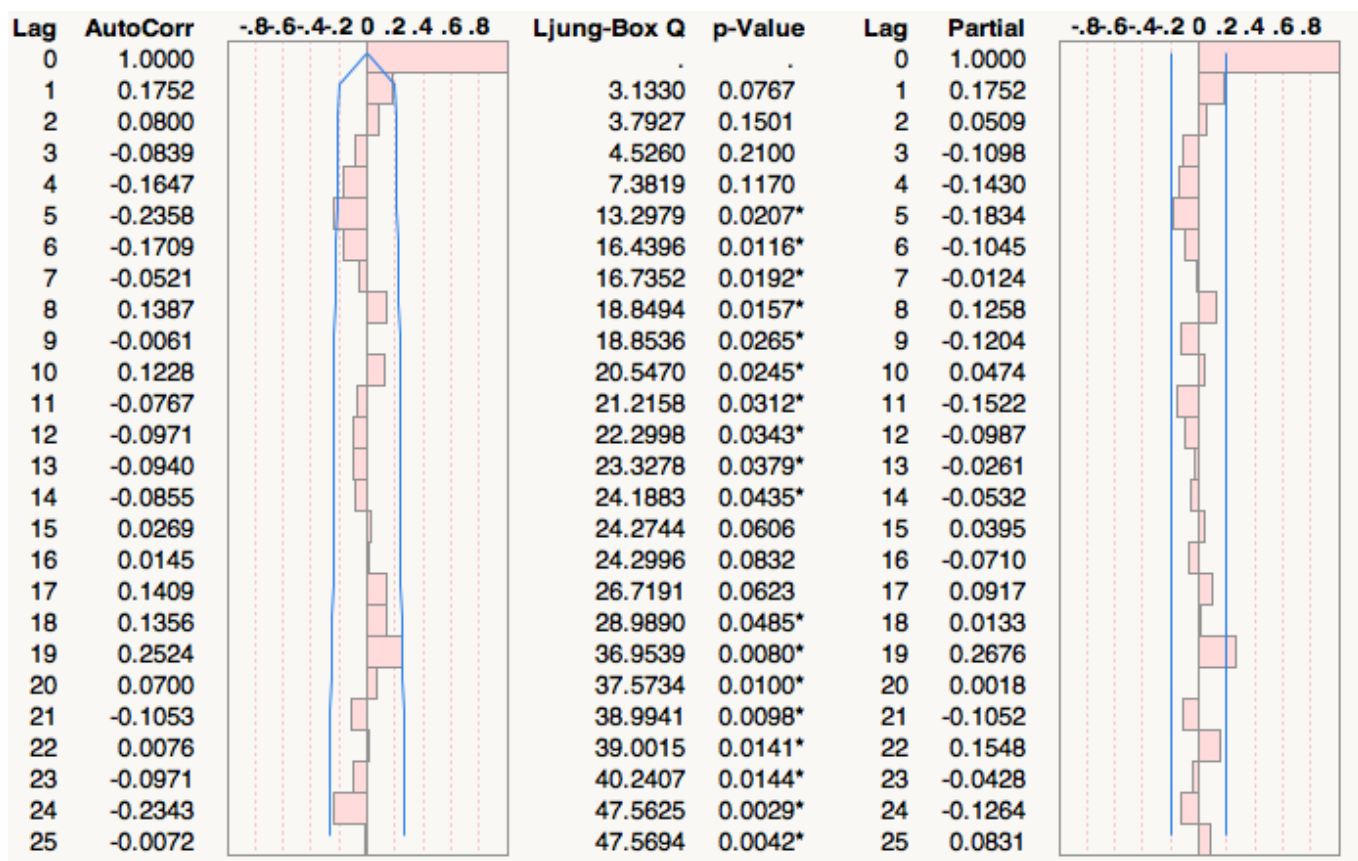


Figura 5.5: Correlograma de los residuos del ARIMA estacional $(1 - B)^9 y_t$ de DLIPC

$ARIMA(p, d, q)$	G. L.	AIC	$-2\log LH$
$ARI(19, 1, 0)$	96	-620.7146	313.3573
$ARIMA(2, 1, 2)$	94	-617.0550	313.5275
$ARIMA(2, 1, 3)$	93	-615.1718	313.5859
$ARIMA(3, 1, 2)$	93	-615.1620	313.5810
$ARIMA(3, 1, 1)$	94	-614.7206	312.3603
$ARI(1, 1, 0)$	97	-613.9086	308.9543
$IMA(0, 1, 1)$	97	-613.4661	308.7330
$ARIMA(3, 1, 3)$	92	-613.1363	313.5682
$ARIMA(1, 1, 3)$	94	-612.6563	311.3281
$IMA(0, 1, 2)$	96	-612.5489	309.2745
$I(0, 1, 0)$	98	-612.1460	307.0730
$ARI(2, 1, 0)$	96	-612.0206	309.0103
$ARIMA(1, 1, 1)$	96	-611.9542	308.9771
$ARI(3, 1, 0)$	95	-611.6125	309.8063
$ARIMA(2, 1, 1)$	95	-610.7334	309.3667
$IMA(0, 1, 3)$	95	-610.6451	309.3225
$ARIMA(1, 1, 2)$	95	-610.5769	309.2885

Después de ajustar diversos modelos, incluyendo entre ellos modelos sólo con autorregresivos (modelos ARI) o inclu-

yendo sólo los proceso de medias móviles (modelos IMA), se determinó que el mejor modelo fue el que incluyó solamente los rezagos autorregresivos 5 y 19, es decir, imponiendo la restricción de que el resto de los coeficientes autorregresivos fuese cero, basándose en los criterios de la FAC y FACP, pues sólo dichos rezagos presentaron una alta correlación. Entre las características que avalan dicha determinación están que dicho modelo fue el que presentó un menor *Akaike*, una menor log verosimilitud y que sus residuos parecen comportarse como un ruido blanco como se muestra en la *Figura 5.6*:

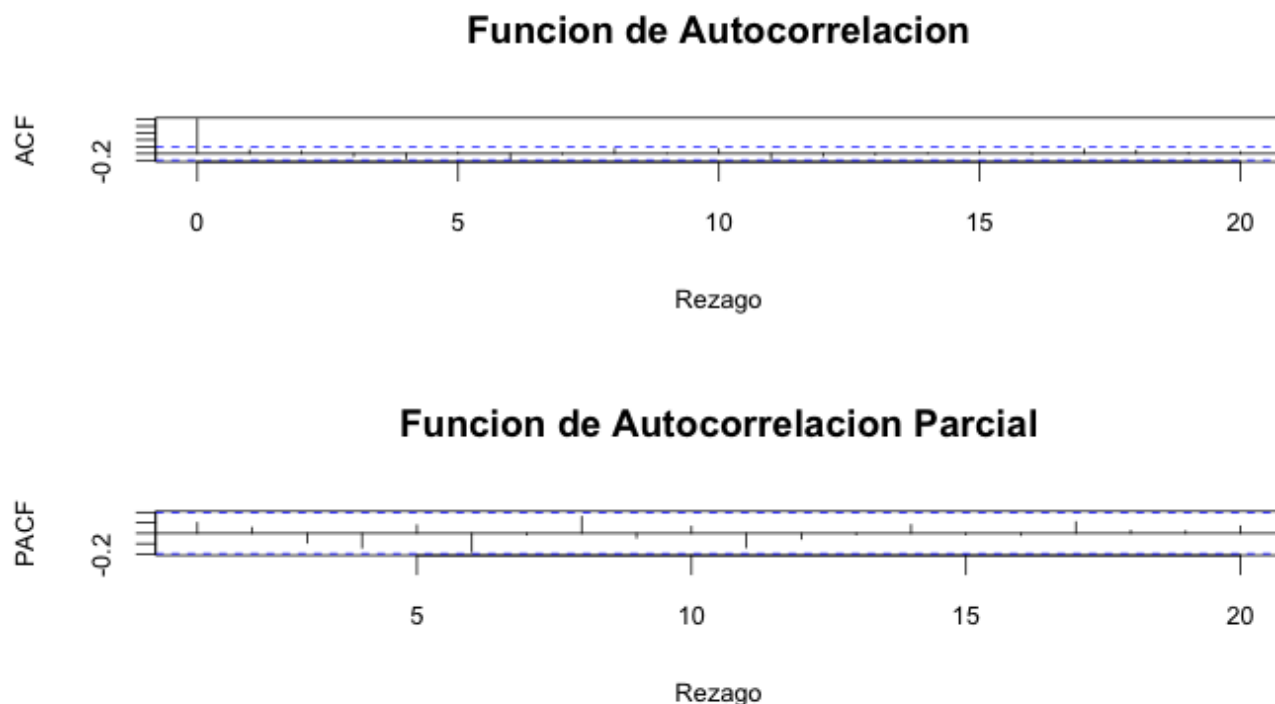


Figura 5.6: Correlograma de los residuos del $ARI(19, 1, 0)$

Asimismo, este modelo cumple con los criterios de *parsimonia*, pues al incluir sólo dos coeficientes resulta un modelo más sencillo en comparación con otros modelos.

5.2.2. Estimación

Los parámetros fueron estimados con su correspondiente significancia y error estándar, siendo el contraste $H_0 : \phi_1 = 0, \theta_1 = 0$ vs $H_a : \phi_1 \neq 0, \theta_1 \neq 0$, donde la regla de decisión es que se rechaza H_0 si el p-value asociado al estadístico es $p < 0.05$:

Parámetro	Estimación	Error Std.	$Prob > t $	
$AR(5)$	-0.2181854	0.09305967	-2.34457	< 0.0001*
$AR(19)$	0.2435912	0.09577357	2.61758	< 0.0001*

Los coeficientes son significativos además de que el modelo ajusta bien al presentar un $R^2_{adj} = 0.8708504$ y proporciona una varianza muy pequeña, $\hat{\sigma}^2 = 0.000102736$. Además, al estar los coeficientes dentro del círculo unitario, el modelo es estable e invertible ³.

Por lo tanto, el modelo queda de la siguiente forma:

$$z_t = \phi_5 z_{t-5} + \phi_{19} z_{t-19} + a_t$$

$$z_t = -0.2181854 z_{t-5} + 0.2435912 z_{t-19} + a_t$$

donde a_t es un proceso de ruido blanco con media cero y varianza constante (como se mostrará en breve).

5.2.3. Validación del modelo

Una parte fundamental de la construcción del modelo es su validación, pues con esta se determina si el modelo es realmente útil para explicar la serie y realizar los debidos pronósticos.

Pruebas de normalidad

Para comprobar si los residuos se distribuyen normal, se construyó la gráfica Q-Q plot y el histograma, al cual se le ajusta una curva normal. En ambos casos resulta evidente que los residuos siguen una distribución normal.

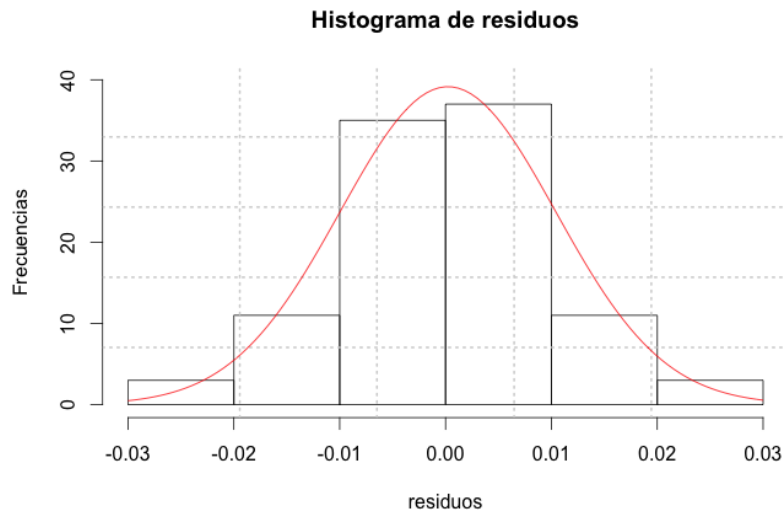


Figura 5.7: Histograma de los residuos

Para reforzar lo visto en las gráficas se realizaron las pruebas de normalidad más utilizadas como la Shapiro-Wilk, la Anderson-Darling y la Lilliefors. En el caso de estas pruebas, la hipótesis es $H_0 : a_t \sim N(0, \sigma^2)$ y como se desea no rechazar H_0 se espera ver un p-value mayor a $\alpha = 0.05$ que hace no rechazar la normalidad. Se tiene que:

³Estas características son de suma importancia, pues la condición de invertibilidad implica que se pueden expresar los errores como la suma ponderada de las observaciones actuales y pasadas, es decir, el los residuos presentes dependes de las observaciones pasadas de forma convergente; mientras que la estabilidad se refiere a la estacionariedad

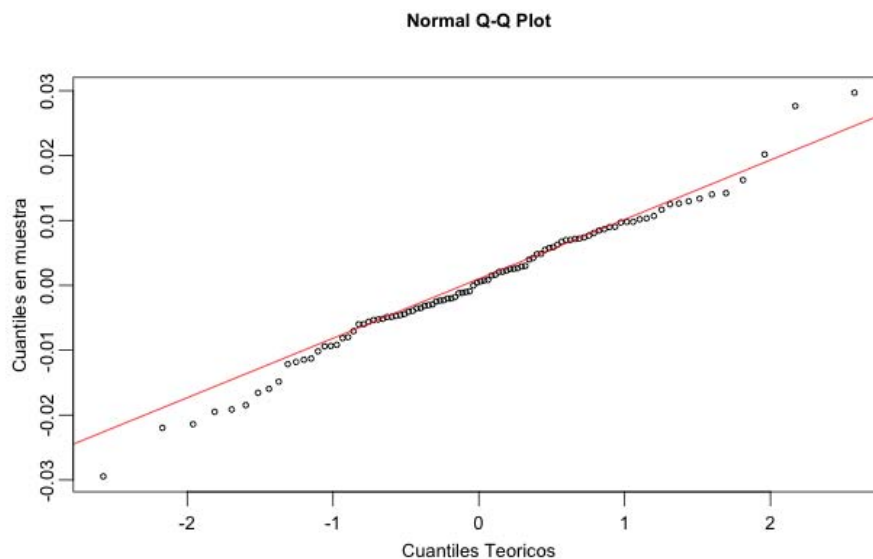


Figura 5.8: Q-Q Plot de los residuos

Shapiro-Wilk normality test
 $W = 0.9866$, $p\text{-value} = 0.4093$

Anderson-Darling normality test
 $A = 0.4076$, $p\text{-value} = 0.3422$

Lilliefors (Kolmogorov-Smirnov) normality test
 $D = 0.0707$, $p\text{-value} = 0.2523$

De acuerdo a las tres pruebas, no se acepta la normalidad de los residuales, por lo que el modelo Gaussiano parece ser un buen modelo para aproximarse a la verdadera distribución.

Pruebas de media y varianza constantes

Para comprobar qué tanto la media como la varianza de los errores es constante, se hace una gráfica de estos mediante la cual se puede constatar que su media es 0.000223970489, es decir muy cercana a cero y su varianza es 0.000103841865 como se observa en la *Figura 5.9*, donde se pueden ver como se comportan los residuos:

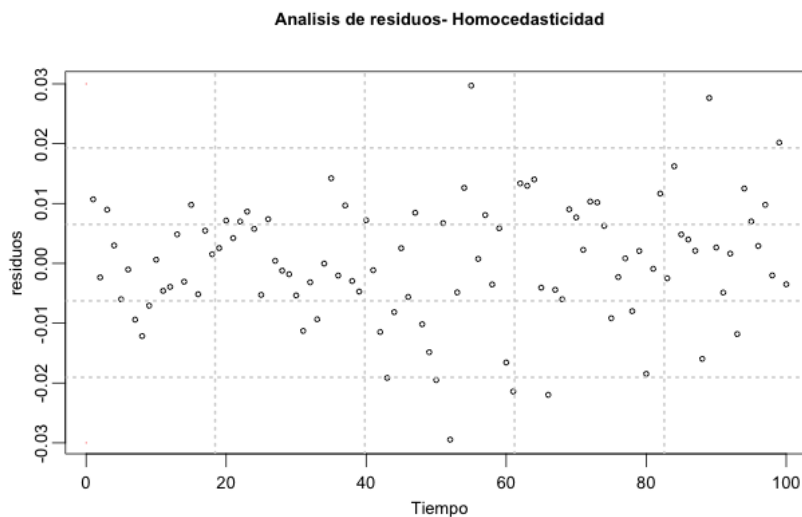


Figura 5.9: Gráfica de los residuos

En el caso de la varianza, se realizan las respectivas pruebas para determinar si esta procede de una varianza homogénea. Como se demostró que los residuos siguen una distribución normal, se realiza la prueba Bartlett y el test de Levene. Para ello se forman grupos de tamaño $n = 15$, con la finalidad de hacer aproximadamente 7 grupos, con lo cual se tiene:

Bartlett test of homogeneity of variances		
Bartlett's K-squared = 13.4669, df = 7, p-value = 0.06152		
Levene's Test for Homogeneity of Variance		
	Df	F value
group	7	1.57213
	92	0.15356

Para ambas pruebas, H_0 es la presencia de varianza homogénea, por lo que de acuerdo a estas no se rechaza la homoscedasticidad, lo cual sugiere que el modelo de varianza homogénea entre los residuos es adecuado. De la misma forma, para descartar que se presente un falso negativo debido a la formación artificial de grupos en estas pruebas, se realiza de forma alternativa la prueba de Breush-Pagan, considerando que si existe varianza homogénea entre los errores, estos no deben ser significativos con respecto del tiempo. Por tanto, se establece una regresión de los errores como variable dependiente, el tiempo como variable independiente y se realiza la prueba:

studentized Breusch-Pagan test
 data: $lm(e \sim t)$
 BP = 2.9621, df = 1, p-value = 0.08524

Como el $P - value = 0.08524$, la prueba sugiere no rechazar la hipótesis nula de que los residuos son homoscedásticos, por lo que los residuos de este modelo cumple con este supuesto.

Pruebas de correlación

Finalmente, para comprobar que los errores no están correlacionados y los residuos se comportan como un proceso de ruido blanco a_t , se realiza el correspondiente correlograma de los residuos.

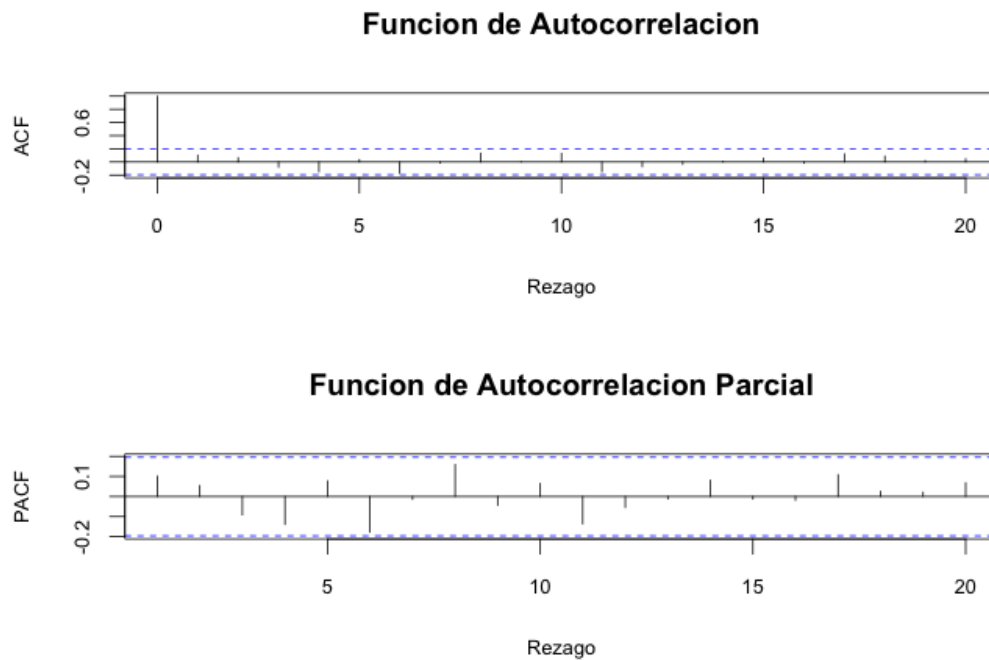


Figura 5.10: Correlograma de la FAC y FACP de los residuos

De la misma forma, se utiliza al estadístico de Ljung-Box para determinar si de manera conjunta, los residuos son independientes y se tiene:

Box-Ljung test
 X-squared = 33.2576, df = 28, p-value = 0.2264

Considerando que se tienen 28 grados de libertad, se calcula la región crítica de una $\chi^2_{28,0.05}$, es decir, a un $\alpha = 0.05$, lo cual da el valor de 41.337. Al ser menor el estadístico Ljung-Box no se rechaza H_0 y entonces el modelo de no correlación en los residuos parece ser adecuado.

5.3. Predicción

Una vez que se han determinado los modelos que mejor ajustan a la serie y se han hecho las pruebas pertinentes para comprobar que el modelo cumple con los requerimientos básicos, se procede a realizar el respectivo pronóstico de las siguientes tres observaciones, es decir, $x_{t+1}, x_{t+2}, x_{t+3}$.

Para ello se emplea el modelo que de acuerdo al análisis de los residuos resultó ser el modelo más adecuado: el $ARI(19, 1, 0)$. Entonces, se calculan los siguientes 3 valores futuros de la primer diferencia del logaritmo del IPC, y para una mayor interpretación y comparabilidad con los datos reales, se regresa la serie a su valor en niveles aplicando los operadores inversos, es decir, de $z_t \rightarrow y_t \rightarrow x_t$, de la misma forma que se realiza el mismo procedimiento con sus respectivos intervalos de donde se tiene:

Periodo	L95	Pronóstico	U95	Dato real
101	43453.12471	44325.01036	45214.39036	43882.71
102	43079.21579	44306.69834	45569.1563	43600.37
103	43018.08569	44524.07716	46082.79087	43296.55

Como se puede apreciar, los intervalos crecen a medida que incrementa el número de pronósticos. Esto se debe a que el error en la estimación es mayor a medida de que se incorporan más pronósticos, razón por la cual es importante moderar el número de pronósticos que se realizan.

Por su parte, a pesar de que la estimación puntual de $t + 1$ del $ARI(19, 1, 0)$ difiere en cerca de 400 unidades respecto del valor real, no es un valor tan lejano, además de que dicho valor real está incluido en el intervalo generado por el modelo.

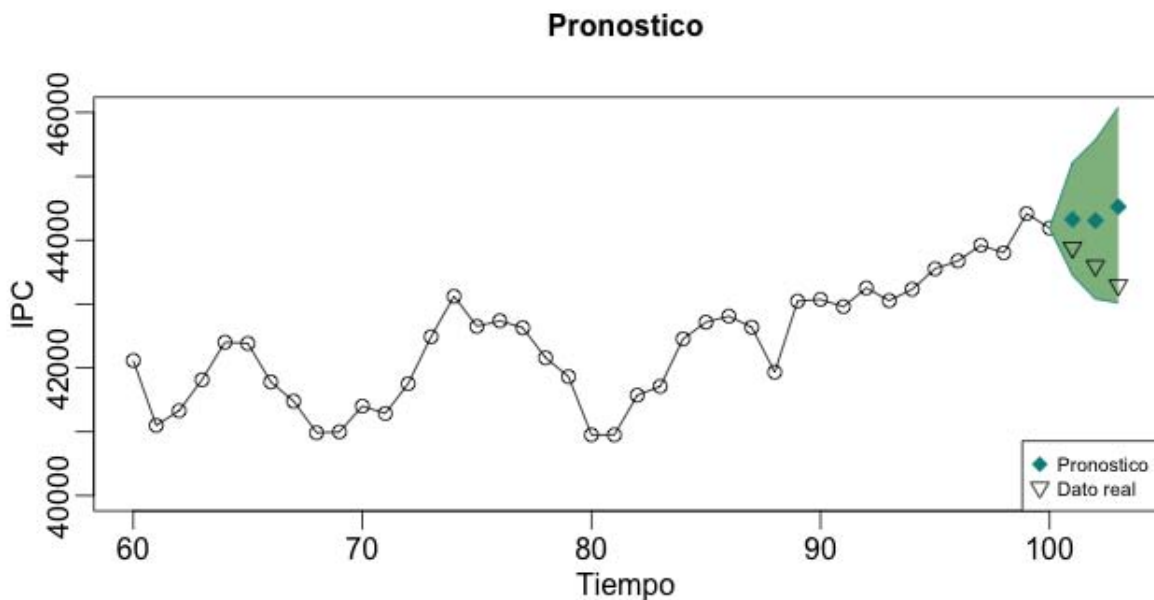


Figura 5.11: Gráfica de los valores de IPC y su predicción

Aunado a esta predicción, se generan bandas de volatilidad con validez predictiva mediante la técnica de *bandas de Bollinger*, el cual es un método empleado en análisis financiero y que es de suma utilidad en el ámbito aplicado de la series de tiempo [11, Politis, p. 4]. Esta herramienta de análisis creada por John Bollinger se trata de un intervalo generado a partir de una media móvil, que puede ser simple o exponencial, y su respectiva desviación estándar. Medante el uso de una media móvil como filtro, Bollinger crea un intervalo superior e inferior que proporcionan información sobre la volatilidad o actividad del mercado [11, Politis, p. 6-7]. De acuerdo al análisis de Bollinger, los parámetros por defecto son una media móvil simple de 20 periodos y 2 desviaciones estándar para el cálculo de las bandas superior e inferior. Sin embargo, por el tamaño de muestra del IPC y por la periodicidad, se ajusta una media móvil de 10 periodos, y con el respectivo ajuste, 1.5 desviaciones [11, Politis, p. 9] de donde se tiene la *Figura 5.12*

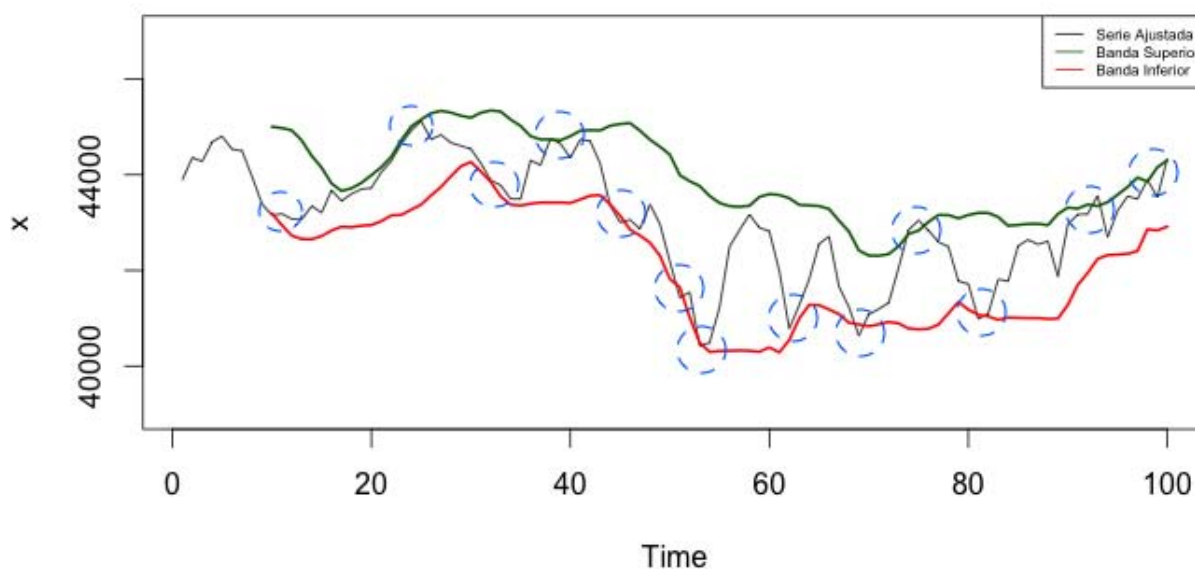


Figura 5.12: Gráfica de la serie ajustada con *las bandas de Bollinger*

Mediante este análisis técnico, se produce el efecto de que las bandas se ensanchan durante los periodos de mayor volatilidad, mientras se contraen durante los periodos menos volátiles. En un sentido económico/ financiero las *bandas de Bollinger* resultan útiles pues permiten identificar la oferta y demanda relativa del IPC. Si el IPC se mueve cerca de la intervalo superior, entonces indica que hay una fuerte demanda de acciones, lo cual produce que en periodo posterior se presente una baja en las cotizaciones producto del exceso de demanda. A la inversa, si el IPC se mueve cerca de la intervalo inferior, se presenta un exceso de oferta de las acciones que componen el IPC, lo cual produce *a posteriori* una aumento en las cotizaciones. De esta forma, el valor predictivo de esta técnica radica en que permite ubicar el rango relativo del IPC basado en su información pasada y también permite observar qué tan volátil es la serie de tiempo.

Conclusiones

Como se mencionó, las *series de tiempo* son conjuntos de datos diferentes a los usualmente empleados en otras áreas de la estadística pues poseen características muy particulares que no permiten que se apliquen los mismos métodos y técnicas. Esta diferencia radica principalmente, en que son datos ordenados y que se encuentran correlacionados al tiempo, lo cual implica que no hay una plena independencia entre las observaciones que componen al conjunto de datos. Debido a ello, surgen métodos y técnicas particulares para este tipo de conjuntos de datos que se conoce como *Análisis de series de tiempo*.

Se describieron y aplicaron dichas metodologías para explicar una serie de tiempo financiera univariada como el Índice de Precios y Cotizaciones de México, el cual es el índice más representativo de actividad bursátil en México, pues como se explicó, en su composición se procura que este índice contenga a las acciones más representativas del mercado de valores, motivo por el cual es un fiel reflejo del “estado de salud” de la actividad financiera en el país.

Mediante este ejercicio, y de acuerdo a lo planteado en los primeros capítulos, se comprobó que esta serie contaba con componentes tendenciales y estacionales, los cuales determinaron que la serie fuese no estacionaria en un principio. Sin embargo, debido a la importancia de que la serie cumpliera con el requisito de estacionariedad para encontrar un modelo capaz de reproducir el efecto estocástico de la serie, se realizaron dos transformaciones en la serie: primero una transformación a logaritmos $y_t = \log(x_t)$ (LIPC) buscando eliminar cualquier tipo de crecimiento exponencial, además de convertir a la serie en aditiva, y en segundo lugar una diferenciación de la serie $z_t = (1 - B)y_t$ (DLIPC) para eliminar cualquier tipo de efecto tendencial.

Una vez obtenida una serie estacionaria, se procedió a aplicar la metodología Box & Jenkins, con la cual se comprobó la gran importancia que posee la Función de Autocorrelación y la Función de Autocorrelación Parcial, pues a partir de estas funciones se pudo encontrar un patrón de correlación sistemático (el $ARI(19, 1, 0)$), que apoyado también en los criterios de akaike y de la función de verosimilitud, demostró ser un buen modelo cumpliendo con los requisitos necesarios, además de ser un modelo con alta *parsimonia*.

Por lo tanto, el modelo general para explicar el comportamiento del IPC fue un $ARI(19, 1, 0)$, de forma que se obtuvo la siguiente ecuación:

$$z_t = -0.2181854z_{t-5} + 0.2435912z_{t-19} + a_t$$

Con este modelo se generó un pronóstico de los siguientes tres valores ($t + 1, t + 2$ y $t + 3$), el cual estuvo cercano en el tiempo $t + 1$ con un intervalo bastante aceptable, sin embargo, conforme éste se alejaba del tiempo t incrementaba de forma bastante considerable el error.

Periodo	L95	Pronóstico	U95	Dato real
101	43453.12471	44325.01036	45214.39036	43882.71
102	43079.21579	44306.69834	45569.1563	43600.37
103	43018.08569	44524.07716	46082.79087	43296.55

Finalmente, se puede destacar que el modelo presentó buen ajuste, pues además de que los residuales cumplieron con todos los requerimientos necesarios, el pronóstico en si tampoco fue tan lejano. El único aspecto del modelo $ARI(19, 1, 0)$ que podría ser sujeto a crítica, en cuanto a la predicción, es la gran amplitud de los intervalos. Sin embargo, mediante *Bandas de Bollinger*, se demostró que esa gran amplitud de los intervalos se debe a la gran volatilidad de la serie.

Bibliografía

- [1] BOWERMAN, BRUCE (2007), *Pronósticos, series de tiempo y regresión: un enfoque aplicado*, Cengage Learning (Ed.), México, D. F.
- [2] BOX, GEORGE & JENKINS, GWILYM (2008), *Time Series Analysis. Forecastings and Control*, Wiley-Interscience (Ed.), 4ta edición, Nueva Jersey, EUA.
- [3] CHAN, NGAI (2002) *Time series: Applications to finance*, Wiley-Interscience (Ed.), Nueva York, EUA.
- [4] CHATFIELD, CHRIS (2004) *The Analysis of Time Series. An Introduction*, Chapman & Hall (Ed.), 6ta Edición, Boca Raton, Florida, EUA.
- [5] CHATFIELD, CHRIS (2000) *Time-Series Forecasting*, Chapman & Hall (Ed.), Boca Raton, Florida, EUA.
- [6] COWPERTWAIT, PAUL (2009) *Introductory time series with R*, Springer (Ed.), Londres, Reino Unido.
- [7] GUERRERO, VICTOR (2003) *Análisis estadístico de series de tiempo económicas*, 2a. edn, Editorial Thomson
- [8] KIRCHGÄSSNER, GEBHARD & WOLTERS, JÜRGEN (2007) *Introduction to Modern Time Series Analysis*, Springer (Ed.), Berlín, Alemania.
- [9] KITAGAWA, GENSHIRO (2010) *Introduction to Time Series Modeling*, Chapman & Hall (Ed.), Boca Raton, Florida, EUA.
- [10] MONTGOMERY, DOUGLAS (2002) *Diseño y análisis de experimentos*, 2a. edn, Editorial Limusa
- [11] POLITIS, DIMITRIS (2007) *Volatility bands with predictive validity*, Working Paper, University of California, San Diego, www.math.ucsd.edu/~politis
- [12] SHUMWAY, ROBERT & STOFFER, DAVID (2006) *Time Series Analysis and Its Applications with R Examples*, Springer (Ed.), Nueva York, EUA.
- [13] TSAY, RUEY (2005) *Analysis of Financial Time Series*, Wiley-Interscience (Ed.), 2da Edición, Nueva Jersey, EUA.
- [14] WOODWARD, WAYNE & GRAY, HENRY (2012). *Applied Time Series Analysis*, Chapman & Hall (Ed.), Boca Raton, Florida, EUA.

Apéndice A

Métodos para corregir la no estacionariedad

Un punto de referencia crucial en la construcción de modelos de Series de Tiempo es el supuesto de estacionariedad, como se describió en el *Capítulo 3*. Esta propiedad implica que la media y la varianza del proceso son constantes, mientras que las autocovarianzas dependen únicamente del rezago y no del tiempo. Sin embargo, en la práctica muchas series de tiempo financieras (y no financieras también por igual) no cumplen con este requisito indispensable para modelarlas bajo las metodologías y técnicas descritas en este trabajo (especialmente la metodología propuesta por Box y Jenkins). Debido a este motivo, en este anexo se presentan alternativas de transformaciones a las series de tiempo de acuerdo al componente que viola el supuesto de estacionariedad (ya sea media o varianza), que les permitirán a estas cumplir con dicho requisito tan importante.

No estacionariedad en varianza

Cuando una serie no es estacionaria en varianza, no se sostiene el supuesto de que ha sido generado por un proceso con varianza constante en el tiempo (varianza homogénea). La solución es transformar la serie de tal forma que se consiga estabilizar la varianza.

Un método propuesto por Guerrero [7]p. 109 sugiere elegir una potencia λ de manera que se cumpla la siguiente relación:

$$\sigma_t \backslash \mu_t^{1-\lambda} = \text{constante} \quad \text{para } t = 1, 2, \dots, N$$

En donde σ_t y μ_t son la respectiva desviación estándar y media de la serie x_t y N es el número de observaciones. En ese sentido, como para cada tiempo t se hace solo una observación de x_t (y por tanto es imposible estimar σ_t), Guerrero describe el siguiente procedimiento:

1. Dividir el total de datos en H grupos con $R = (n - k/H)$ elementos contiguos cada uno, siendo n el total de observaciones que se tienen para la serie x_t , dejando fuera de los cálculos a un total de k observaciones ($0 \leq k < R$) ya sea al final al comienzo, donde se resalta que la formación de grupos debe partir de una inspección visual de los datos con la intención de que en cada grupo la varianza parezca constante.
2. Se calcula la media \bar{x}_h y desviación estándar S_h de cada $h = 1, \dots, H$
3. Sin que exista una regla definida, escoger r -valores posibles para λ , que oscilen entre -1 y 1 .

4. Evaluar un coeficiente de variación para cada λ_i de la siguiente forma:

$$CV(\lambda_i) = \frac{\sigma_{\lambda_i}}{M(\lambda_i)}$$

Donde la idea es escoger a λ de las diferentes λ_i de menor CV.

Otra alternativa para estabilizar la varianza es la *transformación de Box-Cox*, la cual se define a partir de la siguiente regla, sea x_t el proceso que se desea transformar:

$$x_t^{(\lambda)} = \begin{cases} \frac{x_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(x_t) & \lambda = 0 \end{cases}$$

Donde λ es el parámetro de transformación. Además, las transformaciones Box-Cox no sólo estabilizan la varianza sino que también mejoran la aproximación a la distribución normal del proceso x_t .

Cabe señalar que en el caso de las series de tiempo financieras es muy común transformar la serie a base logarítmica por las diversas ventajas que esta transformación posee. A nivel de propiedades, además de convertir la serie en aditiva, esta transformación permite suavizar (linealizar) los crecimientos exponenciales que suelen afectar muy frecuentemente a este tipo de series de tiempo, mientras que a nivel interpretativo, si a esta transformación se le aplica un operador diferencia, permite que los resultados sean muy aproximados a los de una tasa de variación ordinaria.

No estacionariedad en media

El caso de la no estacionariedad en la media generalmente se debe a la presencia de tendencia en una serie de tiempo. La tendencia es un movimiento en el largo plazo asociado principalmente a la evolución de la variable y esta puede tener un comportamiento creciente o decreciente, exponencial o aproximadamente lineal. Sea cual sea el caso, la tendencia es un comportamiento sistemático que no permite a la serie ser estacionaria, pues provoca que esta no evolucione en torno a un nivel constante.

Como se explicó en el *Capítulo 1* y en el *Capítulo 3*, los tipos de tendencia más comunes son la tendencia determinista y la tendencia estocástica. En el caso de ambas, un método eficaz para eliminar este componente es mediante los operadores de retraso y diferencia. El primero, denotado por B , actúa sobre un proceso x_t de la siguiente forma:

$$Bx_t = x_{t-1} \quad B^2x_t = B(Bx_t) = Bx_{t-1} = x_{t-2} \quad \dots \quad B^jx_t = x_{t-j}$$

Asimismo se puede definir que $B^0x_t = x_t$. A partir de este operador retraso, se define al operador diferencia:

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

Donde el operador diferencia actúa como un polinomio, por lo que si se desea diferenciar la serie j veces se tiene:

$$\nabla^j x_t = (1 - B)^j x_t = \sum_{k=0}^j \binom{j}{k} (-1)^k B^k x_t = \sum_{k=0}^j \binom{j}{k} (-1)^k x_{t-k}$$

Apéndice B

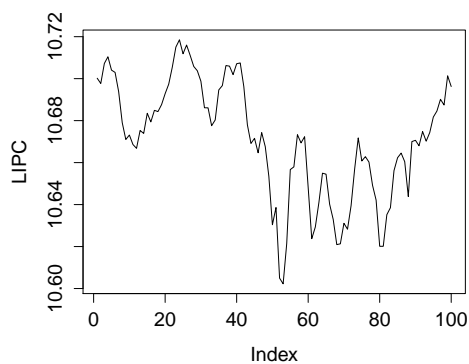
Ajuste del modelo en R

- Se extraen los datos de un archivo .csv y se les nombra como IPC

```
IPC <- read.csv("ipc.csv")
attach(IPC)
IPC<-as.data.frame(IPC)
```

- De acuerdo a la inspección visual, se determinó que lo más conveniente era trabajar la serie en logaritmos

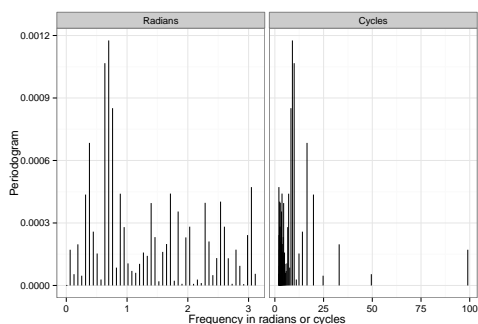
```
LIPC<-log(IPC)
```



Periodograma

- Se calcula el periodograma de donde se encuentra el punto máximo y se obtiene el ciclo

```
library(season)
per<-peri(LIPC)
```



```
max.location <- which.max(per$per)
ciclo<-per$c[max.location]
```

9

- Se controla el número de dígitos, se ajusta el mejor modelo de acuerdo a lo obtenido por JMP, que es el $ARIMA(19, 1, 0)$, y se revisan los coeficientes de dicho modelo

```
od <- options(digits = 8)
TS<-arima(LIPC,order=c(19,1,0))

fixed = c(0, 0, 0, 0, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA))
fixedc(0,0,0, NA,0,0,0,0,0,0,0,0,0,0,0,0,0,0, NA))
```

Call: arima(x = LIPC, order = c(19, 1, 0), transform.pars = FALSE, fixed = c(0, 0, 0, 0, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA))

Coefficients: ar1 ar2 ar3 ar4 ar5 ar6 ar7 ar8 ar9 ar10 ar11 0 0 0 0 -0.218185482 0 0 0 0 0 0 s.e. 0 0 0 0 0.093059671
0 0 0 0 0 0 ar12 ar13 ar14 ar15 ar16 ar17 ar18 ar19 0 0 0 0 0 0 0.243591173 s.e. 0 0 0 0 0 0 0.095773573

σ^2 estimated as 0.000102736027: log likelihood = 313.36, aic = -620.71

- Se obtienen estadísticos para determinar qué tan bueno fue el ajuste como el estadístico t para los coeficientes y el R^2 ajustado para el modelo

```
#Prueba t
coef(TS)/diag(sqrt(vcov(TS)))
```

```
ar5 ar19 -2.34457612327 2.61758042520
```

```
library(forecast)
cor(fitted(TS),LIPC)^2
```

0.870850426038

Diagnóstico del modelo

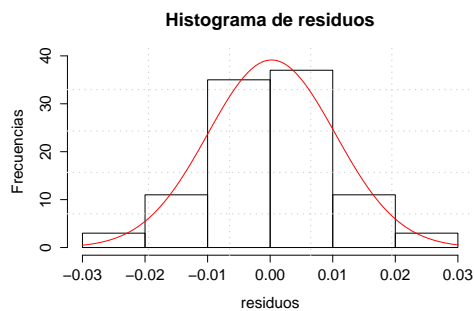
- Para realizar el respectivo diagnóstico del modelo, se extraen los residuales del modelo

```
e<-TS$res
```

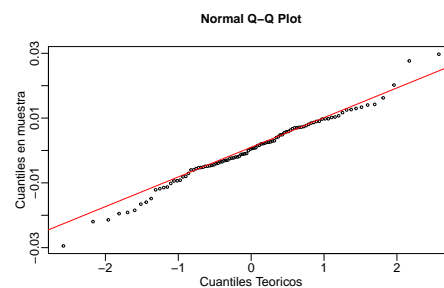
Pruebas de Normalidad

- Se hace una exploración gráfica de los residuales para ver si estos cumplen con el supuesto de normalidad mediante un histograma con una curva normal ajustada y un Q-Q plot

```
hist(e,freq=F,xlab="Residuales", ylab="Frecuencias",
     main="Histograma de Residuales")
grid(5, 5, lwd = 2)
curve(dnorm(x,mean(e),sqrt(var(e))),
     add=TRUE, col=2)
```

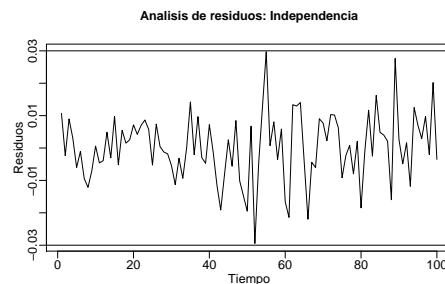


```
qqnorm(e,xlab="Cuantiles Teoricos", ylab="Cuantiles en muestra",
       main="Normal Q-Q Plot",cex=0.5,
       cex.main=0.8,cex.axis=0.8,cex.lab=0.8,
       mgp=c(1.5,0.5,0))
qqline(e, col=2)
```



- Asimismo, se observa el comportamiento de los residuales para verificar, si de acuerdo a los supuestos de normalidad, se distribuyen con $\mu = 0$ y $\sigma^2 = 1$

```
plot(e,xlab="Tiempo",ylab="Residuales")
abline(a=0,b=0)
```



- Se realizan las pruebas de normalidad más utilizadas para corroborar si los residuales se comportan como una distribución normal.

```
#Shapiro Wilks
library(nortest)

shapiro.test(e)
```

Shapiro-Wilk normality test

data: e W = 0.9866, p-value = 0.4093

```
#Anderson Darling
library(nortest)

ad.test(e)
```

Anderson-Darling normality test

data: e A = 0.4076, p-value = 0.3422

```
#Lilliefors
lillie.test(e)
```

Lilliefors (Kolmogorov-Smirnov) normality test

data: e D = 0.0707, p-value = 0.2523

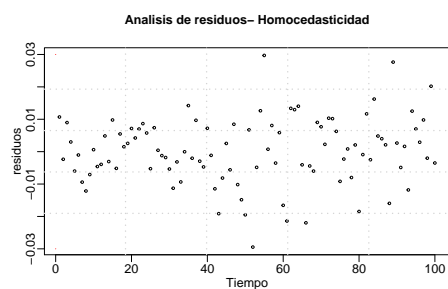
Pruebas de homogeneidad de varianzas

- Se grafican los residuos con respecto al tiempo para comprobar si las varianzas son homogéneas entre los residuales

```

par(mfrow=c(1, 1))
t<-c(1:81)
plot(t,e,xlab="Tiempo", ylab="Residuales",
     main="Análisis de Residuales- Homocedasticidad",
     axes=T,cex=0.5, cex.main=0.8,cex.axis=0.8,
     cex.lab=0.8,mgp=c(1.5,0.5,0))
lines(c(0,40),c(40,40), col=2 )
lines(c(0,40),c(-40,-40), col=2 )
grid(5, 5, lwd = 2)

```



- De la misma forma se hacen las correspondientes pruebas para determinar si hay homocedasticidad en las varianzas de los residuales.

```

#Bartlett
g=seq(1:81)
g=round(g/9,0)
bartlett.test(e,g)

```

Bartlett test of homogeneity of variances

data: e and v Bartlett's K-squared = 13.4669, df = 7, p-value = 0.06152

```

#Levene
library(car)

g=seq(1:81)
g=round(g/9,0)
levene.test(e,g)

```

Levene's Test for Homogeneity of Variance (center = median) Df F value Pr(>F) group 7 1.57213 0.15356 92

```

#Breusch-Pagan
library(lmtest)

bptest(lm(e ~ t))

```

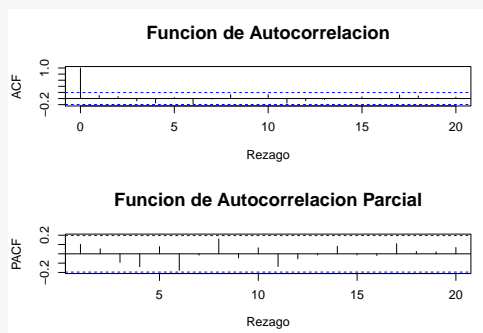
studentized Breusch-Pagan test

data: $\text{lm}(e \sim t)$ BP = 2.9621, df = 1, p-value = 0.08524

Pruebas de Independencia

- Para comprobar si los residuales son independientes conforme evolucionan en el tiempo, se utiliza la función de autocorrelación y la función de autocorrelación parcial

```
par(mfrow=c(2, 1))
acf(e,xlab="Rezago", ylab="ACF",
    main="Funcion de Autocorrelacion", axes=TRUE, cex=0.5,
    cex.main=0.5,cex.axis=0.8,cex.lab=0.8,
    mgp=c(1.5,0.5,0))
```



- Y se realizan las respectivas pruebas de independencia

```
#Ljung-Box
BT = Box.test(e, lag=30, type = "Ljung-Box", fitdf=2)
BT
```

Box-Ljung test

data: e X-squared = 33.2576, df = 28, p-value = 0.2264

```
qchisq(0.95,28)
```

```
[1] 41.337138
```

```
1-pchisq(40.3407,25)
```

```
[1] 0.026897726
```

Predicción de datos futuros

- Una vez validado el modelo, se realiza el respectivo pronóstico de los valores futuros $t + 1$, $t + 2$ y $t + 3$ de la serie agregando el componente heterocedástico de los residuos extraído del modelo $ARCH(1)$. Las predicciones se presenta en términos del IPC, sin diferencia y sin logaritmos

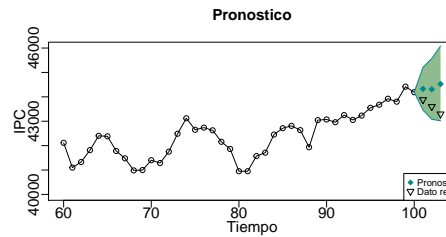
```
pronostico<-predict(auto.arima(LIPC,max.p=5, max.q=5), n.ahead = 3)
Pronostico<-exp(pronostico$pred)
```

- Asimismo, también se obtienen los respectivos intervalos de confianza del pronóstico al 95 %

```
I95 = exp(pronostico$pred- 1.96*pronostico$se)
S95 = exp(pronostico$pred+ 1.96*pronostico$se)
```

- Se grafica el pronóstico para apreciarlo de forma más visual y se agregan los datos reales del IPC para compararlos con los del pronóstico

```
Li<-length(e)
plot(c((Li-40):Li),IPC[(Li-40):Li,], type="o",
     xlim=c((Li-40),(Li+3)), ylim=c(40000,46000),
     xlab="Tiempo", ylab="IPC",
     main="Pronostico",
     axes=T,cex=1, cex.main=1,cex.axis=1,cex.lab=1,
     mgp=c(1.5,0.5,0))
polygon(c(c(Li:(Li+3)),rev(c(Li:(Li+3)))),
        c(c(IPC[Li,],as.integer(I95)),rev(c(IPC[Li,],as.integer(S95)))),
        col = "darkseagreen", border = FALSE)
points(Pronostico, pch=18, col="cyan4",cex=1.3)
lines(c(Li:(Li+3)), c(IPC[Li,],as.integer(I95)),
      cex=2, col="cyan4")
lines(c(Li:(Li+3)), c(IPC[Li,],as.integer(S95)),
      cex=2, col="cyan4")
#Valores reales
R<-c(43882.71,43600.37,43296.55)
R<-ts(R, frequency = 1, start = c(101))
points(R, pch=25, col="black",cex=1)
par(new=TRUE)
legend("bottomright", cex = 0.65, legend = c("Pronostico", "Dato real"),
      col = c("cyan4","black"), pch = c(18,25), pt.cex=c(1.2,1))
```



- Finalmente, se generan las bandas de Bollinger para determinar la volatilidad y puntos de corte del IPC

```
library(TTR)
BB<- BBands(LIPC, hlkwdn = 10, sd = 1.5)

ajuste<-fitted.values(TS)
inf<-BB[,1]
sup<-BB[,3]
plot(exp(ajuste),type='l',ylim=c(39000,47000))
lines(exp(inf),col="red", lwd="2")
lines(exp(sup),col="darkgreen", lwd="2")
legend("topright", cex = 0.5, legend = c("Series Ajustada", "Banda Superior",
"Banda Inferior"), col = c("black","darkgreen","red"), lty = c(1,1,1), pt.cex=c(1.2,1,1))
```

