



Benemérita Universidad Autónoma de Puebla

**Facultad de Ciencias Físico Matemáticas
Licenciatura en Matemáticas Aplicadas**

**Análisis de la radiación solar en el municipio
de Tlaxco-Tlaxcala usando la metodología
de Box-Jenkins**

TESIS

que para obtener el título de

Licenciada en Matemáticas Aplicadas

Presenta

Gloria Aragón Merino

Director de tesis

Dr. Bulmaro Juárez Hernández

Puebla, Pue.

Noviembre 2018.

DEDICATORIA

Agradecimientos

INTRODUCCIÓN

La radiación solar es la energía transferida por el sol a través de ondas electromagnéticas, a través de esta se pueden inferir procesos de transferencia de energía en las diferentes capas atmosféricas que se manifiestan como fenómenos climáticos que pueden medirse.

Uno de los objetivos principales en este trabajo es predecir el comportamiento a corto plazo de la radiación solar, se utilizó una base de datos tomadas de la estación meteorológica de la región de Tlaxco - Tlaxcala. Se encuentra información a partir del 18 de Junio del 2011 al 21 de Junio del 2017, la base de datos cuenta con datos que son recolectados diariamente cada media hora, para los cuales se realizó un filtrado para así obtener promedios semanales. Se realiza el análisis de los datos usando series de tiempo, principalmente se trabaja con la metodología de Box - Jenkins que, mediante este análisis se halla un modelo que mejor se ajuste a los datos para después obtener pronósticos para la radiación solar.

Este trabajo pretende ser autocontenido, por lo que se presentan los conceptos básicos necesarios sobre meteorología y series de tiempo, lo cual se desarrolla en el Capítulo 1 donde se da una breve descripción del clima y tiempo, así como las variables atmosféricas, la descripción de la radiación, además de, el monitoreo atmosférico. En el Capítulo 2 se mencionan los conceptos básicos de series de tiempo, iniciando por el concepto de proceso estocástico para luego introducir la definición serie de tiempo, seguido de esto, está la definición de serie de tiempo estacionaria, la función de autocovarianza, autocorrelación y autocorrelación parcial, donde estas dos últimas son de ayuda para la detección de los modelos de

series de tiempo estacionarias, tales como los modelos autoregresivos, modelos de promedios móviles y autoregresivos de promedios móviles. En el Capítulo 3 se muestran modelos para series de tiempo no estacionarias, tales como el modelo ARIMA, después se mencionan métodos para la estabilización de la varianza y de la media. Inmediatamente, se presentan los modelos estacionales multiplicativos ARIMA, así como la identificación del modelo, estimación de parámetros (donde se presenta estimación por método de momentos y estimación por máxima verosimilitud), diagnóstico del modelo (verificación de que las suposiciones del modelo se cumplan), selección del modelo y por último el pronóstico. En el Capítulo 4 se aplican los conceptos antes mencionados para el análisis de los datos de la radiación solar. Y por último en el Capítulo 5 se presentan las discusiones y conclusiones de este trabajo.

En el Apéndice A se muestra la representación de un proceso $AR(p)$ como un $MA(\infty)$. En el Apéndice B se mencionan las condiciones que deben cumplir las raíces del proceso $AR(2)$ para que este sea estacionario (así como un $MA(2)$ para que este sea invertible). En el Apéndice C se presenta el test de Dickey-Fuller, que sirve para la verificación de la estacionariedad de la serie de tiempo. En el Apéndice D se hace mención del criterio de información de Akaike y BIC, que son utilizados para la selección de un modelo ARIMA. En el Apéndice E se explica sobre cómo se realizó la filtración de los datos, en el Apéndice F, se presentan las partes principales de los diagramas de cajas y se finaliza con el Apéndice G donde se presenta el código en R de la serie de tiempo.

Gloria Aragón Merino
Facultad de Ciencias Físico Matemáticas
Benemérita Universidad Autónoma de Puebla
Puebla, Pue. México
2018

ÍNDICE GENERAL

Introducción	I
1. Dinámica atmosférica	1
1.1. ¿Qué es el clima?	2
1.2. Variables atmosféricas	3
1.3. Radiación solar en la atmósfera	5
1.4. Monitoreo atmosférico	8
2. Conceptos básicos de series de tiempo	11
2.1. Proceso estocástico	11
2.2. Serie de Tiempo	12
2.3. Proceso estacionario y la autocorrelación	14
2.3.1. Serie de tiempo estrictamente y débilmente estacionaria	15
2.3.2. Función de autocovarianza (FACV) y autocorrelación (FAC)	16
2.3.3. Función autocorrelación parcial (FACP)	19
2.3.4. Ejemplos de series de tiempo estacionarias	21
2.3.5. Estimación de la media, varianza, y las autocorrelaciones simple y parcial	29
2.3.6. Operadores de retardo y diferencia	31
2.4. Modelos de series de tiempo estacionarias	33
3. Modelos para series de tiempo no estacionarias	61
3.1. Modelo ARIMA(p,d,q)	62

3.1.1.	No estacionariedad en media	64
3.1.2.	No estacionariedad en varianza	67
3.2.	Modelos estacionales multiplicativos ARIMA	74
3.3.	Identificación, Estimación, Diagnóstico, Selección del modelo y Pronóstico	79
3.4.	Identificación	79
3.5.	Estimación	83
3.5.1.	Estimación por Método de Momentos	83
3.5.2.	Mínimos cuadrados	85
3.5.3.	Método de máxima verosimilitud	87
3.5.4.	Estimación de máxima verosimilitud incondicional	89
3.5.5.	Función exacta de verosimilitud	91
3.6.	Diagnóstico	94
3.7.	Selección del modelo	96
3.8.	Pronóstico	96
4.	Caso de estudio	105
4.1.	Zona de estudio	105
4.2.	Análisis de la serie de tiempo de los datos de radiación	111
5.	Discusión y Conclusiones	143
A.	Representación de un proceso AR(p) como un MA(∞)	145
B.	Condiciones de las raíces del proceso AR(2)	149
C.	Test de Dickey- Fuller	153
D.	Criterios de selección	159
E.	Selección de datos	161
F.	Análisis de diagramas de caja	165
G.	Código en R	169
	Bibliografía	176

Gloria Aragón Merino

Análisis de la radiación solar en el
municipio de Tlaxco-Tlaxcala usando
la metodología de Box-Jenkins

DINÁMICA ATMOSFÉRICA

El planeta está formado por la litosfera, la hidrosfera, la criosfera, la biosfera y la atmósfera.

Esta última es un escudo importante que protege la flora, la fauna y la vida humana de la radiación solar y de distintos meteoros que pueden causar daños a la vida en la Tierra. Es una capa delgada de unos cientos de kilómetros de espesor que rodea a nuestro planeta desde la superficie terrestre hasta aproximadamente 900 km de altura.

En la atmósfera, se encuentran reunidos una infinidad de procesos atmosféricos que dan lugar a los fenómenos meteorológicos y climatológicos: huracanes, tormentas, nevadas, tornados, brisas, inundaciones, sequías, etcétera., los cuales afectan a las actividades diarias y frecuentemente resultan pérdidas de vida y de bienes materiales. Estos fenómenos meteorológicos que son de gran importancia se encuentran comprendidos en los primeros 15 km de altitud.

La atmósfera se divide en cuatro capas: la troposfera, la estratosfera, la mesosfera y la termosfera, y está conformada de una mezcla de tres tipos de gases: gases permanentes, gases variables, constituyentes no gaseosos. [25]

1.1. ¿Qué es el clima?

El clima terrestre es el producto de la constante y compleja interacción entre los efectos físicos de la atmósfera, los océanos, las capas de hielo y nieve, los continentes y sobre todo la vida en el planeta.

Cada día hay variaciones en las condiciones atmosféricas de nuestro planeta, por lo que también a diario se presentan variaciones en las condiciones de la temperatura y lluvia planetaria. Sin embargo, a esas variaciones no se les llama clima, sino **estado del tiempo**, es decir, el tiempo es un estado de la atmósfera que **se presenta en un lugar determinado durante un periodo corto de tiempo determinado**. De ahí que los pronósticos meteorológicos que usted puede ver todos los días en la televisión o escuchar por radio son los cambios en el tiempo más no en el clima.

En algunas partes del planeta, en especial en nuestro país, se espera frío en invierno y calor en verano, que en primavera inicien las lluvias y que sean más intensas y regulares en verano. Eso que esperamos se llama **clima**. Pero, puede suceder que un día invernal sea sumamente caluroso, que las lluvias se retrasen o que se tengan intensos aguaceros en invierno. Esto no es señal para decir que el clima ha cambiado, sino que son parte de una variación del tiempo que se espera.

Para hablar del clima de un país es necesario medir diariamente las condiciones de temperatura, lluvia, humedad y viento, observar las condiciones de nubosidad, la trayectoria de los huracanes, las masas de aire frío que entran por el norte de nuestro país, etcétera. Para ello se requiere conocer durante varios años el estado del tiempo. Por lo que varios años de mediciones y observaciones le permiten a uno decir que el promedio de temperatura en verano de México es de 26 grados centígrados y que llueve en esa estación unos 370 milímetros. Esta cantidad de años debe ser por lo menos de tres décadas, según los especialistas, para poder hablar con seguridad del clima esperado. Entonces podríamos decir que **el clima es el tiempo promedio para un determinado lugar**.

Para medir algunas características que determinan el tiempo existen las estaciones meteorológicas, y para describir el clima existen las estaciones climatológicas (ver Figuras 1.1). [13]



Colégio de Pongas/Instituto Estación Agro-Meteorológica. [Figura]
Recuperado de <http://www.casipos.mz/acceso/latex.html>

(a) Estación COLPOS.



Wheeler, G. (2017). 3 New Resources to Help You Design Your Automated Weather Station. [Figura]. Recuperado de <https://www.campbellsci.com/blog/resources/design-automated-weather-station>

(b) Estación meteorológica.

Figura 1.1: Estaciones.

1.2. Variables atmosféricas

Se ha conocido el clima y el tiempo atmosférico a través del estudio de las variables que los afectan de manera más directa.

Las variables atmosféricas son: **la temperatura atmosférica, las precipitaciones en sus distintas formas (lluvia, nieve, granizo), presión atmosférica, humedad, velocidad y dirección del viento y radiación solar.** Además de estas variables, existen algunos factores que modifican las variables atmosféricas, estos son: **latitud, altitud, la distancia al mar, las corrientes marinas, relieve, etc.**

Temperatura

Como se sabe, toda la materia está compuesta de átomos y moléculas que están en movimiento permanente. El estado en que se encuentra la materia dependerá de si la energía cinética de las partículas aumenta o disminuye, entonces la temperatura del cuerpo (sólido) o fluido (líquido o gas) también aumenta o disminuye, respectivamente. De manera análoga, cuando un cuerpo o fluido se calienta o se enfría, su energía cinética y su temperatura aumenta o disminuye, respectivamente. La sensación de calor o frío es quizás la que representa de manera más fácil el concepto de temperatura.

La temperatura de un fluido también se puede relacionar con el concepto del movimiento de moléculas que componen el fluido. Es decir, mientras mayor sea la energía cinética promedio de las moléculas, mayor será la temperatura del fluido; en otras palabras, la temperatura es proporcional al calor contenido en un cuerpo. La temperatura puede ser medida en diversas escalas, las más usuales son la escala Centígrado o Celsius (C), Fahrenheit (F). La escala centígrado está basada en que la temperatura de congelación del agua es de 0 grados y la temperatura de ebullición del agua es de 100 grados, esta escala es la más usada actualmente. [25]

Precipitación

La precipitación se refiere a partículas de agua en fase sólida o líquida que se forman en la atmósfera y que caen y alcanzan la superficie de la Tierra.

Esto es, las nubes que parecen flotar en el cielo, siempre caen, sus gotitas de agua son arrastradas lentamente por la gravedad. Debido a que sus gotitas de agua son tan pequeñas y livianas, puede tomar 21 días para caer 304.8 metros y las corrientes de viento pueden interrumpir fácilmente su descenso. El agua líquida cae como lluvia o llovizna. Todas las gotas de lluvia se forman alrededor de partículas de sal o polvo. Las gotas de agua o hielo se adhieren a estas partículas, luego las gotas atraen más agua y continúan creciendo hasta que son lo suficientemente grandes como para caer de la nube.

Las gotas de llovizna son más pequeñas que las gotas de lluvia. En muchas nubes, las gotas de lluvia en realidad comienzan como pequeños cristales de hielo que se forman cuando una parte o toda la nube está por debajo del punto de congelación. A medida que los cristales de hielo caen dentro de la nube, pueden colisionar con gotas de agua que se congelan sobre ellos. Los cristales de hielo continúan creciendo, hasta que son lo suficientemente grandes como para caer de la nube. Pasan a través del aire caliente, se derriten y caen como gotas de lluvia.

Generalmente se mide en unidades lineales de agua líquida (milímetros o pulgadas) con el supuesto de que esta medida es por unidad de área. Cuando los cristales de hielo se mueven dentro de una nube muy fría y las gotas de agua se congelan en los cristales de hielo, la nieve caerá de la nube. [23]

Presión

La atmósfera consiste en una variedad de gases que forman una envoltura alrededor de la Tierra y que se mantienen allí por gravedad. La presión de la atmósfera no es la misma en cualquier parte, ésta, depende de la altura.

Es decir, las moléculas de aire son atraídas hacia la Tierra y se compactan más a medida que se acercan al nivel del mar porque cada capa de la atmósfera está siendo comprimida por la masa de aire que está sobre ella, en pocas palabras, a menor altura la cantidad de aire que queda comprimida por encima es mayor, y a mayor altura la cantidad de aire que queda comprimida por encima es menor. Si esta masa se multiplica por la fuerza de la gravedad, el resultado es el peso del aire, que se describe en términos del área sobre la que se mide (es decir, libras por pulgada cuadrada, gramos por centímetro cuadrado). Este peso del aire es la presión atmosférica. [24]

Humedad relativa

La mayor parte del aire que nos rodea tiene agua en forma de gas llamado vapor de agua. La humedad se refiere a la cantidad de vapor de agua del aire.

Los dos tipos de humedad más usados son la **humedad absoluta**, la cual expresa el contenido de vapor de agua por unidad de volumen de aire. Este parámetro indica con relativa precisión la probabilidad de precipitación atmosférica.

La **humedad relativa** es la relación de la cantidad de vapor de agua en el aire en comparación con la cantidad de vapor requerida para la saturación (a una temperatura y presión particulares), medida en porcentaje.

Si el valor de la humedad relativa es del 50 %, quiere decir que la humedad que contiene el aire en ese momento es el 50 % de vapor de agua respecto al máximo que podría contener, que sería el 100 %. Si la humedad relativa es del 100 % quiere decir que el aire está lleno de vapor al máximo. Está saturado de vapor. No admite más vapor. Si se intenta meter más vapor, este se condensará.

Velocidad y dirección del viento

Al movimiento de aire cerca de la superficie de la Tierra y en una dirección horizontal cercana es llamada viento. Los parámetros que afectan al viento son la dirección y la velocidad. La dirección del viento es determinada por una veleta, moviéndose libremente en un eje vertical en una aguja o poste elevado. Por otro lado, la velocidad del viento es la razón del movimiento del viento en unidad de distancia por unidad de tiempo. [22]

1.3. Radiación solar en la atmósfera

La radiación solar se podría definir como la emisión, propagación y absorción de la energía en forma de ondas electromagnéticas y, que es expresado como la

energía emitida por unidad de tiempo que incide en una superficie (Watts/m^2). La radiación que llega del Sol a la Tierra es la fuente principal de energía para la generación de los fenómenos en la atmósfera, en los océanos y, en general para la vida en la Tierra.

El Sol está continuamente radiando energía en forma de ondas electromagnéticas, la cantidad total de energía solar que llega a la Tierra depende de los siguientes factores: emisión de radiación solar, distancia entre el Sol y la Tierra (es la distancia promedio de 1 unidad astronómica (149 597 870.691 kilómetros) del Sol a la Tierra), altitud del Sol (es el ángulo entre los rayos solares y el plano tangente a la superficie terrestre) y longitud de día (es el tiempo en que está expuesta la faz de la Tierra a la radiación solar, también afecta la cantidad de energía solar que recibe la Tierra).

Por otro lado, la atmósfera juega un papel muy importante en cuanto al balance entre la radiación solar que entra a la Tierra y la radiación que sale de la misma (es decir que está sujeta a un mecanismo de absorción y dispersión), manteniendo un promedio global de temperatura superficial de 15 grados C. Si la Tierra careciera de una atmósfera para absorber y disminuir el calor que llega del Sol, entonces gran parte de esa energía solar sería reflejada inmediatamente y la temperatura media sobre la superficie terrestre estaría por debajo de los 0 grados C. Cuando la radiación solar llega a la superficie de la Tierra, es transformada en calor y en radiación infrarroja. Esta energía la absorbe la atmósfera y la emite nuevamente al espacio. Al término de un año, el sistema terrestre devuelve exactamente la misma cantidad de energía que llega del Sol ya que de no ser así la Tierra se calentaría (o se enfriaría) ilimitadamente.

No toda la radiación solar que llega a la atmósfera logra penetrar la superficie de la Tierra (ver Figura 1.2). De esta, una pequeña parte es reflejada al espacio, de la restante una parte es absorbida por la atmósfera (principalmente por el O_2 y el CO_2). Y de la otra parte restante (la cual es absorbida por la superficie terrestre), una parte se transforma en movimiento de la atmósfera y la hidrosfera, las cuales transportan el exceso de calor de los trópicos hacia los polos. Y la otra parte es transformada en radiación de onda larga, la cual es emitida a la atmósfera, reteniéndola por el efecto invernadero, para después remitirla a la superficie y al espacio exterior, manteniendo un balance radiativo en el sistema terrestre. Los principales gases de invernadero son: el vapor de agua, el dióxido de carbono, el metano, el óxido nítrico y los clorofluorocarbonos, los cuales retienen un gran porcentaje de la radiación infrarroja terrestre. [25]

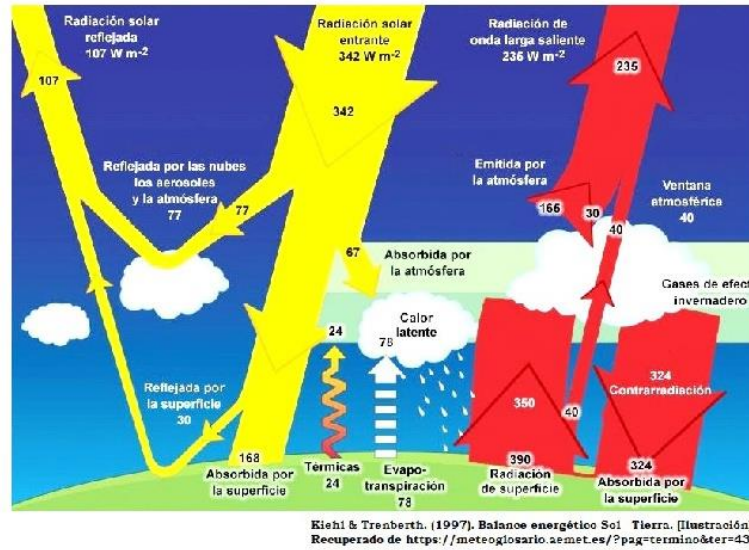


Figura 1.2: Mecanismo de la Radiación Solar.

Pero, si la atmósfera contiene demasiados de estos gases, toda la Tierra se convierte en un invernadero cada vez más caliente. La atmósfera conserva demasiado calor por la noche, en vez de dejarlo salir al espacio. Entonces, al día siguiente, el Sol calienta la superficie de la Tierra todavía más (ver Figura 1.3).

La atmósfera funciona como un invernadero, cada día hace un poco más de calor que el día anterior. No es posible medir este efecto de un día para otro, ni de un año a otro. Sin embargo, a lo largo de los años, apenas unos pocos grados de calentamiento comienzan a provocar cambios. El aumento de temperatura del agua oceánica afecta el clima en todas partes del mundo. Algunos sitios tienen tormentas más intensas, mientras que en otros apenas llueve.

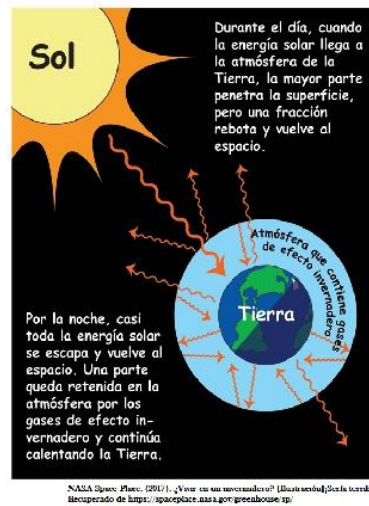


Figura 1.3: Efecto invernadero.

Es de interés mencionar que, la atmósfera en alguna ubicación de la superficie terrestre, es frecuentemente clasificada en atmósfera con nubosidad y sin nubosidad. Los mecanismos de absorción y dispersión son similares en ambos casos. Sin embargo, es evidente que la menor disminución de radiación toma lugar en un cielo sin nubosidad. En consecuencia, la radiación máxima es recibida en la superficie terrestre bajo las condiciones de cielo sin nubosidad. La radiación se puede dividir como sigue:

1. Radiación directa o rayo: Es la radiación solar recibida en la superficie terrestre son cambiar la dirección, es decir, en línea con el sol.
2. Radiación difusa: Es la radiación recibida en la superficie terrestre desde todas las direcciones del hemisferio celeste (después de ser sujeto a la dispersión en la atmósfera).
3. Radiación global o total: Es la suma de la radiación directa y la difusa. [30]
En ésta tesis se trabajó con la radiación total.

1.4. Monitoreo atmosférico

Para llevar a cabo la obtención de los datos meteorológicos es necesario contar con un instrumento que nos indique el valor de cada variable, en este caso, una estación meteorológica la cual permite cuantificar mediciones y hacer registros regulares de diversas variables meteorológicas (por ejemplo la velocidad y dirección del viento, la presión atmosférica, la insolación del suelo, la humedad, la radiación solar, la caída del agua, la temperatura). Una estación meteorológica cuenta con sensores específicos tales como anemómetro, barómetro, heliógrafo, higrómetro, piranómetro, pluviómetro, termómetro entre otros. El flujo de la radiación es usualmente medido mediante la ayuda de un piranómetro o un pirheliómetro. [30]

Los datos registrados se utilizan tanto para la elaboración de predicciones meteorológicas a partir de modelos numéricos como para estudios climáticos. El instrumental meteorológico registrará sus mediciones en función del relieve topográfico y urbano donde se encuentre instalado, ya que es necesario que este se encuentre en un lugar despejado, sin elementos de encajonamiento que introduzcan una alteración en los datos obtenidos y en los procesos de medición. Estos aparatos deben cumplir con regularidad en su funcionamiento, precisión, sencillez en el diseño, comodidad de manejo y solidez de construcción.

Piranómetro

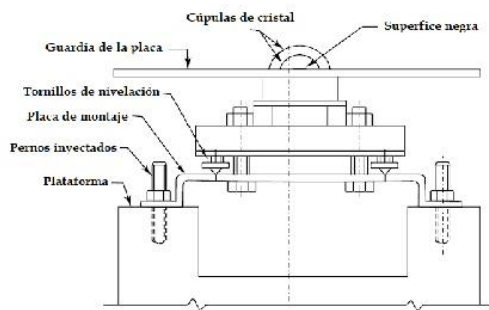


Figura 1.4: Piranometro.

El piranómetro es un instrumento que mide la radiación global o difusa que cae sobre una superficie sobre un campo hemisférico de visión (ver Figura 1.4). Básicamente el piranómetro consiste de una superficie “negra” la cual se calienta cuando es expuesta a la radiación solar. Su temperatura aumenta hasta que la tasa de ganancia de calor por la radiación solar es igual a la tasa de pérdida de calor por convección, la radiación de retorno y la conducción. Las uniones calientes de una termopila están unidos a la superficie de color negro, mientras que las uniones frías se encuentran en una placa de protección de modo que no reciben la radiación directamente. Como resultado, se genera una electromotriz. Esta fuerza electromotriz que está generalmente en el intervalo de 0 a 10 mV puede ser leída, o integrada en un periodo de tiempo y es una medida de la radiación global.

El piranómetro también puede ser utilizado para la medición de la radiación difusa. Esto se hace mediante el montaje en el centro de un anillo de sombreado semicircular. El anillo de sombreado se fija de tal manera que su plano es paralelo al plano de la trayectoria de movimiento diario del sol por el cielo y sombrea el elemento termopila y las dos cúpulas de vidrio del piranómetro en todo el tiempo de la luz del sol directa. En consecuencia, el piranómetro mide sólo la radiación difusa recibida desde el cielo.

Pirheliómetro

El pirheliómetro es un instrumento que mide el haz de radiación que incide sobre una superficie normal a los rayos del sol. En contraste con un piranómetro, la placa de absorción negra (con las uniones calientes de una termopila unido a él) está situado en la base de un tubo de colimación. El tubo está alineado con la dirección de los rayos del sol con la ayuda de un mecanismo de seguimiento en dos ejes y

un indicador de alineación. Así, la placa de negro sólo recibe radiación de haz y una pequeña cantidad de radiación difusa que entra en el “ángulo de aceptación” del instrumento (ver Figura 1.5).

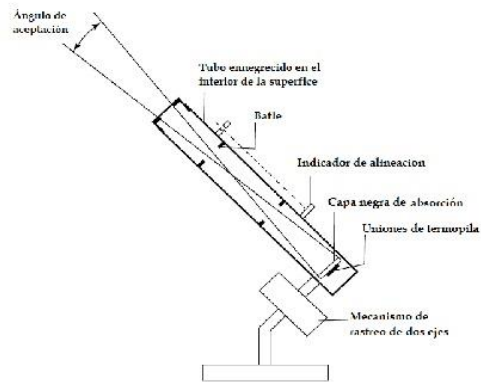


Figura 1.5: Pirheliómetro.

CONCEPTOS BÁSICOS DE SERIES DE TIEMPO

2.1. Proceso estocástico

Antes de dar la definición de una serie de tiempo, se definirá lo que es un proceso estocástico, ya que, una serie de tiempo es un caso particular de un proceso estocástico.

Considérese un sistema que se caracteriza por estar en cualquier conjunto de estados previamente especificados. Ahora suponga que el sistema cambia de un estado a otro a lo largo del tiempo, y sea X_t el estado del sistema al tiempo t . Ahora, considérese que la evolución del sistema es provocada de alguna forma azarosa, entonces se puede pensar que X_t es una variable aleatoria para cada valor del índice t .

A esta colección de variables aleatorias les llamaremos proceso estocástico, el cual sirve para representar la evolución de un sistema de manera secuencial a lo largo del tiempo. Cabe mencionar que las variables aleatorias que conforman un proceso no son independientes entre sí, estas están relacionadas unas con otras de alguna forma particular. [26]

Definición 2.1 (Proceso estocástico). *Un proceso estocástico es una colección o familia de variables aleatorias $\{X_t, \text{ con } t \in T\}$, parametrizada por un conjunto de índices T , llamado espacio parametral, definidas en algún espacio de probabilidad (Ω, \mathcal{F}, P) , y con valores en un conjunto S llamado espacio de estados. De*

tal manera que para cada elemento del conjunto T le corresponde una y sólo una variable aleatoria representada por X_t .

Si T es un intervalo de número reales ya sea abierto o cerrado, se dirá que el proceso estocástico es continuo, por otro lado, si el conjunto es finito o infinito pero numerable el proceso será discreto. Téngase en cuenta que si el proceso es discreto o continuo este no proporcionará información acerca de la variable aleatoria, ya que ésta puede ser discreta o continua.

De acuerdo a esto, podemos clasificar dependiendo de cómo sea el conjunto T y el conjunto de variables aleatorias X . (Ver tabla (2.1))

t \ X	Discreto	Continuo
Discreto	Cadenas de Markov, Procesos de ramificación.	Series de tiempo.
Continuo	Procesos de Poisson, Teoría de colas, Proceso de nacimiento y muerte.	Movimientos Brownianos.

Tabla 2.1: Clasificación de los procesos estocásticos

Nota 2.1. En este trabajo nos enfocaremos en los procesos estocásticos a tiempo discreto con variable aleatoria continua, también conocidas como series de tiempo.

2.2. Serie de Tiempo

Definición 2.2 (Serie de tiempo). Una serie de tiempo es una sucesión cronológica de observaciones x_t , generada por un proceso estocástico, a tiempo discreto y espacio de estados continuo, donde cada observación está asociada a un tiempo en específico $t \in T$.

Definición 2.3. Las funciones $X_t(w)$, $w \in \Omega$ y $t \in T$ son conocidas como las realizaciones de un proceso X_t , $t \in T$ y simplemente se denota por x_t .

Las series de tiempo consisten en observaciones anuales, trimestrales, mensuales o aplicada a cualquier otro periodo. Se examinan los datos de las series de tiempo con el propósito de identificar un patrón que se pueda utilizar para describirlo, luego, este patrón es extrapolado hacia un futuro con el objetivo de preparar un pronóstico.

Existen muchos ejemplos de series de tiempo, algunos de los cuales se mencionan a continuación.

- **Meteorología:** temperaturas máximas, medias o mínimas, precipitaciones mensuales, etc.
- **Economía:** tasa de desempleo, tasa de inflación, alza de comestibles, precio del petróleo, etc.
- **Demografía:** nacimientos anuales, casamientos mensuales, etc.
- **Medio ambiente:** concentración media anual de nitratos en agua, emisiones anuales de CO_2 , abundancia de animales, etc.

Como ya se ha mencionado, las series de tiempo se examinan con la esperanza de hallar algún patrón. Con objeto de identificar dicho patrón, es conveniente muchas veces pensar que la serie de tiempo consta de varios componentes.

1. **Tendencia:** esta componente representa el movimiento hacia arriba o hacia abajo que caracteriza a las series de tiempo. Esta refleja el crecimiento o declinación de larga duración en las series de tiempo. La tendencia puede ser: constante, lineal, cuadrática, exponencial, etc.
 2. **Ciclo:** esta componente representa los movimientos por arriba y por debajo de los niveles de la tendencia. Estas oscilaciones tienen una duración que varía dependiendo el problema a estudiar.
 3. **Variaciones estacionales:** esta componente representa patrones periódicos, es decir, secuencias que debido a la influencia de algún fenómeno se repite de forma frecuente año tras año, ya sea en los mismos meses (o trimestres) del año con aproximadamente la misma intensidad.
 4. **Fluctuaciones irregulares:** esta componente representa los movimientos inesperados en una serie de tiempo que siguen un patrón incierto o irregular.
- [8]

2.3. Proceso estacionario y la autocorrelación

Como se mencionó anteriormente, las series de tiempo son un proceso estocástico cuyo objetivo en este tipo de procesos es determinar el comportamiento de la serie y predecir en el futuro. Para lograr predicciones seguras no es posible usar cualquier tipo de proceso estocástico, es decir, este debe de contar con cierta estabilidad, pero si a cada instante de tiempo se observa un comportamiento diferente e inestable, este proceso no será de gran utilidad para lograr lo que se desea. A este tipo de condiciones se les conoce como estacionariedad. La estacionariedad implica tener un tipo de estabilidad estadística en los datos que permite aprender sobre las propiedades del fenómeno, observándolo por un periodo suficientemente largo de tiempo.

Se dice que un proceso estocástico se puede caracterizar cuando se puede determinar el conjunto de funciones de distribución para cada conjunto finito de variables aleatorias del proceso. Sin embargo, suele ser complejo determinar las características de un proceso por medio de su función de distribución, por ello, la caracterización se suele hacer a través de los dos primeros momentos (media y varianza) de cada X_t .

Definición 2.4 (Media). *El conjunto de medias de un proceso estocástico X_t al tiempo t está definido por*

$$\mu_X(t) = E[X_t] = \int_{-\infty}^{\infty} x f(x) dx \quad \text{si} \quad \int_{-\infty}^{\infty} |x| f(x) dx < \infty. \quad (2.1)$$

Es decir, cuando la integral de valores absolutos es convergente.

En general, $\mu_X(t)$ puede ser diferente en cada instante de tiempo t . Y la varianza está definida como

Definición 2.5 (Varianza). *El conjunto de varianzas de un proceso estocástico X_t al tiempo t está definido por*

$$\sigma_X^2(t) = Var[X_t] = E[(X_t - \mu_x)^2] = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx, \quad (2.2)$$

si

$$\int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx < \infty.$$

2.3.1. Serie de tiempo estrictamente y débilmente estacionaria

Definición 2.6. Sea X_t un proceso estocástico, se dice que este proceso es estrictamente estacionario si para cualesquiera dos colecciones finitas $\{X_t, X_{t+1}, \dots, X_{t+k}\}$ y $\{X_{t+h}, X_{t+h+1}, \dots, X_{t+h+k}\}$, estas tienen la misma función de distribución conjunta para todo entero h . Es decir

$$F[X_t, X_{t+1}, \dots, X_{t+k}] = F[X_{t+h}, X_{t+h+1}, \dots, X_{t+h+k}] \quad (2.3)$$

la función de distribución conjunta de la serie no se ve alterada si esta es desplazada k periodos en el tiempo, es decir, que son invariantes ante una traslación del tiempo.

Como se mencionó, determinar estas funciones suele ser algo muy complicado, ya que se requeriría determinar todas las distribuciones de la serie. Es por ello que se usará una condición más débil, que consiste en restringir los dos primeros momentos.

Definición 2.7. Decimos que un proceso estocástico es débilmente estacionario si

$$E(X_t) = \mu, \quad \forall t \in \mathbb{R}, \mu^2 < \infty \quad (2.4a)$$

$$Var(X_t) = \sigma^2, \quad \forall t \in \mathbb{R}, \sigma^2 < \infty \quad (2.4b)$$

$$Cov(X_t, X_{t+k}) = \gamma_k, \quad \forall t, k \in \mathbb{R}. \quad (2.4c)$$

No pierda de vista que en las expresiones de la media y la varianza, el término t no aparece, pues no dependen del tiempo, esto es, si aún cuando el proceso se separa de la media, este siempre retornara a la vecindad de la misma, por otro lado, si observamos a la covarianza, tampoco existe dependencia del tiempo, pero si de el lapso k que existe entre las variables.

Con lo anterior se puede decir que, dado cualquier instante de tiempo, el comportamiento de la serie será el mismo sin importar el lapso en que esta sea observada, por ejemplo, si se decidiera graficar una cierta cantidad de observaciones contiguas de la serie, la gráfica resultante sería semejante a la gráfica con la misma cantidad de observaciones contiguas pero k lapsos hacia adelante o hacia atrás dependiendo de donde se inicie. [18]

Observación 2.1.

- *Se puede probar que la estacionariedad estricta implica estacionariedad débil, pero, no necesariamente la implicación inversa se cumple, esto ocurriría si, las variables que integran el proceso tienen una función de distribución conjunta dada por una distribución normal multivariada.*

2.3.2. Función de autocovarianza (FACV) y autocorrelación (FAC)

Comúnmente cuando se hace un análisis estadístico se hace el supuesto de la existencia de independencia entre las variables aleatorias. Sin embargo en las series de tiempo existen ciertas correlaciones entre las observaciones, es decir, que los cambios que sufre una de las variables influyen en la otra, es por eso que obtener la función de distribución de cada una de las variables resulta ser complicado.

La **covarianza** entre dos observaciones X_t y X_{t+k} separadas un lapso k es llamada **autocovarianza** y nos indica el grado de asociación lineal que existe entre esas observaciones, es decir, si una de las variable aumenta y la otra también lo hace, entonces la autocovarianza será positiva, pero si cuando una de las variables aumenta la otra disminuye, entonces la autocovarianza será negativa y cuando no hay dependencia entre las variables la autocovarianza es cero.

Dada la definición (2.3), se tiene que la función de distribución bivariada de X_t y X_s , $\forall t, s \in T$, es la misma para un desplazamiento de tiempo k , esto es, X_{t-k} y X_{s-k} , se sigue que

$$\gamma_{t,s} = Cov(X_t, X_s) = Cov(X_{t-k}, X_{s-k}), \quad \forall t, k, s \in T.$$

Cuando la longitud de desplazamiento $k = s$ y luego $k = t$, resulta que

$$\begin{aligned} \gamma_{t,s} &= Cov(X_{t-k}, X_{s-k}) \\ &= Cov(X_{t-s}, X_0) \\ &= Cov(X_0, X_{s-t}) \\ &= Cov(X_0, X_{|t-s|}) \\ &= \gamma_{0,|t-s|}. \end{aligned}$$

Es decir que la covarianza $Cov(X_t, X_s)$, depende solamente del lapso de tiempo $|t - s|$ y no del tiempo en el que se haya calculado, del mismo modo ocurre para

la función de autocorrelación.

Así, para un proceso estacionario se tiene que

$$\gamma_k = Cov(X_t, X_{t-k}) \text{ y } \rho_k = Corr(X_t, X_{t-k})$$

donde solamente dependen del lapso $|k|$. [14]

Definición 2.8 (Coeficiente de Autocovarianza). La autocovarianza entre dos observaciones X_t y X_{t+k} en un lapso k está definido por

$$\gamma_k = Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]. \quad (2.5)$$

El conjunto de valores γ_k , $k = 1, 2, \dots$ es llamado **función de autocovarianza (FACV)**.

Nota 2.2. Si el proceso es estacionario $\mu = E[X_t] = E[X_{t+k}]$, es decir, que para cualquier valor de t y k , la función de valor esperado es la misma para cualquier valor en el tiempo.

Definición 2.9 (Coeficiente de Autocorrelación). El coeficiente de autocorrelación entre dos variables X_t y X_{t+k} en un lapso k mide el grado de relación lineal que existe entre esas observaciones, y está definido por

$$\rho_k = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sqrt{E[(X_t - \mu)^2]E[(X_{t+k} - \mu)^2]}} = \frac{Cov(X_t, X_{t+k})}{Var(X_t)} = \frac{\gamma_k}{\gamma_0}. \quad (2.6)$$

Nota 2.3. Si el proceso es estacionario $\sigma^2 = Var[X_t] = Var[X_{t+k}] = \gamma_0$, es decir, que para cualquier valor de t y k , la varianza también es la misma para cualquier valor en el tiempo.

Este coeficiente toma valores $-1 \leq \rho_k \leq 1$.

- Si $\rho_k > 0$, indica que una de la variables aumenta a medida que la otra también aumenta.
- Si $\rho_k = 1$, existe una autocorrelación lineal perfecta con pendiente positiva.
- Si $\rho_k \leq 0$, indica que al aumentar una de las variables la otra disminuye.
- Si $\rho_k = -1$, existe una autocorrelación lineal perfecta con pendiente negativa.
- Si $\rho_k = 0$, no existe autocorrelación lineal.

	Función de autocovarianza (γ_k)	Función de autocorrelación (ρ_k)
1	$\gamma_0 \geq 0$	$\rho_0 = 1$
2	$ \gamma_k \leq \gamma_0$	$ \rho_k \leq 1$
3	$\gamma_k = \gamma_{-k}$	$\rho_k = \rho_{-k}$
4	$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{ i-j } \geq 0$	$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_{ i-j } \geq 0$

Tabla 2.2: Propiedades de las funciones de autocovarianza y autocorrelación.

El conjunto de valores de ρ_k , $k = 1, 2, \dots$ es llamado **función de autocorrelación (FAC)**.

▪ Propiedades

Obsérvese que en la propiedad **(1)** de la función de autocovarianza, no es otra cosa que la afirmación de que la $Var(X_t) \geq 0$.

Y para la función de autocorrelación, se tiene que $\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$.

En la propiedad **(2)** de la función de autocorrelación, resulta que las correlaciones son menores o iguales a 1 en valor absoluto, de ahí que en la función de autocovarianza se tenga que $|\gamma_k| \leq \gamma_0$, esto es,

$$|\rho_k| \leq 1 \Leftrightarrow \left| \frac{\gamma_k}{\gamma_0} \right| \leq 1 \Leftrightarrow |\gamma_k| \leq |\gamma_0| \Leftrightarrow |\gamma_k| \leq \gamma_0.$$

La propiedad **(3)** en la función de autocovarianza, es establecida observando que

$$\gamma_k = Cov(X_{t+k}, X_t) = Cov(X_t, X_{t+k}) = \gamma_{-k}.$$

Obsérvese también que, la propiedad **(3)** indica que las funciones son simétricas para todo k . Por ello la función de autocorrelación se grafica solamente para lapsos (retardos) positivos. El gráfico de esta función es llamado **correlograma**.

Análogamente ocurre con la función de autocovarianza.

En la propiedad **(4)** ambas funciones son definidas positivas, para cualquier conjunto de puntos t_1, t_2, \dots, t_n y números reales $\alpha_1, \alpha_2, \dots, \alpha_n$.

Así, la función de autocovarianza es definida positiva, ya que,

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{|i-j|} &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j E[(X_i - E[X_i])(X_j - E[X_j])] \\
&= E\left[\left(\sum_{i=1}^n \alpha_i (X_i - E[X_i])\right) \left(\sum_{j=1}^n \alpha_j (X_j - E[X_j])\right)\right] \\
&= E\left[\left(\sum_{i=1}^n \alpha_i (X_i - E[X_i])\right)^2\right] \geq 0.
\end{aligned}$$

Del mismo modo se verifica que la función de autocorrelación es semidefinida positiva.

2.3.3. Función autocorrelación parcial (FACP)

Además de la autocorrelación entre X_t y X_{t+k} , es de interés conocer sobre la correlación que existe entre X_t y X_{t+k} después de eliminar el efecto de las variables aleatorias que intervienen entre ellas, es decir, cuando las variables $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ han sido removidas; dicha correlación está dada por

$$\phi_{k,k} = \text{Corr}(X_t, X_{t+k} | X_{t+1}, X_{t+2}, \dots, X_{t+k-1}). \quad (2.7)$$

La ecuación 2.7 es llamada función de autocorrelación parcial o FACP.

Si se desea hallar el valor de dicha correlación, considere el modelo de regresión donde, la variable dependiente X_{t+k} de un proceso estacionario con media cero es expresada en términos de las k variables retrasadas $X_{t+k-1}, X_{t+k-2}, \dots, X_{t+1}, X_t$, en la forma

$$X_{t+k} = \phi_{k1} X_{t+k-1} + \phi_{k2} X_{t+k-2} + \dots + \phi_{kk} X_t + e_{t+k} \quad (2.8)$$

donde ϕ_{ki} denota el i -ésimo parámetro de regresión y e_{t+k} denota el término de error, con media cero el cual no está correlacionado con X_{t+k-j} para $j = 1, 2, \dots, k$.

Luego, multiplicado en ambos lados de la ecuación 2.8 por X_{t+k-j} y obteniendo la esperanza se tiene,

$$E[X_{t+k-j} X_{t+k}] = \phi_{k1} E[X_{t+k-j} X_{t+k-1}] + \dots + \phi_{kk} E[X_{t+k-j} X_t] + E[X_{t+k-j} e_{t+k}]$$

y dado que X_t es un proceso estacionario, entonces,

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \dots + \phi_{kk}\gamma_{j-k}. \quad (2.9)$$

Dividiendo la ecuación 2.9 por la varianza del proceso (γ_0), se tiene,

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}. \quad (2.10)$$

Haciendo correr el índice de la ecuación 2.10 desde $j=1,2,\dots,k$ obtenemos el siguiente sistema de ecuaciones

$$\begin{aligned} \rho_1 &= \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \dots + \phi_{kk}\rho_{k-1} \\ \rho_2 &= \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \dots + \phi_{kk}\rho_{k-2} \\ &\vdots \\ \rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \dots + \phi_{kk}\rho_0. \end{aligned}$$

Aplicando la regla de Cramer, logramos obtener ϕ_{kk} para $k = 1, 2, \dots$

Para $k = 1$

$$\phi_{11} = \rho_1$$

Para $k = 2$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

Para $k = 3$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

\vdots

Para k

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}} \quad (2.11)$$

Donde ϕ_{kk} es conocida como la Función de Autocorrelación Parcial (FACP). [31]

A continuación se mostrarán algunos ejemplos de procesos estacionarios con su función de autocovarianza y su función de autocorrelación. [31] [10]

2.3.4. Ejemplos de series de tiempo estacionarias

Ejemplo 2.1. Ruido i.i.d. (independiente e idénticamente distribuido)

Se dice que un ruido i.i.d es una sucesión de variables aleatorias ε_t independientes e idénticamente distribuidas, con media cero y varianza σ^2 , denotado de la forma $\varepsilon_t \sim IID(0, \sigma^2)$. Luego, por ser independientes

$$\gamma_{t,t-k} = \begin{cases} \sigma^2, & \text{si } k = 0; \\ 0, & \text{si } k \neq 0. \end{cases}$$

Función de autocorrelación

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ 0, & \text{si } k \neq 0. \end{cases}$$

Y función de autocorrelación parcial

$$\phi_{kk} = \begin{cases} 1, & \text{si } k = 0; \\ 0, & \text{si } k \neq 0. \end{cases}$$

Ejemplo 2.2. Procesos de ruido blanco

Se dice que un proceso $\{\varepsilon_t\}$ es un proceso de **ruido blanco** si es una sucesión de variables aleatorias no correlacionadas a partir de una distribución fija. Este tiene las siguientes características:

- $E[\varepsilon_t] = 0$.
- $Var(\varepsilon_t) = \sigma^2$, varianza constante.
- $\gamma_k = Cov(\varepsilon_t, \varepsilon_{t+k}) = 0, \forall k \neq 0$.

De aquí se sigue que un proceso de ruido blanco es débilmente estacionario. Se denotará al proceso de ruido blanco como $\{\varepsilon_t\} \sim RB(0, \sigma^2)$.

Con función de autocovarianza

$$\gamma_k = \begin{cases} \sigma^2, & \text{si } k = 0; \\ 0, & \text{si } k \neq 0. \end{cases}$$

Función de autocorrelación

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ 0, & \text{si } k \neq 0. \end{cases}$$

Y función de autocorrelación parcial

$$\phi_{kk} = \begin{cases} 1, & \text{si } k = 0; \\ 0, & \text{si } k \neq 0. \end{cases}$$

Cuando se hable de autocorrelación y autocorrelación parcial se hace referencia a ρ_k y ϕ_{kk} para $k \neq 0$, en otro caso $\rho_0 = \phi_{00} = 1$.

En la Figura 2.1 se puede observar la trayectoria para un Ruido blanco y el comportamiento de FAC y FACP.

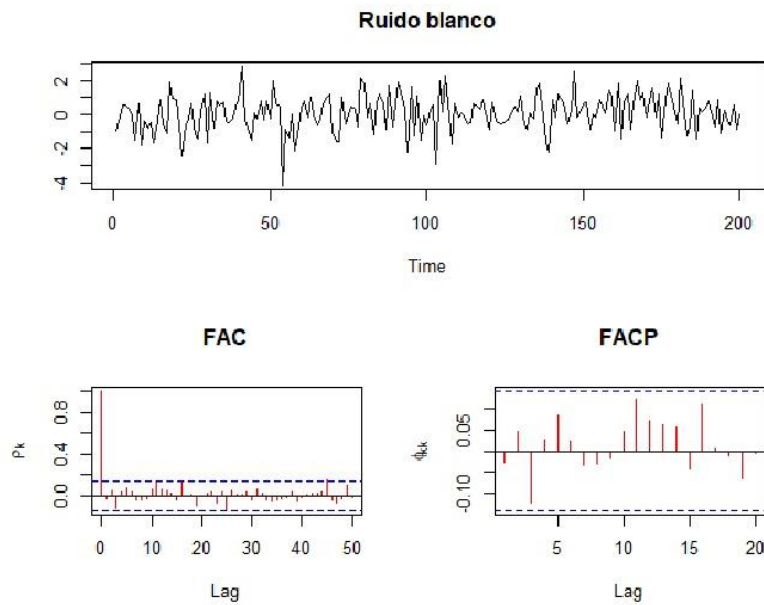


Figura 2.1: Trayectoria, FAC y FACP del Ruido blanco.

Observación 2.2.

- *Es claro que un proceso i.i.d. implica un proceso de ruido blanco, pero un proceso de ruido blanco no necesariamente es un proceso i.i.d. (ya que los ε 's no precisamente son idénticamente distribuidas o son independientes).*
- *Si se cumple que $\varepsilon_t \sim N(0, \sigma^2)$, se dice que ε_t es ruido blanco Gaussiano.*

Ejemplo 2.3. Caminata aleatoria

Sean $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots\}$ una sucesión de variables aleatorias i.i.d con media cero y varianza σ^2 (ruido i.i.d). Una caminata aleatoria $\{X_t, t = 1, 2, \dots\}$ es construida

de la siguiente forma

$$\begin{aligned} X_1 &= \varepsilon_1 \\ X_2 &= \varepsilon_1 + \varepsilon_2 \\ &\vdots \\ X_t &= \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t. \end{aligned}$$

De ahí que la caminata aleatoria X_t , se obtiene de la suma de variables aleatorias i.i.d. Así, una caminata aleatoria con media cero se obtiene cuando

$$X_t = \begin{cases} 0, & \text{si } t = 0; \\ \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t, & \text{si } t = 1, 2, \dots \end{cases}$$

Donde $\{\varepsilon_t\}$ es un ruido i.i.d. con media cero.

Si ε_t es un proceso binario, por ejemplo, $P[\varepsilon_t = 1] = p$, $P[\varepsilon_t = -1] = 1 - p$, entonces $\{X_t, t = 0, 1, 2, \dots\}$, es llamada **caminata aleatoria con simetría simple**.

Ahora bien, dado que las variables $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_t$ son i.i.d. con $\mu = 0$, se tiene

$$E[X_t] = E[\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t] = 0$$

y

$$Var[X_t] = Var[\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t] = t\sigma^2 < \infty, \forall t,$$

y $\forall k \geq 0$ se tiene

$$\begin{aligned} \gamma_X(t, t+k) &= Cov(X_t, X_{t+k}) \\ &= Cov(X_t, X_t + X_{t+1} + \dots + X_{t+k}) \\ &= Cov(X_t, X_t) \\ &= t\sigma^2. \end{aligned}$$

Obsérvese que $\gamma_X(t, t+k)$ depende de t y por definición de estacionariedad, se tiene que la serie X_t es no estacionaria.

Ejemplo 2.4. Proceso de promedios móviles de primer orden o Proceso MA(1)

Consideremos la siguiente serie definida por la ecuación

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}, \quad t = 0, \pm 1, \dots$$

Donde $\{\varepsilon_t\} \sim RB(0, \sigma^2)$ y θ es un valor real constante. De la ecuación anterior se puede ver que

$$E[X_t] = E[\varepsilon_t - \theta_1 \varepsilon_{t-1}] = E[\varepsilon_t] - \theta_1 E[\varepsilon_{t-1}] = 0.$$

$$Var[X_t] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})^2] = E[\varepsilon_t^2 + \theta_1^2 \varepsilon_{t-1}^2 - 2\theta_1 \varepsilon_t \varepsilon_{t-1}] = \sigma^2(1 + \theta_1^2) < \infty.$$

Ahora, para obtener los valores de la función de autocovarianza y función de autocorrelación, consideremos

Para $k = 1$

$$\begin{aligned} \gamma_1 &= Cov(X_t, X_{t-1}) = E[(X_t - E[X_t])(X_{t-1} - E[X_{t-1}])] \\ &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2})] \\ &= E[\varepsilon_t \varepsilon_{t-1} - \theta_1 \varepsilon_t \varepsilon_{t-2} - \theta_1 \varepsilon_{t-1} \varepsilon_{t-1} + \theta_1^2 \varepsilon_{t-1} \varepsilon_{t-2}] \\ &= E[\varepsilon_t \varepsilon_{t-1}] - \theta_1 E[\varepsilon_t \varepsilon_{t-2}] - \theta_1 E[\varepsilon_{t-1}^2] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-2}] \\ &= -\theta_1 \sigma^2. \end{aligned}$$

Para $k = 2$

$$\begin{aligned} \gamma_2 &= Cov(X_t, X_{t-2}) = E[(X_t - E[X_t])(X_{t-2} - E[X_{t-2}])] \\ &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-2} - \theta_1 \varepsilon_{t-3})] \\ &= E[\varepsilon_t \varepsilon_{t-2}] - \theta_1 E[\varepsilon_t \varepsilon_{t-3}] - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-2}] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-3}] \\ &= 0. \end{aligned}$$

Siguiendo el mismo razonamiento, se tiene que para toda $k > 1$

$$\begin{aligned} \gamma_k &= Cov(X_t, X_{t-k}) = E[(X_t - E[X_t])(X_{t-k} - E[X_{t-k}])] \\ &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-k} - \theta_1 \varepsilon_{t-k-1})] \\ &= E[\varepsilon_t \varepsilon_{t-k}] - \theta_1 E[\varepsilon_t \varepsilon_{t-k-1}] - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-k}] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-k-1}] \\ &= 0. \end{aligned}$$

Es decir, que **el proceso tiene autocovarianza cero para los lapsos mayores a dos.**

En resumen, para un **Proceso MA(1)**, la función de autocovarianza (γ_k), esta dada por

$$\gamma_k = \begin{cases} \sigma^2(1 + \theta_1^2), & \text{si } k = 0; \\ -\sigma^2\theta_1, & \text{si } k = \pm 1; \\ 0, & \text{si } |k| > 1. \end{cases}$$

Y la función de autocorrelación (ρ_k), esta dada por

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ -\frac{\theta_1}{1+\theta_1^2}, & \text{si } k = \pm 1; \\ 0, & \text{si } |k| > 1. \end{cases}$$

De esta manera se cumplen las condiciones de estacionariedad, por lo que X_t es estacionario.

Ahora bien, para hallar la función de autocorrelación parcial (FACP), usando (2.11) resulta

$$\begin{aligned} \phi_{11} = \rho_1 &= -\frac{\theta_1}{1 + \theta_1^2} = -\frac{\theta_1(1 - \theta_1^2)}{1 - \theta_1^4} \\ \phi_{22} &= -\frac{\rho_1^2}{1 - \rho_1^2} = -\frac{\theta_1^2}{1 + \theta_1^2 + \theta_1^4} = -\frac{\theta_1^2(1 - \theta_1^2)}{1 - \theta_1^6} \\ \phi_{33} &= \frac{\rho_1^3}{1 - 2\rho_1^2} = -\frac{\theta_1^3}{1 + \theta_1^2 + \theta_1^4 + \theta_1^6} = -\frac{\theta_1^3(1 - \theta_1^2)}{1 - \theta_1^8} \end{aligned}$$

En general

$$\phi_{kk} = -\frac{\theta_1^k(1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}, \quad k \geq 1.$$

Obsérvese que la FACP de un MA(1), contrariamente a su FAC, no se corta después del lapso (retardo) uno, ésta más bien se desvanece exponencialmente dependiendo del signo de θ_1 .

Ejemplo 2.5. Modelo autorregresivo de primer orden o proceso AR(1)

Suponga que la serie $\{X_t\}$ es estacionaria y satisface la siguiente ecuación

$$X_t = \phi_1 X_{t-1} + \varepsilon_t, \text{ para } t = \dots, -1, 0, 1, \dots$$

Donde $\{\varepsilon_t\} \sim RB(0, \sigma^2)$, $|\phi_1| < 1$, y ε_t no está correlacionada con X_s para cada $s < t$.

Luego, tomando la esperanza de cada lado de la ecuación anterior y dado que $E[\varepsilon_t] = 0$, tenemos que $E[X_t] = 0$ (esto se probará más adelante).

Ahora, para hallar la función de autocorrelación de X_t , si se multiplica en cada lado de la ecuación por X_{t-k} , ($k > 0$) y se obtiene su esperanza, entonces,

$$E[X_t X_{t-k}] = E[\phi_1 X_{t-1} X_{t-k} + \varepsilon_t X_{t-k}].$$

Esto es,

$$E[X_t X_{t-k}] = E[(X_t - E(X_t))(X_{t-k} - E(X_{t-k}))] = Cov(X_t, X_{t-k}) = \gamma_k.$$

Así

$$\begin{aligned} \gamma_k &= E[\phi_1 X_{t-1} X_{t-k} + \varepsilon_t X_{t-k}] \\ &= E[\phi_1 X_{t-1} X_{t-k}] + E[\varepsilon_t X_{t-k}] \\ &= \phi_1 E[X_{t-1} X_{t-k}] + 0 \\ &= \phi_1 \gamma_{k-1}. \end{aligned}$$

Por lo que

$$\begin{aligned} \gamma_1 &= \phi_1 \gamma_0 \\ \gamma_2 &= \phi_1 \gamma_1 = \phi_1 (\phi_1 \gamma_0) = \phi_1^2 \gamma_0 \\ \gamma_3 &= \phi_1^3 \gamma_0 \\ &\vdots \\ \gamma_k &= \phi_1^k \gamma_0. \end{aligned}$$

Y como se ha mencionado anteriormente $\gamma_k = \gamma_{-k}$, entonces

$$\gamma_k = \phi_1^{|k|} \gamma_0.$$

Luego, la función de autocorrelación es de la forma

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^{|k|}, \text{ para } k = 0, \pm 1, \pm 2, \dots$$

Como se mencionó anteriormente, la $Cov(X_t, X_t) = \gamma_0$, y que X_t y ε_{t-1} no están correlacionadas, y usando la propiedad de que la $Cov(U + V, W + Z) = Cov(U, W) + Cov(U, Z) + Cov(V, W) + Cov(V, Z)$ se tiene que

$$\gamma_0 = Cov(X_t, X_t) = Cov(\phi X_{t-1} + \varepsilon_t, \phi X_{t-1} + \varepsilon_t) = \phi_1^2 \gamma_0 + \sigma^2.$$

Así despejando γ_0 , resulta

$$\gamma_0 = \frac{\sigma^2}{1 - \phi_1^2}.$$

En resumen, la función de autocovarianza de un proceso AR(1) es

$$\gamma_k = \begin{cases} \frac{\sigma^2}{(1 - \phi_1^2)}, & \text{si } k = 0; \\ \phi_1^{|k|} \gamma_0, & \text{si } k > 0. \end{cases}$$

Y la función de autocorrelación de un proceso AR(1) es

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ \phi_1^{|k|}, & \text{si } k = \pm 1. \end{cases}$$

Por lo que se cumplen las condiciones de estacionariedad, así X_t es estacionario.

Observación 2.3.

- Dado que $|\phi_1| < 1$, la magnitud de la función de autocorrelación disminuye exponencialmente a medida que el número de retardos, k , aumenta.
- Si $0 < \phi_1 < 1$, todas las correlaciones son positivas.
- Si $-1 < \phi_1 < 0$, la autocorrelación en el lapso 1 es negativa y los signos de las autocorrelaciones sucesivas alternan de positivo a negativo, con sus magnitudes van decreciendo exponencialmente.

Luego, para hallar la función de autocorrelación parcial (FACP), usando (2.11), se obtiene que la FACP es de la forma

$$\phi_{kk} = \begin{cases} \rho_1 = \phi_1, & \text{si } k = 1; \\ 0, & \text{si } k > 1. \end{cases}$$

Por consiguiente, la FACP del proceso AR(1), muestra un pico positivo o negativo en el lapso (retardo) 1, dependiendo del signo de ϕ_1 y luego se corta.

2.3.5. Estimación de la media, varianza, y las autocorrelaciones simple y parcial

Una serie de tiempo débilmente estacionaria se caracteriza por su media, varianza, autocorrelación simple y autocorrelación parcial. Para saber los valores exactos de cada uno de los parámetros, se debe conocer el conjunto de todas las realizaciones posibles, en el caso de series de tiempo normalmente sólo se cuenta con una realización, a partir de ésta se desea estimar su media, varianza y autocorrelaciones. De lo contrario, pueden estimarse si se dispone de múltiples realizaciones independientes (lo cuál no sucede en este caso). Sin embargo, es difícil o imposible obtener múltiples realizaciones. Pero dado que los valores de una realización del proceso representan al mismo proceso, es aceptable usar la muestra del proceso para hacer inferencia de los parámetros poblacionales.

Definición 2.10. Sean x_1, x_2, \dots, x_n observaciones de la serie de tiempo, donde $n \in T$ y n es un entero.

La **media muestral** de x_1, x_2, \dots, x_n está definida como

$$\hat{\mu}_x = \bar{x} = \frac{1}{n} \sum_{t=1}^n x_t. \quad (2.12)$$

La **varianza muestral** es

$$\hat{\sigma}_x^2 = \hat{\gamma}_0 = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2. \quad (2.13)$$

La función de autocovarianza muestral es

$$\hat{\gamma}_k = n^{-1} \sum_{t=1}^{n-|k|} (x_{t+|k|} - \bar{x})(x_t - \bar{x}) \quad \text{donde } -n < k < n. \quad (2.14)$$

La función de autocorrelación muestral es

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad \text{donde } -n < k < n. \quad (2.15)$$

La función de autocorrelación parcial muestral (FACPM) es obtenida sustituyendo ρ_k por $\hat{\rho}_k$ en la ecuación (2.11), pero en vez de calcular los determinantes complicados para un k grande, usamos un método recursivo, dado por Durbin(1960) (ver [16]), empieza con $\hat{\phi}_{11} = \hat{\rho}_1$ para llegar a $\hat{\phi}_{kk}$.

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j} \quad (2.16)$$

y

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{k,j} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j}, \quad j = 1, \dots, k.$$

Este método también funciona para el cálculo de la FACP (ϕ_{kk}).

Si se desea más información sobre de los estimadores anteriores ver [31].

Los gráficos de la FAC y de la FACP aportan información acerca de la estacionariedad de la serie de tiempo, a estos gráficos se les conoce como **correlograma**. Los gráficos de estas funciones para los diferentes valores de k , proporcionarán información que, permite proponer un modelo que ajustará la serie de tiempo. Estos pueden presentar un comportamiento de extinción exponencial, extinción exponencial con oscilaciones, extinción lenta, movimientos sinusoidales o truncamiento.

Observación 2.4. *Dependiendo del comportamiento del correlograma correspondiente a la función de autocorrelación:*

- Si en el correlograma sus valores se cortan claramente con rapidez o se desvanecen (decrecen) muy rápidamente, entonces se debe considerar que los valores de la serie temporal son **estacionarios**.
- Si por el contrario, en el correlograma sus valores se desvanecen con lentitud extrema, entonces se debe considerar que los valores de la serie temporal son **no estacionarios**. [8]

2.3.6. Operadores de retardo y diferencia

■ Operador de retardo (B)

Uno de los operadores que se usará frecuentemente es el operador de retardo, expresado por la letra B (del inglés Backward), el cuál actúa de la siguiente forma

$$BX_t = X_{t-1}, \quad \text{para } t = \dots, -1, 0, 1 \dots \quad (2.17)$$

Es decir, que B aplicada a X_t , la desfasa un lapso y la convierte en X_{t-1} . Ahora, si se le aplica el operador dos veces, se obtiene

$$B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}.$$

Y aplicándolo k -veces, se tiene

$$B^k X_t = X_{t-k}.$$

Así, al aplicar B k -veces, se obtiene la variable desfasada k periodos.

Por otro lado $B^0 = 1$, entonces, $B^0 X_t = X_t$. Esto indica que $B^0 = 1$ denota al **operador identidad**, ya que deja intacta a la variable sobre la cual se opera.

Cabe señalar que al hacer uso de operador B , éste modifica a toda la sucesión de valores $\{x_1, x_2, \dots, x_N\}$, transformandola en otra nueva sucesión $\{x_{1-k}, x_{2-k}, \dots, x_{t-k}, \dots, x_{N-k}\}$.

No se pierda de vista que, si se tienen las observaciones $\{x_1, x_2, \dots, x_N\}$, y se les aplica el operador B^k , entonces, las observaciones $\{x_{1-k}, x_{2-k}, \dots, x_0\}$ no se aparecerán en la nueva sucesión, es decir, si la serie tenía N observaciones, esta se reducirá a una serie de simplemente $N - k$ observaciones. [18]

Ejemplo 2.6.

Para entender lo que se dice, daremos un pequeño ejemplo. Sean

$$\begin{aligned} X_1 &= 10, X_2 = 20, X_3 = 30, X_4 = 40, X_5 = 50 \\ X_6 &= 60, X_7 = 70, X_8 = 80, X_9 = 90, X_{10} = 100 \\ X_{11} &= 110, X_{12} = 120 \end{aligned}$$

Construiremos ahora dos series nuevas $\{Z_t\}$ y $\{Y_t\}$, definidas como $Z_t = BX_t = X_{t-1}$ y $Y_t = B^3X_t = X_{t-3}$, entonces las nuevas series quedan de la siguiente forma

$$Z_1 = \emptyset, Z_2 = 10, Z_3 = 20, Z_4 = 30, \dots, Z_{12} = 110$$

y

$$Y_1 = \emptyset, Y_2 = \emptyset, Y_3 = \emptyset, Y_4 = 10, Y_5 = 20, \dots, Y_{12} = 90$$

Nótese que Z_t está retrasado un periodo, tiene los mismos datos que X_t a excepción del último, por otra lado Y_t está retrasado tres periodos, contiene los mismos datos excepto los tres últimos.

■ Operador diferencia (∇)

Otro operador que se usará y que a su vez está relacionado con B es el operador diferencia (∇). El cuál se utiliza para expresar relaciones de la siguiente forma

$$\nabla X_t = X_t - X_{t-1}. \quad (2.18)$$

Y la relación que une a ∇ con B es la siguiente

$$\nabla = 1 - B \quad \text{entonces} \quad \nabla X_t = (1 - B)X_t.$$

Ahora bien, tal como ocurrió con B , tenemos que $\nabla^0 X_t = X_t$. Y así como se consiguió una expresión general para B^k mediante la aplicación sucesiva del operador B , de igual forma es posible obtener una forma general para ∇^k

$$\nabla^k X_t = \sum_{j=0}^k \frac{k!}{j!(k-j)!} (-1)^j X_{t-j} \quad \text{para } k = 0, 1, 2, \dots \text{ y } t = \dots, -1, 0, 1, \dots$$

La expresión anterior no es nada más que el uso del Teorema de Binomio, ya que ∇^k es un binomio elevado a la k -ésima potencia, es decir

$$\nabla^k X_t = (1 - B)^k X_t.$$

En este caso se tendría $\nabla^k = (1 - B)^k$ es una **diferencia de orden k** .

Es necesario determinar que la serie de tiempo que se pretende pronosticar es

estacionaria. Algunas veces se puede transformar a la serie temporal no estacionaria, en una serie temporal estacionaria. Esto es posible gracias a la **técnica de diferenciación**, el cual consiste en obtener las **primeras diferencias** de los valores de la serie temporal no estacionaria, para así obtener una nueva serie a partir de la original $\{x_1, x_2, \dots, x_N\}$, esto es,

$$Y_t = \nabla X_t = X_t - X_{t-1}, \quad t = 2, 3, \dots, N.$$

A veces, cuando se obtienen las primeras diferencias de la serie de tiempo no estacionaria, estas se convierten en valores de una serie temporal estacionaria, pero en ocasiones es necesario usar otras formas de diferenciar para tener valores de series temporales estacionarias, es por eso que se aplicarán **segundas diferencias** (las primeras diferencias de las primeras diferencias), es decir

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

para $t = 3, 4, 5, \dots, N$.

Como ya se mencionó anteriormente, la intención de obtener diferencias de una serie de tiempo es, volver estacionaria a la serie temporal, ahora bien, ¿qué pasa si se toman diferencias de una serie de tiempo estacionaria? pues bien, estas seguirán siendo estacionarias, al hacer esto estaremos cometiendo **sobrediferenciación** de una serie al tratar de estacionalizarla.

Pero ¿sobrediferenciar es malo? Sí, sobrediferenciar a una serie de tiempo trae ciertas complicaciones, como

- Resulta complicado identificar un modelo para representar a la serie.
- Aumenta la varianza de la serie estacionaria.
- Se pierden observaciones, esto se debe al aplicar d veces el operador ∇ .

Por lo que es necesario tener cuidado de NO sobrediferenciar a la serie temporal. [18]

2.4. Modelos de series de tiempo estacionarias

Los modelos de series tiempo se descomponen en dos partes: **la parte sistemática y la parte aleatoria**. La parte sistemática es la parte predecible constituida

por la serie de tiempo, mientras que la parte aleatoria es la parte donde sus valores no tiene relación alguna o dependencia entre si (esta parte también es conocida como ruido blanco). Usualmente cuando se construye un modelo estadístico el problema es formular la parte sistemática.

Deben cumplirse dos condiciones para que un proceso estocástico estacionario admita un modelo, los cuáles son

1. El proceso no debe ser anticipante, esto es, que el futuro no determine el presente. Una manera informal de interpretar esto, es que el valor de X al tiempo t , no puede depender de los valores futuros de X .
2. El proceso debe ser invertible, es decir, que la correlación entre una variable y su pasado debe ir disminuyendo conforme nos alejemos del pasado. [17]

■ **Procesos lineales**

Se dice que una serie de tiempo $\{X_t\}$, es un proceso lineal si se escribe como una combinación lineal infinita de ruidos blancos, es decir

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad \forall t. \quad (2.19)$$

Donde $\{\varepsilon_t\} \sim RB(0, \sigma^2)$ y ψ_j es una sucesión de constantes y $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

La condición de $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ nos garantiza que la suma infinita que define al proceso converge.

Aplicando el **operador de retardo** a la ecuación, es posible escribir al proceso lineal de manera más compacta como:

$$X_t = \psi(B)\varepsilon_t. \quad (2.20)$$

Donde $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$. Esto lleva a pensar que operador $\psi(B)$ es un filtro lineal, que al aplicarlo al proceso de ruido blanco ε_t produce como resultado el proceso X_t .

Decimos que un proceso lineal es un proceso de **promedio móvil** o **MA(∞)** si $\psi_j = 0, \forall j < 0$, es decir, si

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \quad (2.21)$$

Donde $\psi_0 = 1$, $\varepsilon_t \sim RB(0, \sigma^2)$ y $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ (**condición de estacionariedad**).

Otra manera útil de escribir un proceso X_t en un **proceso autorregresivo** o (**AR(∞)**) es

$$X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t. \quad (2.22)$$

Donde $\varepsilon_t \sim RB(0, \sigma^2)$ y $\sum_{j=1}^{\infty} |\pi_j| < \infty$ (**condición de invertibilidad**).

■ Procesos de promedio móvil

Los modelos promedio móvil de orden finito q , también conocido como **MA(q)**, consiste en representar a un proceso estocástico $\{X_t\}$ como una suma finita ponderada de ruidos blancos, que pertenecen a periodos pasados, estas ponderaciones están representadas por θ_j , $j = 1, 2, \dots, q$, es decir, $\psi_1 = -\theta_1$, $\psi_2 = -\theta_2$, \dots , $\psi_q = -\theta_q$ y $\psi_k = 0$ para $k > q$.

Este es representado de la siguiente manera

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

donde $\varepsilon_t \sim RB(0, \sigma^2)$.

Cabe mencionar que el proceso de promedio móvil es estacionario para cualesquiera $\theta_1, \dots, \theta_q$, siempre que $1 + \theta_1^2 + \dots + \theta_q^2 < \infty$.

Se detallará más adelante a este proceso, por el momento se mencionaran casos particulares de los procesos MA(q). [31]

■ Proceso de promedio móvil de primer orden o MA(1)

Como ya se vio anteriormente en el ejemplo 2.4, el proceso de promedio móvil de orden uno, es uno de los casos más simples de un MA(q), en este caso q es igual a uno, y esta dado de la siguiente forma

$$\begin{aligned} X_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &= (1 - \theta_1 B) \varepsilon_t \end{aligned} \quad (2.23)$$

donde $\{\varepsilon_t\} \sim RB(0, \sigma^2)$.

El proceso tiene $E[X_t] = 0$ y varianza $\gamma_0 = Var[X_t] = \sigma^2(1 + \theta_1^2) < \infty$.
Luego, su función de autocovarianza (γ_k) y autocorrelación simple (ρ_k) son

$$\gamma_k = \begin{cases} \sigma^2(1 + \theta_1^2), & \text{si } k = 0; \\ -\sigma^2\theta_1, & \text{si } k = \pm 1; \\ 0, & \text{si } |k| > 1. \end{cases} \quad (2.24)$$

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ -\frac{\theta_1}{1 + \theta_1^2}, & \text{si } k = \pm 1; \\ 0, & \text{si } |k| > 1. \end{cases} \quad (2.25)$$

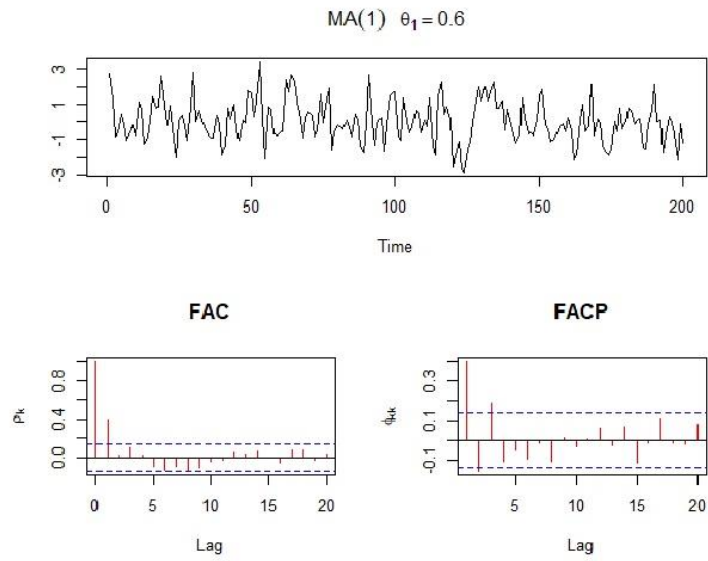
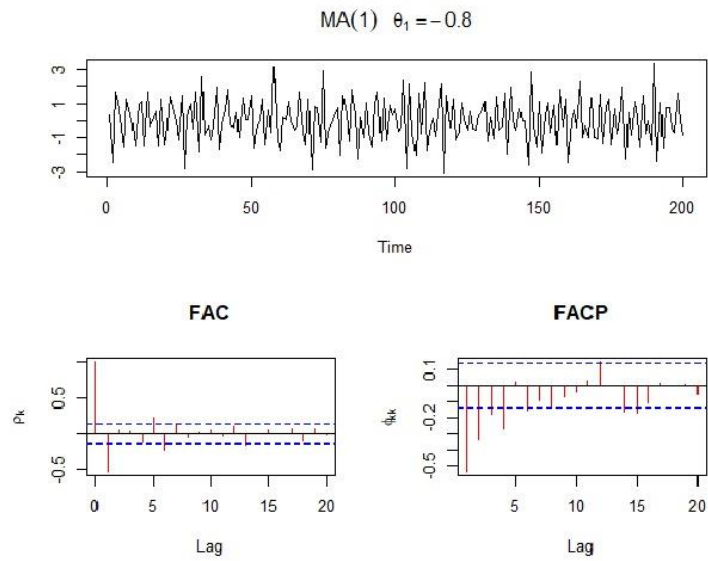
Y como ya se halló en el ejemplo (2.4), la función de autocorrelación parcial (ϕ_{kk}), es

$$\phi_{kk} = -\frac{\theta_1^k(1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}, \quad k \geq 1. \quad (2.26)$$

Obsérvese que la FACP de un MA(1) contrariamente a su FAC, no se corta después del lapso (retardo) 1, esta más bien se desvanece exponencialmente dependiendo del signo de θ_1 .

Ahora bien, para confirmar que el modelo MA(1) cumple con la condición de invertibilidad, considere el polinomio $\theta_1(B) = (1 - \theta_1 B)$, conocido como **polinomio de retardo**, el cual se resolverá con el fin de encontrar las raíces del polinomio, esto es, $1 - \theta_1 B = 0$, así $B = \frac{1}{\theta_1}$, para que el proceso sea invertible la raíz debe estar fuera del círculo unitario, es decir, $\frac{1}{\theta_1} > 1$, así $|\theta_1| < 1$.

En las Figuras 2.2 y 2.3 se puede observar las diferentes trayectorias para el proceso MA(1) y el comportamiento de FAC y FACP.

Figura 2.2: Trayectoria, FAC y FACP del proceso MA(1) con $\theta_1 = 0.6$.Figura 2.3: Trayectoria, FAC y FACP del proceso MA(1) con $\theta_1 = -0.8$.

■ **Proceso de promedios móviles de segundo orden o MA(2)**

Trabajaremos ahora con el proceso de promedios móviles de orden dos, denotado por **MA(2)** y cuyo modelo es

$$\begin{aligned} X_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \\ &= (1 - \theta_1 B - \theta_2 B^2) \varepsilon_t. \end{aligned} \quad (2.27)$$

Donde $\varepsilon \sim RB(0, \sigma^2)$, y el proceso MA(2) es siempre debilmente estacionario. La esperanza de un proceso MA(2) es $E[X_t] = E[\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}] = 0$ y la varianza es

$$\gamma_0 = Var[X_t] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})^2] = \sigma^2(1 + \theta_1^2 + \theta_2^2).$$

Para obtener los valores de la función de autocovarianza y función de autocorrelacion, considere

Para $k = 1$

$$\begin{aligned} \gamma_1 &= Cov(X_t, X_{t-1}) = E[(X_t - E[X_t])(X_{t-1} - E[X_{t-1}])] \\ &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3})] \\ &= E[\varepsilon_t \varepsilon_{t-1} - \theta_1 \varepsilon_t \varepsilon_{t-2} - \theta_2 \varepsilon_t \varepsilon_{t-3} - \theta_1 \varepsilon_{t-1}^2 + \theta_1^2 \varepsilon_{t-1} \varepsilon_{t-2} \\ &\quad + \theta_1 \theta_2 \varepsilon_{t-1} \varepsilon_{t-3} - \theta_2 \varepsilon_{t-2} \varepsilon_{t-1} + \theta_1 \theta_2 \varepsilon_{t-2}^2 + \theta_2^2 \varepsilon_{t-2} \varepsilon_{t-3} \\ &= E[\varepsilon_t \varepsilon_{t-1}] - \theta_1 E[\varepsilon_t \varepsilon_{t-2}] - \theta_2 E[\varepsilon_t \varepsilon_{t-3}] - \theta_1 E[\varepsilon_{t-1}^2] \\ &\quad + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-2}] + \theta_1 \theta_2 E[\varepsilon_{t-1} \varepsilon_{t-3}] - \theta_2 E[\varepsilon_{t-2} \varepsilon_{t-1}] \\ &\quad + \theta_1 \theta_2 E[\varepsilon_{t-2}^2] + \theta_2^2 E[\varepsilon_{t-2} \varepsilon_{t-3}] \\ &= (-\theta_1 + \theta_1 \theta_2) \sigma^2. \end{aligned}$$

Para $k = 2$

$$\begin{aligned} \gamma_2 &= Cov(X_t, X_{t-2}) = E[(X_t - E[X_t])(X_{t-2} - E[X_{t-2}])] \\ &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-2} - \theta_1 \varepsilon_{t-3} - \theta_2 \varepsilon_{t-4})] \\ &= E[\varepsilon_t \varepsilon_{t-2}] - \theta_1 E[\varepsilon_t \varepsilon_{t-3}] - \theta_2 E[\varepsilon_t \varepsilon_{t-4}] - \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-2}] \\ &\quad + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-3}] + \theta_1 \theta_2 E[\varepsilon_{t-1} \varepsilon_{t-4}] - \theta_2 E[\varepsilon_{t-2}^2] \\ &\quad + \theta_1 \theta_2 E[\varepsilon_{t-2} \varepsilon_{t-3}] + \theta_2^2 E[\varepsilon_{t-2} \varepsilon_{t-4}] \\ &= -\theta_2 \sigma^2. \end{aligned}$$

Siguiendo el mismo razonamiento, si $k > 2$

$$\begin{aligned}\gamma_k &= Cov(X_t, X_{t-k}) = E[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})(\varepsilon_{t-k} - \theta_1\varepsilon_{t-k-1} - \theta_2\varepsilon_{t-k-2})] \\ &= 0\end{aligned}$$

siempre que $k > 2$, es decir, que **el proceso no tiene correlación más allá del lapso 2**.

Así, la función de autocovarianza (γ_k) de un proceso **MA(2)**, está dada por

$$\gamma_k = \begin{cases} \sigma^2(1 + \theta_1^2 + \theta_2^2), & \text{si } k = 0; \\ (-\theta_1 + \theta_1\theta_2)\sigma^2, & \text{si } k = \pm 1; \\ -\theta_2\sigma^2, & \text{si } k = \pm 2; \\ 0, & \text{si } |k| > 2. \end{cases} \quad (2.28)$$

Y la función de autocorrelación (ρ_k), esta dada por

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, & \text{si } k = \pm 1; \\ -\frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}, & \text{si } k = \pm 2; \\ 0, & \text{si } |k| > 2. \end{cases} \quad (2.29)$$

Ahora bien, para la FACP del proceso MA(2), dado que ρ_k se tiene

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (2.30)$$

$$\phi_{33} = \frac{\rho_1^3 - \rho_1\rho_2(2 - \rho_2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)}$$

⋮

Note que el proceso MA(2) contiene al proceso MA(1) como un caso particular, así, la FACP se desvanece como un decaimiento exponencial o en forma sinusoidal amortiguada, dependiendo de los signos y magnitudes de θ_1 y θ_2 .

Ahora verifiquemos que el proceso MA(2) cumple la condición de invertibilidad. Para ello considere el polinomio de retardo del modelo MA(2), el cual se resolvera para obtener sus raíces, es decir, $\theta_2(B) = 1 - \theta_1 B - \theta_2 B^2 = 0$. Y sus raíces son

$$B_{1,2} = \frac{-\theta_1 \pm \sqrt{\theta_1^2 + 4\theta_2}}{2\theta_2} \quad (2.31)$$

Estableciendo la condición de invertibilidad se tiene

$$|B_1| = \left| \frac{-\theta_1 + \sqrt{\theta_1^2 + 4\theta_2}}{2\theta_2} \right| > 1, \quad |B_2| = \left| \frac{-\theta_1 - \sqrt{\theta_1^2 + 4\theta_2}}{2\theta_2} \right| > 1$$

Para que esto se satisfaga deben cumplirse las siguientes condiciones

1. $\theta_2 + \theta_1 < 1$.
2. $\theta_2 - \theta_1 < 1$.
3. $-1 < \theta_2 < 1$.

La prueba de las condiciones anteriores se pueden consultar en el Apéndice B. En la Figura 2.4 y 2.5 se puede observar las diferentes trayectorias del proceso MA(2) y el comportamiento de FAC y FACP, dependiendo de los signos de θ_1 y θ_2 .

■ Proceso de promedio móvil de orden q o MA(q)

Inicialmente se mencionó que los procesos de promedio móvil de orden q son de la forma

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (2.32)$$

donde $\varepsilon_t \sim RB(0, \sigma^2)$.

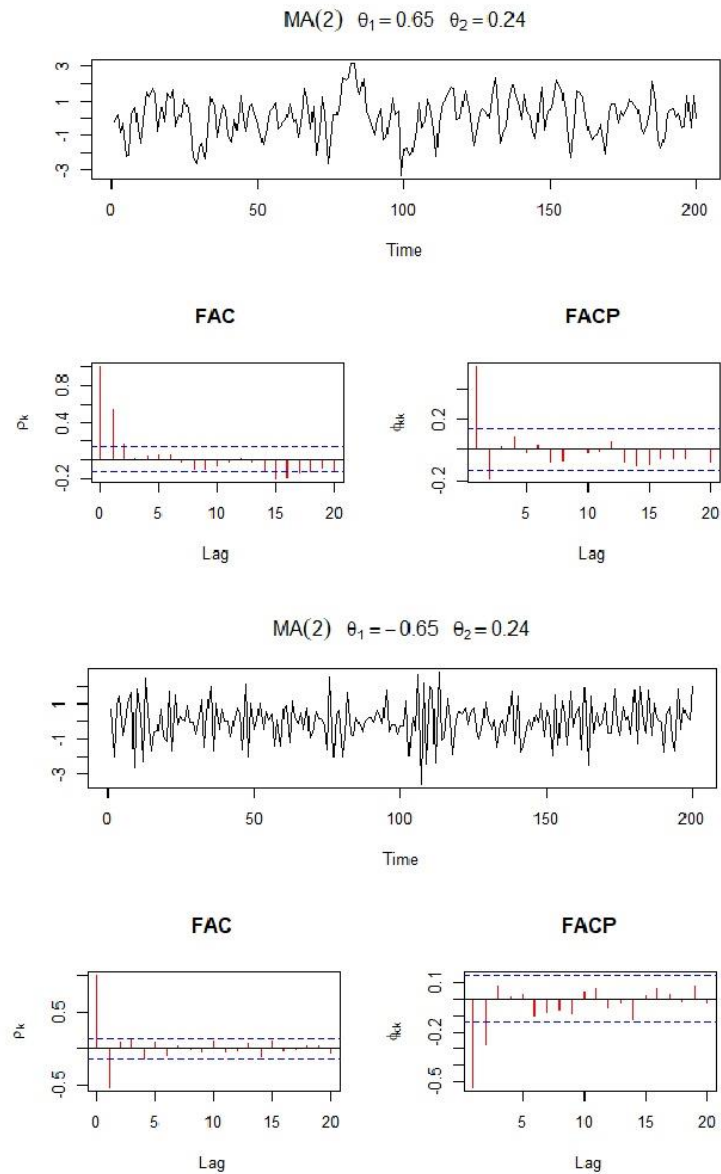
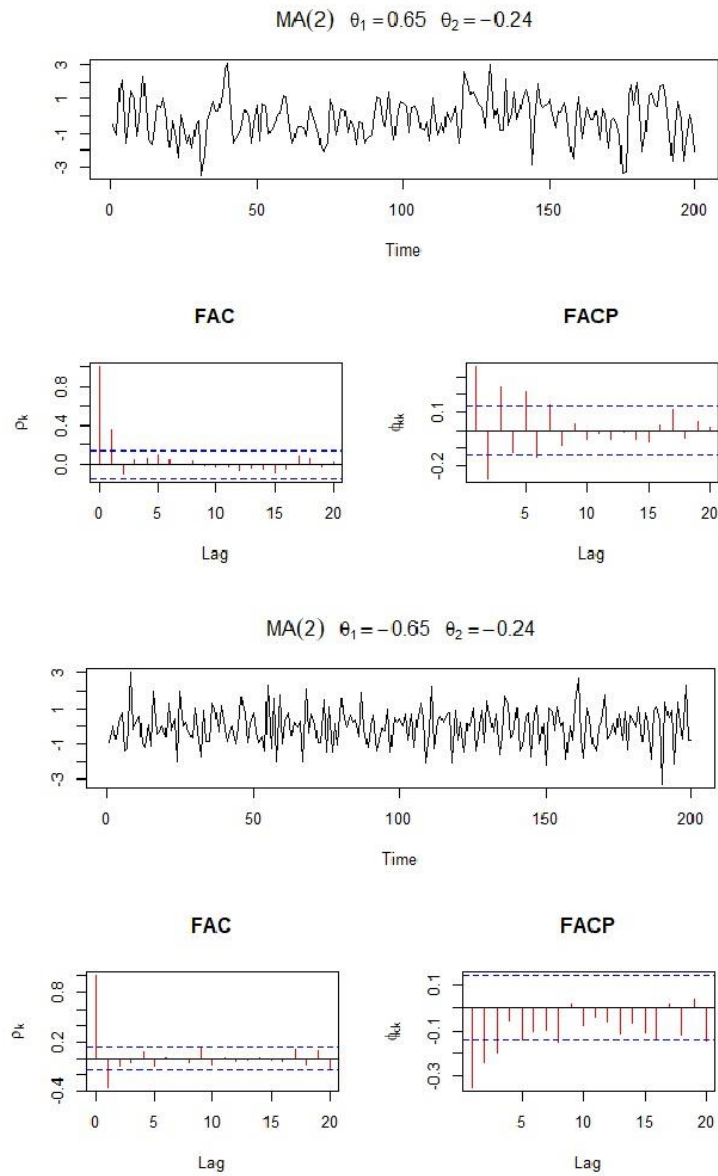


Figura 2.4: Trayectoria, FAC y FACP del proceso MA(2).

Figura 2.5: Trayectoria, FAC y FACP del proceso $MA(2)$.

Y escribiendo a la ecuación anterior en términos del operador de retardo, se tiene

$$\begin{aligned} X_t &= (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \\ &= (1 - \sum_{i=1}^q \theta_i B^i) \varepsilon_t \\ &= \theta(B) \varepsilon_t \end{aligned}$$

donde $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, que es el **polinomio de retardo**.

Cabe mencionar que el proceso de promedio móvil es estacionario para cualesquiera $\theta_1, \dots, \theta_q$, siempre que $1 + \theta_1^2 + \dots + \theta_q^2 < \infty$.

Es necesario verificar que el proceso es invertible, esto ocurre si las raíces del polinomio de retardo $\theta_q(B) = 0$ se encuentran fuera del círculo unitario.

Ahora bien, la media del proceso es

$$E[X_t] = E[\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}] = 0$$

y la varianza

$$\gamma_0 = \text{Var}[X_t] = E[(X_t - E[X_t])^2] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q})^2].$$

Desarrollando los cuadrados, se obtiene finalmente

$$\gamma_0 = \sigma^2 \sum_{i=0}^q \theta_i^2 \text{ donde } \theta_0 = 1.$$

La función de autocovarianza de un modelo MA(q) se define como

$$\gamma_k = \begin{cases} \sigma^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2), & \text{si } k = 0; \\ \sigma^2(-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q), & \text{si } k = 1, 2, \dots, q; \\ 0, & \text{si } |k| > q. \end{cases} \quad (2.33)$$

Y la función de autocorrelación de un modelo MA(q) se define como

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, & \text{si } k = 1, 2, \dots, q; \\ 0, & \text{si } |k| > q. \end{cases} \quad (2.34)$$

La función de autocorrelación de un proceso MA(q) se corta después del lapso q , esto nos permite identificar si se trata de un proceso de promedio móvil.

La FACP del proceso, resulta ser una mezcla de decaimientos exponenciales y ondas sinusoidales amortiguadas dependiendo de las raíces del polinomio de retardo $\theta_q(B) = 0$.

Los procesos de promedios móviles son útiles para representar fenómenos en donde los eventos originan efectos inmediatos que sólo duran periodos cortos de tiempo.

■ Procesos autorregresivos

Consideremos ahora los procesos autorregresivos finitos de orden p , también conocidos como **AR(p)**, éste consiste en representar al valor actual del proceso $\{X_t\}$ como una combinación lineal de los p últimos valores pasados de sí mismo más un término de error (ruido blanco), este último incorpora todo lo nuevo en la serie en el tiempo t que no se explica por los valores pasados. Y es representado de la siguiente forma

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t$$

en este caso $\pi_1 = \phi_1, \pi_2 = \phi_2, \pi_3 = \phi_3, \dots, \pi_p = \phi_p$ y $\pi_k = 0, k > p$.

Cabe mencionar que los procesos autorregresivos de cualquier orden son invertibles siempre que $\sum_{i=1}^p |\phi_i| < \infty$.

Se detallará más adelante a este proceso, por ahora se mencionaran casos particulares de los procesos AR(p). [31]

■ Procesos autorregresivos de primer orden o AR(1)

Puesto que ya se hab  referido con anterioridad a este proceso, se recordará que el proceso AR(1) en el ejemplo 2.5 est  dado de la siguiente manera

$$X_t = \phi_1 X_{t-1} + \varepsilon_t \quad (2.35)$$

donde $\varepsilon_t \sim RB(0, \sigma^2)$, $|\phi_1| < 1$, y ε_t es independiente de X_s para cada $s < t$. A este proceso tambi n se le conoce como proceso de Markov. Expresando la ecuaci n anterior en t rminos del operador de retardo se tiene

$$(1 - \phi_1 B)X_t = \varepsilon_t.$$

Como se indico inicialmente, el proceso autorregresivo de cualquier orden siempre es invertible, por lo que solo faltaría confirmar la estacionariedad del proceso, para ello, debe satisfacerse que la raíz del polinomio de retardo $\phi_1(B) = 1 - \phi_1 B = 0$ este fuera del circulo unitario, es decir que $\frac{1}{\phi_1} > 1$, esto implica que $|\phi_1| < 1$.

Ahora bien, note que si se itera la ecuación del proceso AR(1) k - veces hacia el pasado se obtiene

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \varepsilon_t \\ &= \phi_1(\phi_1 X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 X_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \phi_1^2(\phi_1 X_{t-3} + \varepsilon_{t-2}) + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ &= \phi_1^k X_{t-k} + \sum_{i=0}^{k-1} \phi_1^i \varepsilon_{t-i}. \end{aligned}$$

Continuando con el proceso de iteración, siempre que $|\phi_1| < 1$, se puede observar que es posible representar un modelo AR(1) como un proceso de promedios móviles de orden infinito (MA(∞)), es decir

$$X_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}.$$

De la ecuación anterior, notemos que

$$E[X_t] = E[\varepsilon_t] + \phi_1 E[\varepsilon_{t-1}] + \phi_1^2 E[\varepsilon_{t-2}] + \phi_1^3 E[\varepsilon_{t-3}] + \dots = 0$$

ya que ε_t es un ruido blanco.

Y la varianza calculada en el ejemplo (2.5) es

$$\gamma_0 = Var[X_t] = \frac{\sigma^2}{1 - \phi_1^2}$$

Su función de autocovarianza (γ_k) y autocorrelación simple (ρ_k) son

$$\gamma_k = \begin{cases} \frac{\sigma^2}{(1 - \phi_1^2)}, & \text{si } k = 0; \\ \phi_1^k \gamma_0, & \text{si } k > 0. \end{cases} \quad (2.36)$$

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ \phi_1^k, & \text{si } k > 0. \end{cases} \quad (2.37)$$

Y por último la función de autocorrelación parcial (ϕ_k), es

$$\phi_{kk} = \begin{cases} \phi_1, & \text{si } k = 1; \\ 0, & \text{si } k > 1. \end{cases} \quad (2.38)$$

La FACP del proceso AR(1), muestra un pico positivo o negativo en el lapso (retardo) 1 dependiendo del signo de ϕ_1 y luego se corta.

En las Figuras 2.6 y 2.7 se puede observar las diferentes trayectorias para el proceso AR(1) y el comportamiento de FAC y FACP.

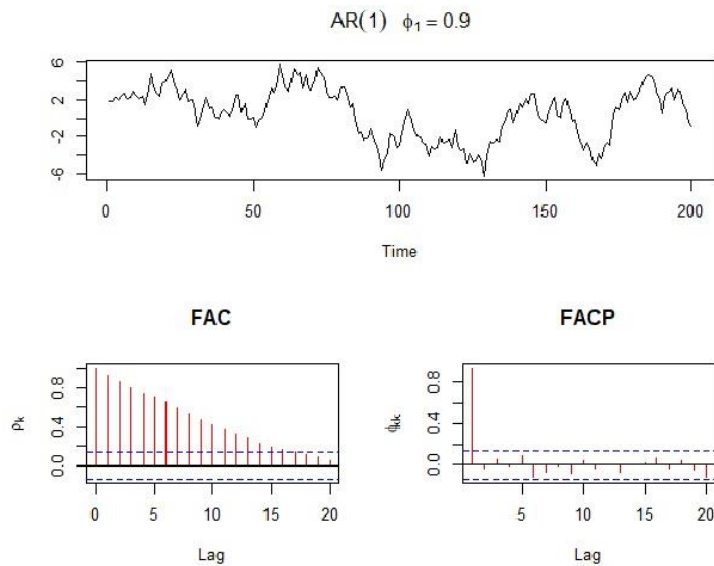


Figura 2.6: Trayectoria, FAC y FACP del proceso AR(1) con $\phi_1 = 0.9$.

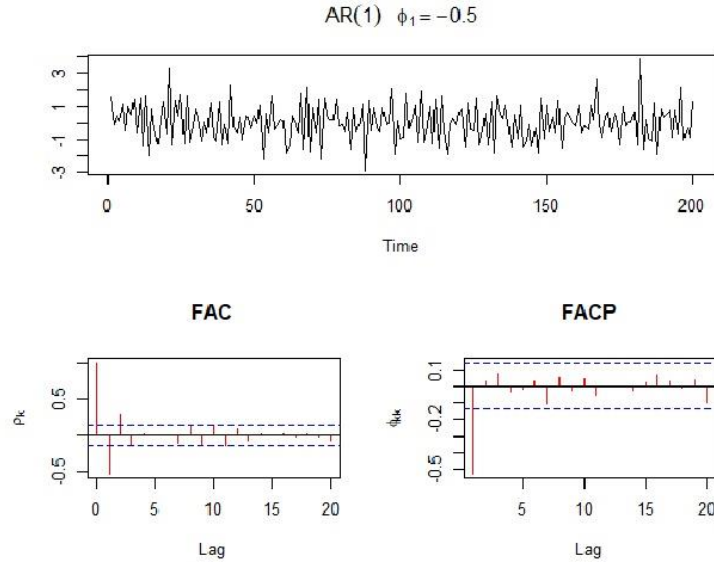


Figura 2.7: Trayectoria, FAC y FACP del proceso AR(1) con $\phi_1 = -0.5$.

■ Procesos autorregresivos de segundo orden o AR(2)

El proceso autorregresivo de orden 2, denotado por AR(2), está representado como

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t. \quad (2.39)$$

Y expresándolo en términos del operador de retardo se tiene

$$(1 - \phi_1 B - \phi_2 B^2)X_t = \varepsilon_t.$$

Y como el proceso autorregresivo de cualquier orden es invertible, sólo restaría comprobar la estacionariedad, para ello se hallaran las raíces del polinomio de retardo $\phi_2(B) = 1 - \phi_1 B - \phi_2 B^2 = 0$, las cuales son

$$B_{1,2} = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \quad (2.40)$$

Estas deben estar fuera del círculo unitario, esto es,

$$|B_1| = \left| \frac{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \right| > 1, \quad |B_2| = \left| \frac{-\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \right| > 1$$

Para que esto sea cierto, deben cumplirse estas condiciones

1. $\phi_2 + \phi_1 < 1$.
2. $\phi_2 - \phi_1 < 1$.
3. $-1 < \phi_2 < 1$.

La prueba de las condiciones anteriores se pueden consultar en el Apéndice B. Y al igual que los proceso AR(1), es posible representar a un proceso de AR(2) como un proceso MA(∞), que es posible consultar en el Apéndice A, por lo que la esperanza de un AR(2) es $E[X_t] = 0$. La varianza del proceso es

$$\begin{aligned}\gamma_0 = Var[X_t] &= E[(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t)^2] \\ &= \phi_1^2 E[(X_{t-1})^2] + \phi_2^2 E[(X_{t-2})^2] + E[\varepsilon_t^2] + 2\phi_1\phi_2 E[X_{t-1}X_{t-2}] \\ &\quad + 2\phi_1 E[X_{t-1}\varepsilon_t] + 2\phi_2 E[X_{t-2}\varepsilon_t] \\ &= \phi_1^2\gamma_0 + \phi_2^2\gamma_0 + \sigma^2 + 2\phi_1\phi_2\gamma_1 \\ (1 - \phi_1^2 - \phi_2^2)\gamma_0 &= \sigma^2 + 2\phi_1\phi_2\gamma_1.\end{aligned}$$

Para los valores de la función de autocovarianza y autocorrelación, considere para $k = 1$

$$\begin{aligned}\gamma_1 = Cov(X_t, X_{t-1}) &= E[(X_t - E[X_t])(X_{t-1} - E[X_{t-1}])] \\ &= E[(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t)X_{t-1}] \\ &= \phi_1 E[X_{t-1}^2] + \phi_2 E[X_{t-2}X_{t-1}] + E[\varepsilon_t X_{t-1}] \\ &= \phi_1\gamma_0 + \phi_2\gamma_1 \\ \gamma_1 &= \frac{\phi_1\gamma_0}{1 - \phi_2}.\end{aligned}$$

Ahora bien, para que γ_0 quede expresada en términos solamente de ϕ_1 , ϕ_2 y σ^2 , se sustituye γ_1 en γ_0 , esto es,

$$\begin{aligned}(1 - \phi_1^2 - \phi_2^2)\gamma_0 &= \sigma^2 + 2\phi_1\phi_2 \left(\frac{\phi_1\gamma_0}{1 - \phi_2} \right) \\ (1 - \phi_2)(1 - \phi_1^2 - \phi_2^2)\gamma_0 &= \sigma^2(1 - \phi_2) + 2\phi_1^2\phi_2\gamma_0 \\ ((1 - \phi_2)(1 - \phi_1^2 - \phi_2^2) - 2\phi_1^2\phi_2)\gamma_0 &= \sigma^2(1 - \phi_2) \\ \gamma_0 &= \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)((1 - \phi_2)^2 - \phi_1^2)}.\end{aligned}$$

Por otro lado, para $k > 1$ y dado que $E[X_t] = 0$, se obtiene

$$\begin{aligned}\gamma_k &= Cov(X_t, X_{t-k}) = E[(X_t - E[X_t])(X_{t-k} - E[X_{t-k}])] \\ &= E[(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t)X_{t-k}] \\ &= \phi_1 E[X_{t-1}X_{t-k}] + \phi_2 E[X_{t-2}X_{t-k}] + E[\varepsilon_t X_{t-k}] \\ &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \quad \forall k > 1.\end{aligned}$$

De esta manera, la función de autocovarianzas (γ_k) de un proceso AR(2) es

$$\gamma_k = \begin{cases} \gamma_0, & \text{si } k = 0; \\ \gamma_1, & \text{si } k = 1; \\ \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, & \text{si } k > 1. \end{cases} \quad (2.41)$$

Y para la función de autocorrelación (ρ_k) se tiene

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}}{\gamma_0} = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 1$$

por lo que la FAC simple es

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} & \text{si } k \geq 1. \end{cases} \quad (2.42)$$

Luego, para hallar la función de autocorrelación parcial (ϕ_{kk}), usando (2.11), tenemos

$$\phi_{11} = \rho_1 = \frac{\phi_1}{1 - \phi_2}$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \phi_2$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho_1 & \phi_1 + \phi_2 \rho_1 \\ \rho_1 & 1 & \phi_1 \rho_1 + \phi_2 \\ \rho_2 & \rho_1 & \phi_1 \rho_2 + \phi_2 \rho_1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = 0$$

ya que la última columna del numerador es una combinación lineal de las dos primeras columnas. Por lo que la función de autocorrelación parcial es

$$\phi_{kk} = \begin{cases} \frac{\phi_1}{1-\phi_2}, & \text{si } k = 1; \\ \phi_2 & \text{si } k = 2; \\ 0, & \text{si } k > 2. \end{cases} \quad (2.43)$$

Por consiguiente, se observa que la FACP del proceso AR(2) se corta después del lapso (retardo) 2.

En la Figura 2.8 y 2.9 se puede observar las diferentes trayectorias para el proceso AR(2) y el comportamiento de FAC y FACP, dependiendo de los signos de ϕ_1 y ϕ_2 .

■ Proceso autorregresivo orden p o AR(p)

Inicialmente se señaló que los procesos autorregresivos son de la forma

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \quad (2.44)$$

donde $\varepsilon \sim RB(0, \sigma^2)$ y ε_t no está correlacionado con X_s para cada $s < t$.

Expresando la ecuación 2.44 en términos del operador de retardo, se obtiene

$$\begin{aligned} \varepsilon_t &= (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t \\ &= \left(1 - \sum_{i=1}^p \phi_i B^i\right) X_t \\ &= \phi_p(B) X_t \end{aligned}$$

donde $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ es el polinomio de retardo de AR(p).

Ahora bien, los procesos autorregresivos de cualquier orden son invertibles siempre que $\sum_{i=1}^p |\phi_i| < \infty$.

Para confirmar la estacionariedad del proceso, se debe satisfacerse que las raíces del polinomio de retardo $\phi_p(B) = 0$ se encuentren fuera del círculo unitario.

En el Apéndice A, se verifica que un proceso AR(p) puede ser presentado por un MA(∞).

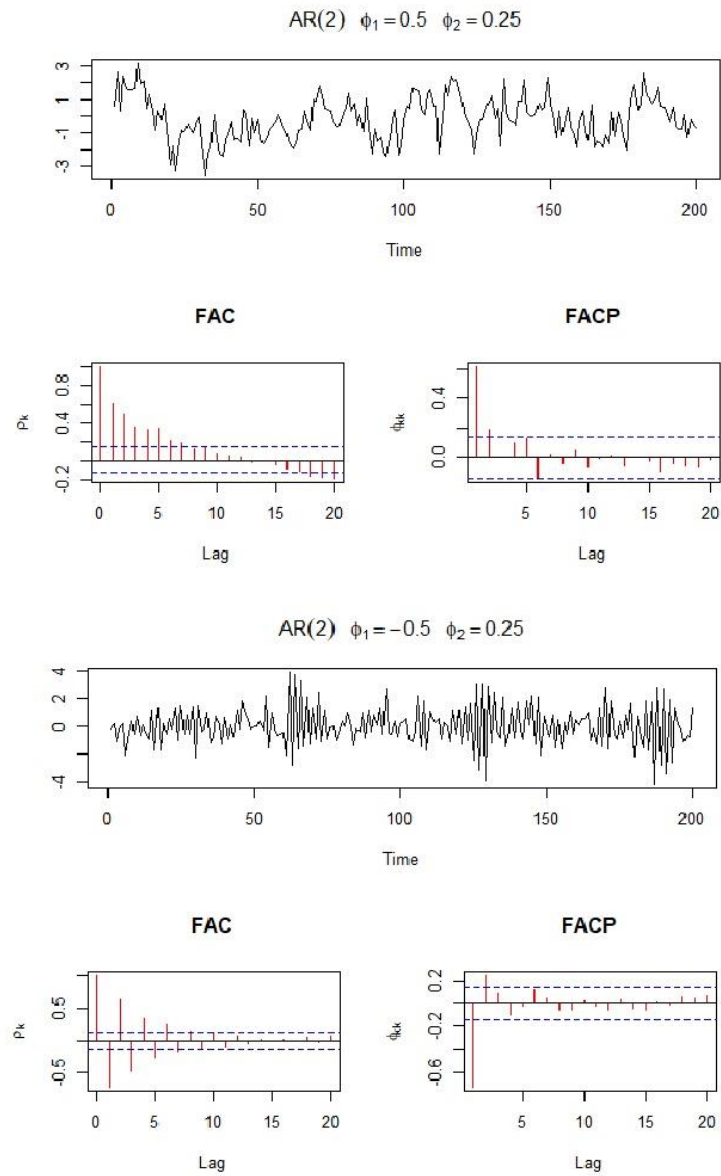


Figura 2.8: Trayectoria, FAC y FACP del proceso AR(2).

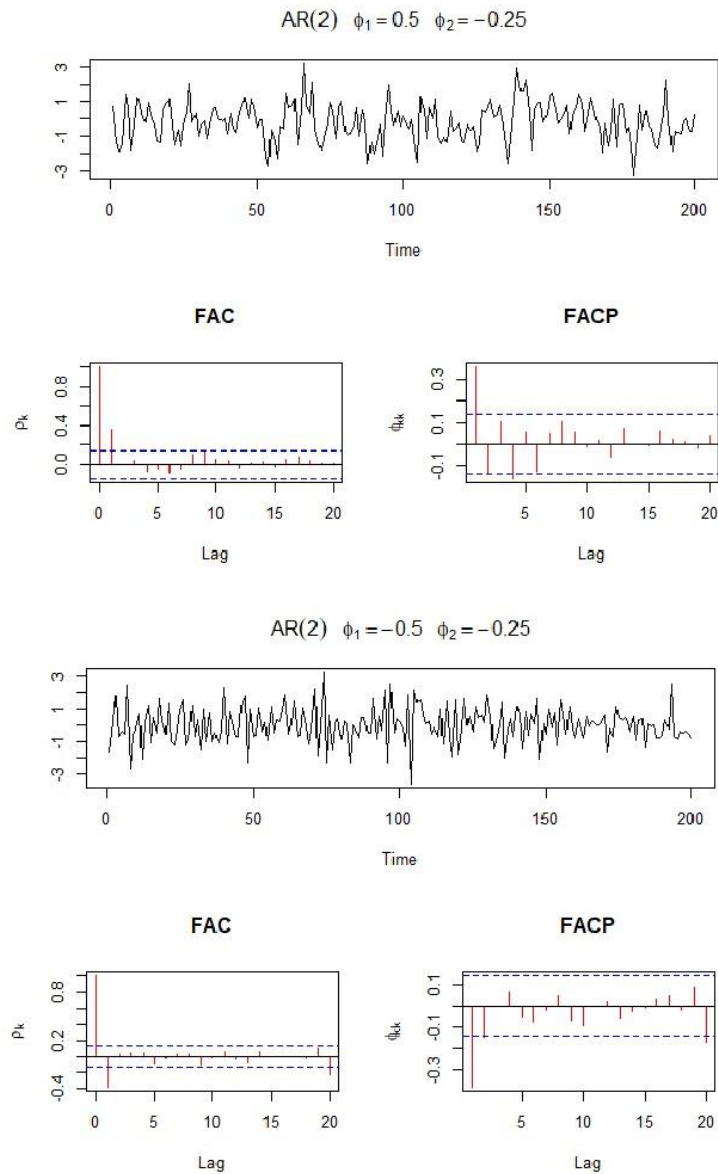


Figura 2.9: Trayectoria, FAC y FACP del proceso AR(2).

Para hallar las funciones de autocovarianza (γ_k) y autocorrelación (ρ_k) del proceso AR(p), se multiplica por X_{t-k} en ambos lados de la ecuación (2.44) y se calcula la esperanza, esto es,

$$\begin{aligned} X_{t-k}X_t &= \phi_1 X_{t-k}X_{t-1} + \phi_2 X_{t-k}X_{t-2} + \cdots + \phi_p X_{t-k}X_{t-p} + X_{t-k}\varepsilon_t \\ E[X_{t-k}X_t] &= \phi_1 E[X_{t-k}X_{t-1}] + \cdots + \phi_p E[X_{t-k}X_{t-p}] + E[X_{t-k}\varepsilon_t] \\ \gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}. \end{aligned} \quad (2.45)$$

Y la función de autocorrelación (ρ_k) del proceso es

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \quad k \geq 1. \quad (2.46)$$

Para la FACP (ϕ_{kk}), dado que $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}$, $k \geq 1$, notar que cuando $k > p$ la última columna del determinante en el numerador de (2.11) puede ser escrita como una combinación lineal de las columnas previas del determinante, de esta manera el determinante es igual a cero y la FACP se corta después del lapso p .

El proceso AR(p) es útil para describir situaciones en la que el valor actual de una serie temporal depende de sus valores anteriores más un choque aleatorio.

■ Proceso Autorregresivo de Promedio Móvil (ARMA(p,q))

Una extensión del proceso autorregresivo y el proceso de promedios móviles, es el proceso mixto autorregresivo en promedios móviles, que incluye el proceso autorregresivo y el proceso de promedios móviles como casos especiales.

Como se demuestra en el Apéndice A, un proceso estacionario e invertible puede representarse como un proceso de promedio móvil o autorregresivo.

Sin embargo, un problema es que puede contener demasiados parámetros en cualquiera de las dos representaciones, incluso para un proceso de promedio móvil de orden finito o proceso autorregresivo de orden finito, pues a menudo se necesita un modelo de mayor orden para una buena aproximación.

En general, una gran cantidad de parámetros reduce la eficiencia en la estimación. Por lo tanto, en la construcción de modelos, puede ser necesario incluir términos

de procesos autorregresivos y de procesos de promedio móvil en un modelo, es decir, la combinación de los procesos AR(p) y MA(q), que dan como resultado a los proceso mixtos, conocidos como procesos ARMA(p,q).

Definición 2.11. Se dice que $\{X_t\}$ es un proceso ARMA(p,q), si $\{X_t\}$ es estacionario y si para cada t , se satisface

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.47)$$

donde $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q} \sim i.i.d. RB(0, \sigma^2)$ y los polinomios $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ y $(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ no tienen factores en común.

Luego, escribiendo a la ecuación 2.47 en términos del operador de retardo

$$\phi_p(B)X_t = \theta_q(B)\varepsilon_t$$

donde $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ es el polinomio de retardo de un proceso autorregresivo y $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ es el polinomio de retardo de un proceso de promedios móviles.

Además, para que el proceso sea invertible, se necesita que las raíces de $\theta_q(B) = 0$ estén fuera del círculo unitario, y para que el proceso sea estacionario, se necesita que las raíces de $\phi_p(B) = 0$ estén fuera del círculo unitario.

También, se asume que $\theta_q(B) = 0$ y $\phi_p(B) = 0$ no tienen raíces en común. De aquí en adelante, nos referimos a este proceso como un proceso ARMA(p,q), en donde p y q indican los órdenes de los polinomios autorregresivos y promedios móviles respectivamente.

Es posible escribir al proceso estacionario e invertible ARMA como un proceso autorregresivo, es decir

$$\pi(B)X_t = \varepsilon_t$$

donde $\pi(B) = \frac{\phi_p(B)}{\theta_q(B)} = (1 - \pi_1 B - \pi_2 B^2 - \dots)$.

También, es posible escribirlo como un proceso de promedios móviles, es decir

$$X_t = \psi(B)\varepsilon_t$$

donde $\psi(B) = \frac{\theta_q(B)}{\phi_p(B)} = (1 + \psi_1 B + \psi_2 B^2 + \dots)$.

Antes de hallar la función de autocovarianza (γ_k) de un proceso ARMA(p,q), se reescribirá la ecuación 2.47 de la siguiente manera

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}.$$

Multiplicando X_{t-k} por ambos lados y obteniendo la esperanza, se llega a

$$E[X_{t-k}X_t] = \phi_1 E[X_{t-k}X_{t-1}] + \cdots + \phi_p E[X_{t-k}X_{t-p}] + E[X_{t-k}\varepsilon_t] - \theta_1 E[X_{t-k}\varepsilon_{t-1}] - \cdots - \theta_q E[X_{t-k}\varepsilon_{t-q}].$$

Y dado que $E[X_{t-k}\varepsilon_{t-i}] = 0, k > i$, tenemos

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, \quad k \geq (q+1). \quad (2.48)$$

Así, la función de autocorrelación es

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \quad k \geq (q+1). \quad (2.49)$$

Obsérvese que, la FAC simple del proceso ARMA es análoga a la del proceso AR(p). Entonces la FAC simple del proceso ARMA(p,q), se corta después del lapso q al igual que un proceso AR (p).

Para la FACP (ϕ_{kk}), dado que el proceso ARMA contiene al proceso MA como un caso particular, la FACP tendrá una mezcla de decaimientos exponenciales u ondas sinusoidales amortiguadas dependiendo de las raíces de $\phi_p(B)$ y $\theta_q(B)$. A continuación, se mencionará un ejemplo del proces ARMA. [31]

■ Proceso ARMA(1,1)

El proceso ARMA(1,1), esta dado de la siguiente forma

$$(1 - \phi_1 B)X_t = (1 - \theta_1 B)\varepsilon_t$$

o bien

$$X_t = \phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}. \quad (2.50)$$

Y para que el proceso sea estacionario de asumirá que $|\phi_1| < 1$ y para la invertibilidad se asumirá $|\theta_1| < 1$.

Para hallar la FAC del proceso ARMA(1,1), se multiplica en ambos lados de la ecuación 2.50 por X_{t-k} y se calcula la esperanza

$$\begin{aligned} E[X_{t-k}X_t] &= \phi_1 E[X_{t-k}X_{t-1}] + E[X_{t-k}\varepsilon_t] - \theta_1 E[X_{t-k}\varepsilon_{t-1}] \\ \gamma_k &= \phi_1 \gamma_{k-1} + E[X_{t-k}\varepsilon_t] - \theta_1 E[X_{t-k}\varepsilon_{t-1}]. \end{aligned}$$

Si $k = 0$

$$\gamma_0 = \phi_1 \gamma_1 + E[X_t \varepsilon_t] - \theta_1 E[X_t \varepsilon_{t-1}]$$

dado que

$$\begin{aligned} E[X_t \varepsilon_t] &= E[\varepsilon_t(\phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] \\ &= \phi_1 E[\varepsilon_t X_{t-1}] + E[\varepsilon_t^2] - \theta_1 E[\varepsilon_t \varepsilon_{t-1}] \\ &= \sigma_\varepsilon^2. \end{aligned}$$

Luego, multiplicando en ambos lados de la ecuación 2.50 por ε_{t-1} y calculando la esperanza observamos que

$$\begin{aligned} E[X_t \varepsilon_{t-1}] &= \phi_1 E[X_{t-1} \varepsilon_{t-1}] + E[\varepsilon_t \varepsilon_{t-1}] - \theta_1 E[\varepsilon_{t-1}^2] \\ &= (\phi_1 - \theta_1) \sigma_\varepsilon^2. \end{aligned}$$

Sustituyendo la ecuación anterior en γ_0

$$\gamma_0 = \phi_1 \gamma_1 + \sigma_\varepsilon^2 - \theta_1 (\phi_1 - \theta_1) \sigma_\varepsilon^2.$$

Si $k = 1$

$$\gamma_1 = \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2.$$

Así sustituyendo γ_1 en γ_0 , se tiene

$$\begin{aligned} \gamma_0 &= \phi_1 (\phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2) + \sigma_\varepsilon^2 - \theta_1 (\phi_1 - \theta_1) \sigma_\varepsilon^2 \\ &= \phi_1^2 \gamma_0 - \phi_1 \theta_1 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 - \phi_1 \theta_1 \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2. \end{aligned}$$

Por lo que γ_0 es de la forma

$$\gamma_0 = \frac{(1 + \theta_1^2 - 2\phi_1\theta_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2. \quad (2.51)$$

De esta manera, para γ_1 se tiene

$$\begin{aligned} \gamma_1 &= \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2 \\ &= \frac{\phi_1(1 + \theta_1^2 - 2\phi_1\theta_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2 - \theta_1 \sigma_\varepsilon^2 \\ &= \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2. \end{aligned} \quad (2.52)$$

Para $k \geq 2$

$$\gamma_k = \phi_1 \gamma_{k-1}. \quad (2.53)$$

Haciendo correr a k y resolviendo de forma recursiva, se observa que

$$\begin{aligned} \gamma_2 &= \phi_1 \gamma_1 \\ \gamma_3 &= \phi_1 \gamma_2 = \phi_1^2 \gamma_1 \\ \gamma_4 &= \phi_1^3 \gamma_1 \\ \gamma_5 &= \phi_1^4 \gamma_1 \\ &\vdots \\ \gamma_k &= \phi_1^{k-1} \gamma_1. \end{aligned}$$

Por lo que, la función de autocorrelación (ρ_k), es de la forma

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ \frac{(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{1 + \theta_1^2 - 2\phi_1 \theta_1}, & \text{si } k = 1; \\ \phi_1^{k-1} \rho_1, & \text{si } k \geq 2. \end{cases} \quad (2.54)$$

Si se desea hallar la FACP de un modelo mixto resulta algo complicado y es no necesario. Es suficiente con observar que, dado que el proceso ARMA(1,1) contiene al proceso MA(1) como un caso particular, la FACP de este proceso decae exponencialmente al igual que FAC, dependiendo de los signos y magnitudes de ϕ_1 y θ_1 . De este modo, la FAC y FACP indicarán que se trata de un modelo ARMA mixto.

En las Figuras 2.10 y 2.11 se puede observar las diferentes trayectorias para el proceso ARMA(1,1) y el comportamiento de FAC y FACP, dependiendo de los signos de ϕ_1 y θ_1 .

Por último, en la Tabla 2.3 se da una descripción de la FAC Y FACP para cada uno de los procesos vistos anteriormente.

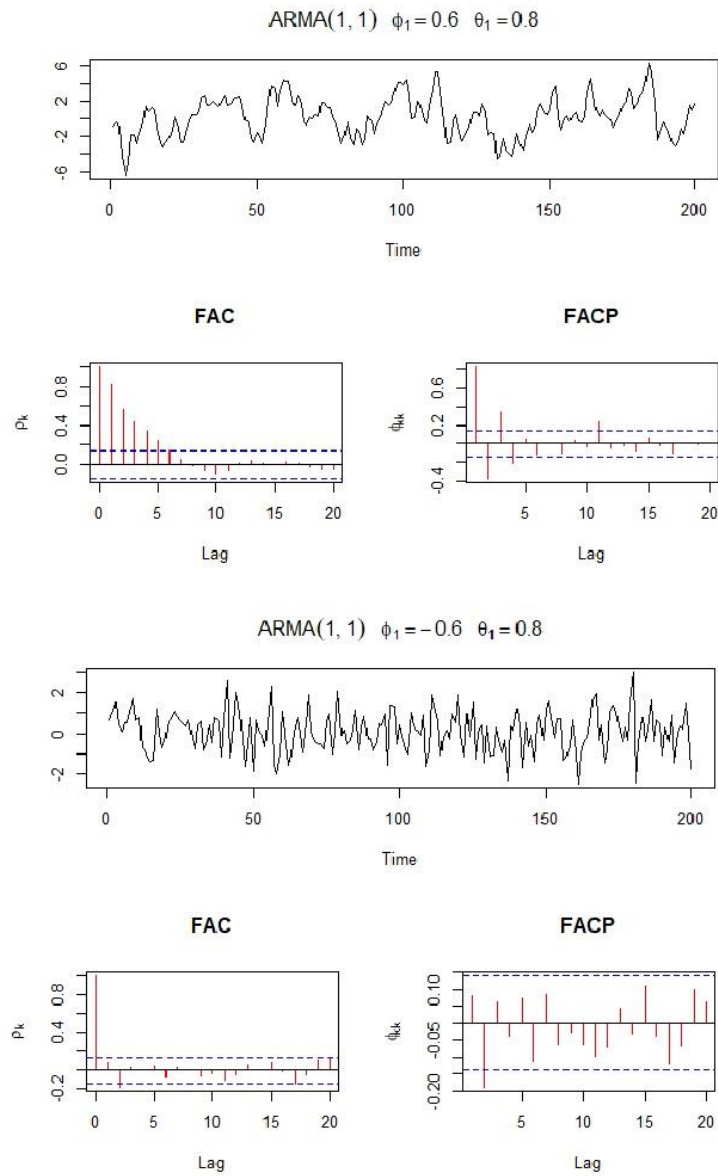


Figura 2.10: Trayectoria, FAC y FACP del proceso ARMA(1,1).

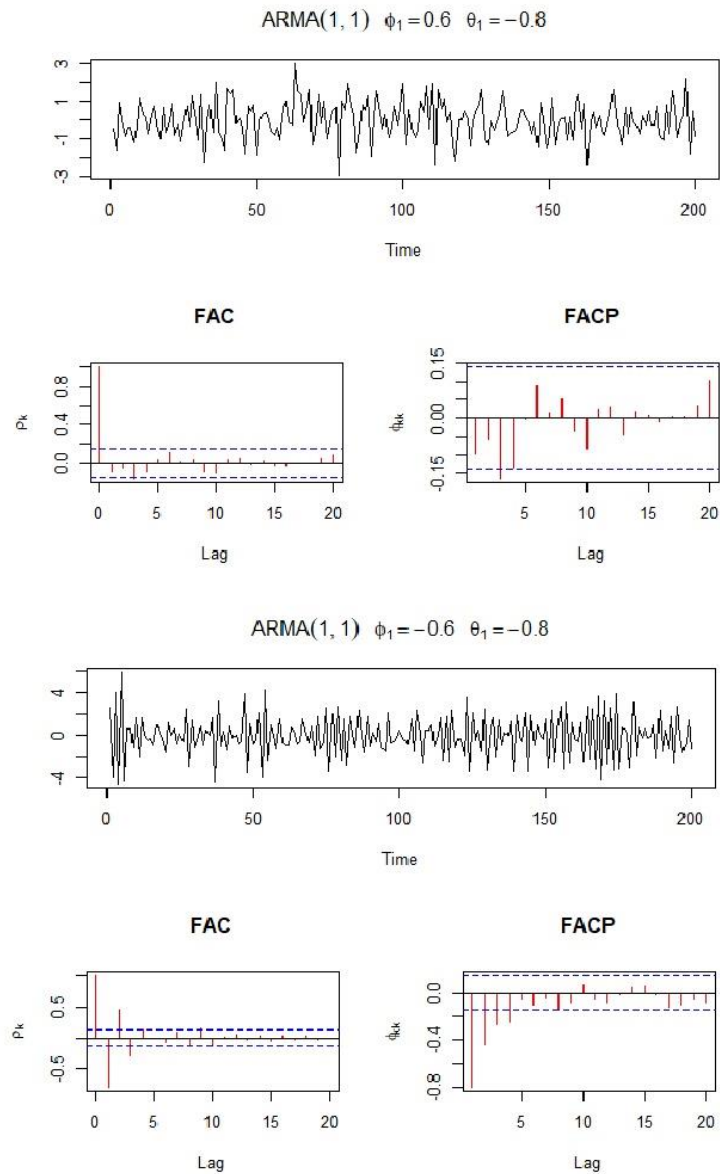


Figura 2.11: Trayectoria, FAC y FACP del proceso ARMA(1,1).

Proceso	FAC	FACP
Promedio móvil de orden uno. $X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$	Se corta después del lapso $k = 1$.	Se desvance exponencialmente dependiendo del signo de θ_1 .
Promedio móvil de segundo orden $X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$	Se corta después del lapso $k = 2$.	Se desvance exponencialmente o en forma sinusoidal amortiguada, dependiendo de los signos de θ_1, θ_2 .
Promedio móvil de orden q $X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$	Se corta después del lapso q .	Resulta una mezcla de decaimientos exponenciales y ondas sinusoidales amortiguadas, dependiendo de los signos de $\theta_q(B) = 0$.
Proceso autorregresivo de orden uno. $X_t = \phi_1 X_{t-1} + \varepsilon_t$	Se desvance exponencialmente.	Se corta después del lapso $k = 1$.
Proceso autorregresivo de segundo orden $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$	Se desvance exponencialmente, en forma sinusoidal amortiguada o en ambos.	Se corta después del lapso $k = 2$.
Proceso autorregresivo de orden p $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$	Se extingue.	Se corta después del lapso p .
Proceso autorregresivo de promedios móviles de orden $(1, 1)$ $X_t = \phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$	Se desvance de manera exponencial.	Se corta de manera que domina el decaimiento exponencial amortiguado.
Proceso autorregresivo de promedios móviles de orden (p, q) $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$	Se extingue.	Se extingue.

Tabla 2.3: Resumen de la FAC y FACP para los diferentes procesos.

MODELOS PARA SERIES DE TIEMPO NO ESTACIONARIAS

Los procesos vistos en el capítulo anterior están establecidos para cuando se tratan de procesos estacionarios, donde la media, la varianza son constantes a través del tiempo y la covarianza no depende del tiempo, ésta mas bien, depende de los lapsos de tiempo.

Usualmente en aplicaciones, específicamente en negocios y economía los procesos se desenvuelven de manera no estacionaria, esto es, ya sea que la media o la varianza del proceso son no constantes o que ocurran ambos a la vez, un ejemplo de serie no estacionaria es el tipo de cambio (peso por dólar) (Figura 3.1).

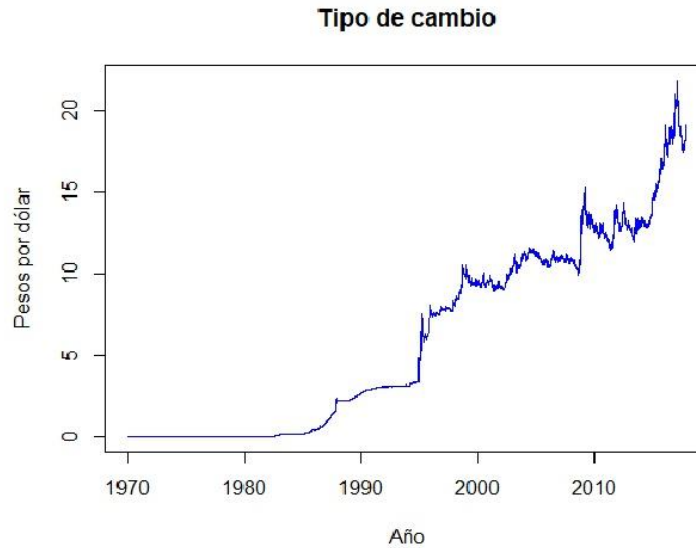


Figura 3.1: Tipo de cambio spot.

3.1. Modelo ARIMA(p,d,q)

Los modelos de series de tiempo desarrollados en la sección anterior, están establecidos para series estacionarias, pero existen casos en los que hay presencia de no estacionariedad, un modelo establecido para este tipo de series es el modelo **Autorregresivo Integrado de Promedios Móviles (ARIMA)**, los cuales son una generalización de los ARMA que resulta de aplicar una diferencia de orden “ d ” a la serie. De esta manera, al aplicar el operador ∇^d se estaría eliminando la posible tendencia polinomial de orden d hallada en la serie que se esté analizando, es decir que, al aplicar el método de diferencias es posible convertir la serie no estacionaria en estacionaria, una vez transformada, se le aplica el modelo ARMA, lo que nos llevaría al modelo ARIMA.

Definición 3.1. Un proceso X_t decimos que es un ARIMA(p,d,q) ($d \geq 0$ entero) si

$$\nabla^d X_t = (1 - B)^d X_t \quad (3.1)$$

es un ARMA(p,q).

De manera general, es posible escribir al modelo ARIMA de la forma

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)\varepsilon_t \quad (3.2)$$

donde $\{\varepsilon_t\} \sim RB(0, \sigma^2)$, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ es el polinomio de retardo de un proceso autorregresivo y $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ es el polinomio de retardo de un proceso de promedios móviles, los cuales no tienen factores en común.

Si $E[\nabla^d X_t] = \mu$, escribiremos al proceso como

$$\phi_p(B)(1-B)^d X_t = \delta + \theta_q(B)\varepsilon_t. \quad (3.3)$$

La constante δ , puede tomar un diferente rol para cuando $d = 0$ y $d > 0$. Cuando $d = 0$ el proceso original es estacionario y $\delta = \mu(1 - \phi_1 - \dots - \phi_p)$. Cuando $d > 0$, la constante δ es llamado término de tendencia determinista, comúnmente es descartado a menos que sea necesario su uso.

Los parámetros (p, d, q) representan lo siguiente

- p , es el orden del polinomio autorregresivo.
- d , es el exponente en el operador de diferencia.
- q , es el orden del polinomio de promedios móviles.

Ahora bien, en cuanto a las condiciones de estacionariedad e invertibilidad para los procesos ARIMA, serán las mismas que se aplican en los procesos ARMA, es decir, para que el proceso sea **estacionario** las raíces de $\phi_p(B) = 0$ deben encontrarse fuera del círculo unitario y para que sea **invertible** las raíces de $\theta_q(B) = 0$ se deben encontrar fuera del círculo unitario.

Cuando $p = 0$ el modelo ARIMA(p,d,q) es también llamado modelo integrado de promedio móvil de orden (d,q), denotado **IMA(d,q)**, cuando $q = 0$ este es llamado autorregresivo integrado denotado por **ARI(p,d)** y si $p = 0$, $d = 1$ y $q = 0$ se tiene lo que se conoce como **modelo de caminata aleatoria**, dichos modelos se presentarán más adelante. [29], [31]

3.1.1. No estacionariedad en media

Una serie de tiempo es no estacionaria en media, cuando esta depende del tiempo, lo que ocasiona la presencia de tendencia o varios tramos con medias diferentes. Si la serie es no estacionaria en media, esta es posible transformarla en estacionaria tomando diferencias de orden d . Así la serie que no era estacionaria en media X_t , se volverá estacionaria por el hecho de aplicar d diferencias $(1 - B)^d X_t$, para $d \geq 1$ enteros. Normalmente d toma los valores de 0, 1, 2. El problema entonces será determinar el número de diferencias que se deben aplicar, ya que pueden surgir los siguientes problemas:

- Raíz autorregresiva próxima a la unidad.
- Sobrediferenciación.

Unas sugerencias para decidir el valor de d , son los siguientes

1. Gráficos de la serie original y las transformaciones correspondientes, para observar si se cumple o no la media constante.
2. Correlograma de las FAC y FACP, para observar si decrecen rápidamente o no.
3. Contrastes de raíces unitarias.

Este último proporciona contrastes estadísticos, y a partir de ello se hace inferencia sobre la existencia o no de raíz unitaria en la serie, es decir, sobre la no estacionariedad de la serie de tiempo. Si se rechaza la hipótesis nula que es la existencia de raíz unitaria, no se diferenciará más la serie. En caso contrario, se tomará una diferencia más de orden 1. El contraste de raíz unitaria más usado es **Dickey-Fuller**, el cuál puede consultarse en el Apéndice C.

▪ Modelo de caminata aleatoria

De la ecuación (3.2), si $p = 0$, $d = 1$ y $q = 0$ se tiene la siguiente ecuación

$$(1 - B)X_t = \varepsilon_t \quad (3.4a)$$

$$X_t = X_{t-1} + \varepsilon_t. \quad (3.4b)$$

Este tipo de modelo es usualmente usado en el comportamiento de series de un precio de acciones.

En la ecuación (3.4b) el valor de X en el tiempo t es igual a su valor en el tiempo $(t-1)$ más un choque aleatorio. Obsérvese que el modelo de la caminata aleatoria es un proceso límite del proceso AR(1) esto es, $(1 - \phi_1 B)X_t = \varepsilon_t$ con $\phi_1 \rightarrow 1$. Ahora bien, dado que la FAC del proceso AR(1) es $\rho_k = \phi^k$, el modelo de caminata aleatoria se puede caracterizar por picos grandes, no desvanecidos de la FAC muestral de la serie original $\{X_t\}$ y la FAC se desvanece completamente para la serie diferenciada $(1 - B)X_t$.

De la ecuación (3.4b), al añadir un término constante el modelo resultará ser el siguiente

$$X_t = X_{t-1} + \delta + \varepsilon_t. \quad (3.5)$$

A la constante δ se le llama **deriva**, de ahí que a la ecuación (3.5) se le llame modelo de **caminata aleatoria con deriva**. Note que cuando $\delta = 0$ se tiene un caminata aleatoria.

Considere ahora, $X_{t-1} = X_{t-2} + \delta + \varepsilon_{t-1}$ que al sustituirla en 3.5, se tiene

$$X_t = X_{t-2} + 2\delta + \varepsilon_{t-1} + \varepsilon_t.$$

Haciendo lo mismo con $X_{t-2} = X_{t-3} + \delta + \varepsilon_{t-2}$, obtenemos

$$X_t = X_{t-3} + 3\delta + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t.$$

Continuado de esta manera, sustituyendo sucesivamente hasta un tiempo t , resulta

$$X_t = X_0 + t\delta + \sum_{j=1}^t \varepsilon_j. \quad (3.6)$$

Dado que $\{\varepsilon_t\} \sim RB(0, \sigma^2)$ y $X_0 = 0$, calculando la esperanza de 3.6, se obtiene

$$E[X_t] = t\delta + \sum_{j=1}^t E[\varepsilon_j] = t\delta.$$

Esto muestra que el **proceso de caminata aleatoria con deriva es no estacionario**, ya que la media del proceso depende del tiempo t .

■ **Modelo ARIMA(0,1,1) o IMA(1,1)**

De la ecuación 3.2, si $p = 0$, $d = 1$ y $q = 1$ se tiene lo siguiente

$$(1 - B)X_t = (1 - \theta_1 B)\varepsilon_t, \quad (3.7)$$

o bien,

$$X_t = X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \quad (3.8)$$

donde $|\theta_1| < 1$ y $\{\varepsilon_t\} \sim RB(0, \sigma^2)$.

Obsérvese que el modelo IMA(1,1) para X_t se reduce al modelo estacionario MA(1) para la serie de primeras diferencias $(1 - B)X_t$. Note que la caminata aleatoria es un caso particular de el modelo IMA(1,1) haciendo $\theta_1 = 0$.

Por lo tanto, el modelo IMA(1,1) se caracteriza por que la FAC muestral de la serie original no se extingue y a FAC muestral de la serie de primeras diferencias muestra un patrón de un fenómeno de MA(1).

Sea $|\theta_1| < 1$, entonces

$$\begin{aligned} \frac{(1 - B)}{(1 - \theta_1 B)} &= (1 - B)(1 + \theta_1 B + \theta_1^2 B^2 + \dots) \\ &= 1 + \theta_1 B + \theta_1^2 B^2 + \dots - B - \theta_1 B^2 - \theta_1^2 B^3 - \dots \\ &= 1 - \alpha B - \alpha(1 - \alpha)B^2 - \alpha(1 - \alpha)^2 B^3 - \dots \end{aligned}$$

donde $\alpha = (1 - \theta_1)$.

Así, de 3.7 se tiene

$$\begin{aligned} (1 - \alpha B - \alpha(1 - \alpha)B^2 - \alpha(1 - \alpha)^2 B^3 - \dots)X_t &= \varepsilon_t \\ X_t - \alpha X_{t-1} - \alpha(1 - \alpha)X_{t-2} - \alpha(1 - \alpha)^2 X_{t-3} - \dots &= \varepsilon_t \\ X_t &= \alpha \sum_{j=1}^{\infty} (1 - \alpha)^{j-1} X_{t-j} + \varepsilon_t. \end{aligned} \quad (3.9)$$

Obsérvese que la ecuación anterior es la representación de un modelo AR(∞). Luego, el pronóstico óptimo \hat{X}_t de X_t es dado por

$$\hat{X}_t = \hat{\alpha} \sum_{j=1}^{\infty} (1 - \hat{\alpha})^{j-1} X_{t-j}. \quad (3.10)$$

Es decir, que el pronóstico óptimo de X_t es un autorregresivo ponderado que

decrece exponencialmente de sus valores pasados. Así, el pronóstico aproximado del periodo de tiempo $t + 1$ es igual a

$$\begin{aligned}
 \hat{X}_{t+1} &= \hat{\alpha} \sum_{j=1}^{\infty} (1 - \hat{\alpha})^{j-1} X_{t+1-j} \\
 &= \hat{\alpha} X_t + (1 - \hat{\alpha}) \hat{\alpha} \sum_{j=2}^{\infty} (1 - \hat{\alpha})^{j-2} X_{t+1-j} \\
 &= \hat{\alpha} X_t + (1 - \hat{\alpha}) \hat{\alpha} \sum_{i=1}^{\infty} (1 - \hat{\alpha})^{i-1} X_{t-i} \\
 &= \hat{\alpha} X_t + (1 - \hat{\alpha}) \hat{X}_t.
 \end{aligned}$$

El cuál es una combinación lineal de la nueva observación disponible y el último pronóstico. El parámetro $1 - \hat{\alpha}$ es llamado parámetro de suavización y es restringido a estar entre cero y uno. Cuando se tienen valores más grandes de $\hat{\alpha}$, estos conducen a previsiones más suaves.

Este método de pronóstico es popular porque es fácil de usar, sólo se necesita retener el valor del pronóstico anterior y la observación actual para pronosticar el próximo periodo de tiempo. [31]

3.1.2. No estacionariedad en varianza

Como se menciona al inicio, una serie de tiempo es no estacionaria en varianza, cuando esta depende del tiempo. Si desea hacer estacionaria a la varianza, la solución es transformar a la serie mediante un método que estabilice la varianza.

La diferenciación por ejemplo, es aplicada para reducir una serie temporal no estacionaria homogénea a una serie temporal estacionaria, pero, muchas series de tiempo no estacionarias no son homogéneas.

La no estacionariedad de estas no se debe a que sus medias dependan del tiempo, es más bien, que sus varianzas y autocovarianza dependen del tiempo. Para reducir este tipo de no estacionariedad, se requieren transformaciones que no sean de diferenciación.

■ **Varianza y autocovarianza de un modelo ARIMA**

Un proceso que es estacionario en media, no precisamente es estacionario en varianza y autocovarianza. Pero un proceso que es no estacionario en media, será no estacionario en varianza y autocovarianza.

Si bien el modelo es no estacionario, la característica completa del proceso se determina para todos los tiempos sólo por un número finito de parámetros, $\phi_i, \theta_j, \sigma_\varepsilon^2$. Así, la evolución completa del proceso se puede desarrollar desde el modelo ARIMA ajustado al conjunto de datos X_1, X_2, \dots, X_n .

Sea el modelo IMA(1,1)

$$X_t = X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}.$$

Haciendo sustituciones sucesivas y dado que $X_0 = 0$ y $\varepsilon_0 = 0$, resulta

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &= X_{t-2} + \varepsilon_t + (1 - \theta_1) \varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} \\ &\vdots \\ &= X_0 + \varepsilon_t + (1 - \theta_1) \varepsilon_{t-1} + \dots + (1 - \theta_1) \varepsilon_1 - \theta_1 \varepsilon_0 \\ &= \varepsilon_t + \sum_{i=1}^{t-1} (1 - \theta_1) \varepsilon_{t-i}. \end{aligned}$$

Siguiendo el mismo proceso para un lapso $t - k$

$$\begin{aligned} X_{t-k} &= X_0 + \varepsilon_{t-k} + (1 - \theta_1) \varepsilon_{t-k-1} + \dots + (1 - \theta_1) \varepsilon_1 - \theta_1 \varepsilon_0 \\ X_{t-k} &= \varepsilon_{t-k} + \sum_{i=1}^{t-k-1} (1 - \theta_1) \varepsilon_{t-k-i} \end{aligned}$$

Es claro que

$$E[X_t] = E[\varepsilon_t] + \sum_{i=1}^{t-1} (1 - \theta_1) E[\varepsilon_{t-i}] = 0.$$

Y la varianza de X_t está dada por

$$Var[X_t] = E \left[\left(\varepsilon_t + \sum_{i=1}^{t-1} (1 - \theta_1) \varepsilon_{t-i} \right)^2 \right]$$

o bien, escrita de forma desarrollada

$$Var[X_t] = E[(\varepsilon_t + (1-\theta_1)\varepsilon_{t-1} + \dots + (1-\theta_1)\varepsilon_1)(\varepsilon_t + (1-\theta_1)\varepsilon_{t-1} + \dots + (1-\theta_1)\varepsilon_1)].$$

Al desarrollar los productos se observa que

$$2(1-\theta_1)E[\varepsilon_t\varepsilon_{t-i}] = 2(1-\theta_1) \cdot 0 = 0$$

y

$$(1-\theta_1)^2 E[\varepsilon_{t-i}\varepsilon_{t-j}] = (1-\theta_1)^2 \cdot 0 = 0, \quad \text{para } i \neq j.$$

Por lo tanto

$$\begin{aligned} Var[X_t] &= E[\varepsilon_t^2] + (1-\theta_1)^2 E[\varepsilon_{t-1}^2] + (1-\theta_1)^2 E[\varepsilon_{t-2}^2] + \dots + (1-\theta_1)^2 E[\varepsilon_1^2] \\ &\quad + (1-\theta_1)^2 E[\varepsilon_1^2] \\ &= \sigma_\varepsilon^2 + (1-\theta_1)^2 \sigma_\varepsilon^2 + (1-\theta_1)^2 \sigma_\varepsilon^2 + \dots + (1-\theta_1)^2 \sigma_\varepsilon^2 + (1-\theta_1)^2 \sigma_\varepsilon^2 \\ &= [1 + (1-\theta_1)^2(t-1)]\sigma_\varepsilon^2. \end{aligned}$$

Nótese que la varianza del proceso depende de t , de esta manera el proceso resulta ser no estacionario en varianza.

Del mismo modo ocurre con $Var[X_{t-k}]$, la cual esta dada por

$$Var[X_{t-k}] = [1 + (1-\theta_1)^2(t-k-1)]\sigma_\varepsilon^2.$$

De esta manera, la covarianza entre X_t y X_{t-k} resulta ser

$$Cov[X_{t-k}, X_t] = [(1-\theta_1) + (1-\theta_1)^2(t-k-1)]\sigma_\varepsilon^2.$$

De lo anterior, es fácil hallar a la función de correlación entre X_{t-k} y X_t

$$\begin{aligned} Corr[X_{t-k}, X_t] &= \frac{Cov[X_{t-k}, X_t]}{\sqrt{Var[X_{t-k}]Var[X_t]}} \\ &= \frac{(1-\theta_1) + (1-\theta_1)^2(t-k-1)}{\sqrt{[1 + (1-\theta_1)^2(t-k-1)][1 + (1-\theta_1)^2(t-1)]}} \end{aligned}$$

Por consiguiente se tienen las siguientes observaciones

Observación 3.1.

- La varianza $Var[X_t]$ de un proceso $IMA(1,1)$ es dependiente del tiempo, y $Var[X_t] \neq Var[X_{t-k}]$ para $k \neq 0$.
- La varianza $Var[X_t]$ es no acotada para $t \rightarrow \infty$.
- La covarianza $Cov[X_{t-k}, X_t]$ y la autocorrelación $Corr[X_{t-k}, X_t]$ también dependen del tiempo y por tanto, no son invariantes respecto a la traslación del tiempo.
- La función de autocorrelación $Corr[X_{t-k}, X_t]$ de un proceso $IMA(1,1)$, se desvanece lentamente a medida que k incrementa.

En general, resulta imposible hacer una inferencia estadística con sólo una realización, cuando esta es no estacionaria tanto en la media como en la autocovarianza o autocorrelación. [31]

- **Transformaciones para estabilizar la varianza**

No todas las series de tiempo pueden ser transformadas a una serie de tiempo estacionaria por medio de diferencias. Por ejemplo, algunas series de tiempo pueden ser estacionarias en media pero no en varianza como se vio anteriormente. Es por ello que se tiene que usar una transformación adecuada para estabilizar a la varianza. Es muy común que la varianza de un proceso no estacionario cambie a medida que su media cambia. Esto es,

$$Var[X_t] = cf(\mu_t).$$

Donde c es una constante positiva y f una función. Ahora bien, la cuestión es ¿Cómo hallar una función T para la serie de tiempo que al aplicarle dicha transformación $T(X_t)$ tenga una variación constante?

Pues bien, para iniciar el método, se debe aproximar a la función por medio de una serie de Taylor de primer orden alrededor de μ_t , es decir

$$T(X_t) \simeq T(\mu_t) + T'(\mu_t)(X_t - \mu_t).$$

Así, la varianza de $T(X_t)$ es proxímadamente igual a

$$\begin{aligned}
\text{Var}[T(X_t)] &\cong \text{Var}[T(\mu_t) + T'(\mu_t)(X_t - \mu_t)] \\
&= \text{Var}[T(\mu_t)] + \text{Var}[T'(\mu_t)(X_t - \mu_t)] \\
&= [T'(\mu_t)]^2 \text{Var}[X_t] \\
&= c[T'(\mu_t)]^2 f(\mu_t).
\end{aligned}$$

Por lo tanto, para que la varianza de $T(X_t)$ sea constante, la transformación de estabilización de la varianza $T(X_t)$ debe elegirse de modo que

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}}. \quad (3.11)$$

Por lo que la ecuación 3.11 implica

$$T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t. \quad (3.12)$$

A continuación se mencionarán unos ejemplos donde la varianza de la serie de tiempo se comporta de diferentes maneras.

- Por ejemplo, si se tiene que la desviación estándar de una serie de tiempo es proporcional a su media, es decir, $\text{Var}[X_t] = c\mu_t^2$, entonces

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^2}} d\mu_t = \ln(\mu_t). \quad (3.13)$$

Por consiguiente, aplicar una transformación logarítmica a la serie de tiempo, $\ln(X_t)$ hará que esta tenga una variación constante.

- Si la varianza de la serie de tiempo es proporcional a su media, esto es, $\text{Var}[X_t] = c\mu_t$, entonces

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t}} d\mu_t = 2\sqrt{\mu_t}. \quad (3.14)$$

Por lo tanto, al aplicar una transformación de raíz cuadrada a la serie de tiempo, $\sqrt{X_t}$, provocará que esta tenga una variación constante.

- Ahora bien, si la desviación estándar de la serie de tiempo es proporcional al cuadrado de la media, es decir, $Var[X_t] = c\mu_t^4$, entonces

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^4}} d\mu_t = -\frac{1}{\mu_t}. \quad (3.15)$$

Consecuentemente, al aplicar la transformación recíproca a la serie de tiempo, $\frac{1}{X_t}$, hará que la serie tenga varianza constante.

De forma más general, es posible usar la transformación potencia

$$y_t = T(X_t) = \frac{X_t^\lambda - 1}{\lambda}. \quad (3.16)$$

$$T(X_t) = y_t = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0; \\ \ln X_t, & \text{si } \lambda = 0. \end{cases} \quad (3.17)$$

Donde 3.16 es conocida generalmente como **transformación de Box-Cox(1964)** (ver [9]), la cual contiene muchas de las transformaciones antes mencionadas como casos especiales. En la tabla 3.1, se aprecian los valores de λ que normalmente son utilizados.

Valores de λ	Transformación
-1.0	$\frac{1}{X_t}$
-0.5	$\frac{1}{\sqrt{X_t}}$
0.0	$\ln(X_t)$
0.5	$\sqrt{X_t}$
1.0	X_t (No hay transformación)

Tabla 3.1: Transformaciones generales.

De la transformación 3.16, note que cuando $\lambda \rightarrow 0$ la transformación se indefine, esto es,

$$\lim_{\lambda \rightarrow 0} T(X_t) = \lim_{\lambda \rightarrow 0} \frac{X_t^\lambda - 1}{\lambda}.$$

Ahora bien, al resolver esto por medio de la regla de L'Hôpital (derivando respecto a λ), resulta

$$\lim_{\lambda \rightarrow 0} T(X_t) = \lim_{\lambda \rightarrow 0} \frac{X_t^\lambda - 1}{\lambda} = \lim_{\lambda \rightarrow 0} \frac{(X_t^\lambda - 1)'}{(\lambda)'} = \lim_{\lambda \rightarrow 0} X_t^\lambda \ln(X_t) = \ln(X_t)$$

resultando la transformación logarítmica.

Lo anterior resultó debido a lo siguiente, como $(X_t^\lambda - 1)$ se derivó con respecto a λ , esto es, $\frac{d(X_t^\lambda - 1)}{d\lambda} = \frac{dX_t^\lambda}{d\lambda}$, luego

$$\begin{aligned} X_t^\lambda &= e^{\ln X_t^\lambda} \\ X_t^\lambda &= e^{\lambda \ln X_t} \\ \Rightarrow (X_t^\lambda)' &= (e^{\lambda \ln X_t})' \\ &= e^{\lambda \ln X_t} (\lambda \ln X_t)' \\ &= e^{\lambda \ln X_t} (\lambda)' \ln X_t \\ &= e^{\lambda \ln X_t} (\ln X_t) \\ &= X_t^\lambda (\ln X_t) \end{aligned}$$

es así como $\frac{dX_t^\lambda}{d\lambda} = X_t^\lambda (\ln X_t)$.

Ahora bien, si se tiene que $\lambda = 0.25$, la transformación resultante es $\sqrt[4]{X_t}$, ya que

$$T(X_t) = \frac{X_t^{0.25} - 1}{0.25} = \frac{X_t^{\frac{1}{4}} - 1}{\frac{1}{4}} = 4(\sqrt[4]{X_t} - 1).$$

Es posible tomar a λ como un parámetro de la transformación y estimar su valor a partir de los datos. Por ejemplo, podemos incluir a λ como un parámetro en el modelo $(1 - \phi_1 B - \dots - \phi_p B^p)(X_t^\lambda - \mu) = (1 - \theta_1 B - \dots - \theta_q B^q)\varepsilon_t$ y elegir el valor de λ de manera que los errores cuadrados medios residuales sean mínimos. Box y Cox mostraron que λ puede ser estimado por el método de máxima verosimilitud que minimiza el error cuadrado medio calculado a partir del modelo ajustado en los datos normalizados.

Observación 3.2.

- Las transformaciones para estabilizar a la varianza se definen solamente para series positivas. Sin embargo estas definiciones no son tan restrictivas como se ve, porque se puede agregar una constante a la serie sin afectar la estructura de correlación de la serie.
- Si es necesario hacer transformaciones para estabilizar a la varianza, esta conviene realizarse antes de cualquier otro análisis, como la diferenciación.
- Normalmente al hacer transformaciones no sólo se está estabilizando a la varianza, esta también mejora la aproximación de la distribución a una distribución normal. [31]

3.2. Modelos estacionales multiplicativos ARIMA

Muchas series de tiempo provienen de procesos con características estacionales, las cuales se repiten en cierto periodo en el tiempo. Donde la dependencia del pasado tiende a ocurrir con mayor, menor o igual fuerza. Al periodo de tiempo en el cual se repite este fenómeno se le conoce como periodo estacional de la serie.

Por ejemplo, en datos económicos mensuales, existe un fuerte componente anual que ocurre en rezagos que son múltiplos de $s = 12$, debido a las fuertes conexiones de toda la actividad anual en el calendario. Los datos trimestrales exhibirán el periodo anual repetitivo en $s = 4$ trimestres. En los fenómenos naturales, también es posible observar periodos fuertes que corresponden a las estaciones. Por lo que, la variabilidad natural de muchos procesos físicos, biológicos y económicos tienden a coincidir con los cambios estacionales naturales.

Es por ello que resulta conveniente introducir polinomios de promedio móvil y autoregresivos que se identifiquen con los rezagos (lapsos) estacionales. [29]

Definición 3.2. Se dice que un modelo es un *autorregresivo de promedio móvil estacional puro*, si este es de la forma

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)\varepsilon_t \quad (3.18)$$

denotado como $ARMA(P,Q)_s$. Donde los operadores

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$$

y

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs}.$$

Son el operador estacional autorregresivo de orden P y el operador estacional de media móvil de orden Q , con un periodo estacional s .

De igual manera que los procesos $ARMA(p,q)$, el modelo $ARMA(P,Q)_s$ es **estacionario** si las raíces de $\Phi_P(B^s)$ se encuentran fuera del círculo unitario y es **invertible** si las raíces de $\Theta_Q(B^s)$ se encuentran fuera del círculo unitario.

También es posible combinar los operadores estacional y no estacional en una expresión multiplicativa dada de la siguiente manera.

Definición 3.3. Se dice que un modelo es un **autorregresivo de promedio móvil estacional multiplicativo**, si este es de la forma

$$\Phi_P(B^s)\phi_p(B)X_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t \quad (3.19)$$

denotado como $ARMA(p,q) \times (P,Q)_s$.

En el proceso $ARMA(p,q) \times (P,Q)_s$ los operadores $\phi_p(B)$ y $\theta_q(B)$ son los operadores AR (estacionario) y MA (invertible) respectivamente, llamados **operadores regulares** que representan la parte no estacional, los cuales permiten modelar la correlación entre pares de componentes de X_t separados entre sí por $k = 1, 2, 3, \dots$ periodos básicos (por ejemplo, meses o trimestres).

Los operadores $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$ y $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs}$ son llamados **operadores anuales** que representan la parte estacional, estos describen la correlación entre pares de componentes de X_t separados entre sí por $k = s, 2s, 3s, \dots$ periodos básicos (o bien $1, 2, 3, \dots$ años).

La presencia de la estacionalidad, normalmente ocurre cuando se tiene un proceso casi periódico en la temporada. Por ejemplo, si se mide la temperatura mensual promedio a lo largo de los años, se observaría que cada enero sería aproximadamente la misma temperatura, cada febrero y así sucesivamente. Este proceso podría modelarse mediante la siguiente expresión.

$$X_t = S_t + v_t$$

donde S_t es la componente estacional que cambia lentamente año con año, y de acuerdo a una caminata aleatoria

$$S_t = S_{t-s} + \varepsilon_t$$

v_t, ε_t son procesos de ruido blanco no correlacionados.

Si se tiene que la serie de tiempo es no estacionaria en media, debido a la estacionalidad de la serie, entonces se propone un número D para realizar una diferencia estacional, de manera que la serie de tiempo sea estacionaria en la parte estacional.

Definición 3.4. El operador de diferencia estacional de periodo s y orden D está definido de la siguiente manera

$$\nabla_s^D X_t = (1 - B^s)^D X_t \quad (3.20)$$

donde $D = 1, 2, 3, \dots$ toma valores enteros positivos.

Frecuentemente si $D = 1$, es suficiente para tener estacionariedad en la parte estacional.

Por ejemplo, dado que S_t es no estacionario, entonces X_t no lo es, por lo que al tomar diferencias estacionales a X_t resulta

$$\begin{aligned} \nabla_s X_t &= (1 - B^s)X_t \\ &= X_t - B^s X_t \\ &= (S_t + v_t) - X_{t-s} \\ &= (S_t + v_t) - (S_{t-s} + v_{t-s}) \\ &= S_t - S_{t-s} + v_t - v_{t-s} \\ &= \varepsilon_t + v_t - v_{t-s}. \end{aligned}$$

Así $\nabla_s X_t$ es estacionario y tiene la FAC de un $MA(1)_s$. Su FAC tendrá una sola espiga en el lapso s .

En general, la diferencia estacional puede ser la indicada cuando la FAC se extingue lentamente en los múltiplos de los lapsos estacionales s y es insignificante entre los periodos.

Ahora bien, conectando todos los conceptos anteriores, llevan a la siguiente definición.

Definición 3.5. Un modelo autorregresivo estacional integrado de promedio móvil multiplicativo o modelo *SARIMA* esta dado de la siguiente manera

$$\Phi_P(B^s)\phi_p(B)\nabla_s^D \nabla^d X_t = \delta + \Theta_Q(B^s)\theta_q(B)\varepsilon_t \quad (3.21)$$

donde ε_t es un proceso de **ruido blanco Gaussiano**. El modelo es denotado por $ARIMA(p,d,q) \times (P,D,Q)_s$, donde

- $\phi_p(B)$, operador del proceso autorregresivo de orden p .
- $\theta_q(B)$, operador del proceso de promedio móvil de orden q .
- $\Phi_P(B^s)$, operador estacional autorregresivo de orden P .
- $\Theta_Q(B^s)$, operador estacional de promedio móvil de orden Q .
- ∇^d , componente de diferencia de orden d .
- ∇_s^D , componente de diferencia estacional de periodo s y orden D .

Ejemplo 3.1.

Considere el proceso $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$, en este ejemplo se tiene que $s = 12$, esto representa el hecho de que las fluctuaciones estacionales suceden cada 12 meses, y $\delta = 0$

$$(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})\varepsilon_t \quad (3.22)$$

o bien, desarrollando el lado derecho de la ecuación 3.22 resulta

$$W_t = (1 - \theta_1 B - \Theta_1 B^{12} + \theta_1 \Theta_1 B^{13})\varepsilon_t$$

donde $W_t = (1 - B - B^{12} + B^{13})X_t$.

Este modelo es muy útil para representar datos de líneas aéreas y series comerciales. El modelo fue mostrado por primera vez por Box-Jenkins para representar los datos de viajes aéreos internacionales.

La esperanza de W_t resulta ser

$$\begin{aligned} E[W_t] &= E[\varepsilon_t - \theta_1 \varepsilon_{t-1} - \Theta_1 \varepsilon_{t-12} + \theta_1 \Theta_1 \varepsilon_{t-13}] \\ &= E[\varepsilon_t] - \theta_1 E[\varepsilon_{t-1}] - \Theta_1 E[\varepsilon_{t-12}] + \theta_1 \Theta_1 E[\varepsilon_{t-13}] \\ &= 0. \end{aligned}$$

Luego, calculando la varianza de W_t se tiene

$$\begin{aligned}
 \gamma_0 &= Var[W_t] = E[(W_t - \mu)^2] \\
 &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \Theta_1 \varepsilon_{t-2} + \theta_1 \Theta_1 \varepsilon_{t-3})^2] \\
 &= E[\varepsilon_t^2] + \theta_1^2 E[\varepsilon_{t-1}^2] + \Theta_1^2 E[\varepsilon_{t-2}^2] + \theta_1^2 \Theta_1^2 E[\varepsilon_{t-3}^2] + 2\theta_1 E[\varepsilon_t \varepsilon_{t-1}] \\
 &\quad + 2\Theta_1 E[\varepsilon_t \varepsilon_{t-2}] + 2\theta_1 \Theta_1 E[\varepsilon_t \varepsilon_{t-3}] + 2\theta_1 \Theta_1 E[\varepsilon_{t-1} \varepsilon_{t-2}] \\
 &\quad + 2\theta_1^2 \Theta_1 E[\varepsilon_{t-1} \varepsilon_{t-3}] + 2\theta_1 \Theta_1^2 E[\varepsilon_{t-2} \varepsilon_{t-3}] \\
 &= (1 + \theta_1^2 + \Theta_1^2 + \theta_1^2 \Theta_1^2) \sigma^2 \\
 &= (1 + \theta_1^2)(1 + \Theta_1^2) \sigma^2.
 \end{aligned}$$

De la misma maera es posible obtener a γ_1 , que resulta ser

$$\gamma_1 = Cov(W_t, W_{t-1}) = -\theta_1(1 + \Theta_1^2) \sigma^2.$$

Con lo anterior, es fácil ver que

$$\gamma_2 = \gamma_3 = \dots = \gamma_{10} = 0.$$

Entonces γ_{11} , γ_{12} , γ_{13} resultan ser

$$\begin{aligned}
 \gamma_{11} &= \theta_1 \Theta_1 \sigma^2 \\
 \gamma_{12} &= -\Theta_1(1 + \theta_1^2) \sigma^2 \\
 \gamma_{13} &= \theta_1 \Theta_1 \sigma^2 \\
 \gamma_j &= 0, \quad j > 13.
 \end{aligned}$$

Dado lo anterior, es posible obtener $FAC(\rho_k)$

$$\begin{aligned}
 \rho_1 &= \frac{-\theta_1}{(1 + \theta_1^2)} \\
 \rho_{11} &= \frac{\theta_1 \Theta_1}{(1 + \theta_1^2)(1 + \Theta_1^2)} = \rho_{13} \\
 \rho_{12} &= \frac{-\Theta_1}{(1 + \Theta_1^2)} \\
 \rho_j &= 0, \quad j > 13.
 \end{aligned}$$

3.3. Identificación, Estimación, Diagnóstico, Selección del modelo y Pronóstico

La metodología de Box - Jenkins consiste en las siguientes etapas: Identificación, Estimación, Comprobación del diagnóstico, Selección del modelo y Pronóstico. En la Figura 3.2 se muestra un diagrama de flujo que representa los pasos que componen la metodología de Box - Jenkins, los cuales se describirán a continuación.

Una vez que se haya identificado el modelo tentativamente, lo siguiente es estimar los parámetros del modelo.

Considérese el proceso ARMA(p,q), es decir, que se realizará la estimación de los parámetros $\phi = (\phi_1, \phi_2, \dots, \phi_p)$, $\mu = E[X_t]$, $\theta = (\theta_1, \theta_2, \dots, \theta_q)$ y $\sigma^2 = E[\varepsilon_t^2]$ en el modelo

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (3.23)$$

donde X_t , $t = 1, \dots, n$ son n observaciones estacionarias y $\{\varepsilon_t\}$ es ruido blanco i.i.d $N(0, \sigma^2)$.

Además, en este capítulo se discutirán algunos de los métodos de estimación normalmente usados. Una vez que se realizó la estimación de los parámetros se verifica que el modelo se ajuste a la serie de tiempo, usualmente varios modelos representan la serie de tiempo dada. Por lo tanto, después de introducir la verificación de diagnóstico, se presentaran algunos criterios de selección del modelo para la representación de la serie de tiempo.

Y por último, una vez conseguido el modelo final, éste se usará para pronosticar los valores futuros de la serie temporal.

3.4. Identificación

La parte más complicada en el análisis de series de tiempo es la identificación y construcción del modelo. Por ello, es necesario que se tenga una buena comprensión del proceso, al igual que de su FAC (ρ_k) y FACP (ϕ_{kk}). En la práctica, la FAC y FACP son desconocidos, por lo que estos deben ser estimados por la FAC muestral ($\hat{\rho}_k$) y la FACP muestral ($\hat{\phi}_{kk}$), de manera que estos coincidan con los patrones de la FAC y la FACP para los modelos ARMA.

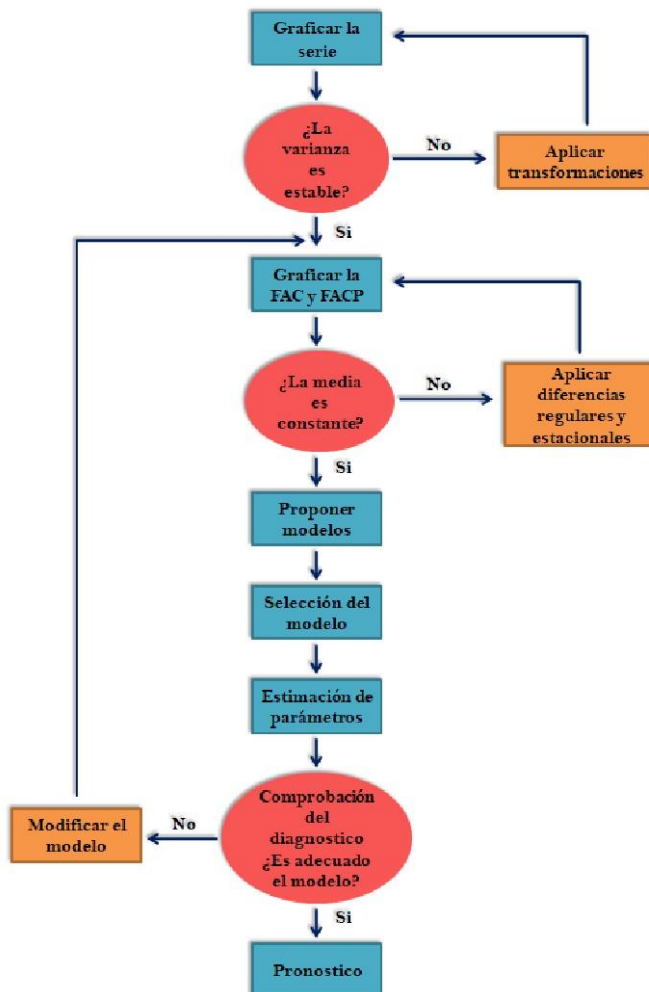


Figura 3.2: Diagrama de flujo, metodología de Box - Jenkins.

■ **Pasos para la identificación del modelo**

Para ilustrar el modelo de identificación, considere el modelo general ARIMA(p , d , q).

$$\phi_p(B)(1 - B)^d X_t = \delta + \theta_q(B)\varepsilon_t. \quad (3.24)$$

Cuando se habla de identificación del modelo, se refiere a la metodología para identificar las transformaciones requeridas, como transformaciones de estabilización de varianza, transformaciones de diferenciación, la decisión de incluir el parámetro determinista δ , cuando $d \geq 1$, y el orden de p y q .

Dada una serie de tiempo, los siguientes pasos son de utilidad para la identificación de un modelo tentativo. [31]

Paso 1

En algunos análisis de series de tiempo, el primer paso es el gráfico de los datos. A través del cual es posible tener una idea sobre los fenómenos que presenta la serie, ya sea: tendencia, estacionalidad, varianza no constante, valores atípicos, no normalidad y no estacionariedad.

Ahora bien, en las series de tiempo, las transformaciones usadas generalmente son las transformaciones para estabilizar a la varianza y la diferenciación para la media. Como se mencionó en el capítulo 2, debe aplicarse primero la transformación necesaria para estabilizar la varianza y después la diferenciación.

Paso 2

Calcular y examinar FAC muestral y la FACP muestral de la serie original para confirmar el grado de diferenciación necesaria para que las series diferenciadas sean estacionarias. Algunas reglas generales son las siguientes

1. Si la FAC muestral decae lentamente y la FACP muestral se corta después del lapso 1, entonces esto indica que es necesario hacer diferenciación, por ejemplo, puede tomarse la primera diferenciación $(1 - B)X_t$. Además es posible usar la prueba de la raíz unitaria, propuesta por Dickey-Fuller(1979) (ver [15]), pero la diferenciación es generalmente recomendado.
2. En general, para eliminar la no estacionariedad es posible que se tenga que considerar una diferenciación de orden superior $(1 - B)^d X_t$, $d > 1$. Normalmente d toma los valores de 0, 1 o 2. Obsérvese que si $(1 - B)^d X_t$ es

estacionaria, entonces $(1 - B)^{d+i}X_t$, $i = 1, 2, \dots$ también es estacionaria y como se mencionó en el capítulo 1, esto trae consecuencias pues se estaría cometiendo sobre diferenciación.

Paso 3

Calcular y examinar la FAC y FACP muestrales de la serie ya transformada y diferenciada, para identificar el orden de p y q , donde p es el orden del polinomio del proceso AR y q es orden del polinomio del proceso MA, normalmente el orden de p y de q es menor o igual a 3.

Existe una fuerte confusión entre los modelos AR y MA en términos de sus FAC y FACP. Si se desea identificar un modelo ARIMA, es necesario un número mínimo de observaciones $n = 50$, y el número de autocorrelaciones muestrales y parciales en un lapso k deben calcularse aproximadamente para $n/4$. Aunque ocasionalmente para datos de buena calidad uno puede identificar un modelo adecuado con un tamaño de muestra más pequeño.

Para identificar el orden de p y q , los patrones de la FAC y FACP muestrales deben coincidir con patrones teóricos del modelo.

La habilidad para identificar modelos para series de tiempo, es como si uno fuese agente del FBI y está en búsqueda de los criminales, en donde la mayoría de estos se disfrazan para evitar ser identificados, lo que es cierto para FAC y la FACP.

Paso 4

Inclusión del término independiente δ .

La media de un proceso estacionario ARMA(p, q), está directamente relacionada con la constante δ . Si $\delta = 0$, entonces la media del proceso es cero. Para determinar si se incluye el término independiente no nulo, se contrastará si la media de la serie estacionaria es cero o no, esto es,

$$H_0 : E[X_t] = 0 \quad vs \quad H_a : E[X_t] \neq 0$$

donde el estadístico de contraste es

$$t = \frac{\bar{X}}{\hat{\sigma}_{\bar{X}}} \sim t(T - d - 1)$$

donde $\hat{\sigma}_{\bar{X}}^2$ es el estimador de la varianza de la media muestral \bar{X} , el cual está dado por

$$\hat{\sigma}_{\bar{X}}^2 = \frac{C_0}{T - d} (1 + 2\hat{\rho}_1 + 2\hat{\rho}_2 + \dots + 2\hat{\rho}_n)$$

donde $C_0 = \sum (X_t - \bar{X})^2 / T - 1$ es la varianza muestral de la serie estacionaria y $(1 + 2\hat{\rho}_1 + 2\hat{\rho}_2 + \dots + 2\hat{\rho}_n)$ representa las n primeras autocorrelaciones muestrales significativas de $\{X_t\}$.

Para calcular la varianza se utiliza la aproximación

$$\hat{\sigma}_{\bar{X}}^2 \simeq \frac{C_0}{T - d}.$$

Y se rechazará la hipótesis nula $H_0 : E[X_t] = 0$ a un nivel de significancia α si $t > t_{\frac{\alpha}{2}}(T - d - 1)$, por lo que se incluirá la constante δ . [17]

3.5. Estimación

Una vez identificado el modelo, lo siguiente es la estimación de los parámetros de dicho modelo. Estos parámetros pueden ser estimados por el método de momentos, mínimos cuadrados o máxima verosimilitud. [31]

3.5.1. Estimación por Método de Momentos

El método de momentos consiste en la resolución de un sistema de ecuaciones, el cual se deriva de la sustitución de los momentos muestrales como la media muestral \bar{X} , la varianza muestral $\hat{\gamma}_0$ y los coeficientes de la FAC muestrales $\hat{\rho}_i$ por sus contrapartes teóricas. Para ser un poco más claros, considérese un proceso AR(p)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

donde la media $\mu = E[X_t]$ es estimada por la media muestral \bar{X} . Para estimar los parámetros $\phi_1, \phi_2, \dots, \phi_p$, se hace uso de su FAC(ρ_k), esto es, $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}$ para $k \geq 1$, y haciendo correr a k resulta el siguiente sistema de ecuaciones conocido como sistema de **Yule-Walker**

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1}. \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2}. \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p. \end{aligned} \tag{3.25}$$

Ahora bien, sustituyendo ρ_k por $\hat{\rho}_k$ y resolviendo el sistema de ecuaciones, se obtiene la estimación de los parámetros $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$. Esto es,

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \dots & \hat{\rho}_{p-2} & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{p-3} & \hat{\rho}_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \dots & \hat{\rho}_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}$$

Estos estimadores son usualmente llamados estimadores de Yule-Walker.

Una vez obtenidos los parámetros $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$, para obtener γ_0 se multiplica X_t por ambos lados de la ecuación del proceso y se obtiene la esperanza, esto es,

$$\begin{aligned} \gamma_0 = E[X_t X_t] &= \phi_1 E[X_t X_{t-1}] + \phi_2 E[X_t X_{t-2}] + \dots + \phi_p E[X_t X_{t-p}] + E[X_t \varepsilon_t] \\ &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + E[X_t \varepsilon_t]. \end{aligned}$$

Obsérvese que

$$\begin{aligned} E[X_t \varepsilon_t] &= E[(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t) \varepsilon_t] \\ &= \phi_1 E[X_{t-1} \varepsilon_t] + \phi_2 E[X_{t-2} \varepsilon_t] + \dots + \phi_p E[X_{t-p} \varepsilon_t] + E[\varepsilon_t \varepsilon_t] \\ &= \sigma_\varepsilon^2. \end{aligned}$$

Dado que X_t y ε_t no están correlacionados $E[X_{t-k} \varepsilon_t] = 0$ y $E[\varepsilon_t \varepsilon_t] = \sigma_\varepsilon^2$, así

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2. \quad (3.26)$$

Despejando σ_ε^2 de 3.26, se obtiene

$$\begin{aligned} \sigma_\varepsilon^2 &= \gamma_0 - \phi_1 \gamma_1 - \phi_2 \gamma_2 - \dots - \phi_p \gamma_p \\ &= \gamma_0 \left(1 - \phi_1 \frac{\gamma_1}{\gamma_0} - \phi_2 \frac{\gamma_2}{\gamma_0} - \dots - \phi_p \frac{\gamma_p}{\gamma_0} \right). \end{aligned}$$

Y dado que $\rho_k = \frac{\gamma_k}{\gamma_0}$, se tiene

$$\sigma_\varepsilon^2 = \gamma_0 (1 - \phi_1 \rho_1 - \phi_2 \rho_2 - \dots - \phi_p \rho_p). \quad (3.27)$$

Y sustituyendo por los estimadores, se obtiene el estimador para σ_ε^2

$$\hat{\sigma}_\varepsilon^2 = \hat{\gamma}_0 (1 - \phi_1 \hat{\rho}_1 - \phi_2 \hat{\rho}_2 - \dots - \phi_p \hat{\rho}_p). \quad (3.28)$$

3.5.2. Mínimos cuadrados

En el análisis de regresión es muy probable que se haga uso del método de mínimos cuadrados. Este método, también se puede usar para la obtención de los estimadores de los parámetros en el análisis de serie de tiempo.

Considere el modelo de regresión lineal simple

$$Z_t = \phi X_t + e_t, \quad t = 1, \dots, n. \quad (3.29)$$

Para el término de error e_t se consideran las siguientes suposiciones básicas

1. Media cero: $E[e_t] = 0$.
2. Varianza constante: $E[(e_t)^2] = \sigma_e^2$.
3. No hay autocorrelación: $E[e_t e_k] = 0$, para $t \neq k$.
4. No correlacionado con la variable X_t : $E[X_t e_t] = 0$.

Para obtener el estimador ϕ de mínimos cuadrados, es necesario minimizar

$$\sum_{t=1}^n (Z_t - \phi X_t)^2.$$

Para minimizar la ecuación anterior, se deriva con respecto a ϕ y se iguala a 0, esto es,

$$\begin{aligned} \frac{d(\sum_{t=1}^n (Z_t - \phi X_t)^2)}{d\phi} &= -2 \sum_{t=1}^n (Z_t - \phi X_t) X_t = 0 \\ &= \sum_{t=1}^n (Z_t X_t - \phi X_t X_t) = 0 \end{aligned}$$

así

$$\hat{\phi} = \frac{\sum_{t=1}^n X_t Z_t}{\sum_{t=1}^n X_t^2} \quad (3.30)$$

donde $\hat{\phi}$ es un estimador consistente y el mejor estimador linealmente insesgado de ϕ . Más aún, note que la suposición 4 es importante para este resultado, ya que

esta suposición continúa si las variables explicativas son no estocásticas. En un estudio no controlable, particularmente cuando los datos de series de tiempo son involucrados, las variables explicativas son usualmente variables aleatorias.

Ahora bien, considere el siguiente modelo de serie de tiempo

$$Z_t = \phi Z_{t-1} + e_t, \quad t = 1, \dots, n. \quad (3.31)$$

El estimador de mínimos cuadrados ordinarios de ϕ , basado en los datos disponibles es

$$\hat{\phi} = \frac{\sum_{t=2}^n Z_{t-1} Z_t}{\sum_{t=2}^n Z_{t-1}^2}. \quad (3.32)$$

Se desea saber si $\hat{\phi}$ continúa siendo insesgado y consistente, cuando la variable explicativa es una variable dependiente retrasada. Pues bien, esto dependerá de la naturaleza estocástica del término de error e_t . Para que esto sea un poco más visible, se reescribe $\hat{\phi}$ como

$$\begin{aligned} \hat{\phi} &= \frac{\sum_{t=2}^n Z_{t-1} Z_t}{\sum_{t=2}^n Z_{t-1}^2} = \frac{\sum_{t=2}^n Z_{t-1} (\phi Z_{t-1} + e_t)}{\sum_{t=2}^n Z_{t-1}^2} \\ &= \phi + \frac{\sum_{t=2}^n Z_{t-1} e_t}{\sum_{t=2}^n Z_{t-1}^2}. \end{aligned} \quad (3.33)$$

Y considere los siguientes casos.

- **Caso1:** $e_t = \varepsilon_t$. Donde e_t es un $RB(0, \sigma_\varepsilon^2)$. En este caso es fácil ver que $\hat{\phi}$ en (3.32) es equivalente al primer lapso de autocorrelación muestral $\hat{\rho}_1$ para la serie Z_t . Si $|\phi| < 1$ entonces X_t llega a ser un proceso AR(1) con una función de autocorrelación absolutamente sumable. Entonces $\hat{\rho}_1$ es un estimador consistente de ρ_1 , que es igual a ϕ . Así, $\hat{\phi}$ es un estimador asintóticamente insesgado y consistente de ϕ .
- **Caso2:** $e_t = (1 - \theta B)\varepsilon_t$, donde ε_t es $RB(0, \sigma_\varepsilon^2)$, por lo tanto e_t es un proceso de MA(1). Bajo estas condiciones la serie Z_t llega a ser un proceso ARMA(1,1)

$$Z_t = \phi Z_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \quad (3.34)$$

y

$$E[Z_{t-1} e_t] = E[Z_{t-1} (\varepsilon_t - \theta \varepsilon_{t-1})] = -\theta \sigma_\varepsilon^2. \quad (3.35)$$

Este resultado prueba que la autocorrelación en el término de error no sólo viola la suposición 3, sino también la suposición 4, cuando las variables

explicativas contienen una variable dependiente retrasada.

En este caso $\hat{\phi}$ no es un estimador consistente de ϕ , ya que $\hat{\phi} \simeq \hat{\rho}_1$ es un estimador consistente de ρ_1 , y para un proceso ARMA(1,1)

$$\phi_1 = \frac{(\phi - \theta)(1 - \phi\theta)}{1 + \theta^2 - 2\phi\theta} \neq \phi.$$

En resumen, el estimador de mínimos cuadrados ordinarios para el parámetro de una variable explicativa en un modelo de regresión será inconsistente a menos que el término de error esté no correlacionado con la variable explicativa. Para un modelo ARMA(p,q) esta condición generalmente no se cumple, excepto cuando $q = 0$.

3.5.3. Método de máxima verosimilitud

El método de estimación por máxima verosimilitud consiste en la elección de un estimador de un determinado parámetro que tiene la propiedad de maximizar el valor de la probabilidad de la muestra aleatoria observada. Es decir, el método de máxima verosimilitud consiste en hallar el valor del parámetro que maximiza la función de verosimilitud.

Definición 3.6. *La función de verosimilitud de n v.a's, (es decir la serie de tiempo), se define como la función de densidad de probabilidad conjunta de de las n v.a's, esto es, $f_{X_1, \dots, X_n}(x_1, \dots, x_n | \vartheta)$, que se considera como una función de ϑ . En particular si X_1, \dots, X_n es una m.a de una población con densidad $f(x; \vartheta)$, la función.*

$$L(x_1, \dots, x_n | \vartheta) = f(x_1, \dots, x_n | \vartheta) = \prod_{i=1}^n f(x_i | \vartheta) \quad (3.36)$$

se le llama **función de verosimilitud**.

Aquí ϑ puede ser un parámetro real o un vector de parámetros.

Como se desconoce el valor de ϑ , si se considera a $L(x_1, \dots, x_n | \vartheta)$ como una función de ϑ , entonces nos podemos hacer la siguiente pregunta: ¿para que valor de ϑ la función $L(x_1, \dots, x_n | \vartheta)$ tendrá su valor máximo, al considerar a (x_1, \dots, x_n) como un vector de valores fijos?

Nótese que, si $\hat{\vartheta}$ es el valor para el cual la función de verosimilitud se maximiza, entonces, $\hat{\vartheta}$ depende de (x_1, \dots, x_n) . Es decir

$$\hat{\vartheta} = \hat{\vartheta}(x_1, \dots, x_n),$$

para cada valor (x_1, \dots, x_n) de (X_1, \dots, X_n) .

Considerando lo anterior, se tiene la siguiente definición.

Definición 3.7. Sea (X_1, \dots, X_n) , una muestra aleatoria de una población con densidad $f(x; \vartheta)$, se dice que $\hat{\vartheta} = \hat{\vartheta}(x_1, \dots, x_n)$, el estimador de máxima verosimilitud (EMV) de ϑ si, $\hat{\vartheta}$ maximiza a $L(x_1, \dots, x_n | \vartheta)$.

Considérese un proceso ARMA(p,q),

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}.$$

Donde ε_t son ruidos blancos idénticamente distribuidos $N(0, \sigma_\varepsilon^2)$ y la densidad de probabilidad conjunta de $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ está dada por

$$P(\varepsilon | \phi, \mu, \theta, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-\frac{n}{2}} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n \varepsilon_t^2 \right] \quad (3.37)$$

en donde $\phi = (\phi_1, \phi_2, \dots, \phi_p)$, $\theta = (\theta_1, \theta_2, \dots, \theta_q)$. Despejando ε_t de el proceso ARMA(p,q), resulta

$$\varepsilon_t = \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}.$$

Se puede escribir la función de verosimilitud en términos de los parámetros $(\phi, \mu, \theta, \sigma_\varepsilon^2)$, esto es, $L(\phi, \mu, \theta, \sigma_\varepsilon^2)$. Ahora bien, aplicando el logaritmo natural a la función de verosimilitud condicional resulta

$$\ln[L(\phi, \mu, \theta, \sigma_\varepsilon^2)] = -\frac{n}{2} \ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n \varepsilon_t^2. \quad (3.38)$$

Sea $X = (X_1, X_2, \dots, X_n)'$ y sustituyendo las siguientes condiciones iniciales $X_* = (X_{1-p}, \dots, X_{-1}, X_0)'$ y $\varepsilon_* = (\varepsilon_{1-q}, \dots, \varepsilon_{-1}, \varepsilon_0)'$, en 3.38 resulta

$$\ln[L_*(\phi, \mu, \theta, \sigma_\varepsilon^2)] = -\frac{n}{2} \ln(2\pi\sigma_\varepsilon^2) - \frac{S_*(\phi, \mu, \theta)}{2\sigma_\varepsilon^2} \quad (3.39)$$

donde

$$S_*(\phi, \mu, \theta) = \sum_{t=1}^n \varepsilon_t^2(\phi, \mu, \theta | X_*, \varepsilon_*, X) \quad (3.40)$$

y $S_*(\phi, \mu, \theta)$ es la suma de cuadrados condicionales. En lo anterior, los subíndices estrella, en las funciones de verosimilitud y suma de cuadrados se utilizan para resaltar que están condicionados a la elección de los valores iniciales.

A los estimadores $\hat{\phi}, \hat{\mu}, \hat{\theta}$ los cuales maximizan a la función de verosimilitud, se llaman estimadores máxima verosimilitud condicional.

Obsérvese que la probabilidad condicional $\ln[L_*(\phi, \mu, \theta, \sigma_\varepsilon^2)]$ involucra a los datos sólo a través de la suma condicional de cuadrados $S_*(\phi, \mu, \theta)$, estos estimadores son los mismos que los estimadores de mínimos cuadrados condicionales obtenidos al minimizar la función de suma de cuadrados condicional S_* .

Con base en los supuestos de que $\{X_t\}$ es estacionario y $\{\varepsilon_t\}$ es una serie i.i.d $N(0, \sigma_\varepsilon^2)$, es posible reemplazar a X_t desconocida por la media muestral \bar{X} y ε_t desconocida por su valor esperado 0.

Para el modelo, también se puede asumir que $\varepsilon_p = \varepsilon_{p-1} = \dots = \varepsilon_{p+1-q} = 0$ y calcular ε_t para $t \geq p+1$. La función de suma de cuadrados condicional resulta

$$S_*(\phi, \mu, \theta) = \sum_{t=p+1}^n \varepsilon_t^2(\phi, \mu, \theta | X). \quad (3.41)$$

Después de obtenidas las estimaciones de los parámetros $\hat{\phi}, \hat{\mu}, \hat{\theta}$, el estimador $\hat{\sigma}_\varepsilon^2$ de σ_ε^2 resulta ser

$$\hat{\sigma}_\varepsilon^2 = \frac{S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})}{d.f.}, \quad (3.42)$$

donde *d.f.* es el número de grados de libertad, que es igual al número de términos usados en la suma de $S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})$ menos el número de parámetros estimados. Si se emplea 3.41 para calcular $\hat{\sigma}_\varepsilon^2$, entonces $d.f. = (n-p) - (p+q+1) = n - (2p+q+1)$.

3.5.4. Estimación de máxima verosimilitud incondicional

Probablemente, uno se pregunte si es posible retroceder el pronóstico. Como se mostró anteriormente, es necesario el uso de los valores $X_* = (X_{1-p}, \dots, X_{-1}, X_0)'$ y $\varepsilon_* = (\varepsilon_{1-q}, \dots, \varepsilon_{-1}, \varepsilon_0)'$ para el cálculo de la suma de cuadrados y la obtención

de la función de verosimilitud. La siguiente forma del modelo ARMA consiste en una representación para este pronóstico.

$$(1 - \phi_1 B - \dots - \phi_p B^p)X_t = (1 - \theta_1 B - \dots - \theta_q B^q)\varepsilon_t. \quad (3.43)$$

O bien, en forma de valores pasados

$$(1 - \phi_1 B - \dots - \phi_p B^p)X_t = (1 - \theta_1 B - \dots - \theta_q B^q)e_t, \quad (3.44)$$

donde $B^j X_t = X_{t-j}$. Debido a la estacionariedad (3.43) y (3.44) deben tener justamente la misma estructura de autocovarianza, esto implica que $\{e_t\}$ es una serie de ruidos blancos con media cero y varianza σ_e^2 .

Así, de la misma manera que se realizan los pronósticos de los valores futuros X_{n+j} para $j > 0$ en base a los valores (X_1, X_2, \dots, X_n) , también es posible usar el pronóstico de los valores pasados (3.44) X_j y calcular ε_j para $j \leq 0$ basado en los valores $(X_n, X_{n-1}, \dots, X_1)$.

La función de verosimilitud logarítmica incondicional es

$$\ln L(\phi, \mu, \theta, \sigma_\varepsilon^2) = -\frac{n}{2} \ln 2\pi\sigma_\varepsilon^2 - \frac{S(\phi, \mu, \theta)}{2\sigma_\varepsilon^2}, \quad (3.45)$$

en donde $\phi = (\phi_1, \phi_2, \dots, \phi_p)$, $\theta = (\theta_1, \theta_2, \dots, \theta_q)$. Y $S(\phi, \mu, \theta)$ es la función de suma de cuadrados incondicional dada por

$$S(\phi, \mu, \theta) = \sum_{t=-\infty}^n [E(\varepsilon_t | \phi, \mu, \theta, X)]^2. \quad (3.46)$$

Y $E[\varepsilon_t | \phi, \mu, \theta, X]$ es la esperanza incondicional de ε_t dado ϕ, μ, θ y X .

A los estimadores $\hat{\phi}, \hat{\mu}, \hat{\theta}$ que maximizan la función (3.45) son llamados estimadores de máxima verosimilitud incondicional.

Nuevamente, debido a que, $\ln L(\phi, \mu, \theta, \sigma_\varepsilon^2)$ involucra a los datos a través de $S(\phi, \mu, \theta)$, así, obtener los estimadores de máxima verosimilitud incondicional es equivalente a obtener los estimadores de mínimos cuadrados incondicionales, obtenidos al minimizar $S(\phi, \mu, \theta)$. Normalmente (3.46) es aproximado por

$$S(\phi, \mu, \theta) = \sum_{t=-M}^n [E(\varepsilon_t | \phi, \mu, \theta, X)]^2 \quad (3.47)$$

donde M es un entero suficientemente grande tal que

$$|E(X_t|\phi, \mu, \theta, X) - E(X_{t-1}|\phi, \mu, \theta, X)| < \xi$$

donde ξ es menor que cualquier valor pequeño arbitrario $t \leq -(M+1)$. Esto implica que $E(X_t|\phi, \mu, \theta, X) \simeq \mu$, por lo que $E(\varepsilon_{t-1}|\phi, \mu, \theta, X)$ es insignificante para $t \leq -(M+1)$.

Después de haber obtenido la estimación de los parámetros $\hat{\phi}, \hat{\mu}, \hat{\theta}$, se calcula el estimador de σ_ε^2 , el cuál está dado por

$$\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\phi}, \hat{\mu}, \hat{\theta})}{n}. \quad (3.48)$$

3.5.5. Función exacta de verosimilitud

Como se observó anteriormente, las funciones de verosimilitud condicional e incondicional son aproximaciones. Para ilustrar la función de verosimilitud exacta para una serie de tiempo, considere el proceso AR(1)

$$X_t = \phi_1 X_{t-1} + \varepsilon_t \quad (3.49)$$

donde $|\phi_1| < 1$ y ε_t son i.i.d $N(0, \sigma_\varepsilon^2)$. Ahora, dados X_1, X_2, \dots, X_n la función de verosimilitud esta dada por

$$L(\phi_1, \mu, \sigma_\varepsilon^2) = f(X_1, X_2, \dots, X_n | \phi_1, \mu, \sigma_\varepsilon^2). \quad (3.50)$$

En el caso de el modelo AR(1), la función de verosimilitud resulta ser

$$L(\phi_1, \mu, \sigma_\varepsilon^2) = f(X_1) \cdot f(X_2|X_1) \cdots f(X_n|X_{n-1}). \quad (3.51)$$

Obsérvese que se ha eliminado los parámetros de las densidades, esto para facilitar la notación. Luego, reescribiendo al proceso como un proceso de media móvil, resulta

$$X_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}. \quad (3.52)$$

Donde X_t está distribuido como $N(0, \sigma_\varepsilon^2/(1-\phi_1^2))$ (ver capítulo 1, modelo AR(1)). Sin embargo X_t está altamente correlacionada.

Para determinar la función de verosimilitud, considere lo siguiente

$$\begin{aligned}
X_1 &= \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{1-j}; \\
\varepsilon_2 &= X_2 - \phi_1 X_1; \\
\varepsilon_3 &= X_3 - \phi_1 X_2; \\
&\vdots \\
\varepsilon_t &= X_t - \phi_1 X_{t-1}; \\
&\vdots \\
\varepsilon_n &= X_n - \phi_1 X_{n-1}.
\end{aligned} \tag{3.53}$$

donde X_1 sigue una distribución $N(0, \sigma_\varepsilon^2 / (1 - \phi_1^2))$, y ε_t para $2 \leq t \leq n$ sigue una distribución $N(0, \sigma_\varepsilon^2)$, y son independientes uno del otro. Esto indicaría que $f(X_t | X_{t-1}) = f_\varepsilon(X_t - \phi_1 X_{t-1})$, donde f_ε es la función de densidad de ε_t , que es también la función de densidad $N(0, \sigma_\varepsilon^2)$. De esta manera, es posible escribir la función de verosimilitud como

$$L(\phi_1, \mu, \sigma_\varepsilon^2) = f(X_1) \prod_{t=2}^n f_\varepsilon(X_t - \phi_1 X_{t-1}). \tag{3.54}$$

Así, la función de de verosimilitud resulta

$$\begin{aligned}
L(\phi_1, \mu, \sigma_\varepsilon^2) &= \left(\frac{1 - \phi_1^2}{2\pi\sigma_\varepsilon^2} \right)^{\frac{1}{2}} \exp\left(\frac{-X_1^2(1 - \phi_1^2)}{2\sigma_\varepsilon^2} \right) \prod_{t=2}^n \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{1}{2}} \\
&\quad \exp\left(\frac{-(X_t - \phi_1 X_{t-1})^2}{2\sigma_\varepsilon^2} \right) \\
&= \left(\frac{1 - \phi_1^2}{2\pi\sigma_\varepsilon^2} \right)^{\frac{1}{2}} \exp\left(\frac{-X_1^2(1 - \phi_1^2)}{2\sigma_\varepsilon^2} \right) \cdot \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{n-1}{2}} \\
&\quad \exp\left(\frac{-\sum_{t=2}^n (X_t - \phi_1 X_{t-1})^2}{2\sigma_\varepsilon^2} \right) \\
&= (1 - \phi_1^2)^{\frac{1}{2}} (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2}} \exp\left(\frac{-X_1^2(1 - \phi_1^2)}{2\sigma_\varepsilon^2} \right) \cdot (2\pi\sigma_\varepsilon^2)^{-\frac{(n-1)}{2}} \\
&\quad \exp\left(\frac{-\sum_{t=2}^n (X_t - \phi_1 X_{t-1})^2}{2\sigma_\varepsilon^2} \right) \\
&= (1 - \phi_1^2)^{\frac{1}{2}} (2\pi\sigma_\varepsilon^2)^{-\frac{n}{2}} \exp\left(-\frac{S(\phi_1, \mu)}{2\sigma_\varepsilon^2} \right)
\end{aligned} \tag{3.55}$$

donde

$$S(\phi_1, \mu) = X_1^2(1 - \phi_1^2) + \sum_{t=2}^n (X_t - \phi_1 X_{t-1})^2, \quad (3.56)$$

$S(\phi_1, \mu)$ es la suma de los términos cuadrados que está en función solamente de ϕ_1 y μ , que es la suma de cuadrados no condicional.

Luego, tomando el logaritmo de 3.55, resulta

$$\ln L(\phi_1, \mu, \sigma_\varepsilon^2) = \frac{1}{2} \ln(1 - \phi_1^2) - \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma_\varepsilon^2 - \frac{S(\phi_1, \mu)}{2\sigma_\varepsilon^2}. \quad (3.57)$$

Derivando parcialmente con respecto a σ_ε^2 , se obtiene

$$\frac{\partial \ln L(\phi_1, \mu, \sigma_\varepsilon^2)}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{S(\phi_1, \mu)}{2(\sigma_\varepsilon^2)^2}.$$

Igualando a cero para hallar el máximo

$$\begin{aligned} -\frac{n}{2\sigma_\varepsilon^2} + \frac{S(\phi_1, \mu)}{2(\sigma_\varepsilon^2)^2} &= 0 \\ \frac{-n\sigma_\varepsilon^2 + S(\phi_1, \mu)}{2(\sigma_\varepsilon^2)^2} &= 0 \\ \sigma_\varepsilon^2 &= \frac{S(\phi_1, \mu)}{n} \end{aligned}$$

el cual es el valor que maximiza a la función de verosimilitud. Así

$$\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\phi}_1, \hat{\mu})}{n}. \quad (3.58)$$

que es el estimador de máxima verosimilitud de σ_ε^2 , y donde $\hat{\phi}_1, \hat{\mu}$ son estimadores de máxima verosimilitud de ϕ_1 y μ .

3.6. Diagnóstico

Una vez realizada la identificación del modelo y la estimación de parámetros, sigue el diagnóstico del modelo, es decir, verificar la bondad del ajuste del modelo.

La suposición básica es que los $\{\varepsilon_t\}$ son ruido blanco (media cero, varianza constante y están no correlacionados).

Para cualquier modelo estimado, los residuales $\hat{\varepsilon}_t$'s son estimaciones de ruido blanco ε_t 's no observados. Por lo que, la comprobación del modelo de diagnóstico se logra a través de un análisis de la serie residual $\{\hat{\varepsilon}_t\}$. Ya que esta serie residual es el producto de la estimación de los parámetros, la comprobación del diagnóstico del modelo usualmente está contenida en la fase de estimación de un paquete de series de tiempo.

Para verificar si los errores se distribuyen normalmente, es posible construir un histograma de los residuos estandarizados $\hat{\varepsilon}_t/\hat{\sigma}_\varepsilon$ y compararlo con la distribución normal estándar utilizando la prueba de bondad de ajuste Chi-cuadrada ¹ o el resumen de los cinco números de Tukey.

Para verificar si la varianza es constante, se puede examinar la gráfica de residuos. Es probable que se consideren diferentes formas de los gráficos de residuos, esto es,

- Una gráfica de residuos en forma de abanico, indica que los términos de error se están dispersando más cuando el valor horizontal de la gráfica se incrementa y se transgrede la suposición de varianza constante. Es decir, que existe una varianza de error creciente.
- Una gráfica en forma de embudo, indica que la dispersión de los términos de error decrece cuando el valor horizontal de la gráfica se incrementa y nuevamente se transgrede la suposición de varianza constante.
- Y una gráfica de residuos en forma de banda horizontal, indica que la dispersión de los términos de error con respecto a cero no cambia mucho cuando el valor de la gráfica horizontal aumenta. Esto indica que la suposición de la varianza constante se conserva. Ver Figura 3.3.

¹La prueba de bondad de ajuste se emplea para decidir cuando un conjunto de datos de apega a una distribución de probabilidad dada.

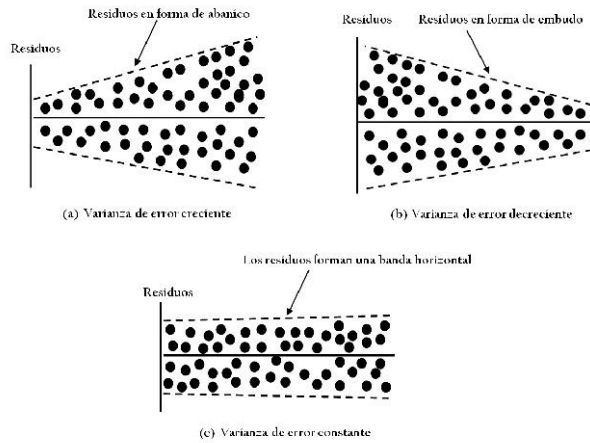


Figura 3.3: Gráficas de residuos y la suposición de varianza constante.

Y para verificar si los residuos son aproximadamente ruido blanco, se calcula la FAC y FACP muestral de los residuos, esto con la finalidad de verificar que los residuos no muestren ningún patrón que revele correlación, esto es, que no sean estadísticamente significantes.

Así, otra prueba es realizar un contraste de significancia sobre los coeficientes de la FAC, teniendo como hipótesis nula

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

vs

$$H_a : \text{al menos algún } \rho_k \text{ es diferente de cero}$$

con el estadístico

$$Q = n(n+2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{(n-k)} \quad (3.59)$$

donde Q es también conocido como estadístico de **Ljung-Box(1978)** (ver [21]), y **Ansley-Newbold(1979b)** (ver [6]), en dichos artículos probaron que bajo la hipótesis nula ésta estadística sigue aproximadamente una distribución chi-cuadrado

χ^2 con $(K - m)$ grados de libertad, donde $m = p + q$ es el número total de parámetros estimados. [31] [8]

3.7. Selección del modelo

En el análisis de series de tiempo o mejor dicho en cualquier análisis de datos, existen varios modelos que pueden representar adecuadamente a un conjunto de datos. Para seleccionar un modelo que mejor se ajuste a los datos se hace uso de los criterios de comparación o selección. Estos criterios pueden consultarse en el Apéndice (D).

3.8. Pronóstico

Uno de los objetivos más importantes en el análisis de series de tiempo es generar pronósticos óptimos para las observaciones futuras, esto es, sea t el valor actual de la serie de tiempo denotada por X_t y se desea pronosticar los valores futuros X_{t+l} , (donde l representa un periodo posterior al actual) a partir de las observaciones previas X_t, X_{t-1}, \dots , y se denotará por $\hat{X}_t(l)$ a la estimación de X_{t+l} .

El objetivo es generar pronósticos óptimos de manera que tengan los mínimos errores posibles. El criterio usado para la obtención de pronósticos óptimos es el error cuadrado medio, para el cual se espera que los valores de los errores cuadrados medios

$$E[(X_{t+l} - \hat{X}_t(l))^2] = E[e_t(l)^2], \quad (3.60)$$

sean minimizados. [31]

Pronóstico para proceso estacionario ARMA(p,q)

Considere el proceso ARMA(p,q)

$$\phi_p(B)X_t = \theta_q(B)\varepsilon_t \quad (3.61)$$

donde $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ y $\varepsilon_t \sim RB(0, \sigma^2)$.

Dado que el modelo es estacionario e invertible, se tiene que

$$X_t = \psi(B)\varepsilon_t \quad (3.62)$$

donde $\psi(B) = \frac{\theta_q(B)}{\phi_p(B)} = (1 + \psi_1 B + \psi_2 B^2 + \dots) = \sum_{j=0}^{\infty} \psi_j B^j$, con $\psi_0 = 1$. Así, (3.62) se puede reescribir como

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \quad (3.63)$$

Ahora bien, considere al proceso a un tiempo $t+l$, entonces

$$X_{t+l} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t+l-j} = \sum_{j=0}^{l-1} \psi_j \varepsilon_{t+l-j} + \sum_{j=l}^{\infty} \psi_j \varepsilon_{t+l-j}. \quad (3.64)$$

Luego, suponga que se encuentra en el tiempo t , uno desearía hacer un pronóstico $\hat{X}_t(l)$ de X_{t+l} que es una combinación lineal de las observaciones X_t, X_{t-1}, \dots . Dado que X_t , puede ser escrito de la forma (3.63), podemos dejar al pronóstico del error cuadrado medio $\hat{X}_t(l)$ de X_{t+l} como

$$\hat{X}_t(l) = \psi_l^* \varepsilon_t + \psi_{l+1}^* \varepsilon_{t-1} + \psi_{l+2}^* \varepsilon_{t-2} + \dots \quad (3.65)$$

Pero ¿Cómo saber el valor de ψ_j^* ? Pues bien, el criterio para hallar los valores de ψ_j^* es

$$\min_{\psi_j^*} E[(X_{t+l} - \hat{X}_t(l))^2], \quad (3.66)$$

lo que resulta ser

$$E[(X_{t+l} - \hat{X}_t(l))^2] = E[(\varepsilon_{t+l} + \psi_1 \varepsilon_{t+l-1} + \psi_2 \varepsilon_{t+l-2} + \dots + \psi_{l-1} \varepsilon_{t+1} + \psi_l \varepsilon_t + \psi_{l+1} \varepsilon_{t-1} + \psi_{l+2} \varepsilon_{t-2} + \dots) - (\psi_l^* \varepsilon_t + \psi_{l+1}^* \varepsilon_{t-1} + \psi_{l+2}^* \varepsilon_{t-2} + \dots)]^2]. \quad (3.67)$$

Agrupando términos en común

$$E[(\varepsilon_{t+l} + \psi_1 \varepsilon_{t+l-1} + \psi_2 \varepsilon_{t+l-2} + \dots + \psi_{l-1} \varepsilon_{t+1} + (\psi_l - \psi_l^*) \varepsilon_t + (\psi_{l+1} - \psi_{l+1}^*) \varepsilon_{t-1} + (\psi_{l+2} - \psi_{l+2}^*) \varepsilon_{t-2} + \dots)^2].$$

Desarrollando los cuadrados y tomando el valor esperado de cada término se obtiene

$$E[\varepsilon_{t+l}^2] + \psi_1^2 E[\varepsilon_{t+l-1}^2] + \psi_2^2 E[\varepsilon_{t+l-2}^2] + \dots + \psi_{l-1}^2 E[\varepsilon_{t+1}^2] + (\psi_l - \psi_l^*)^2 E[\varepsilon_t^2] + (\psi_{l+1} - \psi_{l+1}^*)^2 E[\varepsilon_{t-1}^2] + (\psi_{l+2} - \psi_{l+2}^*)^2 E[\varepsilon_{t-2}^2] + \dots$$

ya que

$$E[\varepsilon_t \varepsilon_{t-k}] = \begin{cases} 0, & \text{si } k \neq 0; \\ \sigma^2, & \text{si } k = 0. \end{cases}$$

Así la ecuación 3.67 se simplifica a

$$\begin{aligned} E[(X_{t+l} - \hat{X}_t(l))^2] &= \sigma^2 + \psi_1^2 \sigma^2 + \psi_2^2 \sigma^2 + \dots + \psi_{l-1}^2 \sigma^2 + (\psi_l - \psi_l^*)^2 \sigma^2 + \\ &\quad (\psi_{l+1} - \psi_{l+1}^*)^2 \sigma^2 + (\psi_{l+2} - \psi_{l+2}^*)^2 \sigma^2 + \dots \\ &= \sigma^2 \sum_{j=0}^{l-1} \psi_j^2 + \sigma^2 \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2. \end{aligned}$$

Luego,

$$\begin{aligned} \frac{\partial E[(X_{t+l} - \hat{X}_t(l))^2]}{\partial \psi^*} &= \sigma^2 \sum_{j=0}^{\infty} 2(\psi_{l+j} - \psi_{l+j}^*)(-1) = 0 \\ &= -2\sigma^2 \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*) = 0 \Leftrightarrow \psi_{l+j} = \psi_{l+j}^*. \end{aligned}$$

En consecuencia, la ecuación (3.65) resulta ser de la forma

$$\hat{X}_t(l) = \psi_l \varepsilon_t + \psi_{l+1} \varepsilon_{t-1} + \psi_{l+2} \varepsilon_{t-2} + \dots \quad (3.68)$$

Considere la ecuación (3.64)

$$X_{t+l} = \varepsilon_{t+l} + \psi_1 \varepsilon_{t+l-1} + \dots + \psi_{l-1} \varepsilon_{t+1} + \psi_l \varepsilon_t + \psi_{l+1} \varepsilon_{t-1} + \psi_{l+2} \varepsilon_{t-2} + \dots$$

Y tomando la esperanza condicional de la ecuación anterior se tiene

$$\begin{aligned} E_t[X_{t+l}] &= E[X_{t+l} | X_t, X_{t-1}, \dots] = E[\varepsilon_{t+l} | X_t, X_{t-1}, \dots] + \psi_1 E[\varepsilon_{t+l-1} | X_t, X_{t-1}, \dots] + \dots \\ &= \psi_l \varepsilon_t + \psi_{l+1} \varepsilon_{t-1} + \psi_{l+2} \varepsilon_{t-2} + \dots \end{aligned}$$

ya que

$$E[\varepsilon_{t+j} | X_t, X_{t-1}, \dots] = \begin{cases} 0, & \text{si } j > 0; \\ \varepsilon_{t+j}, & \text{si } j \leq 0. \end{cases}$$

Lo anterior es posible ya que en el Apéndice A, se prueba que ε_t puede ser expresado como combinación lineal de $X_t, X_{t-1}, X_{t-2}, \dots$

Obsérvese que

$$\hat{X}_t(l) = E[X_{t+l} | X_t, X_{t-1}, \dots].$$

Así, el pronóstico del mínimo error cuadrado medio, $\hat{X}_{t+l}(l)$, no es más que la esperanza condicional de todas las observaciones al tiempo t .

Una vez conocidos X_{t+l} y $\hat{X}_t(l)$, se tiene que

$$e_t(l) = X_{t+l} - \hat{X}_t(l) = \sum_{j=0}^{l-1} \psi_j \varepsilon_{t+l-j}, \quad \psi_0 = 1,$$

$e_t(l)$ es el error de pronóstico de mínimo error cuadrado medio $\hat{X}_t(l)$.

Como $e_t(l)$ es combinación lineal de ruidos blancos, se tiene

$$E[e_t(l)] = E[\varepsilon_{t+l} + \psi_1 \varepsilon_{t+l-1} + \psi_2 \varepsilon_{t+l-2} + \cdots + \psi_{l-1} \varepsilon_{t+1}] = 0,$$

donde $E[\varepsilon_{t+j}] = 0$, $j > 0$.

La varianza del error de pronóstico es

$$\begin{aligned} \text{Var}[e_t(l)] &= E[(e_t(l))^2] = E[(\varepsilon_{t+l} + \psi_1 \varepsilon_{t+l-1} + \psi_2 \varepsilon_{t+l-2} + \cdots + \psi_{l-1} \varepsilon_{t+1})^2] \\ &= \sigma^2 \sum_{j=0}^{l-1} \psi_j^2. \end{aligned}$$

Los errores de pronóstico para un lapso adelante no están correlacionados. Esto es fácil de ver ya que

$$e_t(1) = X_{t+1} - \hat{X}_t(1) = \varepsilon_{t+1}.$$

Pues los ruidos blancos son independientes y, por lo tanto, no están correlacionados, los errores de pronóstico de un lapso adelante tampoco deben estar correlacionados.

Los errores de pronóstico $e_t(l)$ para $l > 1$ están correlacionados.

Y por último, la función lineal de pronóstico de mínimo error cuadrado medio es también un pronóstico de mínimo error cuadrado medio de la función lineal correspondiente de las observaciones futuras. Por ejemplo, suponga que $\hat{X}_t(1)$, $\hat{X}_t(2)$, $\hat{X}_t(3)$ y $\hat{X}_t(4)$ son pronósticos de mínimo error cuadrado medio. Entonces, $10\hat{X}_t(1) + 8\hat{X}_t(2) + 6\hat{X}_t(3) + 4\hat{X}_t(4)$ es un pronóstico de mínimo error

cuadrado medio de $10X_{t+1} + 8X_{t+2} + 6X_{t+3} + 4X_{t+4}$.

Por otro lado, el intervalo de confianza para X_{t+l} , se construirá a partir de la distribución del error de predicción, bajo el supuesto de $\varepsilon_t \sim RBN(0, \sigma^2)$, entonces los errores de pronóstico, $e_t(l)$, también se distribuyen $N(0, Var[e_t(l)])$. Es decir

$$e_t(l) = X_{t+l} - \hat{X}_t(l) \sim N(0, Var[e_t(l)]).$$

Estandarizando se obtiene

$$\frac{X_{t+l} - \hat{X}_t(l) - 0}{\sqrt{Var[e_t(l)]}} \sim N(0, 1).$$

Por lo tanto

$$P \left[-N_{\frac{\alpha}{2}} < \frac{X_{t+l} - \hat{X}_t(l)}{\sqrt{Var[e_t(l)]}} < N_{\frac{\alpha}{2}} \right] = 1 - \alpha.$$

Haciendo los calculos necesarios se obtiene

$$P \left[\hat{X}_t(l) - N_{\frac{\alpha}{2}} \sqrt{Var[e_t(l)]} < X_{t+l} < \hat{X}_t(l) + N_{\frac{\alpha}{2}} \sqrt{Var[e_t(l)]} \right] = 1 - \alpha,$$

donde $N_{\frac{\alpha}{2}}$ es tal que $P[N > N_{\frac{\alpha}{2}}] = \frac{\alpha}{2}$, $N \sim N(0, 1)$.

Por lo tanto, el intervalo de confianza de nivel $100(1 - \alpha)$ por ciento para X_{t+l} es

$$\left[\hat{X}_t(l) - N_{\frac{\alpha}{2}} \sqrt{Var[e_t(l)]}, \hat{X}_t(l) + N_{\frac{\alpha}{2}} \sqrt{Var[e_t(l)]} \right].$$

Reglas de pronóstico

1. $E_t[X_{t-j}] = X_{t-j}$, $j = 0, 1, 2, \dots$

Debido a que se conoce una observación en o antes del tiempo t , la esperanza condicional de este valor conocido o constante es simplemente la observación de la misma.

2. $E_t[X_{t+j}] = \hat{X}_t(j)$, $j = 1, 2, \dots$

La esperanza condicional de un valor de la serie de tiempo después del tiempo t , a saber $t + j$, es el pronóstico de mínimo error cuadrado medio que se desea calcular para el tiempo j de origen t .

3. $E_t[\varepsilon_{t-j}] = \varepsilon_{t-j}, j = 0, 1, 2, \dots$

Debido a que se conoce el error en el tiempo t o antes del tiempo t , la esperanza condicional de este valor conocido es simplemente el error del mismo.

4. $E_t[\varepsilon_{t+j}] = 0, j = 1, 2, \dots$

Los supuestos de ε_t 's es que se distribuye i.i.d y $N(0, \sigma^2)$. En consecuencia, el valor esperado de ε_t desconocido después del tiempo t , a saber $t + j$ es cero porque todavía no han tenido lugar. [20]

En las tablas (3.2) y (3.3), se observan los pronósticos para los proceso AR y MA.

$X_t = \phi_1 \hat{X}_{t-1} + \varepsilon_t$ $X_{t+l} = \phi_1 X_{t+l-1} + \varepsilon_{t+l}$													
Proceso AR(1)	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">$l = 1$</td> <td style="text-align: center;">$\hat{X}_{t+1} = \phi_1 \hat{X}_t + \varepsilon_{t+1}$</td> <td style="text-align: center;">l</td> <td style="text-align: center;">Ecuación de pronóstico para \hat{X}_{t+l}:</td> </tr> <tr> <td style="text-align: center;">$l = 2$</td> <td style="text-align: center;">$\hat{X}_{t+2} = \phi_1 \hat{X}_{t+1} + \varepsilon_{t+2}$</td> <td style="text-align: center;">l</td> <td style="text-align: center;">Ecuación de pronóstico para \hat{X}_{t+l}:</td> </tr> <tr> <td colspan="2" style="text-align: center;">Ecuación de pronóstico $\hat{X}_t(1)$: $E_t[X_{t+1}] = \phi_1 E_t[X_t] + E_t[\varepsilon_{t+1}]$ $= \phi_1 \hat{X}_t$</td> <td colspan="2" style="text-align: center;">Ecuación de pronóstico $\hat{X}_t(2)$: $E_t[X_{t+2}] = \phi_1 E_t[X_{t+1}] + E_t[\varepsilon_{t+2}]$ $= \phi_1 \hat{X}_t(1)$ $= \phi_1^2 \hat{X}_t$</td> </tr> </table>	$l = 1$	$\hat{X}_{t+1} = \phi_1 \hat{X}_t + \varepsilon_{t+1}$	l	Ecuación de pronóstico para \hat{X}_{t+l}:	$l = 2$	$\hat{X}_{t+2} = \phi_1 \hat{X}_{t+1} + \varepsilon_{t+2}$	l	Ecuación de pronóstico para \hat{X}_{t+l}:	Ecuación de pronóstico $\hat{X}_t(1)$: $E_t[X_{t+1}] = \phi_1 E_t[X_t] + E_t[\varepsilon_{t+1}]$ $= \phi_1 \hat{X}_t$		Ecuación de pronóstico $\hat{X}_t(2)$: $E_t[X_{t+2}] = \phi_1 E_t[X_{t+1}] + E_t[\varepsilon_{t+2}]$ $= \phi_1 \hat{X}_t(1)$ $= \phi_1^2 \hat{X}_t$	
$l = 1$	$\hat{X}_{t+1} = \phi_1 \hat{X}_t + \varepsilon_{t+1}$	l	Ecuación de pronóstico para \hat{X}_{t+l}:										
$l = 2$	$\hat{X}_{t+2} = \phi_1 \hat{X}_{t+1} + \varepsilon_{t+2}$	l	Ecuación de pronóstico para \hat{X}_{t+l}:										
Ecuación de pronóstico $\hat{X}_t(1)$: $E_t[X_{t+1}] = \phi_1 E_t[X_t] + E_t[\varepsilon_{t+1}]$ $= \phi_1 \hat{X}_t$		Ecuación de pronóstico $\hat{X}_t(2)$: $E_t[X_{t+2}] = \phi_1 E_t[X_{t+1}] + E_t[\varepsilon_{t+2}]$ $= \phi_1 \hat{X}_t(1)$ $= \phi_1^2 \hat{X}_t$											
Proceso AR(p)	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td colspan="4" style="text-align: center;">$X_t = \phi_1 \hat{X}_{t-1} + \phi_2 \hat{X}_{t-2} + \dots + \phi_p \hat{X}_{t-p} + \varepsilon_t$ $X_{t+l} = \phi_1 \hat{X}_{t+l-1} + \phi_2 \hat{X}_{t+l-2} + \dots + \phi_p \hat{X}_{t+l-p} + \varepsilon_{t+l}$</td> </tr> <tr> <td style="text-align: center;">$l = 1$</td> <td style="text-align: center;">$X_{t+1} = \phi_1 \hat{X}_t + \phi_2 \hat{X}_{t-1} + \dots + \phi_p \hat{X}_{t-p+1} + \varepsilon_{t+1}$</td> <td style="text-align: center;">l</td> <td style="text-align: center;">Ecuación de pronóstico para \hat{X}_{t+l}:</td> </tr> <tr> <td colspan="2" style="text-align: center;">Ecuación de pronóstico $\hat{X}_t(1)$: $E_t[X_{t+1}] = \phi_1 E_t[X_t] + \phi_2 E_t[X_{t-1}] + \dots + \phi_p E_t[X_{t-p+1}] + E_t[\varepsilon_{t+1}]$ $= \phi_1 \hat{X}_t + \phi_2 \hat{X}_{t-1} + \dots + \phi_p \hat{X}_{t-p+1}$</td> <td colspan="2" style="text-align: center;">Ecuación de pronóstico para \hat{X}_{t+l}: $\hat{X}_t(l) = \phi_1 \hat{X}_t(l-1) + \phi_2 \hat{X}_t(l-2) + \dots + \phi_p \hat{X}_t(l-p)$</td> </tr> </table>	$X_t = \phi_1 \hat{X}_{t-1} + \phi_2 \hat{X}_{t-2} + \dots + \phi_p \hat{X}_{t-p} + \varepsilon_t$ $X_{t+l} = \phi_1 \hat{X}_{t+l-1} + \phi_2 \hat{X}_{t+l-2} + \dots + \phi_p \hat{X}_{t+l-p} + \varepsilon_{t+l}$				$l = 1$	$X_{t+1} = \phi_1 \hat{X}_t + \phi_2 \hat{X}_{t-1} + \dots + \phi_p \hat{X}_{t-p+1} + \varepsilon_{t+1}$	l	Ecuación de pronóstico para \hat{X}_{t+l}:	Ecuación de pronóstico $\hat{X}_t(1)$: $E_t[X_{t+1}] = \phi_1 E_t[X_t] + \phi_2 E_t[X_{t-1}] + \dots + \phi_p E_t[X_{t-p+1}] + E_t[\varepsilon_{t+1}]$ $= \phi_1 \hat{X}_t + \phi_2 \hat{X}_{t-1} + \dots + \phi_p \hat{X}_{t-p+1}$		Ecuación de pronóstico para \hat{X}_{t+l}: $\hat{X}_t(l) = \phi_1 \hat{X}_t(l-1) + \phi_2 \hat{X}_t(l-2) + \dots + \phi_p \hat{X}_t(l-p)$	
$X_t = \phi_1 \hat{X}_{t-1} + \phi_2 \hat{X}_{t-2} + \dots + \phi_p \hat{X}_{t-p} + \varepsilon_t$ $X_{t+l} = \phi_1 \hat{X}_{t+l-1} + \phi_2 \hat{X}_{t+l-2} + \dots + \phi_p \hat{X}_{t+l-p} + \varepsilon_{t+l}$													
$l = 1$	$X_{t+1} = \phi_1 \hat{X}_t + \phi_2 \hat{X}_{t-1} + \dots + \phi_p \hat{X}_{t-p+1} + \varepsilon_{t+1}$	l	Ecuación de pronóstico para \hat{X}_{t+l}:										
Ecuación de pronóstico $\hat{X}_t(1)$: $E_t[X_{t+1}] = \phi_1 E_t[X_t] + \phi_2 E_t[X_{t-1}] + \dots + \phi_p E_t[X_{t-p+1}] + E_t[\varepsilon_{t+1}]$ $= \phi_1 \hat{X}_t + \phi_2 \hat{X}_{t-1} + \dots + \phi_p \hat{X}_{t-p+1}$		Ecuación de pronóstico para \hat{X}_{t+l}: $\hat{X}_t(l) = \phi_1 \hat{X}_t(l-1) + \phi_2 \hat{X}_t(l-2) + \dots + \phi_p \hat{X}_t(l-p)$											

Tabla 3.2: Pronósticos de los procesos AR(1) y AR(p).

$\hat{X}_l = \varepsilon_l - \theta_1 \varepsilon_{l-1}$ $\hat{X}_{l+1} = \varepsilon_{l+1} - \theta_1 \varepsilon_{l+1-1}$	
$l = 1$ $X_{l+1} = \varepsilon_{l+1} - \theta_1 \varepsilon_l$	$l = 2$ $X_{l+2} = \varepsilon_{l+2} - \theta_1 \varepsilon_{l+1}$
Ecuación de pronóstico $\hat{X}_l(1)$: $E_l[X_{l+1}] = E_l[\varepsilon_{l+1}] - \theta_1 E_l[\varepsilon_l]$ $= -\theta_1 \varepsilon_l$	Ecuación de pronóstico $\hat{X}_l(2)$: $E_l[X_{l+2}] = E_l[\varepsilon_{l+2}] - \theta_1 E_l[\varepsilon_{l+1}]$ $= 0$
Ecuación de pronóstico para X_{l+i} : $\hat{X}_l(i) = E_l[X_{l+i}] = 0$	
Proceso MA(1)	
$X_l = \varepsilon_l - \theta_1 \varepsilon_{l-1} - \theta_2 \varepsilon_{l-2} - \dots - \theta_q \varepsilon_{l-q}$ $\hat{X}_{l+i} = \varepsilon_{l+i} - \theta_1 \varepsilon_{l+i-1} - \theta_2 \varepsilon_{l+i-2} - \dots - \theta_q \varepsilon_{l+i-q}$	
$l = 1$ $X_{l+1} = \varepsilon_{l+1} - \theta_1 \varepsilon_l - \theta_2 \varepsilon_{l-1} - \dots - \theta_q \varepsilon_{l-q+1}$	$l = 2$ $X_{l+2} = \varepsilon_{l+2} - \theta_1 \varepsilon_{l+1} - \theta_2 \varepsilon_l - \dots - \theta_q \varepsilon_{l-q+2}$
Ecuación de pronóstico $\hat{X}_l(1)$: $E_l[X_{l+1}] = E_l[\varepsilon_{l+1}] - \theta_1 E_l[\varepsilon_l] - \theta_2 E_l[\varepsilon_{l-1}] - \dots - \theta_q E_l[\varepsilon_{l-q+1}]$ $= \theta_1 \varepsilon_l - \theta_2 \varepsilon_{l-1} - \dots - \theta_q \varepsilon_{l-q+1}$	Ecuación de pronóstico $\hat{X}_l(2)$: $E_l[X_{l+2}] = E_l[\varepsilon_{l+2}] - \theta_1 E_l[\varepsilon_{l+1}] - \theta_2 E_l[\varepsilon_l] - \dots - \theta_q E_l[\varepsilon_{l-q+2}]$ $= -\theta_2 \varepsilon_l - \theta_3 \varepsilon_{l-1} - \dots - \theta_q \varepsilon_{l-q+2}$
Obsérvese que para $l = 3$, la ecuación de pronóstico resulta: $= -\theta_3 \varepsilon_l - \theta_4 \varepsilon_{l-1} - \dots - \theta_q \varepsilon_{l-q+3}$	
Ecuación de pronóstico para X_{l+i} : $\hat{X}_l(i) = 0, \forall i = q + 1, q + 2, \dots$	
Proceso MA(q)	

Tabla 3.3: Pronósticos de los procesos MA(1) y MA(p).

CASO DE ESTUDIO

En este capítulo se presenta la parte principal del trabajo, la cual es una aplicación de la de la teoría expuesta anteriormente la, que consiste en un análisis del comportamiento de la radiación solar en la región de Tlaxco-Tlaxcala, considerando una base de datos tomadas en la estación meteorológica instalada en dicho lugar.

El proceso de filtración de los datos se puede consultar en el apéndice (E). Antes de iniciar el análisis se dará una breve descripción de la zona de estudio y el clima en la región de Tlaxco, lo que son algunos efectos meteorológicos como “El niño”, “La niña” y la canícula, para dar paso al análisis de los datos.

4.1. Zona de estudio

El municipio de Tlaxco está ubicado en el altiplano central mexicano a 2 540 metros sobre el nivel del mar, se sitúa en un eje de coordenadas geográficas entre los 19°37' latitud norte y 98°07' longitud oeste. Localizado al norte del estado, el municipio de Tlaxco colinda al norte con el estado de Puebla, al sur colinda con los municipios de Atlangatepec, Tetla de la Solidaridad y Muñoz de Domingo Arenas, al oriente se establecen linderos con los municipios de Emiliano Zapata y Lázaro Cárdenas, asimismo al poniente colinda con el estado de Hidalgo y el municipio de Benito Juárez.

El municipio de Tlaxco comprende una superficie de 573.39 kilómetros cuadrados, lo que representa el 14.37 % del total del territorio estatal, el cual asciende a

3991.14 kilómetros cuadrados.

En el municipio de Tlaxco el clima se considera templado subhúmedo, con régimen de lluvias en los meses de junio a septiembre. Los meses más calurosos son de marzo a mayo. La dirección de los vientos en general es de norte a sur, igualmente la temperatura promedio máxima anual registrada es de 22.9 grados centígrados y la mínima de 5.3 grados centígrados. La precipitación promedio máxima registrada es de 122.5 milímetros y la mínima de 7.6 milímetros. [7]

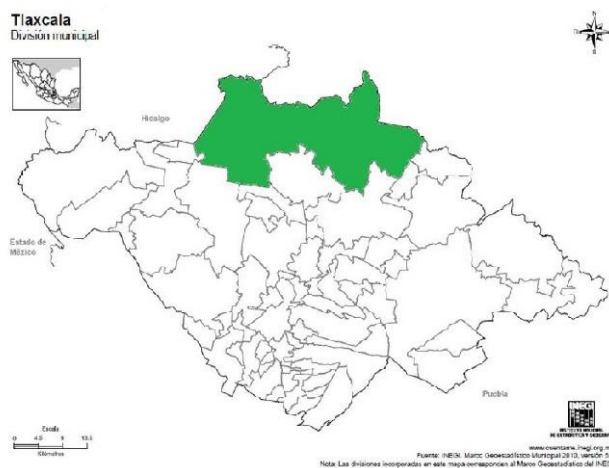


Figura 4.1: Ubicación geográfica del municipio de Tlaxco, Tlaxcala. Mapa extraído de INEGI.

Canícula

Cuando la temporada de lluvia se encuentra en auge y se presenta una reducción de la precipitación, a dicho fenómeno meteorológico se le conoce como canícula (también se le conoce como 'sequía intraestival o de medio verano', 'sequía de agosto', 'veranillo').

Se trata de un fenómeno climático que se presenta año con año en diferentes regiones de México en los meses de julio y agosto, este fenómeno se mantiene aproximadamente 40 días, y se presenta a la mitad de las temporadas de lluvias. En este periodo se registra la disminución de precipitación en el centro y noreste

del país, cielos despejados y un aumento en la temperatura, sobre todo en la región norte de la República Mexicana. Es considerada la época más calurosa del año.

La canícula se genera por la presencia de los vientos denominados “alisios”, los cuales soplan fuertemente en el Golfo de México, impidiendo el desarrollo de nubes convectivas en el océano y, en consecuencia la formación de lluvia en la parte continental de México. El término canícula nace de la palabra latina ‘canis’, que significa perro y hace referencia a la estrella de Sirio (apodada ‘La Abrasadora’) de la constelación del Can Mayor, esta es la más brillante del cielo nocturno durante ciertos días de verano en el hemisferio norte, y su aparición en el cielo, coincidía con los días más calurosos del año. [12] [27]

“El niño”

Es un fenómeno climatológico que se produce en el océano Pacífico, pero como es muy grande afecta al clima en el mundo entero. Donde el océano está tibio, se forman más nubes y cae más lluvia en esa región del mundo. En el océano Pacífico, cerca del ecuador, el Sol hace que el agua en la superficie sea especialmente tibia.

El evento de “El niño”, se caracteriza por un calentamiento en las aguas superficiales en el océano Pacífico central y oriental. Ese cambio está ligado a la atmósfera y a los vientos, que en condiciones normales, los vientos alisios tienen una dirección este-oeste, cuando esto sucede, el agua más fría inferior sube a la superficie del océano cerca de Sudamérica. Sin embargo, durante las condiciones de “El niño”, los vientos alisios se relajan y pueden soplar en dirección contraria. Esto permite que grandes masas de agua caliente se desplacen desde el Pacífico occidental hacia las Américas.

Los cambios en los vientos y las corrientes de agua en el Pacífico ecuatorial pueden causar inundaciones en el desierto de América del Sur y se seca el monzón en Indonesia y la India. Los eventos de “El niño” ocurren aproximadamente cada dos a siete años, ya que el ciclo cálido alterna irregularmente con su hermano “La niña”, un patrón de enfriamiento en el Pacífico oriental, y con condiciones neutras. “El niño” normalmente alcanza su punto máximo entre noviembre y enero, aunque la acumulación se puede detectar con meses de anticipación y sus efectos pueden tardar meses en propagarse en todo el mundo (ver Figura 4.2). [1]



NASA Space Place. (2014) ¿Qué es El Niño de todos modos? [Ilustración]. Recuperado de <https://spaceplace.nasa.gov/el-nino/sp/>

(a) Condiciones normales.



NASA Space Place. (2014) ¿Qué es El Niño de todos modos? [Ilustración]. Recuperado de <https://spaceplace.nasa.gov/el-nino/sp/>

(b) Condiciones de "El niño".

Figura 4.2: Ilustración del fenómeno climatológico denominado "El niño".

“La niña”

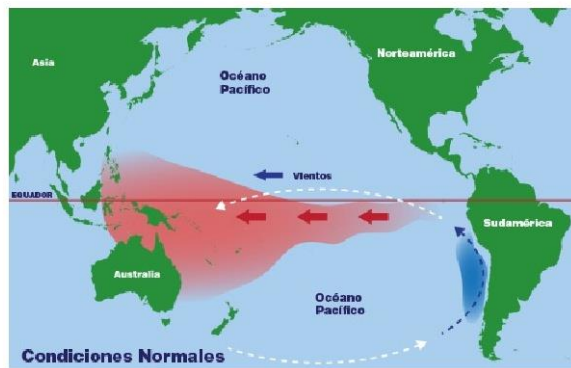
Las nubes de lluvia se forman normalmente sobre el agua caliente del océano. Cuando los vientos fuertes hacen que esta agua caliente se mueva, las nubes y las tormentas también se mueven.

En condiciones normales, los vientos sobre el océano Pacífico empujan suavemente el agua caliente hacia el oeste. Esa agua caliente viaja desde la costa oeste de Sudamérica hasta Indonesia. A medida que el agua caliente se mueve, el agua fría desde el fondo del océano se levanta lentamente para ocupar su lugar.

Sin embargo, durante “La niña”, los vientos sobre el océano Pacífico son mucho más fuertes que en un año normal. Esto suele suceder una vez cada pocos años. Los vientos son tan fuertes durante “La niña” que empujan al cálido océano hacia Indonesia. Esto permite que mucha agua fría suba a la superficie cerca de Sudamérica. Por lo que el agua en el océano Pacífico oriental está unos pocos grados más frío de lo normal.

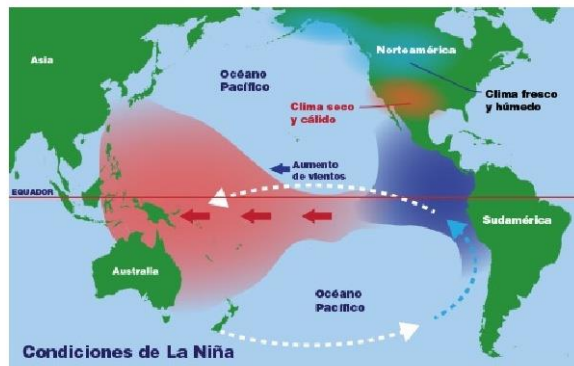
El agua caliente se mueve hacia el oeste durante La niña, así como las nubes. Lo que origina que lugares como Indonesia y Australia llueva más de lo normal. Y lugares como el suroeste de los Estados Unidos puedan estar muy secos. Afortunadamente, los científicos pueden predecir “La niña”, hasta un año antes de que ocurra (ver Figura 4.3).

Entonces ¿Cuál es la diferencia entre “El niño” y “La niña”? Ambos fenómenos comienzan en el océano Pacífico, pero son opuestos, es decir, “La niña” causa que el agua en el Pacífico oriental esté más fría de lo normal. En la misma región, “El niño” puede hacer que el agua esté más caliente de lo normal. Por lo tanto, las zonas afectadas por la sequía durante “La niña” pueden tener mucha lluvia en los años de “El niño”. [2]



NASA Space Place. (2016) ¿Qué es La Niña? [Ilustración].
Recuperado de <https://spaceplace.nasa.gov/la-nina/sp/>

(a) Condiciones normales.



NASA Space Place. (2016) ¿Qué es La Niña? [Ilustración].
Recuperado de <https://spaceplace.nasa.gov/la-nina/sp/>

(b) Condiciones de “La niña”.

Figura 4.3: Ilustración del fenómeno climatológico denominado “La niña”.

4.2. Análisis de la serie de tiempo de los datos de radiación

En la Figura (4.4) se muestra el gráfico de la serie de tiempo de la radiación, es posible observar que la serie presenta estacionalidad, al parecer no hay presencia de tendencia o no es muy visible, pero esto no quiere decir que la serie sea estacionaria (lo cual se analizará más adelante), note que la serie no oscila alrededor de cero, note que la variabilidad parece ser constante. Sin embargo, se realizaron transformaciones para estabilizar la varianza usando logaritmos, calculando raíces a los datos y se usó la transformación Box-Cox, se observaron sus respectivas gráficas e incluso sus FAC y FACP, no observando cierta diferencia comparadas con la serie original. Por lo que se decidió usar los datos originales de la serie.

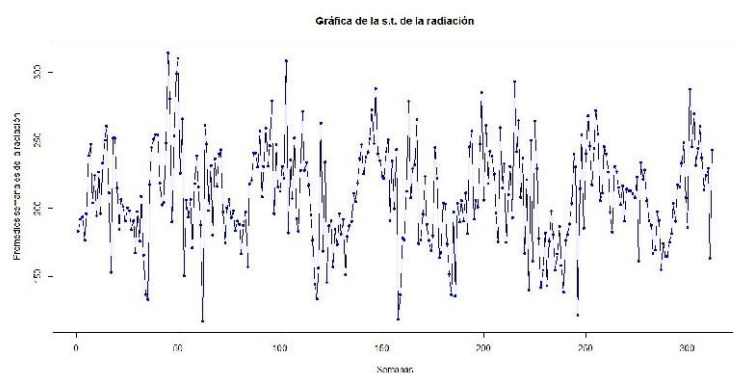


Figura 4.4: Serie de tiempo de la radiación de los datos filtrados.

Dado que lo primero que se debe hacer al analizar una serie de tiempo es verificar si la serie es, al menos, débilmente estacionaria, se graficó la FAC y la FACP de la serie de tiempo, ver Figuras (4.5) y (4.6). Note que la FAC se corta con cierta lentitud y se extingue en forma sinusoidal amortiguada y en cuanto a la FACP es posible observar que esta se extingue a partir del segundo rezago y de nuevo vuelve a ser significativo en los rezagos 4, 12, 15, 28 y 29.

Por otro lado, se realizó el análisis de cajas y bigotes, ver Figura (4.7), se

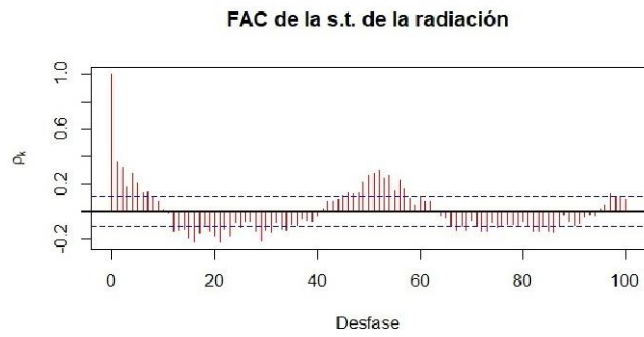


Figura 4.5: Función de autocorrelación de la serie de tiempo.

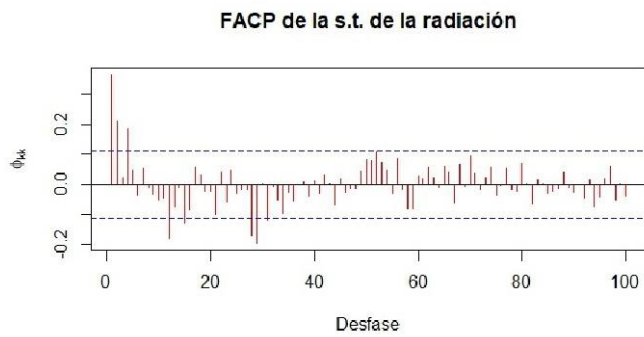


Figura 4.6: Función de autocorrelación parcial de la serie de tiempo.

graficó agrupando las mismas semanas año por año, de ahí que en el diagrama de cajas tenga 52 cajas. Este análisis de cajas se realizó con la finalidad de verificar la estacionalidad, que a criterio nuestro resulta ser estacional.

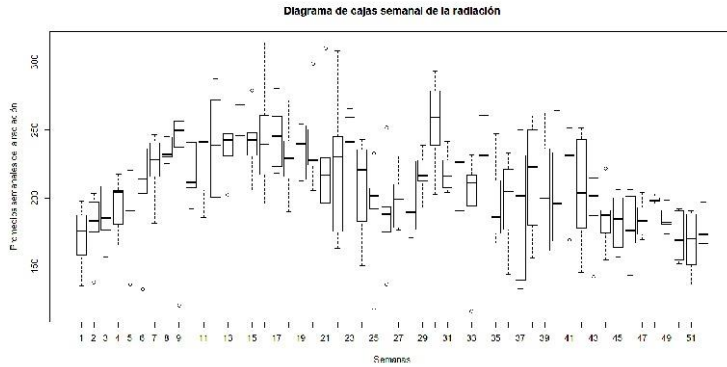


Figura 4.7: Diagrama de cajas de la serie de tiempo.

A continuación se realizará un análisis del gráfico (4.7). Donde cada una de las justificaciones mencionadas fueron halladas en los reportes anuales del clima de la república mexicana. [11]

Estos reportes contienen un resumen sobre el comportamiento de los eventos climáticos como la canícula, el fenómeno “El niño”, “La niña”, ciclones tropicales, huracanes, sequías, ondas de calor, frentes fríos y heladas, entre otros. La mayoría de estos fenómenos están ilustrados con mapas que muestran su distribución, como ejemplo de ello son las anomalías de la lluvia y la temperatura.

Los fenómenos de “El niño” y “La niña” claramente influyen en la República Mexicana, pero enfocándonos en el estado de Tlaxcala, algo que es importante mencionar es que, en ningún momento se hace alusión al municipio de Tlaxco en los reportes, ya que, en los reportes solamente se menciona a los estados de la República Mexicana. Así, a partir de la información obtenida se realiza un análisis descriptivo de algunos aspectos del clima que pudieron presentarse en el Muni-

cipio de Tlaxco - Tlaxcala. Tal análisis, inicia al considerar la distribución del diagrama de cajas de los datos obtenidos en una estación ubicada en tal municipio. Por otro lado, a continuación se presentan algunas observaciones respecto al comportamiento de los fenómenos de “El niño” y “La niña”.

En el año 2011 se observó una combinación de la fase neutral y negativa de “El niño” desde julio de 2010 hasta diciembre de 2011.

Para el año 2012, durante los primeros tres meses dominaron condiciones de “La niña” (temperaturas más frescas de lo normal). Y a partir de abril de 2012 dominaron condiciones neutrales en gran parte del año.

En el año 2013, persistieron condiciones neutrales del fenómeno “El niño” que se observaron desde finales del 2012. No hubo presencia de “La niña”.

Para el año 2014 la falta de acoplamiento entre la atmosfera y el océano favoreció el predominio de la fase neutral de “El niño”. No hubo presencia de “La niña”.

En el año 2015, “El niño” incremento su intensidad pasando a ser moderado del 15 de junio y fuerte a partir del 1 de septiembre. Este año se clasificó como el décimo segundo más lluvioso.

Para el 2016, las condiciones de “El niño” intenso favorecieron valores récord de temperatura en los meses de enero, febrero, marzo, abril, junio, julio y agosto. Los meses de enero, febrero y marzo presentaron anomalías. Y para los meses de noviembre y diciembre se mantuvo “La niña” débil.

Por último en el 2017, los meses de enero a junio se observó “El niño” neutro.

Análisis

Mayor valor máximo

En el gráfico de cajas y bigotes ver Figura (4.7), se observa que el mayor valor máximo se presentó en la caja 16 cuyos datos pertenecen aproximadamente a la penúltima semana de abril (2012-2017) y cuyo valor de radiación se encuentra por arriba de 300, probablemente se deba a que abril se destacó por registrar caída de lluvia que se considera como debajo de lo normal a normal, a excepción de abril 2015 que se registro con lluvias por arriba de lo normal y abril 2017 estuvo entre los menos lluviosos. Más aún, se registraron sequías que iban desde anormalmen-

te seco hasta sequía moderada, registrando también temperaturas altas.

Respecto a la caja 22, cuyos datos pertenecen a las últimas fechas de mayo (2012-2017) y primeras fechas de junio (2012-2017), se observa que, en el rango de tiempo que representan los valores de esta caja, se alcanza un valor máximo de radiación mayor a 300 pero sin superar el valor máximo de la caja 16, cabe señalar que mayo y junio presentaron las mismas condiciones que abril, a excepción de junio 2015, ya que 2015 se clasificó como un año muy lluvioso para Tlaxcala y más aún que en dicho año hubo presencia del fenómeno de “El niño” con intensidad fuerte, para junio 2017 se registró muy seco.

Para la caja 30, que tiene un valor máximo menor a 300, y cuyos datos pertenecen a las últimas fechas de julio (2011-2016) y primeras de agosto (2011-2016), donde se registró en estos meses el fenómeno de la canícula, y de acuerdo a los anuarios (ver [11]), Tlaxcala presentó canículas fuertes, lluvias por debajo de lo normal y temperaturas récord en los meses de julio y agosto.

Finalmente, en este análisis para los valores máximos, las cajas 12 y 14 parecen alcanzar el mismo nivel de radiación pero sin rebasar el valor máximo de la caja 30. Los datos de la caja 12 corresponden a la penúltima semana de marzo (2012-2017), donde en marzo 2012, 2013, 2014, 2017 se registraron lluvias debajo de lo normal, marzo 2015 se registró como el más lluvioso y marzo 2016 las lluvias se registraron normales y temperaturas récord. En marzo 2017 aumento la superficie de sequía registrándose anormalmente seco hasta sequía moderada; se registraron también temperaturas cálidas. Y en los meses de marzo 2013, 2014, 2015 y 2016 se registraron de 1 a 2 días temperaturas menores a 0 °C. Y a la caja 14 corresponden a las primeras lluvias de abril (2012-2017).

Menor valor mínimo

El valor mínimo se presentó en la caja 37, cuyos datos corresponden a la segunda semana de septiembre (2011-2016), donde se registraron lluvias arriba de lo normal en septiembre 2011, en septiembre 2012 y 2015 lluvias debajo de lo normal, septiembre 2013 destaco por ser muy lluvioso debido a la entrada del los huracanes Ingrid y Manuel, en septiembre 2014 también se registro como uno muy lluvioso y más aún, ese año fue para Tlaxcala su séptimo año más lluvioso. En septiembre 2016 se registró como un mes muy cálido. Es posible que por esta situación la caja haya registrado el menor valor mínimo, ya que la cantidad de lluvia fue “constante” y la formación de nubes inhibió el paso de los rayos solares.

Respecto a valores mínimos, el valor mínimo en la caja 37, le sigue el valor mínimo de la caja 1, que corresponde a las primeras semanas de enero (2012-2017), donde se registró presencia de lluvia debajo de lo normal en enero 2012, 2013, 2014, y 2017; mientras que en enero 2015 se registraron lluvias por arriba de lo normal, para enero 2016 se registro lluvia normal y temperaturas récord. En enero 2013, 2014, 2015, 2016 y 2017 se registraron al menos 1 o 2 días con temperaturas menores a 0°C, esto ocasionado por los frentes fríos y masas de aire frío.

Le sigue la caja 51, correspondiente a la penúltima semana de diciembre (2011-2016), donde se registraron 5 días con temperaturas menores a 0°C en diciembre 2011 y 3 días en diciembre 2013. Se registraron también lluvias por debajo de lo normal en diciembre 2011, 2012, 2014, 2015; a excepción de diciembre 2013 que presentó lluvias por arriba de lo normal y diciembre 2016 registró como un mes muy cálido con lluvias normales.

Y por último la caja 36 que corresponde a las primeras semanas de septiembre (2011-2016), donde los valores de radiación de las cajas resultaron menores a 150, pero la caja 37 contiene el registro del menor valor mínimo.

Cajas largas

Se observa que la caja más larga fue la 37, seguido de las cajas 38, 39 (donde las cajas 37, 38 y 39 corresponden al mes de septiembre (2011-2016)), luego sigue la caja 40 y 42, pertenecientes al mes de octubre (2011-2016), donde se presentaron lluvias normales a por arriba de lo normal en octubre 2011, 2013, 2014 y 2015, y en octubre 2012 se presentó la disminución de lluvias, se observaron temperaturas menores a 0 °C probablemente sólo un día y octubre 2016 resulto ser muy cálido.

Después le sigue la caja 12, que corresponde a la penúltima semana de abril (2012-2017), seguido de la caja 22, que corresponde a las últimas fechas de mayo y primeras de junio (2012-2017), y ligeramente la caja 23, que corresponde a la primera semana de junio (2012-2017). Obsérvese que los datos de la caja 12 parecen presentar simetría.

Y en cuanto a la caja 37, es probable que debido a la cantidad de lluvias que se registraron en esos años, la caja se hallara entre las más largas e incluso más larga que las demás.

Caja pequeña

Mientras que la caja más pequeña resulto ser la caja 48, cuyos datos corresponden a los últimos días de noviembre y primeros días de diciembre (2011-2016).

En noviembre 2011 y 2013 se registraron lluvias normal a por arriba de lo normal, en noviembre 2012 lluvias debajo de lo normal, noviembre 2014 y 2015 lluvias arriba de lo normal y noviembre 2016 se registró muy cálido. También se observaron temperaturas menores o iguales a 0°C en noviembre 2012, 2013, 2014 posiblemente sólo 2 días. Y en cuanto al las condiciones del mes de diciembre, éstas ya se mencionaron arriba.

Podría decirse incluso que la variabilidad en esta caja fue nula o bien que los valores de radiación fueron casi siempre constantes, es decir que podríamos atrevernos a decir que la radiación en los últimos días de noviembre y primeros días de diciembre es casi siempre la misma. Más aún, observe que la media coincide con el primer cuartil por lo que la caja presenta un sesgo a la derecha.

Cajas ligeramente pequeñas

Las cajas levemente pequeñas pero no más que pequeñas que la caja 48, resultaron ser la 25, 26, 29, 31, 33, 41, 43, 44, 47, 49, 52, 2, 7, 8, 9, 13, 15, y tal vez podrían considerarse las cajas 18 y 20. Podemos apreciar que debido al tamaño de estas cajas, la variabilidad es pequeña comparada con el resto (a excepción de la caja 48), los bigotes resultan ser ligeramente cortos, y observe que algo muy curioso se presenta en la mayoría de estas cajas es que no presentan ambos bigotes o no presentan un bigote, presentan valores atípicos ya sean mínimos o máximos o ambos. Las cajas estuvieron registradas así:

Caja 25: corresponde a la penúltima semana del mes de junio.

Caja 26: corresponde a los últimos días de junio y primeros de julio.

Caja 29: corresponde a la penúltima semana de julio.

Cajas 31 y 33: corresponden a la primera semana de agosto, y a la penúltima semana de agosto respectivamente.

Cajas 41 y 43: corresponden a la segunda semana de octubre, y a la penúltima semana de octubre respectivamente.

Caja 44: corresponde a los últimos días de octubre y primeros de noviembre.

Caja 47: corresponde a la penúltima semana de noviembre.

Caja 49: corresponde a la primera semana de diciembre.

Caja 52: corresponde a los últimos días de diciembre y primeros de enero.

Caja 2: corresponde a la segunda semana de enero.

Cajas 7 y 8: corresponden a la penúltima y última semana del mes de febrero res-

pectivamente.

Caja 9: corresponde a los últimos días de febrero y primeros de marzo.

Caja 13: corresponde a los últimos días de marzo y primeros de abril.

Caja 15: corresponde a la segunda semana de abril.

Cajas 18 y 20: corresponden a la primera semana de mayo, y a la penúltima semana de mayo respectivamente.

Mayor mediana

La caja que registró la mayor mediana resultó ser la caja 14, como abril se destacó por la presencia de sequías y poca precipitación, es probable que debido a esto, la caja se haya encontrado más arriba que las demás. Seguida de la caja 14 está la caja 30, pues en esta caja como se mencionó arriba, se registraron las últimas fechas de julio y primeras de agosto, es probable que en estos días la canícula se haya encontrado en su pleno apogeo, y posiblemente por ello la caja 30 se encuentra arriba. Incluso puede observarse un ligero sesgo a la izquierda en ambas cajas, lo que nos dice que los mayores niveles de radiación se encuentran concentrados cerca del Q_3 .

Menor mediana

La menor mediana se encuentra en la caja 50 y 51, seguido de la caja 52 correspondientes al mes de diciembre (2011-2016) y la caja 1 que corresponde a la primera semana de enero (2012-2017). Probablemente esto se debió a la presencia de frentes fríos o masas de aire e incluso a la considerable actividad de lluvia.

Otras

Ahora, véase la caja 4 que corresponden al los últimos días del mes de enero (2012-2017) y primeros días del mes de febrero (2012-2017) y la caja 5 correspondiente a la primera semana de febrero (2012-2017), donde se registraron lluvias debajo de lo normal en febrero 2012, 2013, 2014, 2015, 2017, y en febrero 2016 lluvias normales. En febrero 2014, se presentó muy seco y se registraron temperaturas récord, siendo también muy cálido. En febrero 2017 aumento la superficie de sequía registrándose anormalmente seco hasta sequía moderada y se registraron también temperaturas cálidas. También se observaron temperaturas menores o iguales a 0°C aproximadamente de 1 a 2 días en febrero 2013-2017.

Note que algo interesante ocurre en estas cajas, pues la mediana de la caja 4 está muy cerca del cuartil 3, mientras que la mediana de la caja 5 está muy cerca del cuartil 1, dado que ambas ocurrieron en el mes de febrero podría pensarse que en

la caja 4 se vio más afectada por las de bajas temperaturas o lluvias, a comparación de la caja 5, de ahí el cambio tan drástico de la mediana.

Valores atípicos

Las cajas que registraron valores atípicos fueron la caja 25, 26, 33, 41, 43, 44, 2, 5, 6, 9, 13, 15, 20, 21. Donde el menor valor atípico se presentó en la caja 33, seguido de la caja 25 y la caja 9. Y el mayor valor atípico se presentó en la caja 21 y 20. Es probable que estos valores de radiación atípicos, se debieran a las muy bajas o altas temperaturas o bien al error de captura e incluso la a la estimación de los datos faltantes.

Ahora bien, regresando de nuevo a las autocorrelaciones, estas indicarían que se trata de una serie no estacionaria, por lo que se decidió realizar la prueba de estacionariedad Dickey-Fuller, ver apéndice (C); en este caso, note que la serie (ver Figura 4.4) parece estar fluctuando alrededor de una media que no es cero, por lo que se hará uso de la prueba 2 de Dickey-Fuller, esta prueba se aplica cuando la serie posee término constante pero sin tendencia.

En la Tabla (4.1) se muestran los resultados de la prueba Dickey-Fuller y el contraste de hipótesis, donde resulta que la hipótesis nula se rechaza a cualquier nivel de significancia, en este trabajo se tomará un nivel de significancia del %5. De acuerdo a los resultados de la prueba, la serie resulta ser estacionaria y además de que es estacional (visto en el diagrama de cajas). Es aquí donde se presento

H ₀ : La serie es no estacionaria vs H _a : La serie es estacionaria				
	Estadístico de prueba	Valor crítico al 1 %	Valor crítico al 5 %	Valor crítico al 10 %
Dickey-Fuller con intercepto	-8.0062	-3.44	-2.87	-2.57

Tabla 4.1: Prueba de estacionariedad para la serie.

un problema, pues al momento de diferenciar la serie estacionalmente (con $s=52$), resultó complicado e incluso imposible proporcionar un modelo que describiera a la serie de tiempo. Así, se decidió realizar una diferencia regular a la serie original y una diferencia estacional a la serie diferenciada regularmente, ver Figura (4.8).

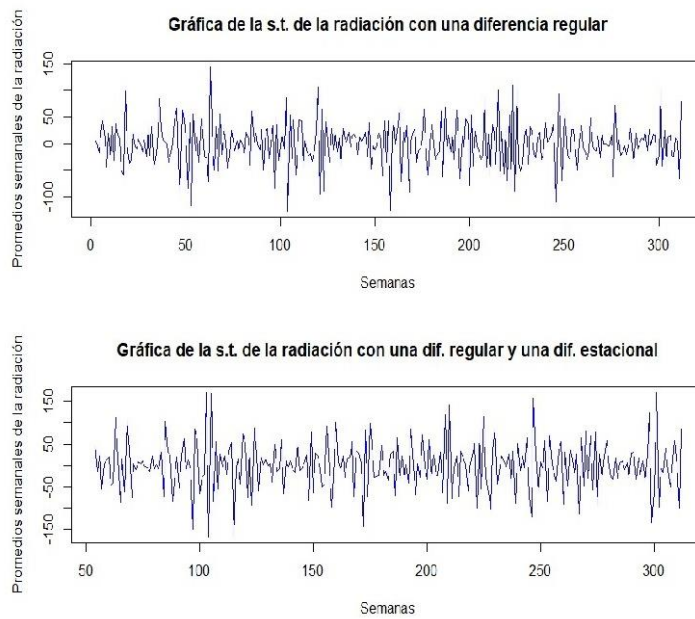


Figura 4.8: Diferencia regular y estacional de la serie de tiempo.

Dado que la serie es estadísticamente estacionaria, al diferenciarla regularmente la serie esta sigue siendo estacionaria. Para ratificar esto, en la Tabla (4.2) se muestran los resultados de la prueba Dickey-Fuller de la serie diferenciada regularmente una vez, se le aplicó la prueba 1 debido a que la serie diferenciada regularmente esta vez si oscila alrededor de una media cero. Y en la Tabla (4.3) se muestran los resultados de la prueba Dickey-Fuller para la serie ya diferenciada pero ahora con una diferencia estacional. Por lo tanto, se descarta la existencia de alguna raíz unitaria, es decir, sigue siendo estacionaria a cualquier nivel de significancia.

Graficando las FAC y FACP de la serie con una diferencia regular y la FAC y FACP de la serie diferenciada estacionalmente, resulta lo siguiente, ver Figuras (4.9) y (4.10).

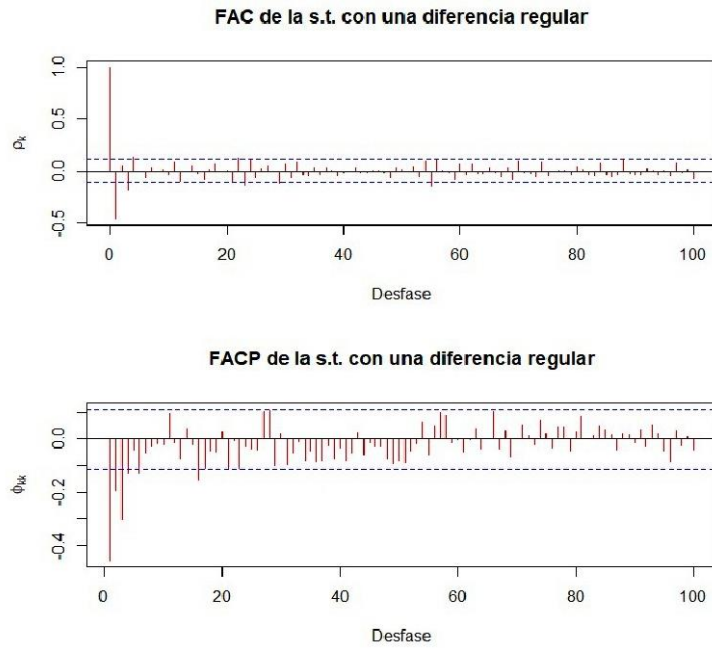


Figura 4.9: FAC y FACP de la serie diferenciada regularmente.

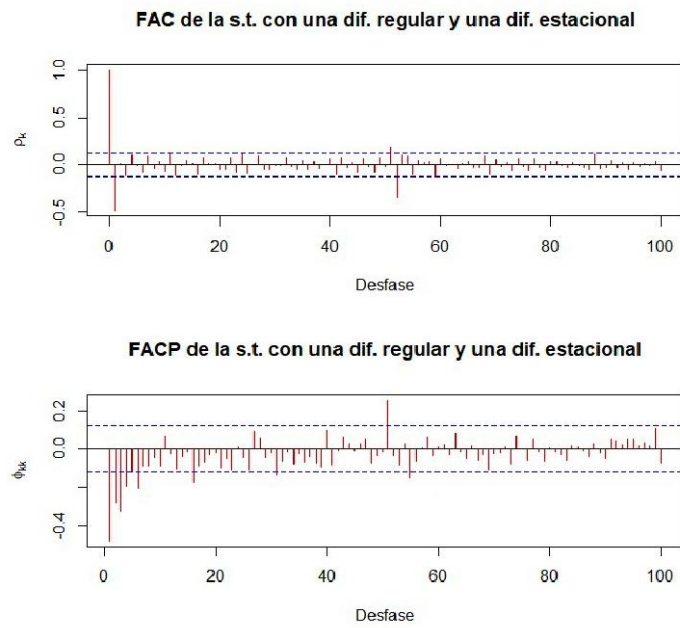


Figura 4.10: FAC y FACP de la serie diferenciada regularmente con una diferencia estacional.

H ₀ : La serie es no estacionaria vs H _a : La serie es estacionaria				
	Estadístico de prueba	Valor crítico al 1 %	Valor crítico al 5 %	Valor crítico al 10 %
Dickey-Fuller con intercepto	-18.1388	-2.58	-1.95	-1.62

Tabla 4.2: Prueba de estacionariedad para la serie diferenciada regularmente.

H ₀ : La serie es no estacionaria vs H _a : La serie es estacionaria				
	Estadístico de prueba	Valor crítico al 1 %	Valor crítico al 5 %	Valor crítico al 10 %
Dickey-Fuller con intercepto	-18.2666	-2.58	-1.95	-1.62

Tabla 4.3: Prueba de estacionariedad para la serie diferenciada estacionalmente.

Las FAC y FACP en 4.9 son de ayuda para definir los órdenes de p y q para el modelo que describa la serie de tiempo en la parte no estacional.

Obsérvese que en el gráfico 4.9, la FAC tiene una espiga sobresaliente en el primer rezago, sin embargo sobresale otra espiga en el rezago 3 y 4, siendo estas dos últimas pequeñas comparadas con la primera y a partir del cuarto rezago se extingue. Por lo que q puede tomar el valor de 1, 2 ó 3.

Para la FACP se tienen 4 espigas sobresalientes en los primeros cuatro rezagos y se corta a partir del rezago 4, siendo la espiga del cuarto rezago un tanto pequeña, junto con la del rezago 6 y 16 que también sobresalen. Por lo que el valor de p puede ser 1, 2, 3 ó 0. Así, se propuso un modelo $ARIMA(0, 1, 3)$ para la parte no estacional.

Por otro lado, se analizó la FAC y FAC del gráfico en la Figura 4.10. Los cuales también son de ayuda para definir los órdenes de P y Q para el modelo que describa a la serie de tiempo en la parte estacional.

Obsérvese que en el gráfico de la Figura 4.10, la FAC tiene una espiga sobresaliente en el primer rezago, y a partir del primer rezago se extingue. Por lo que el valor de Q sería igual a 1.

Para la FACP se tienen 4 espigas sobresalientes en los primeros cuatro rezagos y se corta a partir del rezago 4, siendo la espiga del cuarto rezago un tanto pequeña, junto con la del rezago 6 que también sobresale. Por lo que el valor de P puede ser 1, 2, 3 ó 0. Así, se propuso un modelo para la parte estacional con $s = 52$, de la forma $ARIMA(0, 1, 1)_{52}$. En resumen, el modelo propuesto es $ARIMA(0, 1, 3) \times (0, 1, 1)_{52}$.

Se propusieron otros 4 modelos y junto con el modelo que el software R propone se hizo el ajuste de acuerdo al valor del criterio de Akaike y el valor del criterio BIC (Bayesian Information Criterion) (ver Apéndice (D)). Los resultados se muestran en la Tabla (4.4). Luego de observar los resultados obtenidos de acuerdo

Modelo	AIC	BIC
$ARIMA(0, 1, 1) \times (0, 1, 1)_{52}$	2561.068	2571.68
$ARIMA(0, 1, 2) \times (0, 1, 1)_{52}$	2562.033	2576.182
$ARIMA(3, 1, 1) \times (0, 1, 1)_{52}$	2563.037	2584.261
$ARIMA(0, 1, 3) \times (0, 1, 1)_{52}$	2563.522	2581.209
$ARIMA(3, 1, 2) \times (0, 1, 1)_{52}$	2563.671	2588.433
$ARIMA(1, 0, 1)$	3027.037	3041.944

Tabla 4.4: AIC y BIC para diferentes modelos.

al criterio de Akaike y BIC, el modelo a elegir fue el $ARIMA(0, 1, 1) \times (0, 1, 1)_{52}$, ya que, éste toma el valor más pequeño conforme al los criterios. En este trabajo se analizarán los modelos $ARIMA(0, 1, 3) \times (0, 1, 1)_{52}$, $ARIMA(1, 0, 1)$ y el modelo $ARIMA(0, 1, 1) \times (0, 1, 1)_{52}$, donde, el primero fue el que se propuso de acuerdo a lo que mostraban las FAC y FACP, el segundo fue dado por el software R, y el tercero fue el que obtuvo menor valor en los criterios de Akaike y BIC. Se renombrará al modelo $ARIMA(0, 1, 3) \times (0, 1, 1)_{52}$ como modelo 1, al modelo $ARIMA(1, 0, 1)$ como modelo R y al modelo $ARIMA(0, 1, 1) \times (0, 1, 1)_{52}$ como modelo 2.

Para evaluar el ajuste del modelo y del pronóstico se omitirán del análisis las últimas 5 semanas de la serie original.

Modelo 1

Escribiendo al modelo $ARIMA(0, 1, 3) \times (0, 1, 1)_{52}$ en términos de polinomios de retardo y operadores de diferencia (como en 3.21) se obtiene

$$\nabla_{52}^1 \nabla^1 X_t = \Theta_1(B^{52})\theta_3(B)\varepsilon_t \quad (4.1)$$

donde $\nabla_{52}^1 = (1 - B^{52})^1$ y $\nabla^1 = (1 - B)^1$. Desarrollando 4.1 resulta

$$\begin{aligned}(1 - B^{52})(1 - B)X_t &= (1 - \Theta_1 B^{52})(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)\varepsilon_t \\ (1 - B - B^{52} + B^{53})X_t &= (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \Theta_1 B^{52} + \Theta_1 \theta_1 B^{53} \\ &\quad + \Theta_1 \theta_2 B^{54} + \Theta_1 \theta_3 B^{55})\varepsilon_t.\end{aligned}$$

Lo que puede escribirse como

$$\begin{aligned}X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} - \Theta_1 \varepsilon_{t-52} + \Theta_1 \theta_1 \varepsilon_{t-53} + \Theta_1 \theta_2 \varepsilon_{t-54} \\ + \Theta_1 \theta_3 \varepsilon_{t-55} + X_{t-1} + X_{t-52} - X_{t-53}.\end{aligned}\quad (4.2)$$

Luego, se realizó la estimación de parámetros, utilizando el método de máxima verosimilitud, los resultados se muestran en la Tabla(4.5). Y una vez estimados

Parámetro	Parámetro estimado	Error estándar
MA(1)	-0.925809	0.070328
MA(2)	-0.031454	0.074059
MA(3)	-0.042722	0.059964
SMA(1)	-0.801326	0.156283

Tabla 4.5: Estimación de parámetros para el modelo 1.

los parámetros, el modelo queda escrito de la siguiente forma

$$\begin{aligned}X_t = \varepsilon_t + 0.925809\varepsilon_{t-1} + 0.031454\varepsilon_{t-2} + 0.042722\varepsilon_{t-3} + 0.801326\varepsilon_{t-52} \\ + 0.741875\varepsilon_{t-53} + 0.025205\varepsilon_{t-54} + 0.034234\varepsilon_{t-55} + X_{t-1} + X_{t-52} - X_{t-53}.\end{aligned}\quad (4.3)$$

Como se menciono antes, para la realización del pronóstico (esto con la ayuda de la función *forecast* de la librería *forecast* del software R) se omitieron las 5 últimas semanas, esto con la finalidad de comprobar la eficiencia del modelo propuesto mediante el error, a continuación se comparan los resultados de las últimas 5 semanas que se omitieron contra los valores pronosticados, ver Tabla (4.6), y se muestran los intervalos de confianza para cada valor pronosticado, ver Tabla (4.7). En la Figura(4.11) se observan los datos pronosticados.

Semana	Valor real	Valor pronosticado	Error absoluto $e_a = Vr - Vp $	Error relativo $e_r = \frac{e_a}{Vr}$
308	213.9137	238.2096	24.2959	0.1136
309	224.2083	234.2178	10.0095	0.0446
310	229.6845	218.2615	11.4230	0.0497
311	163.3690	231.3004	67.9314	0.4158
312	242.7173	226.0112	16.7061	0.0688

Tabla 4.6: Valores reales y pronosticados de 5 semanas.

Valor pronosticado	Intervalo de confianza al 80 %		Intervalo de confianza al 95 %	
238.2096	194.7926	281.6265	171.8091	304.6100
234.2178	190.6561	277.7795	167.5959	300.8397
218.2615	174.6446	261.8784	151.5552	284.9679
231.3004	187.6820	274.9188	164.5919	298.0090
226.0112	182.3911	269.6314	159.3000	292.7225

Tabla 4.7: Intervalo de confianza al 80 % y al 95 %.

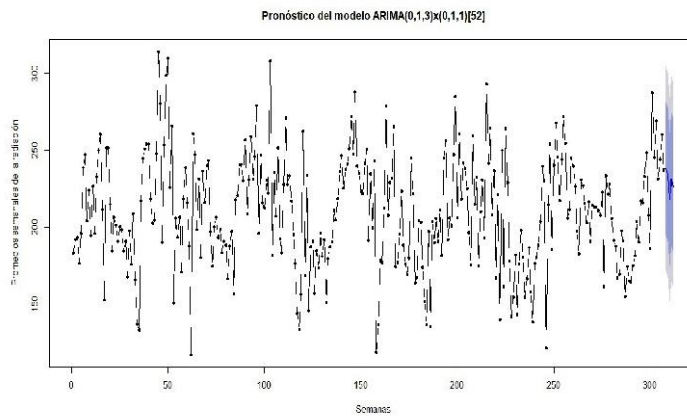


Figura 4.11: Datos pronosticados.

Otro de los planteamientos iniciales para determinar que un modelo es el adecuado, es el diagnóstico del modelo, donde la suposición básica es que los residuales son ruido blanco, es decir, presentan media nula, varianza constante y están no correlacionados.

La validación del modelo es una parte importante para evidenciar la idoneidad del mismo. Para el análisis de residuos, se hizo uso de las herramientas: Q-Q plot, para la comprobación de normalidad de los residuos junto con un histograma, y se usó el comando *tsdiag* y el test de Ljung-Box de la librería *tseries* del software R.

En la Figura (4.12) se puede ver el diagnóstico del modelo 1.

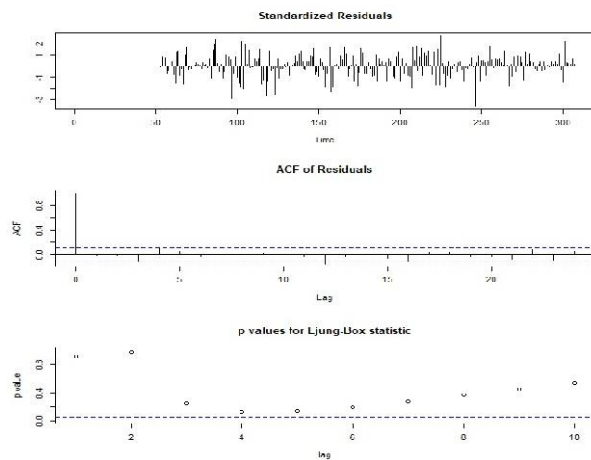


Figura 4.12: Diagnóstico del modelo 1.

Obsérvese que en la primera imagen de la Figura (4.12) se muestran los residuales estandarizados del modelo, donde se observa que están distribuidos de forma aleatoria alrededor del cero, sin mostrar tendencia, por lo que se consideran independientes.

En la segunda imagen, la FAC de los residuales se asemeja a la FAC de un ruido

blanco, por lo que parece que se está ante un modelo adecuado (también vea Figura (4.13), donde se muestra la FAC y FACP).

Finalmente, para la tercera imagen, se observa el test de Ljung-Box, que como se describió en 3.6 contrasta el siguiente juego de hipótesis

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

vs

$$H_a : \text{al menos algún } \rho_k \text{ es diferente de cero}$$

Como puede deducirse del tercer gráfico, se tiene que no se rechaza la hipótesis nula, ya que el valor p es mayor que 0.05. Tenga en cuenta lo siguiente:

- El test de Ljung-Box está probando la no correlación más no la independencia. Ya que no correlación no siempre indica independencia.

El test de Ljung-Box de manera numérica arrojó un p valor de 0.9076, que implica que el p valor es mayor que 0.05, por lo que no se rechaza H_0 .

El Q-Q plot es un método gráfico para determinar si los datos siguen una distribución normal. Si el conjunto de datos es “aceptablemente” normal, el gráfico que se observará será próximo a una recta. Tanto más parecido sea el conjunto de datos a una recta, mayor será la compatibilidad con la distribución normal. En la Figura 4.14(c) el conjunto de datos está distribuido “aceptablemente” normal, nótese también que tiene las colas muy pesadas, por otro lado, en el histograma 4.14(a) note que hay dos curvas, la curva roja representa la estimación de la densidad y la curva azul representa la densidad teórica que, al compararlas parece haber una leve presencia de asimetría, para notar esto, vea la Figura 4.14(b), el diagrama de caja parece estar ligeramente sesgado a la derecha, hay presencia de valores atípicos, de ahí que las colas de la distribución sean muy pesadas, por lo tanto, dado que en este modelo el supuesto de normalidad es robusto, se puede seguir considerando dicho supuesto.

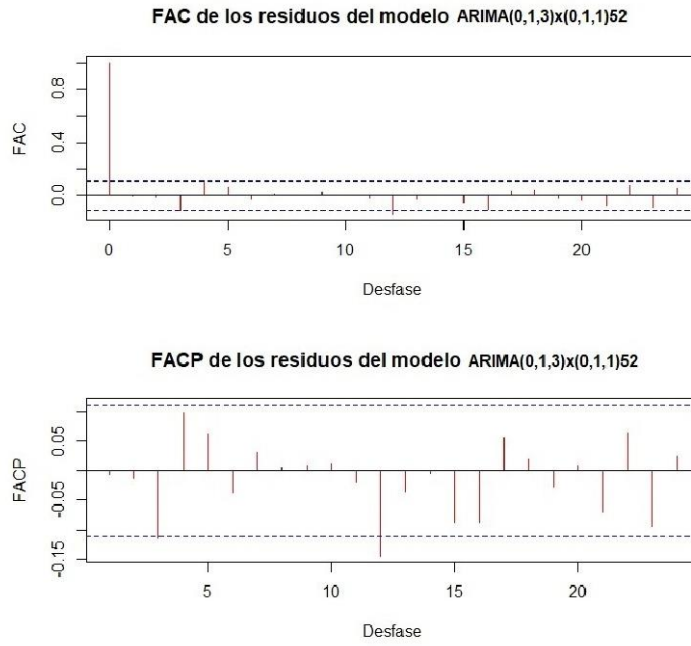


Figura 4.13: FAC y FACP de los residuos del modelo 1.

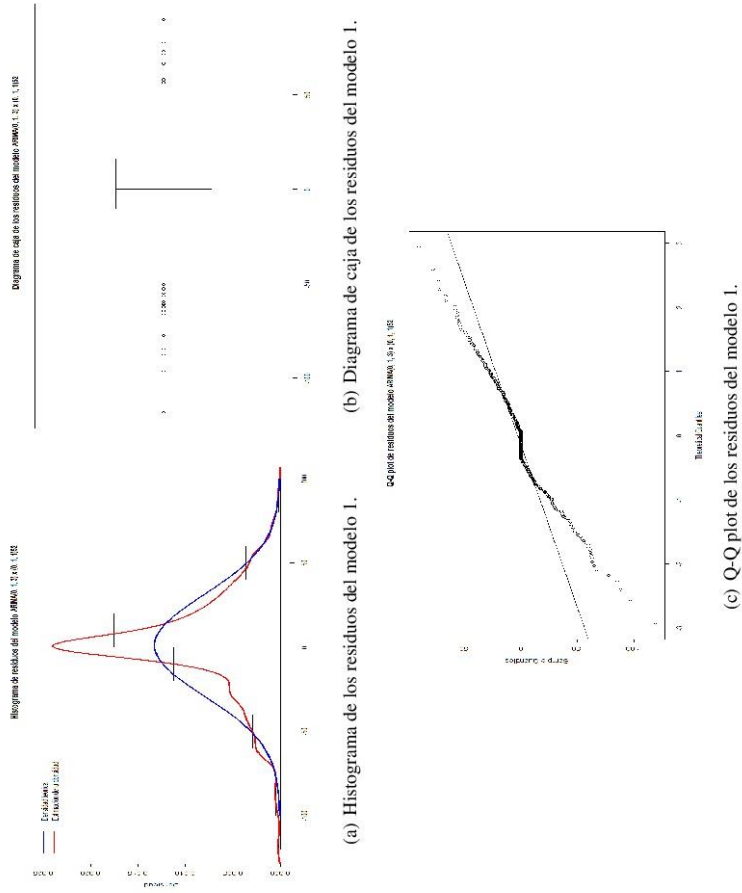


Figura 4.14: Análisis de normalidad de los residuos del modelo I.

Modelo R

Escribiendo al modelo $ARIMA(1, 0, 1)$ en términos de polinomios de retardo (como en 2.47) se obtiene

$$\phi_1(B)X_t = \theta_1(B)\varepsilon_t \quad (4.4)$$

Desarrollando 4.4 resulta

$$\begin{aligned} (1 - \phi_1 B)X_t &= (1 - \theta_1 B)\varepsilon_t \\ X_t - \phi_1 X_{t-1} &= \varepsilon_t - \theta_1 \varepsilon_{t-1}. \end{aligned}$$

Lo que puede escribirse como

$$X_t = \phi_1 X_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t. \quad (4.5)$$

Posteriormente, se realizó la estimación de parámetros, utilizando el método de máxima verosimilitud, los resultados se muestran en la Tabla(4.8). Y una

Parámetro	Parámetro estimado	Error estándar
AR(1)	0.8323	0.0604
MA(1)	-0.5668	0.0879
intercepto	210.1999	4.8141

Tabla 4.8: Estimación de parámetros para el modelo R.

vez estimados los parámetros, el modelo queda escrito de la siguiente forma

$$X_t = 210.1999 + 0.8323_1 X_{t-1} + 0.5668 \varepsilon_{t-1} + \varepsilon_t. \quad (4.6)$$

Como se menciona antes, para la realización del pronóstico (esto con la ayuda de la función *forecast* de la librería *forecast* del software R) se omitieron las 5 últimas semanas, esto con la finalidad de comprobar la eficiencia del modelo propuesto mediante el error, a continuación se comparan los resultados de las últimas 5 semanas que se omitieron contra los valores pronosticados, ver Tabla (4.9), y se muestran los intervalos de confianza para cada valor pronosticado, ver Tabla (4.10). Y la Figura(4.15) muestra los datos pronosticados.

Semana	Valor real	Valor pronosticado	Error absoluto $e_a = Vr - Vp $	Error relativo $e_r = \frac{e_a}{Vr}$
308	213.9137	231.6546	17.7409	0.0829
309	224.2083	228.0563	3.848	0.0171
310	229.6845	225.0615	4.623	0.0201
311	163.3690	222.5690	59.2	0.3623
312	242.7173	220.4945	22.2228	0.0915

Tabla 4.9: Valores reales y pronosticados de 5 semanas.

Valor pronosticado	Intervalo de confianza al 80 %		Intervalo de confianza al 95 %	
231.6546	189.3179	273.9914	166.9061	296.4032
228.0563	184.2531	271.8596	161.0650	295.0476
225.0615	180.2706	269.8524	156.5597	293.5633
222.5690	177.1065	268.0315	153.0401	292.0979
220.4945	174.5726	266.4164	150.2629	290.7261

Tabla 4.10: Intervalo de confianza al 80 % y al 95 %.

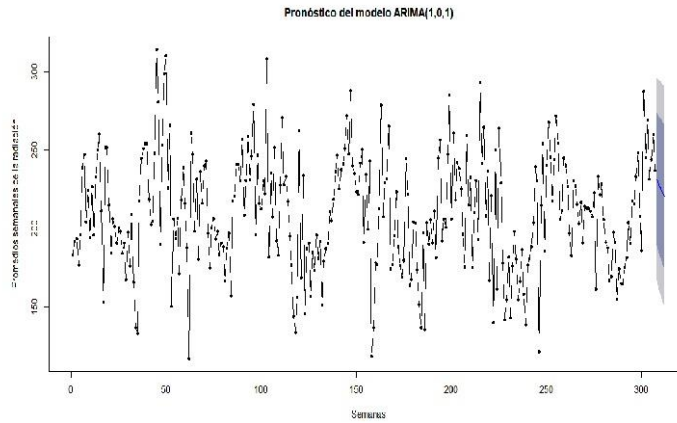


Figura 4.15: Datos pronosticados.

Otro de los planteamientos iniciales para determinar que un modelo es el adecuado, es el diagnóstico del modelo, donde la suposición básica es que los residuales son ruido blanco, es decir, presentan media nula, varianza constante y están no correlacionados.

La validación del modelo es una parte importante para evidenciar la idoneidad del mismo. Para el análisis de residuos, se hizo uso de las herramientas: Q-Q plot, para la comprobación de normalidad de los residuos junto con un histograma, y se usó el comando *tsdiag* y el test de Ljung-Box de la librería *tseries* del software R.

En la Figura (4.16) se puede ver el diagnóstico del modelo R.

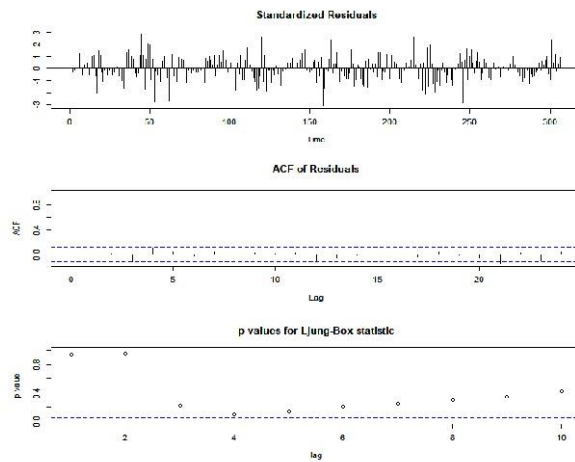


Figura 4.16: Diagnóstico del modelo R.

Obsérvese que en la primera imagen de la Figura (4.16) se muestran los residuales estandarizados del modelo, donde se observa que están distribuidos de forma aleatoria alrededor del cero, sin mostrar tendencia, por lo que se consideran independientes.

En la segunda imagen, la FAC de los residuales se asemeja a la FAC de un ruido

blanco, por lo que parece que se está ante un modelo adecuado (también vea Figura (4.17), donde se muestra la FAC y FACP).

Finalmente, para la tercera imagen, se observa el test de Ljung-Box que, como puede deducirse del tercer gráfico, se tiene que no se rechaza la hipótesis nula, ya que el valor p es mayor que 0.05. La prueba de Ljung-Box de manera numérica arrojó un p valor de 0.9417, que implica que el p valor es mayor que 0.05, por lo que no se rechaza H_0 .

En la Figura 4.18(c) el conjunto de datos está distribuido “muy aceptablemente” normales, nótese que también tiene las colas muy pesadas, por otro lado, en el histograma 4.18(a) note que hay dos curvas, la curva roja representa la estimación de la densidad y la curva azul representa la densidad teórica que, al compararlas, note que el gráfico rojo se asemeja a la normal donde, es muy visible la presencia de simetría, para notar esto, vea la Figura 4.18(b), en este se observa que el diagrama de caja si presenta simetría además de contar con la presencia de valores atípicos, de ahí que las colas de la distribución sean muy pesadas, por lo tanto, dado que en este modelo el supuesto de normalidad es robusto, se puede seguir considerando dicho supuesto.

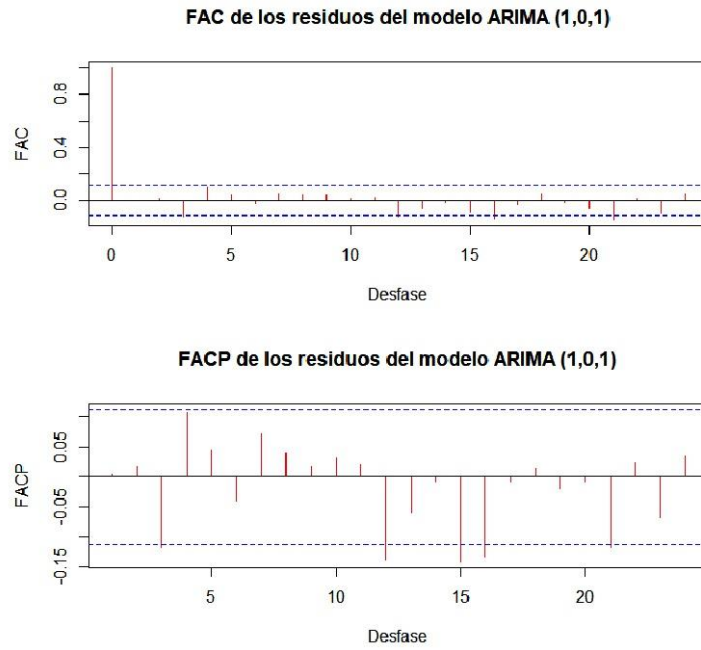


Figura 4.17: FAC y FACP de los residuos del modelo R.

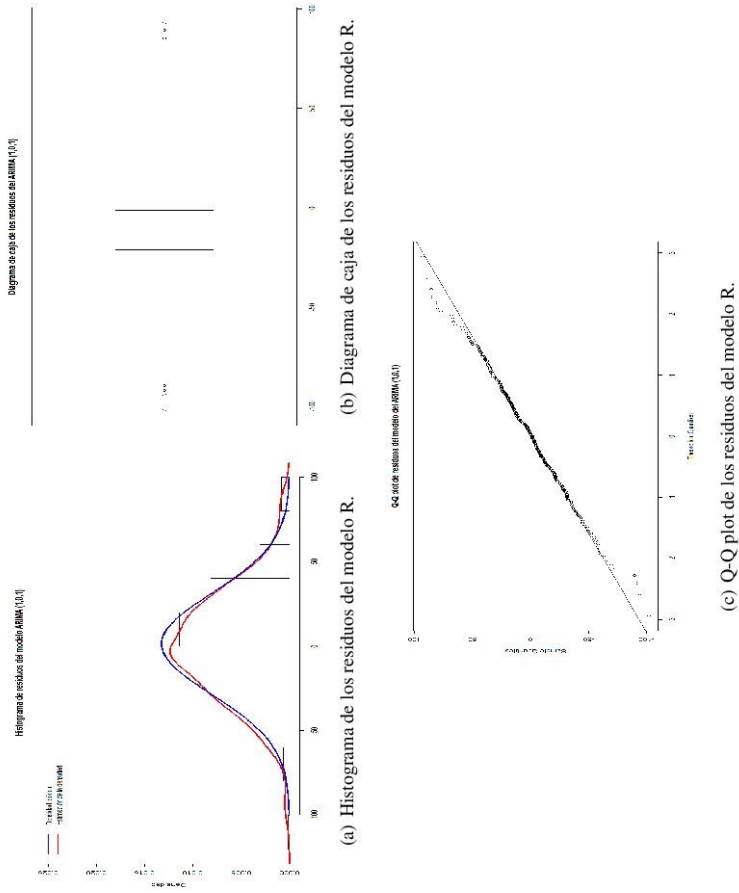


Figura 4.18: Análisis de normalidad de los residuos del modelo R.

Modelo 2

Escribiendo al modelo $ARIMA(0, 1, 1) \times (0, 1, 1)_{52}$ en términos de polinomios de retardo y operadores de diferencia (como en 3.21) se obtiene

$$\nabla_{52}^1 \nabla^1 X_t = \Theta_1 (B^{52}) \theta_1 (B) \varepsilon_t \quad (4.7)$$

donde $\nabla_{52}^1 = (1 - B^{52})^1$ y $\nabla^1 = (1 - B)^1$. Desarrollando 4.7 resulta

$$\begin{aligned} (1 - B^{52})(1 - B)X_t &= (1 - \Theta_1 B^{52})(1 - \theta_1 B)\varepsilon_t \\ (1 - B - B^{52} + B^{53})X_t &= (1 - \theta_1 B - \Theta_1 B^{52} + \Theta_1 \theta_1 B^{53})\varepsilon_t. \end{aligned}$$

Que finalmente puede quedar expresado como

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \Theta_1 \varepsilon_{t-52} + \Theta_1 \theta_1 \varepsilon_{t-53} + X_{t-1} + X_{t-52} - X_{t-53}. \quad (4.8)$$

Se realizó la estimación de parámetros por medio del método de máxima verosimilitud, los resultados se muestran en la Tabla(4.11).

Parámetro	Parámetro estimado	Error estándar
MA(1)	-1.000000	0.050331
SMA(1)	-0.805794	0.161295

Tabla 4.11: Estimación de parámetros para el modelo 2.

Ya estimados los parámetros, el modelo queda escrito de la siguiente forma

$$X_t = \varepsilon_t + \varepsilon_{t-1} + 0.805794\varepsilon_{t-52} + 0.805794\varepsilon_{t-53} + X_{t-1} + X_{t-52} - X_{t-53} \quad (4.9)$$

Para el pronóstico se realizó la omisión de las 5 últimas semanas, esto se hizo con el objetivo de comprobar la eficiencia del modelo propuesto mediante el error, en la Tabla (4.12), se observa la comparación de los valores obtenidos en las últimas 5 semanas que se omitieron contra los valores pronosticados, se observan también los intervalos de confianza para cada valor pronosticado, ver Tabla (4.13).

Por otro lado, en la Tabla (4.14) se muestra la comparación de los errores relativos de los modelos 1, R y 2, note que el modelo R tiene menor error relativo desde la semana 308 a la 311 (comparado con los modelos 1 y 2), excepto la semana 312, que muestra un error del 9% que es un poco mayor comparado con los otros dos. Esto hace pensar lo siguiente:

1. El modelo R sería una buena elección a considerar como buen modelo debido al error relativo, pues es menor comparado con los otros dos modelos.
2. Dado que en la semana 312 se observó un error mayor en el modelo de R, se especularía que el error de los pronósticos seguiría aumentando, lo que conllevará a que los pronósticos no sean muy acertados.
3. Seleccionar al modelo 2 como el mejor modelo. Debido a que en este si se toma en cuenta la estacionalidad, la cual se observó en el diagrama de cajas, pero que, el modelo de R no toma en cuenta.

Semana	Valor real	Valor pronosticado	Error absoluto $e_a = Vr - Vp $	Error relativo $e_r = \frac{e_a}{Vr}$
308	213.9137	236.5758	22.6621	0.1059
309	224.2083	234.0264	9.8181	0.0438
310	229.6845	218.3629	11.3216	0.0493
311	163.3690	231.4973	68.1283	0.4170
312	242.7173	226.2062	16.5111	0.0680

Tabla 4.12: Valores reales y pronosticados de 5 semanas.

Valor pronosticado	Intervalo de confianza al 80 %		Intervalo de confianza al 95 %	
236.5758	193.0705	280.0810	170.0402	303.1113
234.0264	190.5212	277.5317	167.4909	300.5620
218.3629	174.8576	261.8682	151.8273	284.8985
231.4973	187.9920	275.0025	164.9617	298.0328
226.2062	182.7009	269.7114	159.6706	292.7417

Tabla 4.13: Intervalo de confianza al 80 % y al 95 %.

Y en la Figura(4.19) se observan los datos pronosticados.

En cuanto al análisis de residuos, el diagnóstico del modelo 2 puede verse en la Figura (4.20). Obsérvese que en la primera imagen de la Figura (4.20) se muestran los residuales estandarizados del modelo, donde se observa que están distribuidos de forma aleatoria alrededor del cero, sin mostrar tendencia, por lo que se consideran independientes.

Semana	Error relativo modelo 1	Error relativo modelo R	Error relativo modelo 2
308	0.1136	0.0829	0.1059
309	0.0446	0.0171	0.0438
310	0.0497	0.0201	0.0493
311	0.4158	0.3623	0.4170
312	0.0688	0.0915	0.0680

Tabla 4.14: Comparación de errores relativos de los modelos 1, R y 2.

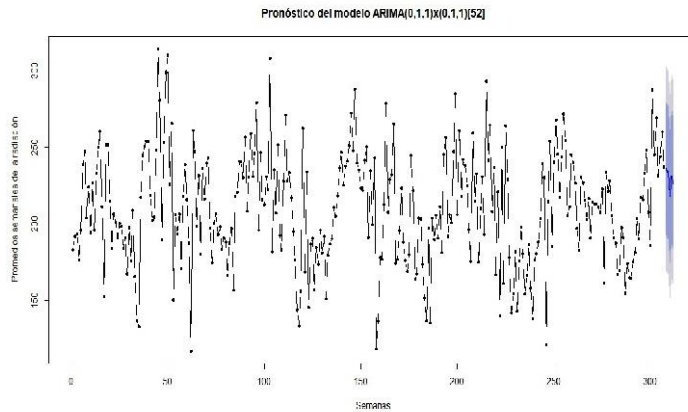


Figura 4.19: Datos pronosticados.

En la segunda imagen, la FAC de los residuales se asemeja a la FAC de un ruido blanco, por lo que parece que se está ante un modelo adecuado (también vea Figura (4.21), donde se muestra la FAC y FACP).

Finalmente, para la tercera imagen, se observa el test de Ljung-Box, que como se describió en 3.6, se puede deducir del tercer gráfico, que no se rechazar la hipótesis nula, ya que el valor p es mayor que 0.05. Tenga en cuenta lo siguiente.

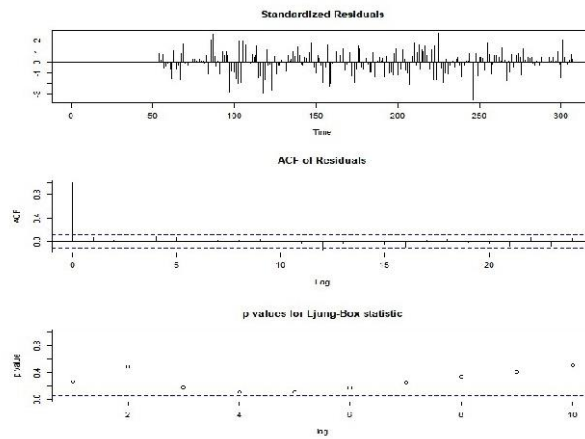


Figura 4.20: Diagnóstico del modelo 2.

Como se dijo antes el test de Ljung-Box está probando la no correlación más no la independencia.

El test de Ljung-Box de manera numérica arrojó un p valor de 0.259, donde en efecto el p valor es mayor que 0.05, por lo que no se rechaza H_0 .

En la Figura 4.22(c), se observa en el Q-Q plot que el conjunto de datos está distribuido “aceptablemente” normal, se observa que también tiene las colas muy largas (de forma similar que en 4.14(c)), por otra parte, note que en el histograma 4.22(a) hay presencia de leve asimetría, y en el diagrama de caja 4.22(b) es más visible el ligero sesgo a la derecha a comparación del diagrama de caja del modelo 1 (4.14(b)), de igual forma que en el modelo 1, dado que el supuesto de normalidad es robusto se puede seguir considerando dicho supuesto.

Existen otros métodos más exactos de tipo numérico para saber si los datos se ajustan a la distribución normal, como por ejemplo: Shapiro- Wilk, Kolmogorov-Smirnov, etc.

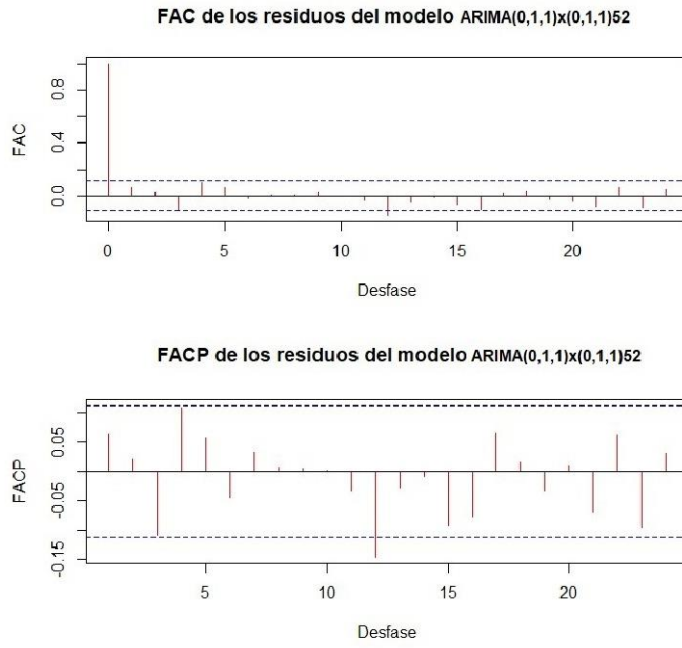
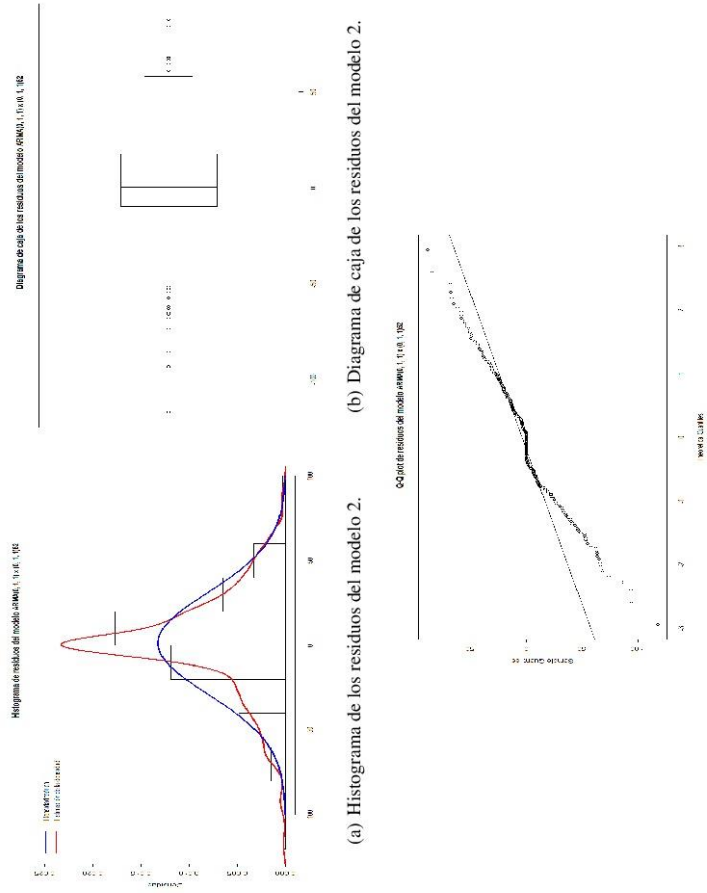


Figura 4.21: FAC y FACP de los residuos del modelo 2.



(a) Histograma de los residuos del modelo 2. (b) Diagrama de caja de los residuos del modelo 2.

(c) Q-Q plot de los residuos del modelo 2.

Figura 4.22: Análisis de normalidad de los residuos del modelo 2.

DISCUSIÓN Y CONCLUSIONES

El desarrollo de este trabajo, comenzó explicando cuál es la diferencia que hay entre la definición de el clima y el tiempo, posteriormente se presentó la definición de las variables atmosféricas, que gracias a ellas es posible el estudio del tiempo y el clima, entre las variables mencionadas en este trabajo se encuentra la radiación solar, siendo esta la variable meteorológica de interés primordial para este trabajo.

Se presentaron los conceptos básicos de series de tiempo, donde fueron de gran importancia los conceptos de estacionariedad débil, función de autocorrelación y autocorrelación parcial; se mostraron modelos para series de tiempo estacionarias y no estacionarias, así como estacionales y no estacionales. En seguida, se procedió a la identificación, estimación, diagnóstico y selección del modelo, para así dar paso al pronóstico.

Se menciona también en que región fueron tomados los datos de la radiación solar, e incluso algunos fenómenos meteorológicos como “El niño”, “La niña” y la canícula, que son fenómenos que influyen en el análisis de los datos. Además, se explicó el proceso de la filtración de los datos, y una vez filtrada se procedió a verificar la estacionariedad de la serie, ya que es un supuesto importante para aplicar la metodología de Box-Jenkins. En este análisis al aplicar Dickey-Fuller se concluye que la serie inicial ya es estacionaria, sin embargo, partiendo de este hecho no se pudo formular un modelo que se ajustara a los datos. De forma que se procedió a diferenciar la serie en la parte no estacional y posteriormente se aplicaron los pasos correspondientes de la teoría de Box-Jenkins expuesta en este trabajo.

Así, ya analizados los resultados de los modelos 1, 2 y el modelo de R, el modelo que mejor se ajusta a los datos de la radiación solar, tomados en una estación meteorológica ubicada en la Ciudad de Tlaxco - Tlaxcala, es el modelo dos, pues a consideración nuestra tiene mejor pronóstico además de tener menor cantidad de variables, su error relativo es menor en comparación al modelo uno y sobre todo que en este modelo si se toma en cuenta la estacionalidad, por lo tanto, debido al criterio de parsimonia se decide considerar que el mejor modelo es el $ARIMA(0, 1, 1) \times (0, 1, 1)_{52}$. El cual después de estimar sus parámetros por máxima verosimilitud queda de la siguiente forma:

$$X_t = \varepsilon_t + \varepsilon_{t-1} + 0.805794\varepsilon_{t-52} + 0.805794\varepsilon_{t-53} + X_{t-1} + X_{t-52} - X_{t-53}$$

Como un posible trabajo a futuro se tienen las siguientes propuestas:

- Agregar más variables meteorológicas al modelo que influyen en la radiación solar y junto con esta, hacer más exacto el análisis usando la teoría de Box-Jenkins.
- De manera externa, con la teoría de Box-Jenkins realizar un análisis sobre las emisiones contaminantes de los autos, como por ejemplo el dióxido de carbono, el monóxido de carbono, óxido de nitrógeno, etc.

REPRESENTACIÓN DE UN PROCESO AR(p) COMO UN MA(∞)

En los procesos AR(1), se observó que estos pueden ser establecidos como MA(∞), uno podría preguntarse si los procesos AR(2) y los de orden p también pueden ser expresados como MA de orden infinito. Pues la respuesta es sí, por ejemplo, dado un proceso invertible AR(p)

$$\phi_p(B)X_t = \varepsilon_t \quad (\text{A.1})$$

donde $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, puede ser escrito de la forma

$$X_t = \frac{1}{\phi_p(B)} \varepsilon_t = \psi(B) \varepsilon_t \quad (\text{A.2})$$

donde $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots$, tal que

$$\phi_p(B)\psi(B) = 1. \quad (\text{A.3})$$

Por ejemplo, considere el proceso AR(2), esto es,

$$X_t = \frac{1}{1 - \phi_1 B - \phi_2 B^2} \varepsilon_t = (1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots) \varepsilon_t \quad (\text{A.4})$$

esto implica que

$$(1 - \phi_1 B - \phi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots) = 1.$$

Entonces

$$\begin{aligned} & 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots \\ & - \phi_1 B - \psi_1 \phi_1 B^2 - \psi_2 \phi_1 B^3 - \psi_3 \phi_1 B^4 - \dots \\ & - \phi_2 B^2 - \psi_1 \phi_2 B^3 - \psi_2 \phi_2 B^4 - \psi_3 \phi_2 B^5 - \dots \end{aligned}$$

Luego, para obtener los ψ'_j s se tiene

$$\begin{aligned} B^1 : \psi_1 - \phi_1 &= 0 \Rightarrow \psi_1 = \phi_1 \\ B^2 : \psi_2 - \psi_1 \phi_1 - \phi_2 &= 0 \Rightarrow \psi_2 = \psi_1 \phi_1 + \phi_2 \\ B^3 : \psi_3 - \psi_2 \phi_1 - \psi_1 \phi_2 &= 0 \Rightarrow \psi_3 = \psi_2 \phi_1 + \psi_1 \phi_2 \\ & \vdots \end{aligned}$$

por lo que para $j \geq 2$ se tiene

$$\psi_j = \psi_{j-1} \phi_1 + \psi_{j-2} \phi_2, \quad \psi_0 = 1. \quad (\text{A.5})$$

Un caso especial resulta cuando $\phi_2 = 0$, entonces el proceso se convierte en un proceso AR(1), es decir

$$\begin{aligned} \psi_1 &= \phi_1 \\ \psi_2 &= \psi_1 \phi_1 = \phi_1^2 \\ \psi_3 &= \psi_2 \phi_1 = \phi_1^3 \\ & \vdots \\ \psi_j &= \phi_1^j, \quad j \geq 0 \\ & \vdots \end{aligned}$$

así

$$X_t = \frac{1}{1 - \phi_1 B} \varepsilon_t = (1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots) \varepsilon_t. \quad (\text{A.6})$$

Ahora bien, si este mismo procedimiento se sigue para un proceso AR(3), se obtendrá

$$\begin{aligned}
B^1 : \psi_1 - \phi_1 = 0 &\Rightarrow \psi_1 = \phi_1 \\
B^2 : \psi_2 - \psi_1\phi_1 - \phi_2 = 0 &\Rightarrow \psi_2 = \psi_1\phi_1 + \phi_2 \\
B^3 : \psi_3 - \psi_2\phi_1 - \psi_1\phi_2 - \phi_3 = 0 &\Rightarrow \psi_3 = \psi_2\phi_1 + \psi_1\phi_2 + \phi_3 \\
&\vdots
\end{aligned}$$

así para $j \geq 3$ se tiene

$$\psi_j = \psi_{j-1}\phi_1 + \psi_{j-2}\phi_2 + \psi_{j-3}\phi_3, \quad \psi_0 = 1. \quad (\text{A.7})$$

Siguiendo este mismo razonamiento, se tendrá para un AR(p)

$$\psi_j = \psi_{j-1}\phi_1 + \psi_{j-2}\phi_2 + \dots + \psi_{j-(p-1)}\phi_{p-1} + \psi_{j-p}\phi_p, \quad \psi_0 = 1. \quad (\text{A.8})$$

Por lo tanto, todo proceso AR de orden finito, se puede expresar de forma equivalente como un proceso MA de orden infinito.

Ahora, también es posible escribir a los procesos MA como proceso AR de orden infinito.

Dado el proceso estacionario MA(q)

$$X_t = \theta_q(B)\varepsilon_t \quad (\text{A.9})$$

con $\theta_q = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$, es posible escribirlo de la siguiente manera

$$\pi(B)X_t = \frac{1}{\theta_q(B)}X_t = \varepsilon_t \quad (\text{A.10})$$

en este caso $\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \dots)$, tal que

$$\pi(B)\theta_q(B) = 1. \quad (\text{A.11})$$

Por ejemplo, para el proceso MA(2)

$$\varepsilon_t = \frac{X_t}{1 - \theta_1(B) - \theta_2 B^2} = (1 - \pi_1 B - \pi_2 B^2 - \dots)X_t \quad (\text{A.12})$$

esto implica que

$$(1 - \theta_1 B - \theta_2 B^2)(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \dots) = 1.$$

Entonces

$$\begin{aligned} & 1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \dots \\ & - \theta_1 B + \pi_1 \theta_1 B^2 + \pi_2 \theta_1 B^3 + \pi_3 \theta_1 B^4 + \dots \\ & - \theta_2 B^2 + \pi_1 \theta_2 B^3 + \pi_2 \theta_2 B^4 + \pi_3 \theta_2 B^5 + \dots \end{aligned}$$

Luego, para obtener los π'_j s se tiene

$$\begin{aligned} B^1 : -\pi_1 - \theta_1 &= 0 \Rightarrow \pi_1 = -\theta_1 \\ B^2 : -\pi_2 + \pi_1 \theta_1 - \theta_2 &= 0 \Rightarrow \pi_2 = -\theta_1^2 - \theta_2 \\ B^3 : -\pi_3 + \pi_2 \theta_1 + \pi_1 \theta_2 &= 0 \Rightarrow \pi_3 = \pi_2 \theta_1 + \pi_1 \theta_2 \\ & \vdots \end{aligned}$$

en general $j \geq 2$ se tiene

$$\pi_j = \pi_{j-1} \theta_1 + \pi_{j-2} \theta_2. \quad (\text{A.13})$$

Un caso particular es cuando $\theta_2 = 0$, entonces el proceso se convierte a un proceso MA(1), esto es, $\pi_j = -\theta_1^j$, $j \geq 1$, y

$$(1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \dots) X_t = \frac{X_t}{1 - \theta_1 B} = \varepsilon_t. \quad (\text{A.14})$$

Por lo tanto un proceso MA estacionario de orden finito es equivalente a un proceso AR de orden infinito. [31]

CONDICIONES DE LAS RAÍCES DEL PROCESO AR(2)

Como se menciona en el Capítulo 1, en los procesos AR(2) las raíces del polinomio de segundo orden son de la forma

$$B_1 = \frac{-\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}, \quad B_2 = \frac{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}. \quad (\text{B.1})$$

Para que el proceso AR(2) sea estacionario, se requiere que las raíces sean mayores a 1 en valor absoluto. Lo cual se cumple si las siguientes condiciones se satisfacen ¹

$$\phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1, \quad |\phi_2| < 1. \quad (\text{B.2})$$

Prueba

Tomando el recíproco de B_1 y haciendo las operaciones correspondientes se tiene

¹Las raíces de un proceso MA(2) son de forma similar a las del proceso AR(2), al igual que la prueba para las condiciones. Pero en lugar de ϕ_i , $i = 1, 2$ se sustituirá por θ_i , $i = 1, 2$.

$$\begin{aligned}
B'_1 = \frac{1}{B_1} &= \frac{2\phi_2}{-\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}} = \frac{2\phi_2}{-\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}} \left[\frac{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}} \right] \\
&= \frac{2\phi_2(-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})}{\phi_1^2 - (\phi_1^2 + 4\phi_2)} = \frac{2\phi_2(-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})}{-4\phi_2} \\
&= \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}.
\end{aligned} \tag{B.3}$$

De forma análoga se tiene que el recíproco de B_2 es de la forma

$$B'_2 = \frac{1}{B_2} = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}. \tag{B.4}$$

La condición que se requiere es $|B_i| > 1$ esto implica que $|B'_i| = \left| \frac{1}{B_i} \right| < 1$ para $i = 1, 2$.

Ahora bien, la prueba se divide en dos casos: raíces reales y raíces complejas.

Caso 1 : Raíces reales

Las raíces serán reales si y sólo si $\phi_1^2 + 4\phi_2 \geq 0$. Dado que $|B'_i| < 1$ para $i = 1, 2$, se tiene

$$-1 < \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} < \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} < 1$$

o bien

$$-2 < \phi_1 - \sqrt{\phi_1^2 + 4\phi_2} < \phi_1 + \sqrt{\phi_1^2 + 4\phi_2} < 2.$$

Luego, considérese la primera desigualdad $-2 < \phi_1 - \sqrt{\phi_1^2 + 4\phi_2}$, si y sólo si, $\sqrt{\phi_1^2 + 4\phi_2} < \phi_1 + 2$, si y sólo si, $\phi_1^2 + 4\phi_2 < \phi_1^2 + 4\phi_1 + 4$, si y sólo si, $\phi_2 < \phi_1 + 1$, esto implica que $\phi_2 - \phi_1 < 1$.

Siguiendo el mismo procedimiento para la segunda desigualdad $\phi_1 + \sqrt{\phi_1^2 + 4\phi_2} < 2$, se llega a $\phi_2 + \phi_1 < 1$.

Por lo tanto, estas ecuaciones junto con $\phi_1^2 + 4\phi_2 \geq 0$, definen la condición de estacionariedad para el caso raíz real.

Caso 2 : Raíces complejas

Las raíces serán complejas si $\phi_1^2 + 4\phi_2 < 0$. Obsérvese que el complejo conjugado de G_1 es G_2 y $|G_1| = |G_2| < 1$ si y sólo si $|G_1|^2 < 1$. Pero

$$\begin{aligned} |G_1|^2 &= G_1 \bar{G}_1 = G_1 \cdot G_2 = \left(\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} \right) \left(\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} \right) \\ &= \frac{1}{4} \left(\phi_1^2 + \phi_1 \sqrt{\phi_1^2 + 4\phi_2} - \phi_1 \sqrt{\phi_1^2 + 4\phi_2} - (\phi_1^2 + 4\phi_2) \right) \\ &= -\phi_2. \end{aligned}$$

Por lo que $\phi_2 > -1$ y junto con la desigualdad $\phi_1^2 + 4\phi_2 < 0$ definen la condición de estacionariedad para el caso de raíz compleja. [14]

TEST DE DICKEY- FULLER

Existen muchas pruebas para determinar si una serie es estacionaria o no es estacionaria. La prueba más popular es la prueba de **Dickey-Fuller**. Hay tres variaciones de la prueba Dickey- Fuller. Esta fue diseñada para tener en cuenta el papel del término constante y la tendencia. Se inicia describiendo las ecuaciones de prueba y las hipótesis para estos tres casos y después describir el procedimiento de la prueba. [19]

- **Prueba 1: sin término constante y sin tendencia**

Considere el proceso AR(1)

$$X_t = \phi_1 X_{t-1} + \varepsilon_t \quad (\text{C.1})$$

el cual será débilmente estacionario si se cumple que, $|\phi_1| < 1$, pero cuando $\phi_1 = 1$ el proceso se convierte en un proceso de caminata aleatoria y el proceso no será estacionario, es decir

$$X_t = X_{t-1} + \varepsilon_t.$$

Por lo que, una forma de probar la estacionariedad es examinar el valor de ϕ_1 . Es decir, se probará si ϕ_1 es igual a uno o significativamente menor a uno. A este tipo de pruebas se le conoce como **pruebas de raíz unitaria para estacionariedad**.

Para tener una idea de esto, considere de nuevo el proceso (C.1). Donde ε_t es un $RB(0, \sigma_\varepsilon^2)$. Es posible justificar la no estacionariedad probando

$$H_0 : \phi_1 = 1 \quad vs \quad H_a : |\phi_1| < 1 \quad (\text{o simplemente } \phi_1 < 1).$$

Esta prueba unilateral (cola izquierda) se puede expresar en una forma más conveniente al restar X_{t-1} en (C.1), esto es,

$$\begin{aligned} X_t - X_{t-1} &= \phi_1 X_{t-1} - X_{t-1} + \varepsilon_t \\ \nabla X_t &= (\phi_1 - 1)X_{t-1} + \varepsilon_t \\ \nabla X_t &= \beta X_{t-1} + \varepsilon_t. \end{aligned} \quad (\text{C.2})$$

Donde $\beta = \phi_1 - 1$. Así, es posible escribir la prueba de hipótesis en términos de ϕ_1 o de β , es decir

$$\begin{aligned} H_0 : \phi_1 = 1 &\Leftrightarrow H_0 : \beta = 0. \\ H_a : \phi_1 < 1 &\Leftrightarrow H_a : \beta < 0. \end{aligned} \quad (\text{C.3})$$

No perder de vista que la hipótesis nula es que se trata de una serie no estacionaria. En otras palabras, si no se rechaza H_0 , se concluiría que se trata de un proceso no estacionario, de lo contrario, es decir si se rechaza H_0 , entonces se concluye que la serie es estacionaria.

■ Prueba 2: con término constante pero sin tendencia

En la segunda prueba de Dickey-Fuller en la ecuación C.1 se incluye el término constante, esto es,

$$X_t = c + \phi_1 X_{t-1} + \varepsilon_t, \quad \text{donde } \varepsilon_t \sim RB(0, \sigma_\varepsilon^2). \quad (\text{C.4})$$

Expresando (C.4) en una forma más conveniente al restar X_{t-1} resulta

$$\begin{aligned} X_t - X_{t-1} &= c + \phi_1 X_{t-1} - X_{t-1} + \varepsilon_t \\ \nabla X_t &= c + (\phi_1 - 1)X_{t-1} + \varepsilon_t \\ \nabla X_t &= c + \beta X_{t-1} + \varepsilon_t. \end{aligned} \quad (\text{C.5})$$

Entonces, el contraste de hipótesis es escrito de la siguiente forma

$$H_0 : \beta = 0 \quad vs \quad H_a : \beta < 0. \quad (\text{C.6})$$

De la misma manera, si no se rechaza H_0 se concluye que la serie no es estacionaria. Si se rechaza H_0 se concluye que la serie es estacionaria.

■ **Prueba 3: con término constante y con tendencia**

En la tercera prueba de Dickey-Fuller en la ecuación C.1 se incluye el término constante y tendencia, esto es,

$$\nabla X_t = c + \beta X_{t-1} + \lambda t + \varepsilon_t, \quad \text{donde } \varepsilon_t \sim RB(0, \sigma_\varepsilon^2). \quad (\text{C.7})$$

La hipótesis nula y alternativa son igual como antes $H_0 : \beta = 0$ vs $H_a : \beta < 0$. Si no se rechaza H_0 se concluye que la serie no es estacionaria. En caso contrario se dirá que la serie es estacionaria.

Valores críticos de Dickey-Fuller

Para probar la hipótesis de los tres casos, simplemente se estima la ecuación de prueba por **mínimos cuadrados** y se examina el estadístico t bajo H_0 . Por desgracia, este estadístico t bajo H_0 no tiene la distribución t que se conoce comúnmente. Esto se debe porque cuando la hipótesis nula es verdadera X_t es no estacionaria y tiene una varianza que aumenta a medida que el tamaño de la muestra aumenta. Esta varianza crece y altera la distribución del estadístico t usual cuando H_0 es verdadera.

A este estadístico t se le denomina **estadístico de Dickey-Fuller** y se denota por τ y su valor debe compararse con los valores críticos. Esto es,

- Se rechaza H_0 (la hipótesis de no estacionariedad), a un nivel de significancia α si

$$\tau < \tau_c. \quad (\text{C.8})$$

Donde τ bajo H_0 es de la forma

$$\tau = t = \frac{\hat{\beta} - \beta}{\sqrt{\hat{V}[\hat{\beta}]}} \quad (\text{C.9})$$

donde $\hat{\beta}$ es el estimador de mínimos cuadrados de β y $\hat{V}[\hat{\beta}]$ el de su varianza. Y τ_c es el valor crítico de las tablas de Dickey-Fuller.

- Y si $\tau > \tau_c$ no se rechaza la hipótesis nula.

En la tabla (C.1) se muestran algunos de los valores críticos más usados para el estadístico τ para los tres casos.¹

Ahora bien, se han visto tres pruebas diferentes de Dickey-Fuller, pero la cuestión es ¿Cómo se va a decidir que prueba usar? Para ello se debe observar lo siguiente

- Si la serie parece estar fluctuando alrededor de una muestra de media cero, usar la prueba Dickey-Fuller 1.
- Si la serie parece estar fluctuando alrededor de una muestra de media que no es cero, usar la prueba Dickey-Fuller 2.
- Si la serie parece estar fluctuando alrededor de una tendencia lineal, usar la prueba Dickey-Fuller 3.

Modelo	1 %	5 %	10 %
$\nabla X_t = \beta X_{t-1} + \varepsilon_t$	-2.56	-1.94	-1.62
$\nabla X_t = c + \beta X_{t-1} + \varepsilon_t$	-3.43	-2.86	-2.57
$\nabla X_t = c + \beta X_{t-1} + \lambda t + \varepsilon_t$	-3.96	-3.41	-3.13
Valores criticos estándar	-2.33	-1.65	-1.28

Tabla C.1: Valores críticos para las pruebas Dickey-Fuller

¹Obsérvese que los valores críticos de Dickey-Fuller (los de la tercera fila de la tabla (C.1)) son más negativos si se comparasen con los valores críticos estándar. Esto implica que la estadística τ debe tomar valores más grandes (negativos) para que la hipótesis nula H_0 sea rechazada a favor de la alternativa.

■ **Prueba Dickey-Fuller aumentado**

Considere el proceso AR(p)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2). \quad (\text{C.10})$$

La expresión anterior se puede reparametrizar de la siguiente forma

$$\nabla X_t = \beta X_{t-1} + a_1 \nabla X_{t-1} + \cdots + a_{p-1} \nabla X_{t-p+1} + \varepsilon_t. \quad (\text{C.11})$$

Donde

$$\beta = \sum_{i=1}^p \phi_i - 1 \quad \text{y} \quad a_i = \sum_{j=1}^i \phi_{p-i+j}, \quad i = 1, \dots, p-1.$$

Entonces, el contraste de hipótesis resulta ser

$$H_0 : \beta = 0 \quad \text{vs} \quad H_a : \beta < 0. \quad (\text{C.12})$$

A este contraste de raíz unitaria se le conoce como **Dickey-Fuller Aumentado (DFA)** y se basa en la estimación de mínimos cuadrados ordinarios del parámetro β y el correspondiente estadístico τ , el cual tiene la misma distribución que en el caso del proceso AR(1), por lo que, es posible usar los mismos valores críticos de la tabla (C.1). En ocasiones el modelo DFA también puede ser presentado con la inclusión del término constante y/o tendencia, esto es,

$$\begin{aligned} \nabla X_t &= c + \beta X_{t-1} + a_1 \nabla X_{t-1} + \cdots + a_{p-1} \nabla X_{t-p+1} + \varepsilon_t. \\ \nabla X_t &= c + \lambda t + \beta X_{t-1} + a_1 \nabla X_{t-1} + \cdots + a_{p-1} \nabla X_{t-p+1} + \varepsilon_t. \end{aligned} \quad (\text{C.13})$$

Y se realizan los contrastes de hipótesis de manera similar al proceso AR(1).

CRITERIOS DE SELECCIÓN

Criterio de Akaike's y BIC ¹

Se asume que un modelo de M parámetros se ajusta a un conjunto de datos para evaluar la calidad del modelo apropiado. **Akaike (1973,1974b)** (ver [3]) introduce un criterio de información llamado criterio **AIC (Akaike's Information Criterion)** dado de la siguiente manera

$$AIC(M) = -2 \ln [L] + 2M$$

donde M es número de parámetros en el modelo y L la función de verosimilitud evaluada en las estimaciones de los parámetros. Dado un modelo ARMA y n el número de observaciones, la función de log-verosimilitud es

$$\ln[L(\phi, \theta, \mu, \sigma_\varepsilon^2)] = -\frac{n}{2} \ln (2\pi\sigma_\varepsilon^2) - \frac{S(\phi, \mu, \theta)}{2\sigma_\varepsilon^2} \quad (D.1)$$

maximizando respecto a $\phi, \theta, \mu, \sigma_\varepsilon^2$ se tiene

$$\ln[\hat{L}] = -\frac{n}{2} \ln (\hat{\sigma}_\varepsilon^2) - \frac{n}{2}(1 + \ln 2\pi). \quad (D.2)$$

Dado que el segundo término de (D.2) es una constante, el criterio se reduce a

$$AIC(M) = n \ln (\hat{\sigma}_\varepsilon^2) + 2M. \quad (D.3)$$

¹Estos criterios suelen emplearse para decidir acerca de el orden de p, P, q y Q.

El valor de M depende de p y q los cuáles deben ser seleccionados de modo que $AIC(M)$ sea mínimo.

Shibata(1976) (ver [28]) demostró que el AIC tiende a sobreestimar el orden de la autorregresión. **Akaike(1978,1979)** (ver [4], [5]) desarrollaron una extensión Bayesiana del procedimiento conocido como criterio **BIC (Bayesian Information Criterion)**, que es de la forma

$$BIC(M) = n \ln \hat{\sigma}_\varepsilon^2 - (n - M) \ln \left(1 - \frac{M}{n} \right) + M \ln n + M \ln \left[\left(\frac{\hat{\sigma}_X^2}{\hat{\sigma}_\varepsilon^2} - 1 \right) / M \right] \quad (D.4)$$

donde $\hat{\sigma}_\varepsilon^2$ es el estimador de máxima verosimilitud de σ_ε^2 , M es el número de parámetros y $\hat{\sigma}_X^2$ es la varianza muestral de la serie.

Akaike(1978) realizó un estudio de simulación, donde afirma que es menos probable que el BIC sobre estime el orden de la autorregresión.

Existen otros criterios como el criterio de Schwartz's SBC (Schwartz's Bayesian Criterion) y el criterio de Parzen's CAT (Criterion for Autoregressive Transfer functions) que pueden consultarse en [31].

SELECCIÓN DE DATOS

Para la selección de datos se realizó lo siguiente.

1. En la base de datos de de Excel **Tlaxco 2011-2017** se encuentra información a partir del 18 de Junio del 2011 al 21 de Junio del 2017. En la base de datos, se tiene que estos datos son recolectados diariamente cada media hora, por ejemplo, para el 18/06/2011 se tiene lo siguiente.

- De la Figura (E.1), note que la recolección de los datos empieza 12:00 a.m. y a partir de ahí cada media hora, hasta las 11:30 p.m.

Siguiendo sobre el mismo ejemplo, se prosiguió a hallar el promedio de la columna radiación solar de la fecha 18/06/2011 tomando todos los valores e incluso los ceros que aparecen. Así, se obtuvo que el promedio del 18/06/2011 fue 231.8958333.

De manera análoga se hizo lo mismo para toda la base de datos, siempre que estuviesen completos.

2. Ahora bien, se observó una falta de datos de ciertos meses en los años 2011, 2014, 2016 y 2017, pero los años 2012, 2013 y 2015 se encontraban completos. Así que, para a completar los datos faltantes se realizó lo siguiente.
 - Primero se comenzó a completar los datos faltantes del año 2014 de cierto mes, esto se hizo tomando la misma fecha de los dos años anteriores y uno siguiente y promediarlos (ver Figura (E.2)).

De la Figura (E.2), en el año 2014 hacían falta los datos del 16 de Julio, así que, se tomo la misma fecha de los dos años anteriores, es decir: 16/07/2012, 16/07/2013 y la del año siguiente al 2014, que es 16/07/2015, (esto fue porque se tenían únicamente esos años completos). Y en cuanto al 16/07/2012 note que tiene un promedio de 189.708333, este fue calculado como se explicó en I (lo mismo sucede para los años 2013 y 2015).

Una vez juntados los respectivos promedios (como se puede ver en la Figura (E.2)), se promediaron nuevamente y ese respectivo promedio que fue de 183.395833 se le asigno al 16/07/2014. Se continuo con este proceso hasta a completar el año 2014.

Una vez completos los datos del año 2014, se completo el año 2011 de manera análoga, pero tomando los años 2012, 2013 y 2014.

Para completar el año 2016, se tomaron los años 2013, 2014 y 2015.

Y para el año 2017 se tomaron los años 2014, 2015 y 2016.

3. Una vez filtrados y completos los años de la base de datos, se prosiguió a juntar todos los días de los años y se empezó a promediar semanalmente. Es decir (ver Figura (E.3)).
 - Obsérvese que a partir del día 18/06/2011 al 24/06/2011 empieza la primera semana, así que, el promedio de la semana del 18 al 24 de Junio de 2011 resulta ser 183.3571429, la segunda semana empieza a partir del 25 de Junio de 2011 al 01 de Julio de 2011 y su respectivo promedio es 192.140873, se continua sucesivamente con este proceso hasta obtener una base de datos de promedios semanales que es con lo que se realiza el análisis.

Fecha	Hora	Rad. Solar
18/06/2011	12:00 a. m.	0
18/06/2011	12:30 a. m.	0
18/06/2011	01:00 a. m.	0
18/06/2011	01:30 a. m.	0
18/06/2011	02:00 a. m.	0
18/06/2011	02:30 a. m.	0
18/06/2011	03:00 a. m.	0
18/06/2011	03:30 a. m.	0
18/06/2011	04:00 a. m.	0
18/06/2011	04:30 a. m.	0
18/06/2011	05:00 a. m.	0
18/06/2011	05:30 a. m.	0
18/06/2011	06:00 a. m.	0
18/06/2011	06:30 a. m.	0
18/06/2011	07:00 a. m.	1
18/06/2011	07:30 a. m.	21
18/06/2011	08:00 a. m.	113
18/06/2011	08:30 a. m.	244
18/06/2011	09:00 a. m.	365
18/06/2011	09:30 a. m.	481
18/06/2011	10:00 a. m.	580
18/06/2011	10:30 a. m.	691
18/06/2011	11:00 a. m.	369
18/06/2011	11:30 a. m.	281
18/06/2011	12:00 p. m.	431
18/06/2011	12:30 p. m.	564
18/06/2011	01:00 p. m.	474
18/06/2011	01:30 p. m.	607
18/06/2011	02:00 p. m.	401
18/06/2011	02:30 p. m.	726
18/06/2011	03:00 p. m.	793
18/06/2011	03:30 p. m.	540
18/06/2011	04:00 p. m.	779
18/06/2011	04:30 p. m.	693
18/06/2011	05:00 p. m.	739
18/06/2011	05:30 p. m.	200
18/06/2011	06:00 p. m.	197
18/06/2011	06:30 p. m.	361
18/06/2011	07:00 p. m.	337
18/06/2011	07:30 p. m.	109
18/06/2011	08:00 p. m.	32
18/06/2011	08:30 p. m.	2
18/06/2011	09:00 p. m.	0
18/06/2011	09:30 p. m.	0
18/06/2011	10:00 p. m.	0
18/06/2011	10:30 p. m.	0
18/06/2011	11:00 p. m.	0
18/06/2011	11:30 p. m.	0

Figura E.1: Datos del 18 de Junio de 2011

	Fecha	Hora	Promedio
16/07/2014	16/07/2012	11:30 p. m.	189.708333
	16/07/2013	11:30 p. m.	161.1875
	16/07/2015	11:30 p. m.	199.291667
	Promedio		183.395833

Figura E.2: Datos faltantes

Fecha	Hora	Promedio	Promedio semanal
18/06/2011	11:30 p. m.	231.8958333	
19/06/2011	11:30 p. m.	250.9583333	
20/06/2011	11:30 p. m.	228.1458333	
21/06/2011	11:30 p. m.	99.66666667	
22/06/2011	11:30 p. m.	179.4722222	
23/06/2011	11:30 p. m.	128.0902778	
24/06/2011	11:30 p. m.	165.2708333	183.3571429 ←
25/06/2011	11:30 p. m.	204.1527778	
26/06/2011	11:30 p. m.	207.9236111	
27/06/2011	11:30 p. m.	176.625	
28/06/2011	11:30 p. m.	191.0347222	
29/06/2011	11:30 p. m.	178.8333333	
30/06/2011	11:30 p. m.	201.8680556	
01/07/2011	11:30 p. m.	184.5486111	192.140873 ←

Figura E.3: Promedio semanal

ANÁLISIS DE DIAGRAMAS DE CAJA

Los elementos que conforman la caja son: **primer cuartil (Q_1)**, **segundo cuartil (Q_2)**, **tercer cuartil (Q_3)**, **caja**, **mediana**, **bigotes**, **valores mínimo y máximo** y **valores atípicos**, ver Figura (F.1(a)). A través de estos elementos es posible analizar el comportamiento de los datos, su simetría, los valores extremos, la mediana, etcétera. Las partes del diagrama de caja se describen de la siguiente manera:

1. Los cuartiles dividen en cuatro partes iguales a la cantidad de datos, en cada parte se concentra un 25 % de la totalidad de los datos.
 - Cuartil 1: coincide con el borde inferior de la caja y es el valor por debajo del cual ocurre el 25 % de las observaciones.
 - Cuartil 2: coincide con la mediana e indica que el 50 % de los valores están por debajo de la mediana y el otro 50 % está por encima.
 - Cuartil 3: coincide con el borde superior de la caja y es el valor por debajo del cual ocurre el 75 % de las observaciones.
2. Caja: es un rectángulo donde el borde inferior de la caja representa el Q_1 y el borde superior el Q_3 . Y se le llama rango intercuartílico (RIQ) a la diferencia entre Q_3 y Q_1 , es decir, $RIQ = Q_3 - Q_1$, y es ahí donde se concentra el 50 % de los valores centrales.
3. Mediana: es la línea horizontal que está dentro de la caja y que coincide con el segundo cuartil. Divide a la distribución en dos partes iguales.

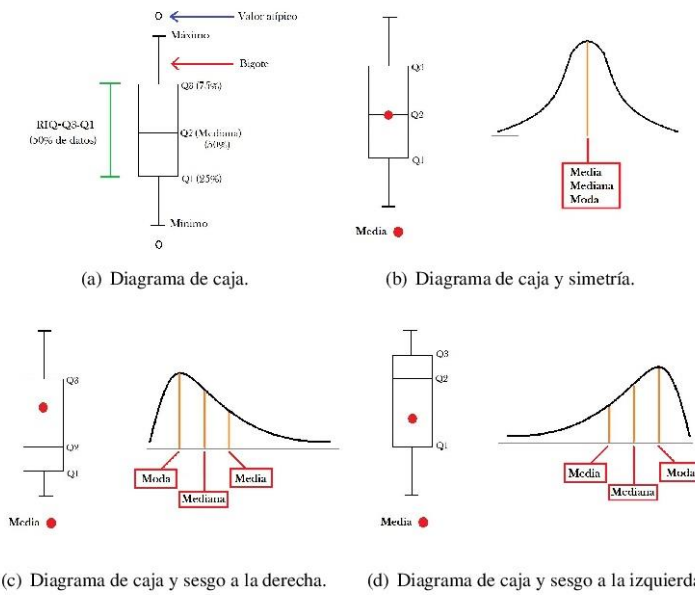


Figura F.1: Diagramas de caja y sesgo.

4. Bigotes: son los segmentos que salen de la caja e indican el intervalo en que los valores se consideran normales.
5. Valores mínimo y máximo: son el extremo inferior y superior que alcanzan el mínimo y máximo (respectivamente) de los datos, están marcados con una pequeña línea horizontal al final del bigote.
6. Valores atípicos: son aquellos valores que se encuentran fuera de los bigotes ya sea por arriba o por debajo. Pueden representar causas extrañas, errores de medición o de captura. Están representados por un (°).

Se debe tener en cuenta lo siguiente para la interpretación de las cajas:

1. Entre más larga sea la caja y los bigotes más variabilidad central habrá en

los datos. Y por el contrario, entre más estrecha la caja y los bigotes habrá menos dispersión central o bien menos variabilidad central.

2. La distancia entre los cuartiles puede variar, pero a pesar de ello, la cantidad de elementos en cada cuartil es aproximadamente la misma.
3. La posición de la mediana nos indica la simetría, que a través de la simetría es posible saber en donde existe mayor concentración de datos, es decir:
 - Si la mediana se encuentra en el centro de la caja, se dice que la distribución es **simétrica**, esto indica que los datos están distribuidos de igual forma a la izquierda y a la derecha de la media. En este caso la media, la mediana y la moda coinciden. Ver Figura (F.1(b)).
 - Si la mediana se encuentra más cerca del Q_1 , se dice que la distribución es **sesgada a la derecha o bien tiene una asimetría positiva**. Es decir, que la cola derecha de la distribución es más larga que la izquierda, esto indica que los valores a la derecha de la media están más dispersos que en el lado contrario. Ver Figura (F.1(c)).
 - Si la mediana se encuentra más cerca del Q_3 , se dice que la distribución es **sesgada a la izquierda o bien tiene una asimetría negativa**. Es decir, que la cola izquierda de la distribución es más larga que la derecha, esto indica que los valores a la izquierda de la media están más dispersos que en el lado contrario. Ver Figura (F.1(d)).
4. Puede darse el caso en que algunas cajas no tengan bigotes, es decir, el extremo inferior coincide con el Q_1 y el extremo superior coincide con el Q_3 , en dicha caja habría evidencia de que hay poca variabilidad o es casi nula.
5. También es posible que la caja no presente valores atípicos.

CÓDIGO EN R

```
#Cargar librerias
> library(readxl)
> library(urca) #Dickey-Fuller
> library(tseries)
> library(forecast)

> Tlaxco_2011_2017 <- read_excel("E:/Gloria/Tlaxco_2011-2017.
  xlsx", sheet = "Prom.sem.", > col_types = c("date", "numeric"
  ))
View(Tlaxco_2011_2017)

> Tlaxco_2011_2017$Fecha<-NULL #Se elimina la primera columna
> View(Tlaxco_2011_2017)

> str(Tlaxco_2011_2017) #Muestra la clase o estructura que
  tienen los datos

> series<-ts(Tlaxco_2011_2017$'Promedio semanal') #Convierte
  los datos a serie de tiempo (ts)
> str(series)
> series

#Graficos de: los datos originales, su FAC y su FACP
> plot(series, col='blue', type='o', pch=20, xlab="Semanas",
  ylab="Promedios_semanales_de_la_radiacion", main="Grafica_de
  _la_s.t._de_la_radiacion")
```

```

> acf(seriets , xlab="Desfase" , ylab=expression(rho[k]) , main="
  FAC_de_la_s.t._de_la_radiacion" , col="2" , lag=110)

> pacf(seriets , xlab="Desfase" , ylab=expression(phi[kk]) , main="
  FACP_de_la_s.t._de_la_radiacion" , col="2" , lag=110)

#Diagramas de cajas
> Tlaxco11_17_BP <- read_excel("E:/Gloria/Tlaxco11-17-BP.xlsx" ,
  sheet = "Hoja3" , col_names = FALSE , col_types = c("numeric" ,
  "numeric" , "numeric" , "numeric" , "numeric" , "
  numeric"))
> View(Tlaxco11_17_BP)
> Tlaxco11_17_BP$X_1<-NULL
> View(Tlaxco11_17_BP)

> str(Tlaxco11_17_BP)
> Tlaxco11_17_BP
> serieTBP<-ts(Tlaxco11_17_BP)
> str(serieTBP)
> serieTBP

#Muestra los boxplot de las filas
> boxplot(t(serieTBP) , xlab="Semanas" , ylab="Promedios_semanales_
  de_la_radiacion" , main="Diagrama_de_cajas_semanal_de_la_
  radiacion")

#Prueba Dickey-Fuller para la serie original
> prueba <- ur.df(seriets , lags=1 , type='drift')
> summary(prueba)

#Diferencias regulares y estacionales
> difer1=diff(seriets , diff=1) #Diferencia regular de orden 1
> ddestacional=diff(difer1 , lag=52 , diff=1) #Diferencia regular y
  diferencia estacional de orden 1

> win.graph()
> par(mfrow=c(3,1))

#Graficos de la serie diferenciada regularmente
> plot(difer1 , col='blue' , type='l' , xlab="Semanas" , ylab="
  Promedios_semanales_de_la_radiacion" , main = "Grafica_de_la_s.t
  .de_la_radiacion_con_una_diferencia_regular")
> acf(difer1 , xlab="Desfase" , ylab=expression(rho[k]) , main="FAC_
  de_la_s.t._con_una_diferencia_regular" , col="2" , lag=100)
> pacf(difer1 , xlab="Desfase" , ylab=expression(phi[kk]) , main="

```

```

FACP_de_la_s.t._con_una_diferencia_regular", col="2", lag=100)

#Graficos de la serie diferenciada regularmente y
  estacionalmente
> plot(ddestacional, col='blue', type='l', xlab="Semanas", ylab="
  Promedios_semanales_de_la_radiacion", main = "Grafica_de_la_s.t
  .de_la_radiacion_con_una_dif.regular_y_una_dif.estacional"
  )
> acf(ddestacional, xlab="Desfase", ylab=expression(rho[k]), main=
  "FACP_de_la_s.t._con_una_dif.regular_y_una_dif.estacional",
  col="2", lag=100)
> pacf(ddestacional, xlab="Desfase", ylab=expression(phi[kk]),
  main="FACP_de_la_s.t._con_una_dif.regular_y_una_dif.
  estacional", col="2", lag=100)

#Prueba Dickey-Fuller para la serie diferenciada regularmente y
  la estacionalmente
> prueba2 <- ur.df(difer1, lags=1, type='none')
> summary(prueba2)

> prueba3 <- ur.df(ddestacional, lags=1, type='none')
> summary(prueba3)

#####Modelos ARIMA#####
#Para el los modelos y el forcast se uso la serie2
> serietsf=Tlaxco_2011_2017[-c(308:312), ] #Se eliminan las
  ultimas 5 semanas
> str(serietsf)
> series2=ts(serietsf$'Promedio semanal') #Se convierte a serie
  de tiempo
> str(series2)

> modelo1<-arima(series2, order=c(0,1,3), seasonal=list(order=c
  (0,1,1), period=52))

> modelo2<-arima(series2, order=c(0,1,2), seasonal=list(order=c
  (0,1,1), period=52))

> modelo3<-arima(series2, order=c(0,1,1), seasonal=list(order=c
  (0,1,1), period=52))

> modelo4<-arima(series2, order=c(3,1,2), seasonal=list(order=c
  (0,1,1), period=52))

> modelo5<-arima(series2, order=c(3,1,1), seasonal=list(order=c

```

```

(0,1,1),period=52))
> modeloR<-arima(series2,order=c(1,0,1)) #Este modelo salio al
  aplicar auto.arima()

#===Analisis de los modelos usando los criterios Akaike y BIC
===#

> AIC(modelo1) #Primer modelo propuesto
> AIC(modelo2)
> AIC(modelo3) #Mejor modelo segun el criterio Akaike
> AIC(modelo4)
> AIC(modelo5)
> AIC(modeloR)

> BIC(modelo1) #Primer modelo propuesto
> BIC(modelo2)
> BIC(modelo3) #Mejor modelo segun el criterio BIC
> BIC(modelo4)
> BIC(modelo5)
> BIC(modeloR)

> summary(modelo1) #Muestra los parametros estimados del modelo1
> summary(modelo3) #Muestra los parametros estimados del modelo3
summary(modeloR) #Muestra los parametros estimados del modeloR

#=====Pronosticos=====#
> pronsMod1<- forecast (modelo1 ,h=5)
> summary(pronsMod1)
> plot(pronsMod1,type='o',pch=20, xlab="Semanas", ylab="
  Promedios_semanales_de_la_radiacion",main =" Pronostico_del_
  modelo_ARIMA(0,1,3)x(0,1,1)[52]")

> pronsMod3<- forecast (modelo3 ,h=5)
> summary(pronsMod3)
> plot(pronsMod3,type='o',pch=20, xlab="Semanas", ylab="
  Promedios_semanales_de_la_radiacion",main =" Pronostico_del_
  modelo_ARIMA(0,1,1)x(0,1,1)[52]")

> pronsModR<- forecast (modeloR ,h=5)
> summary(pronsModR)
> plot(pronsModR,type='o',pch=20, xlab="Semanas", ylab="
  Promedios_semanales_de_la_radiacion",main =" Pronostico_del_
  modelo_ARIMA(1,0,1)")

```

```

#####Residuos#####
#Modelo1: ARIMA(0,1,3)x(0,1,1)[52]

#Diagnostico del modelo 1
> tsdiag(modelo1)

> residuosM1<-residuals(modelo1)
> win.graph()
> par(mfrow=c(2,1))
> acf(residuosM1, xlab="Desfase", ylab="FAC", main="FAC_de_los_
  residuos_del_modelo_1", col="2")
> pacf(residuosM1, xlab="Desfase", ylab="FACP", main="FACP_de_los_
  residuos_del_modelo_1", col="2")

#Histograma
> hist(residuosM1, main="Histograma_de_residuos_del_modelo_1",
  freq = F, ylim=c(0,0.025), xlab="", ylab="Densidad")
> dz1 <- density(residuosM1)
> lines(dz1, col = "red", lwd = 3)
> curve(dnorm(x, mean(residuosM1), sd(residuosM1)), col = "blue"
  , lwd = 3, from=-100, to=100, add = TRUE)
> legend("topleft", col=c("blue", "red"), legend =c("Densidad_
  teorica", "Estimacion_de_la_densidad"), lwd=2, bty = "n")

#Diagrama de caja
> boxplot(residuosM1, main="Diagrama_de_caja_de_los_residuos_del_
  modelo_1", horizontal =T)

#Q-Q plot de residuos
> qqnorm(residuosM1, main="Q-Q_plot_de_residuos_del_modelo_1")
> qqline(residuosM1)

#Test Ljung-Box
> Box.test(residuosM1, lag=1, type = "Ljung")

#Modelo3: ARIMA(0,1,1)x(0,1,1)[52]

#Diagnostico del modelo 3
> tsdiag(modelo3)

> residuosM3<-residuals(modelo3)
> win.graph()
> par(mfrow=c(2,1))
> acf(residuosM3, xlab="Desfase", ylab="FAC", main="FAC_de_los_

```

```

    residuos_del_modelo_3", col="2")
> pacf(residuosM3, xlab="Desfase", ylab="FACP", main="FACP_de_los_
residuos_del_modelo_3", col="2")

#Histograma
> hist(residuosM3, main="Histograma_de_residuos_del_modelo_3",
      freq = F, ylim=c(0,0.025), xlab="", ylab="Densidad")
> dz3 <- density(residuosM3)
> lines(dz3, col = "red", lwd = 3)
> curve(dnorm(x, mean(residuos), sd(residuos)), col = "blue",
      lwd = 3, from=-100, to=100, add = TRUE)
> legend("topleft", col=c("blue", "red"), legend =c("Densidad_
teorica", "Estimacion_de_la_densidad"), lwd=2, bty = "n")

#Diagrama de caja
> boxplot(residuosM3, main="Diagrama_de_caja_de_los_residuos_del_
modelo_3", horizontal =T)

#Q-Q plot de residuos
> qqnorm(residuosM3, main="Q-Q_plot_de_residuos_del_modelo_3")
> qqline(residuosM3)

#Test Ljung-Box
> Box.test(residuosM3, type = "Ljung")

#Modelo: ARIMA(0,1,1)

#Diagnostico del modelo R
> tsdiag(modeloR)

> residuosMR<-residuals(modeloR)
> win.graph()
> par(mfrow=c(2,1))
> acf(residuosMR, xlab="Desfase", ylab="FAC", main="FAC_de_los_
residuos_del_modelo_ARIMA_(1,0,1)", col="2")
> pacf(residuosMR, xlab="Desfase", ylab="FACP", main="FACP_de_los_
residuos_del_modelo_ARIMA_(1,0,1)", col="2")

#Histograma
> hist(residuosMR, main="Histograma_de_residuos_del_modelo_ARIMA_
(1,0,1)", freq = F, ylim=c(0,0.025), xlab="", ylab="Densidad")
> dzR <- density(residuosMR)
> lines(dzR, col = "red", lwd = 3)
> curve(dnorm(x, mean(residuos), sd(residuos)), col = "blue",
      lwd = 3, from=-100, to=100, add = TRUE)

```

```
> legend("topleft", col=c("blue", "red"), legend = c("Densidad_
teorica", "Estimacion_de_la_densidad"), lwd=2, bty = "n")

#Diagrama de caja
> boxplot(residuosMR, main="Diagrama_de_caja_de_los_residuos_del_
ARIMA_(1,0,1)", horizontal =T)

#Q-Q plot de residuos
> qqnorm(residuosMR, main="Q-Q_plot_de_residuos_del_modelo_del_
ARIMA_(1,0,1)")
> qqline(residuosMR)

#Test Ljung-Box
> Box.test(residuosMR, type = "Ljung")
```


BIBLIOGRAFÍA

- [1] What is El Niño anyway? <https://spaceplace.nasa.gov/el-nino/sp/>, 2014.
- [2] What is La Niña? <https://spaceplace.nasa.gov/la-nina/sp/>, 2016.
- [3] AKAIKE, H. A new look at the statistical model identification. *Ieee Transactions on Automatic Control* 19, 6 (1974), 716–723.
- [4] AKAIKE, H. A bayesian analysis of the minimum aic procedure. *Annals of the Institute of Statistical Mathematics* 30 (1978), 9–14.
- [5] AKAIKE, H. A bayesian extension of the minimum aic procedure of autoregressive model fitting. *Biometrika* 66, 2 (Aug. 1979), 237–242.
- [6] ANSLEY, C. F., AND NEWBOLD, P. On the finite sample distribution of residual autocorrelations in autoregressive-moving average models. *Biometrika* 66, 3 (Dec. 1979), 547–553.
- [7] AYUNTAMIENTO DEL MUNICIPIO DE TLAXCO, H. Plan municipal de desarrollo Tlaxco.
- [8] BOWERMAN, B. L., O’CONNELL, R. T., AND KOEHLER, A. B. *Pronósticos, series de tiempo y regresión: Un enfoque aplicado*, 4th ed. Thomson, México, D.F., 2007.

- [9] BOX, G. E. P., AND COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26, 2 (1964), 211–252.
- [10] BROCKWELL, P. J., AND DAVIS, R. A. *Introduction to time series and forecasting*, 3rd ed. Springer Science Business Media, New York, NY, 2016.
- [11] COMISIÓN NACIONAL DEL AGUA. Reporte del clima en México. <http://smn.cna.gob.mx/es/climatologia/diagnostico-climatico/reporte-del-clima-en-mexico>.
- [12] COMISIÓN NACIONAL DEL AGUA. Un calor de ‘perros’. <https://www.gob.mx/conagua/articulos/un-calor-de-perros?idiom=es>, 2016.
- [13] CONDE, C., S., AND UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO. *México y el cambio climático global*. Secretaría de Medio Ambiente y Recursos Naturales : UNAM, México, 2011.
- [14] CRYER, J. D., AND CHAN, K.-S. *Time series analysis: with applications in R*, 2nd ed. Springer, New York, 2008.
- [15] DICKEY, D. A., AND FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*. 74 (1979), 427–431.
- [16] DURBIN, J. The fitting of time - series models. *Revue de l’Institut International de Statistique / Review of the International Statistical Institute* 28, 3 (1960), 233–244.
- [17] GONZÁLEZ, C. M. P. Análisis de series temporales : Modelos arima. *SARRIKO-ON* (2009), 169.
- [18] GUERRERO, V., AND DEL DEPTO. DE MATEMÁTICAS, C. *Modelos estadísticos para series de tiempo univariadas*. México: CINVESTAV-IPN, Pátzcuaro-Michoacán, 1987.
- [19] HILL, R. C., G. W. E., AND LIM, G. C. *Principles of econometrics*, 4th ed. Wiley, Hoboken, NJ, 2011.

- [20] HIPEL, K. W., AND MCLEOD, A. I. Time series modelling of water resources and environmental systems. *Elsevier Science Developments in Water Science*, 45 (1994), 1054.
- [21] LJUNG, G. M., AND BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*. 65, 2 (1978), 297–303.
- [22] MORRIS, D. W. *Elementary Meteorology*. Ginn and Company, Boston U.S.A., 1894.
- [23] NASA.GOV. Glossary. <https://earthobservatory.nasa.gov/Glossary/index.php?mode=all>, 2018.
- [24] OLIVER, J. E. *Encyclopedia of world climatology*, 1st ed. Springer, Great Britain, 2005.
- [25] REYES, C. S. *Introducción a la Meteorología*, 1st ed. UABC, Baja California, 2002.
- [26] RINCÓN, L. Introducción a los procesos estocásticos. *Universidad Nacional Autónoma de México, Facultad de Ciencias* (2011), 291.
- [27] SERVICIO METEOROLÓGICO NACIONAL. Inicia el periodo de la canícula en México:SMN. <http://www.cmic.org.mx/comisiones/Sectoriales/infraestructurahidraulica/normatividad/Boletines%20de%20CONAGUA/2010/Julio%202010/BOLETIN%20174-10.pdf>, 2010.
- [28] SHIBATA, R. Selection of the order of an autoregressive model by akaike's information criterion. *Biometrika* 63 (1976), 117–126.
- [29] SHUMWAY, R. H., AND STOFFER, D. S. *Time series analysis and its applications: with r examples*, 4th ed. Springer Science+Business Media, New York, NY, 2017.
- [30] SUKHATME, S., AND NAYAK, J. *Solar Energy: Principles of Thermal Collection and Storage*, 3rd ed. McGraw-Hill, 2008.
- [31] WEI, W. *Time series analysis: Univariate and multivariate methods*, 2nd ed. Pearson Addison Wesley, United States of America, 2006.