



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico-Matemáticas

Análisis en series de tiempo de los registros de la humedad
relativa en la zona de Texcoco, México

Tesis presentada al

Colegio de Matemáticas

como requisito parcial para la obtención del grado de

LICENCIADO EN MATEMÁTICAS

por

Pedro Pérez Cortez

dirigido por el

Dr. Bulmaro Juárez Hernández

Puebla Pue.
24 de octubre de 2012

Dedicatoria

A mis padres que todo lo dan y que todo lo merecen.

Agradecimientos

A mis padres Pedro y María de los Angeles, por permitirme Ser, a mis hermanos por ser además de hermanos amigos; a mis abuelos y tíos por todo su apoyo.

A mi asesor, el profesor Bulmaro Juárez por presentarme el tema preciso que captara mi interés. Por su orden y planificación contagiantes, y por supuesto, por su guía y orientación en la realización de este trabajo.

A todos los profesores que ayudaron en mi formación matemática durante la carrera. Todos merecen mi admiración y aprecio.

A mis compañeros y amigos, conseguidos del convivir en la universidad, por todos los momentos compartidos y por todos los recuerdos.

Finalmente a la vida, que hoy en una de sus buenas caras me permite culminar esta carrera presentando este trabajo.

Índice general

1. Introducción	1
2. Materiales y métodos	3
2.1. Breve descripción de la atmósfera	3
2.2. Definiciones meteorológicas	4
2.3. Monitoreo atmosférico	7
2.4. Descripción de los datos	8
3. Análisis de series de tiempo	9
3.1. Series de tiempo	9
3.2. Regresión de series de tiempo	10
3.2.1. El modelo de la tendencia mediante funciones polinomiales	10
3.2.2. La prueba de D-W para la autocorrelación positiva de primer orden	12
3.2.3. Aplicación del modelo de regresión	16
3.2.4. Aplicación de variables ficticias al análisis de la humedad relativa	25
3.3. Modelos de descomposición	27
3.3.1. Aplicación de los modelos de descomposición al análisis de la humedad relativa	28
4. Suavización exponencial	33
4.1. Suavización exponencial simple	33
4.2. Indicios de error	34
4.3. Método de Holt, de la suavización exponencial corregida de la tendencia	35
4.4. Método aditivo de Holt-Winters	36
4.5. Método multiplicativo de Holt-Winters	36
4.6. Tendencia amortiguada, y otros métodos de suavización exponencial	37
4.6.1. Método de la tendencia amortiguada	37
4.6.2. Método aditivo de Holt-Winters con tendencia amortiguada	38
4.6.3. Método multiplicativo de Holt-Winters con tendencia amortiguada	39
4.6.4. Método multiplicativo de Holt-Winters sin tendencia	40
4.7. Modelos para suavización exponencial, e intervalos de predicción	40
4.8. Aplicación de los métodos de suavización	41
4.8.1. Aplicación del método de la suavización exponencial simple	41
4.8.2. Aplicación del método de Holt, de la suavización exponencial corregida de la tendencia	47
4.8.3. Aplicación del método multiplicativo de Holt-Winters	54

5. Metodología de Box-Jenkins	63
5.1. Modelos estacionarios y la función de autocorrelación	63
5.2. Modelo estacionario lineal general	67
5.3. Procesos ARMA	69
5.3.1. Función de autocovarianza de un proceso ARMA(p,q)	71
5.3.2. Función de autocorrelación de un proceso ARMA(p,q)	72
5.3.3. Función de autocorrelación parcial de un proceso ARMA(p,q)	73
5.4. Modelos autorregresivos AR	73
5.5. Modelos de promedios móviles MA	77
5.6. Modelos ARIMA	81
5.7. Aplicación de la metodología de Box-Jenkins	85
6. Discusión y conclusiones	93
A. Anexo de datos	97
B. Suavización exponencial	103
C. Herramienta Solver	107
D. Estadístico de Box-Pierce y Ljung-Box	113
E. Criterio de información de Akaike	115

Capítulo 1

Introducción

El objetivo principal de este trabajo es establecer un modelo estadístico paramétrico para realizar la predicción sobre un conjunto de datos los cuales corresponden a los registros meteorológicos obtenidos en la zona de Texcoco proporcionados por la estación Meteorológica del Departamento de Irrigación de la U.A.CH., usando la metodología de Box y Jenkins para series de tiempo.

La meteorología es la ciencia encargada del estudio de la atmósfera, de sus propiedades y de los fenómenos que en ella tienen lugar. El estudio de la atmósfera se basa en el conocimiento de una serie de magnitudes, o variables meteorológicas, como la temperatura, la presión atmosférica o la humedad, las cuales varían tanto en el espacio como en el tiempo. En el presente trabajo se estudia y analiza la humedad de manera específica y su impacto sobre cierta región de la superficie terrestre, que se manifiesta como la cantidad de vapor de agua que contiene el aire, esa cantidad no es constante, sino que dependerá de diferentes factores, como si ha llovido recientemente, si estamos cerca del mar, si hay plantas, etc.

En este trabajo de tesis se realiza un estudio de la humedad contándose por ejemplo con una base de datos del clima en la zona de Texcoco, obtenida en la estación Meteorológica del Departamento de Irrigación de la U.A.CH. (Universidad Autónoma de Chapingo), los datos fueron tomados cada 30 minutos, con las fechas del 9 de Mayo del 2003 al 28 de Octubre del 2010. En la base de datos se cuenta con información meteorológica de registros como lo es, la humedad, la temperatura del aire, temperatura máxima, temperatura mínima, humectación de hoja, barómetro, velocidad del viento, velocidad máxima, sensación térmica, recorrido del viento, lluvia, intensidad máxima, punto rocío y evapotranspiración. Para el desarrollo del trabajo, se realizó, en principio, una filtración de los datos, enseguida se obtuvieron promedios diarios de todas las variables meteorológicas observadas y posteriormente se realiza en primera instancia un análisis de regresión lineal simple tomando como variable dependiente la humedad y como variables independientes a las demás variables meteorológicas, con el propósito de analizar el comportamiento de la humedad en términos de cada una de las otras variables meteorológicas. Después se realizó un análisis de la humedad usando el método de descomposición aditiva, para este análisis se obtuvieron promedios de cada estación del año (es decir

promedios del verano, otoño, invierno y primavera). Luego se realizó un análisis usando el método de la suavización exponencial simple para los registros diarios de la humedad ya que los datos mostraban que no había tendencia ni algún patrón estacional bien definido.

Como uno de los objetivos principales es predecir el comportamiento a corto plazo de la humedad para pronosticar el comportamiento de otras variables meteorológicas que dependen de ella, se realiza un análisis de series de tiempo determinando el modelo **ARIMA** que mejor ajuste a la serie de tiempo manejada para los datos así como el orden del modelo.

A continuación se describen de manera general los puntos importantes de cada capítulo que abarca este trabajo de tesis, como este trabajo pretende ser autocontenido, es necesario un conocimiento mínimo sobre meteorología y series de tiempo de manera que en el capítulo 2 se presentan algunos conceptos meteorológicos que son usados como variables en el análisis de regresión. En el capítulo 3 se revisa la teoría del análisis de series de tiempo como son procesos estocásticos, la definición formal de series de tiempo, posteriormente se presentan los principales modelos de regresión de series de tiempo. En el capítulo 4 se estudian los métodos de la suavización exponencial y en el capítulo 5 se explica la metodología de Box-Jenkins, se presentan los modelos **ARMA** (autorregresivos y promedios móviles) los cuales juegan un papel importante en el desarrollo del análisis, se estudian los modelos de series de tiempo estacionarios tales como los modelos de promedios móviles y autorregresivos. También el concepto de los modelos de primer orden y los de orden mayor, ya por último se introduce el concepto de modelos **ARIMA**, y la aplicación de la metodología de Box-Jenkins al análisis de los datos de la humedad.

Objetivos

Objetivo general: Realizar el análisis de los datos de la humedad obtenidos en la Estación Meteorológica del Departamento de Irrigación de la U.A.CH. ubicada en el Municipio de Texcoco México.

Objetivos Particulares:

1. Presentar los conceptos básicos de una serie de tiempo.
2. Realizar la filtración de los datos.
3. Realizar un análisis de regresión de los datos tomando a la humedad como variable dependiente, para obtener un modelo que permita hacer predicción.
4. Obtener un modelo de predicción para la humedad, usando métodos de suavización exponencial.
5. Obtener un modelo que permita realizar predicción para la humedad, usando la metodología de Box-Jenkins para series de tiempo.

Capítulo 2

Materiales y métodos

La meteorología es la ciencia encargada del estudio de la atmósfera, de sus propiedades y de los fenómenos que en ella tienen lugar, los llamados metéoros ([5],[9]).

El estudio de la atmósfera se basa en el conocimiento de una serie de magnitudes, o variables meteorológicas, como la temperatura, la presión atmosférica o la humedad, las cuales varían tanto en el espacio como en el tiempo.

Para el análisis de este trabajo se cuenta con una base de datos meteorológicos. En este capítulo se introducen estos conceptos extendiéndose un poco más en los aspectos relacionados con la humedad, dado que constituye la variable respuesta para los demás datos climáticos.

2.1. Breve descripción de la atmósfera

La atmósfera es la capa gaseosa que envuelve la Tierra, y que se adhiere a ella gracias a la acción de la gravedad. Es difícil determinar exactamente su espesor, puesto que los gases que la componen se van haciendo menos densos con la altura, hasta prácticamente desaparecer a unos pocos cientos de kilómetros de la superficie.

Entre los gases que componen la atmósfera, hay que destacar el Nitrógeno (N_2), el Oxígeno (O_2), el Argón (Ar), el Dióxido de Carbono (CO_2) y el vapor de agua, es importante recordar que la concentración de estos gases varía con la altura, siendo especialmente acusadas las variaciones del vapor de agua, que se concentra sobre todo en las capas próximas a la superficie.

La atmósfera terrestre está dividida principalmente en cuatro capas: la troposfera, la estratosfera, la mesosfera y la termosfera. Tiene funciones importantes, una de ellas es proteger la biosfera, que es la capa donde se sustenta la vida en el planeta.

2.2. Definiciones meteorológicas

Temperatura

La temperatura es una de las magnitudes más utilizadas para describir el estado de la atmósfera, sabemos que la temperatura del aire varía entre el día y la noche, entre una estación y otra, y también entre una ubicación geográfica y otra. En invierno puede llegar a estar bajo los $0^{\circ}C$ y en verano superar los $40^{\circ}C$.

Formalmente, la temperatura es una magnitud relacionada con la rapidez del movimiento de las partículas que constituyen la materia. Cuanta mayor agitación presenten éstas, mayor será la temperatura.

Para medir la temperatura, tenemos que basarnos en propiedades de la materia que se ven alteradas cuando ésta cambia: la resistencia eléctrica de algunos materiales, el volumen de un cuerpo, el color de un objeto, etc.[9] El instrumento que se utiliza para medir la temperatura se llama termómetro y fue inventado por Galileo en 1593. Hay muchos tipos distintos de termómetros. El modelo más sencillo consiste en un tubo graduado de vidrio con un líquido en su interior que puede ser, por ejemplo, alcohol o mercurio. Como estos líquidos se expanden más que el vidrio, cuando aumenta la temperatura, asciende por el tubo y cuando disminuye la temperatura se contrae y desciende por el tubo.

Como toda magnitud física, la temperatura tiene asociadas unas unidades de medida, diferentes en función de la escala que elijamos:

Escala Celsius ($^{\circ}C$): Fue propuesta en 1742 por el astrónomo Anders Celsius. Consiste en una división regular en 100 intervalos, donde el 0 corresponde al punto de congelación del agua y el 100 al punto de ebullición del mismo. Se expresa en grados centígrados y es la que utilizamos habitualmente y en este trabajo.

Escala Fahrenheit ($^{\circ}F$): Fue introducida en 1714 por Gabriel D. Fahrenheit y se utiliza habitualmente en Estados Unidos. El termómetro se gradúa entre $32^{\circ}F$ (correspondiente a los $0^{\circ}C$) y $212^{\circ}F$ (correspondientes a los $100^{\circ}C$).

Escala Kelvin ($^{\circ}K$): Fue introducida por Lord Kelvin en 1848 y es la escala más usada por los científicos. Es una escala que no tiene valores negativos de la temperatura y su cero se sitúa en el estado en el que las partículas que forman un material no se mueven. El punto de ebullición del agua corresponde a $373^{\circ}K$ y el de congelación a $273^{\circ}K$. Por tanto, una variación de 1 grado en la escala Kelvin es igual que una variación de 1 grado en la escala Celsius (ver Figura 2.1).

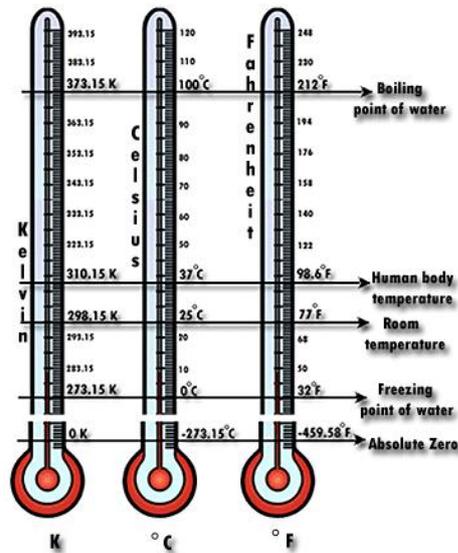


Figura 2.1: Escalas usadas para la medición de la temperatura.

Punto de Rocío

Es la temperatura a la que debera enfriarse la masa de aire para alcanzar su punto de saturación, con relación a su contenido de vapor. Se puede llegar por dos caminos distintos:

1. Añadiendo más vapor de agua a la masa de aire.
2. Enfriando la masa de aire.

También puede definirse como la temperatura a la cual el aire se satura [8].

Sensación Térmica

Se puede definir como la sensación de calor o frío que sentimos en presencia de viento o humedad, o la temperatura aparente que creemos que hace en un momento determinado bajo estas circunstancias [7].

El cuerpo humano mantiene una temperatura corporal de 37°C aproximadamente y para defenderse de temperaturas extremas aumenta o disminuye el diámetro de los vasos sanguíneos encargados de la circulación subcutánea. Cuando se combina temperatura y viento o temperatura y humedad, estamos hablando de sensación térmica.

Nuestro cuerpo mantiene en contacto con nuestra piel una capa aislante de milímetros de espesor que nos aísla de la temperatura exterior. Cuando sopla viento esta capa se reduce y se acelera la pérdida de calor de nuestro cuerpo, es la sensación térmica por efecto del viento.

CAPÍTULO 2. MATERIALES Y MÉTODOS

2.2. DEFINICIONES METEOROLÓGICAS

Cuando las temperaturas son altas la humedad aumenta la sensación de calor, dándonos la impresión de que la temperatura es mucho mayor, esto se debe a que la humedad alta dificulta la transpiración corporal. Si la humedad es baja se facilita la evaporación del sudor, nuestro cuerpo pierde calor más rápidamente y la sensación térmica es de que la temperatura es inferior a la real.

Presión Atmosférica

El aire que nos rodea, aunque no lo notemos, pesa y, por tanto, ejerce una fuerza sobre todos los cuerpos debida a la acción de la gravedad. Esta fuerza por unidad de superficie es la denominada presión atmosférica, cuya unidad de medida en el Sistema Internacional es el Pascal ($1\text{Pascal} = 1\text{N}/\text{m}^2$).

Por lo tanto, la presión atmosférica es la fuerza por unidad de área ejercida por la atmósfera sobre cualquier superficie en virtud de su peso. Equivale al peso de una columna de aire de sección transversal unitaria que se extiende desde un nivel dado hasta el límite superior de la atmósfera [7].

La presión atmosférica depende de muchas variables, sobre todo de la altitud. Cuanto más arriba en la atmósfera nos encontremos, la cantidad de aire por encima nuestro será menor, lo que hará que también sea menor la presión que éste ejerza sobre un cuerpo ubicado allí.

Pero la presión atmosférica, además de la altitud, depende de muchas otras variables. La situación geográfica, la temperatura, la humedad y las condiciones meteorológicas son sus principales condicionantes. Precisamente la relación que existe entre la presión atmosférica y el tiempo en un lugar hace de ésta una variable fundamental en la información meteorológica.

Velocidad y Dirección del Viento

El viento consiste en el movimiento de aire desde una zona hasta otra. Existen diversas causas que pueden provocar la existencia del viento, pero normalmente se origina cuando entre dos puntos se establece una cierta diferencia de presión o de temperatura.

Para poder disponer de medidas directas de velocidad y dirección del viento, se utilizan distintos instrumentos de medida:

1. **Medida de la velocidad horizontal del viento:** el instrumento más utilizado es el anemómetro de cazoletas en el que el giro de las mismas es proporcional a la velocidad del viento. La unidad de medida es el km/h o el m/s .
2. **Medida de la dirección:** para ello se utilizan las veletas, que indican la procedencia geográfica del viento. Hablamos de viento norte, noreste, suroeste, etc. en función de dónde provenga éste [5].

Humedad

El agua es uno de los principales componentes de la atmósfera, en la que puede existir como gas, como líquido, y como sólido. La presencia del agua en los tres estados de agregación se debe a que las condiciones físicas (temperatura y presión) necesarias para que se produzcan dichos cambios de estado se dan normalmente en la atmósfera.

La humedad es la cantidad de vapor de agua que contiene el aire [9]. Esa cantidad no es constante, sino que dependerá de diversos factores, como si ha llovido recientemente, si estamos cerca del mar, si hay plantas, etc. Existen diversas maneras de referirnos al contenido de humedad en la atmósfera:

1. **Humedad absoluta:** masa de vapor de agua, en gramos, contenida en $1m^3$ de aire seco.
2. **Humedad específica:** masa de vapor de agua, en gramos, contenida en $1kg$ de aire.
3. **Razón de mezcla:** masa de vapor de agua, en gramos, que hay en $1kg$ de aire seco.

Sin embargo, la medida de humedad que más se utiliza es la denominada humedad relativa, que se expresa en tanto por ciento y se calcula según la siguiente expresión:

$$h = \frac{e}{E}100.$$

En ella, e representa el contenido de vapor de la masa de aire y E su máxima capacidad de almacenamiento de éste, llamada presión de vapor saturante. Este valor nos indica la cantidad máxima de vapor de agua que puede contener una masa de aire antes de transformarse en agua líquida (esto es lo que se conoce como saturación). De alguna forma, la humedad relativa nos da una idea de lo cerca que está una masa de aire de alcanzar la saturación. Una humedad relativa del 100 por ciento es indicativo de que esa masa de aire ya no puede almacenar más vapor de agua en su seno, y a partir de ese momento, cualquier cantidad extra de vapor se convertirá en agua líquida o en cristalitas de hielo, según las condiciones ambientales.

2.3. 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 éste 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 humedad, la temperatura entre otras). En una estación meteorológica se encuentran los sensores meteorológicos que permiten hacer el pronóstico del tiempo. Los mas importantes son: veleta y anemómetro, barómetro, termómetros de máxima y de mínima, higrómetro, heliógrafo, piranómetro, pluviómetro, entre otros.

Los datos registrados se utilizan tanto para la elaboración de predicciones meteorológicas a partir de modelos numéricos como para estudios climáticos.

CAPÍTULO 2. MATERIALES Y MÉTODOS

2.4. DESCRIPCIÓN DE LOS DATOS

El instrumental meteorológico registrará sus mediciones en función del relieve topográfico y urbano donde se encuentre instalado, ya que es necesario que éste 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.

2.4. Descripción de los datos

Para el desarrollo de este trabajo se cuenta con una base de datos que contiene información meteorológica de las siguientes variables: temperatura del aire, temperatura máxima, temperatura mínima, humedad, humectación de hoja, barómetro, velocidad del viento, velocidad máxima, sensación térmica, recorrido del viento, lluvia, intensidad máxima, punto rocío y evapotranspiración. Esta información fue obtenida en la estación Meteorológica del Departamento de Irrigación de la U.A.CH.

Al trabajar con datos es muy frecuente encontrar omisiones que afectan a las bases de datos, en éste caso se tomaron originalmente los datos del Departamento de Irrigación de la U.A.CH., los datos de cada variable son tomados cada 30 minutos, en este trabajo se decidió tomar, en principio, promedios diarios de cada una de las variables. Los datos tomados del Departamento de Irrigación de la U.A.CH. están registrados con las fechas del 9 de Mayo del 2003 al 28 de Octubre del 2010.

Capítulo 3

Análisis de series de tiempo

El objetivo del análisis de una serie de tiempo es modelar mecanismos estocásticos que se presentan en una serie observada con el propósito de predecir o pronosticar valores futuros de la serie basándose en la historia de ésta y posiblemente otros factores relacionados con la serie. Para abordar la definición de una serie de tiempo, es necesario conocer antes la definición de un proceso estocástico, cuya definición se menciona a continuación.

3.1. Series de tiempo

Definición 3.1. *Un proceso estocástico es una colección o familia de variables aleatorias $\{X_t; t \in T\}$ las cuales están asociadas a un conjunto de índices T , de tal forma que a cada elemento del conjunto le corresponda una y sólo una variable aleatoria. Al conjunto de todos los posibles valores que la variable aleatoria pueda tomar se le llama espacio de estados el cual se denota por S .*

Si T es un intervalo de números reales ya sea cerrado o abierto, se dirá que el proceso estocástico es continuo, por otro lado, si el conjunto de índices es un conjunto finito o infinito numerable el proceso será llamado proceso estocástico en tiempo discreto. Además de la clasificación que se ha dado anteriormente, existe otro tipo de división en donde los procesos estocásticos se clasifican de acuerdo a la cardinalidad del espacio de estados, así se tendrán procesos estocásticos en tiempo continuo con espacio de estado discreto o continuo y procesos estocásticos en tiempo discreto con espacio de estado discreto o continuo. Dada la definición de un proceso estocástico se procede a definir una serie de tiempo.

Definición 3.2. *Una serie de tiempo es una sucesión de observaciones generadas por un proceso estocástico $\{X_t; t \in T\}$ donde el conjunto de índices T está en función del tiempo.*

Definición 3.3. *Un modelo de serie de tiempo para un conjunto de datos observados $\{x_t\}$, es la especificación de la distribución de probabilidad conjunta de una sucesión de variables aleatorias $\{X_t\}$, para las cuales cada $\{x_t\}$ se toma como una realización del proceso.*

Definición 3.4. *Las funciones $\{X_t(\omega); \omega \in \Omega\}$ en T son conocidas como las realizaciones de un proceso estocástico $\{X_t; t \in T\}$, siendo denotadas usualmente como $\{x_t\}$.*

Componentes de una serie de tiempo

Una serie de tiempo frecuentemente es examinada con la intención de descubrir patrones históricos que puedan ser útiles en la predicción ([2] y [6]). Para identificar estos patrones es conveniente pensar que una serie de tiempo consiste de varios componentes:

1. **Componente Tendencial:** Una serie de tiempo tiene tendencia cuando por largos periodos los valores crecen o decrecen. También puede definirse como cambios en la media.
2. **Componente cíclico:** Se refiere a movimientos hacia arriba y hacia abajo alrededor del nivel de la tendencia. Estas fluctuaciones, medidas de pico a pico, pueden tener una duración larga.
3. **Componente Estacional:** Son oscilaciones cuasiperiódicas de media cero, que tienen periodicidad anual o de un submúltiplo del año (trimestrales, mensuales, etc.) y se conocen como oscilaciones estacionales.
4. **Fluctuaciones irregulares:** Son movimientos erráticos en una serie de tiempo que no siguen un patrón regular, ni reconocible. Tales movimientos representan “lo que queda” en una serie de tiempo después de que la tendencia, ciclos y variaciones estacionales han sido explicadas.

3.2. Regresión de series de tiempo

En esta sección se tratan algunos modelos de regresión de series de tiempo, los cuales relacionan la variable dependiente y_t con variables independientes o predictoras x_t que están en función del tiempo.

Cuando se tiene una serie de tiempo es importante observar la dispersión de los datos, ya que mediante estos se pueden notar ciertas características que son llamadas de variación como lo son la tendencia, la estacionalidad, la cíclica y las fluctuaciones irregulares.

3.2.1. El modelo de la tendencia mediante funciones polinomiales

En ocasiones una serie de tiempo y_t se describe usando un modelo de la tendencia, el cual se define como:

$$y_t = T_t + \epsilon_t \quad (3.1)$$

donde

y_t es el valor de la serie de tiempo en el periodo t .

T_t es la tendencia de la serie de tiempo en el periodo t .

ϵ_t es el error de la serie de tiempo en el periodo t .

Este modelo señala que la serie de tiempo y_t se puede representar mediante un nivel promedio que cambia con el tiempo, el término de error ϵ_t representa las fluctuaciones aleatorias que ocasionan que los valores de la serie de tiempo y_t se desvíen del nivel promedio T_t .

Algunas tendencias útiles son:

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

1. **Sin tendencia:** Que se modela como $T_t = \beta_0$ quiere decir que no hay declinación o crecimiento durante un largo periodo de la serie de tiempo.
2. **Tendencia lineal:** Que se modela como $T_t = \beta_0 + \beta_1 t$, significa que hay un crecimiento lineal durante un largo periodo ($\beta_1 > 0$) o declinación ($\beta_1 < 0$) en el tiempo.
3. **Tendencia cuadrática:** Que se modela como $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$, quiere decir que hay un cambio cuadrático o curvilíneo durante un largo periodo en el tiempo.

Estas tendencias son las mas usuales, pero existen otras, por ejemplo:

El modelo de la tendencia polinomial de p-ésimo orden es:

$$y_t = T_t + \epsilon_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_p t^p + \epsilon_t. \quad (3.2)$$

Las estimaciones puntuales de mínimos cuadrados para los parámetros en estos modelos de la tendencia se podrían obtener mediante la aplicación de técnicas de regresión. En este caso se supone que el término de error ϵ_t cumple con las suposiciones de normalidad, independencia y varianza constante.

Sin embargo, al analizar datos de series de tiempo a menudo se transgrede la suposición de independencia, es muy común en los términos de error ordenados en el tiempo que éstos sean autocorrelacionados.

Se dice que existe autocorrelación positiva cuando los términos de error positivos tienden a ser seguidos en el tiempo por términos de error positivos y términos de error negativos tienden a ser seguidos en el tiempo por términos de error negativos y existe autocorrelación negativa cuando los términos de error positivos tienden a ser seguidos en el tiempo por términos de error negativos y términos de error negativos tienden a ser seguidos en el tiempo por términos de error positivos.

La suposición de independencia plantea que los términos de error ordenados en el tiempo no manifiestan autocorrelación positiva o negativa, esto quiere decir que los términos de error ocurren en un patrón aleatorio en el tiempo.

Existe un tipo de autocorrelación positiva o negativa que se llama autocorrelación de primer orden, la cual plantea que el término de error en el periodo de tiempo t , se relaciona con el término de error en el periodo $t - 1$ mediante la ecuación

$$\epsilon_t = \phi_1 \epsilon_{t-1} + a_t. \quad (3.3)$$

Se supone que ϕ_1 es el coeficiente de correlación entre términos de error separados por un periodo y a_1, a_2, \dots son valores seleccionados al azar y en forma independiente de entre una distribución normal que tiene media cero y una varianza constante independiente del tiempo.

La prueba de Durbin-Watson (D-W), es una prueba formal para la autocorrelación de primer orden, positiva o negativa. La estadística para la prueba de Durbin-Watson, denominada estadística de D-W es:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (3.4)$$

donde e_1, e_2, \dots, e_n son los residuos ordenados en el tiempo [12].

3.2.2. La prueba de D-W para la autocorrelación positiva de primer orden

Considere probar la hipótesis nula

H_0 : los términos de error no están autocorrelacionados positivamente

contra la hipótesis alternativa

H_a : los términos de error están autocorrelacionados positivamente.

D-W demostrarán que hay puntos (denotados como $d_{L,\alpha}$ y $d_{U,\alpha}$) tales que si α es el tamaño de la prueba (máximo tamaño del error Tipo I), entonces

1. Si $d < d_{L,\alpha}$, entonces se rechaza H_0 .
2. Si $d > d_{U,\alpha}$, entonces no se rechaza H_0 .
3. Si $d_{L,\alpha} \leq d \leq d_{U,\alpha}$, entonces la prueba no es concluyente.

Para que se pueda llevar a cabo esta prueba se han elaborado tablas que contienen los puntos $d_{L,\alpha}$ y $d_{U,\alpha}$ apropiados para varios valores de α (tamaño de la prueba), k (número de variables independientes) y n (número de observaciones).

La prueba de Durbin-Watson también se puede usar en el caso de la autocorrelación negativa de primer orden. Considere probar la hipótesis nula

H_0 : los términos de error no están autocorrelacionados negativamente

contra la hipótesis alternativa

H_a : los términos de error están autocorrelacionados negativamente.

D-W demostrarán que con base en el establecimiento de la probabilidad de un error Tipo I igual a α , los puntos $d_{L,\alpha}$ y $d_{U,\alpha}$ son tales que

1. Si $(4 - d) < d_{L,\alpha}$, entonces se rechaza H_0 .
2. Si $(4 - d) > d_{U,\alpha}$, entonces no se rechaza H_0 .
3. Si $d_{L,\alpha} \leq (4 - d) \leq d_{U,\alpha}$, entonces la prueba no es concluyente.

Para finalizar, se puede usar la estadística de D-W para probar la autocorrelación positiva o negativa de primer orden.

Considere probar la hipótesis nula

H_0 : los términos de error no están autocorrelacionados positiva o negativamente

contra la hipótesis alternativa

H_a : los términos de error están autocorrelacionados positiva o negativamente.

D-W demostrarán que con base en el establecimiento de la probabilidad de un error tipo I igual a α , los puntos $d_{L,\alpha}$ y $d_{U,\alpha}$ son tales que

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

1. Si $d < d_{L, \frac{\alpha}{2}}$ o si $(4 - d) < d_{L, \frac{\alpha}{2}}$, entonces se rechaza H_0 .
2. Si $d > d_{U, \frac{\alpha}{2}}$ o si $(4 - d) > d_{U, \frac{\alpha}{2}}$, entonces no se rechaza H_0 .
3. Si $d_{L, \frac{\alpha}{2}} \leq d \leq d_{U, \frac{\alpha}{2}}$ o si $d_{L, \frac{\alpha}{2}} \leq (4 - d) \leq d_{U, \frac{\alpha}{2}}$, entonces la prueba no es concluyente.

Observaciones:

1. La validez de la prueba de Durbin-Watson depende del supuesto de que la población de todos los posibles residuos en cualquier tiempo t tiene una distribución normal.
2. La autocorrelación positiva se encuentra con más frecuencia en la práctica que la autocorrelación negativa.
3. No hay dependencia histórica entre las variables y_t .

Se dice que una serie de tiempo manifiesta **variación estacional constante** cuando la magnitud del cambio estacional no depende del nivel de la serie de tiempo, si la magnitud del cambio estacional depende del nivel de la serie de tiempo, se dice que la serie de tiempo muestra una **variación estacional creciente**.

Cuando una serie de tiempo manifiesta una variación estacional creciente, lo común es aplicar una transformación a los datos con el objeto de generar una serie transformada que muestre variación estacional constante. Con frecuencia se usa una transformación de la forma

$$y_t^* = y_t^\lambda, \text{ donde, } 0 < \lambda < 1.$$

En ocasiones una serie de tiempo muestra variación estacional constante, por lo cual a menudo se usa el modelo:

$$y_t = T_t + S_t + \epsilon_t \quad (3.5)$$

donde

y_t es el valor de la serie de tiempo en el periodo t .

T_t es la tendencia de la serie de tiempo en el periodo t .

S_t es el factor estacional de la serie de tiempo en el periodo t .

ϵ_t es el error de la serie de tiempo en el periodo t .

Una manera de modelar los patrones estacionales es utilizar variables ficticias.

El factor estacional expresado mediante variables ficticias es

$$S_t = \beta_{s1}x_{s1,t} + \beta_{s2}x_{s2,t} + \dots + \beta_{s(L-1)}x_{s(L-1),t} \quad (3.6)$$

donde se supone que hay L estaciones al año y $x_{s1,t}, x_{s2,t}, \dots, x_{s(L-1),t}$ son variables ficticias que se definen como sigue

$$x_{s1,t} = \begin{cases} 1, & \text{si el periodo } t \text{ es la estación 1;} \\ 0, & \text{si no es así.} \end{cases}$$
$$x_{s2,t} = \begin{cases} 1, & \text{si el periodo } t \text{ es la estación 2;} \\ 0, & \text{si no es así.} \end{cases}$$
$$\vdots$$
$$\vdots$$
$$x_{s(L-1),t} = \begin{cases} 1, & \text{si el periodo } t \text{ es la estación } L-1; \\ 0, & \text{si no es así.} \end{cases}$$

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

El objetivo de la variable ficticia es asegurar que se incluye un parámetro estacional apropiado en el modelo de la regresión en cada periodo, también observe que el modelo de variable ficticia supone que la serie de tiempo manifiesta variación estacional constante. Cuando una serie muestra variación estacional creciente, casi siempre se aplica una transformación para igualar la variación estacional antes de ajustar el modelo con variable ficticia.

En algunas ocasiones los modelos de regresión que contienen términos trigonométricos se usan para pronosticar series de tiempo que muestran variación estacional constante o creciente.

Los modelos trigonométricos para la variación estacional constante son

$$1. y_t = \beta_0 + \beta_1 t + \beta_2 \text{sen}\left(\frac{2\pi t}{L}\right) + \beta_3 \text{cos}\left(\frac{2\pi t}{L}\right) + \epsilon_t.$$

$$2. y_t = \beta_0 + \beta_1 t + \beta_2 \text{sen}\left(\frac{2\pi t}{L}\right) + \beta_3 \text{cos}\left(\frac{2\pi t}{L}\right) + \beta_4 \text{sen}\left(\frac{4\pi t}{L}\right) + \beta_5 \text{cos}\left(\frac{4\pi t}{L}\right) + \epsilon_t.$$

Donde L es el número de estaciones en el año, los cuales asumen una tendencia lineal, pero se pueden modificar para representar otras tendencias. Algunas veces, los modelos trigonométricos de series de tiempo proporcionan predicciones útiles. Sin embargo en general, la regresión con variable ficticia y otros métodos son superiores a los modelos trigonométricos para modelar la variación estacional [12]. La razón es que los modelos con variable ficticia (y otras técnicas) utilizan un parámetro diferente para modelar el efecto de cada estación diferente en un año.

No todos los modelos útiles son lineales en los parámetros, por ejemplo el modelo de la curva de crecimiento en donde $\beta_0, \beta_1 \in \mathbb{R}$, representado por

$$y_t = \beta_0(\beta_1^t)\epsilon_t. \quad (3.7)$$

Para aplicar técnicas de estimación y predicción es necesario transformar tal modelo no lineal en uno que sea lineal en los parámetros.

Si $\beta_0 > 0$ y $\beta_1 > 0$, aplicando una transformación logarítmica al modelo $y_t = \beta_0(\beta_1^t)\epsilon_t$ se tiene

$$\log(y_t) = \log(\beta_0(\beta_1^t)\epsilon_t) = \log(\beta_0) + \log(\beta_1^t) + \log(\epsilon_t) = \log(\beta_0) + (\log(\beta_1))t + \log(\epsilon_t).$$

Si hacemos $\alpha_0 = \log(\beta_0)$, $\alpha_1 = \log(\beta_1)$, y $a_t = \log(\epsilon_t)$, la versión transformada del modelo es

$$\log(y_t) = \alpha_0 + \alpha_1 t + a_t. \quad (3.8)$$

Por consiguiente, el modelo con la variable dependiente $\log(y_t)$ es lineal en los parámetros α_0 y α_1 .

Observaciones sobre los modelos:

- Es común que mediante las graficas de los datos de y_t contra t , identificar la curva de crecimiento.
- Cuando los términos de error que pertenecen a un modelo de regresión de una serie de tiempo están autocorrelacionados, el modelo es inadecuado. En tal caso se debe remediar el problema modelando la autocorrelación. Si ignoramos los términos de error autocorrelacionados tendremos que pagar una penalización en términos de intervalos de predicción más amplios.

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO
3.2. REGRESIÓN DE SERIES DE TIEMPO

Si tomamos en cuenta la autocorrelación podremos obtener intervalos de predicción más precisos.

Una estructura de error autocorrelacionada que con frecuencia se encuentra es el proceso autorregresivo de primer orden, el cual se define mediante

$$\epsilon_t = \phi_1 \epsilon_{t-1} + a_t. \quad (3.9)$$

En este caso, se supone que a_t es un término de error (con frecuencia llamado choque aleatorio o perturbación aleatoria) con media cero que cumple con las suposiciones de varianza constante, independencia y normalidad. Además a ϕ_1 se le conoce como el coeficiente de correlación entre ϵ_t y ϵ_{t-1} .

Si $\phi_1 > 0$, esto quiere decir que los términos de error están autocorrelacionados positivamente, por otro lado si $\phi_1 < 0$, los términos de error están autocorrelacionados negativamente.

Sean $x_{t1}, x_{t2}, x_{t3}, \dots, x_{tk}$ los valores de las variables independientes $x_1, x_2, x_3, \dots, x_k$ que se han observado o serán observadas en el periodo de tiempo t , considere el modelo de regresión de series de tiempo

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \epsilon_t \quad (3.10)$$

donde ϵ_t se describe mediante un proceso autorregresivo de primer orden. Es decir,

$$\epsilon_t = \phi_1 \epsilon_{t-1} + a_t. \quad (3.11)$$

Una predicción hecha en el tiempo T para el valor futuro $y_{T+\tau}$ es

$$\hat{y}_{T+\tau} = \hat{\beta}_0 + \hat{\beta}_1 x_{T+\tau,1} + \hat{\beta}_2 x_{T+\tau,2} + \dots + \hat{\beta}_k x_{T+\tau,k} + \hat{\phi}_1 \hat{\epsilon}_{T+\tau-1}. \quad (3.12)$$

Si $\tau = 1$ (un periodo hacia adelante), entonces

$$\hat{\epsilon}_{T+\tau-1} = \hat{\epsilon}_T = y_t - [\hat{\beta}_0 + \hat{\beta}_1 x_{T1} + \hat{\beta}_2 x_{T2} + \dots + \hat{\beta}_k x_{Tk}] \quad (3.13)$$

ya que y_t ha sido observada.

Si $\tau > 1$ (la predicción es para más de un periodo de tiempo adelante), entonces $y_{T+\tau-1}$ no ha sido observada. Por lo tanto si $\tau > 1$

$$\hat{\epsilon}_{T+\tau-1} = \hat{y}_{T+\tau-1} - [\hat{\beta}_0 + \hat{\beta}_1 x_{T+\tau-1,1} + \hat{\beta}_2 x_{T+\tau-1,2} + \dots + \hat{\beta}_k x_{T+\tau-1,k}] \quad (3.14)$$

1. Si $\tau = 1$, entonces un intervalo de predicción aproximado $100(1 - \alpha)$ por ciento para y_{T+1} es

$$\left[\hat{y}_{T+1} \pm z_{[\frac{\alpha}{2}]} s \right].$$

2. Si $\tau = 2$, entonces un intervalo de predicción aproximado $100(1 - \alpha)$ por ciento para y_{T+2} es

$$\left[\hat{y}_{T+2} \pm z_{[\frac{\alpha}{2}]} s \sqrt{1 + (\hat{\phi}_1)^2} \right].$$

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

3. Si $\tau \geq 3$, entonces un intervalo de predicción aproximado $100(1 - \alpha)$ por ciento para $y_{T+\tau}$ es

$$\left[\hat{y}_{T+\tau} \pm z_{\left[\frac{\alpha}{2}\right]} s \sqrt{1 + (\hat{\phi}_1)^2 + \dots + (\hat{\phi}_1)^{2(\tau-1)}} \right].$$

Además del proceso autorregresivo de primer orden, existen otras estructuras del error autocorrelacionado. Por ejemplo, el proceso autorregresivo de orden p se define como

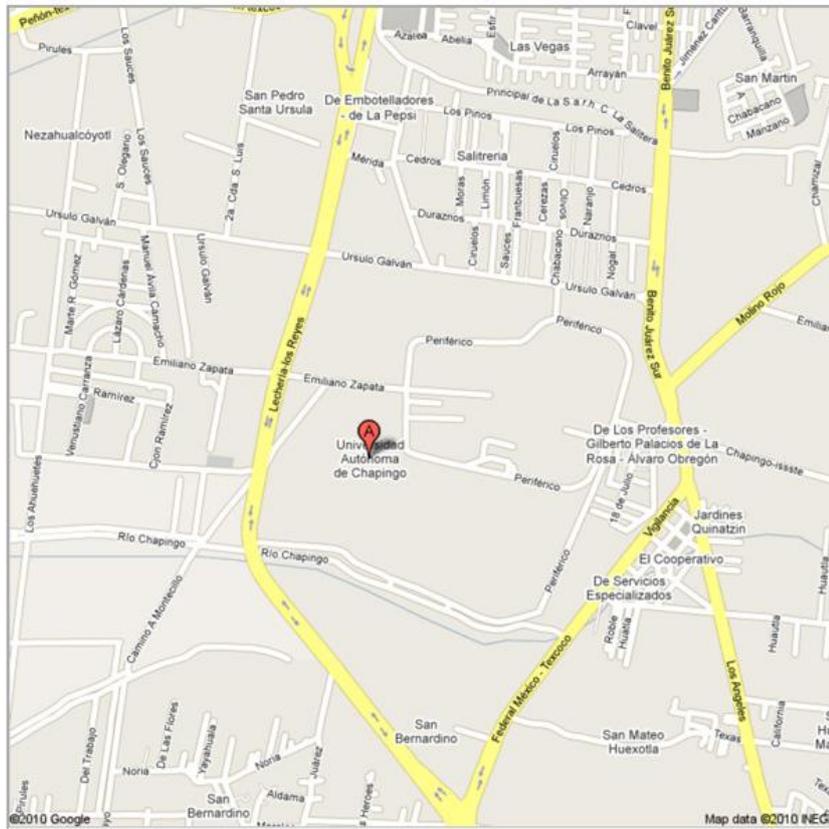
$$\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_p \epsilon_{t-p} + a_t. \quad (3.15)$$

Este proceso relaciona ϵ_t , el término de error en el periodo t , con los términos de error anteriores $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-p}$. Aquí $\phi_1, \phi_2, \dots, \phi_p$ son los parámetros desconocidos y a_t es un término de error (choque aleatorio) con media cero que cumple con las suposiciones [12].

3.2.3. Aplicación del modelo de regresión

El objetivo en esta sección es determinar un modelo que permita realizar pronósticos de la humedad relativa en base a los datos proporcionados por los registros de la humedad relativa diaria en la zona de Texcoco en la estación Meteorológica del Departamento de Irrigación de la U.A.CH. (ver Figura 3.1).

Figura 3.1: Imagen geográfica de la Universidad Autónoma de Chapingo.



CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO
3.2. REGRESIÓN DE SERIES DE TIEMPO

La base de datos de los registros antes mencionados se han filtrado generando otra base de datos que se han ordenado como en la Tabla 1 (ver Apéndice A).

Al graficar los datos, éstos parecen fluctuar en forma aleatoria con respecto a un nivel promedio constante (ver Figura 3.2).

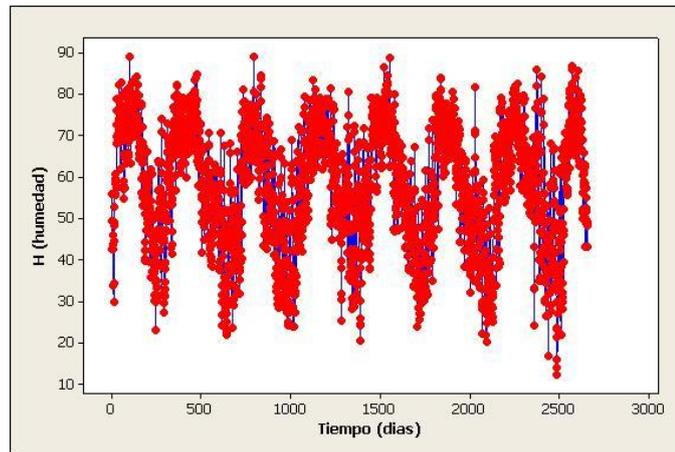


Figura 3.2: Gráfica obtenida mediante MINITAB para la humedad contra el tiempo en días.

Puesto que se cree en forma subjetiva que este patrón de datos continuará en el futuro, es razonable utilizar el modelo de regresión

$$y_t = T_t + \epsilon_t = \beta_0 + \epsilon_t$$

para predecir la humedad en los días futuros.

Se puede demostrar que la estimación puntual de mínimos cuadrados de β_0 es igual a

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_{2657}}{2657} = 57.87726$$

la media de la humedad observada. Por consiguiente,

$$\hat{y} = \bar{y} = 57.87726$$

es la predicción puntual de la humedad (y_t) en cualquier día futuro. Además un intervalo de predicción de $100(1 - \alpha)$ por ciento para y_t es

$$\left[\bar{y} \pm t_{\alpha/2}^{n-1} s \sqrt{1 + \frac{1}{n}} \right] \quad (3.16)$$

donde

$$s = 14.648.$$

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

Si se desea determinar un intervalo de predicción del 95 por ciento, usamos

$$t_{\alpha/2}^{n-1} = t_{.025}^{2656} = 1.96$$

y se encontró que el intervalo es, de acuerdo a (3.16) igual a

$$[29.1614, 86.5931].$$

Este intervalo, se encuentra en los resultados de MINITAB presentados en la Tabla 3.1, interpretándose que en una región limítrofe de la estación Meteorológica del Departamento de Irrigación de la U.A.CH. indica con un 95 por ciento de confianza que la humedad y_t en algún día futuro estará entre 29.16 y 86.59.

Tabla 3.1: Resultados que se obtienen con MINITAB para un análisis de regresión de los datos de la humedad.

Predictor	Coef	SE Coef	T	P
Noconstant				
Intercep	57,8773	0,2842	203,67	0,000

S = 14,6477 Durbin-Watson statistic = 0,212203

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	57,877	0,284	(57,320; 58,434)	(29,150; 86,605)

En la Figura 3.3 se ilustra la gráfica de los residuos ordenados en el tiempo para el modelo sin tendencia para la humedad.

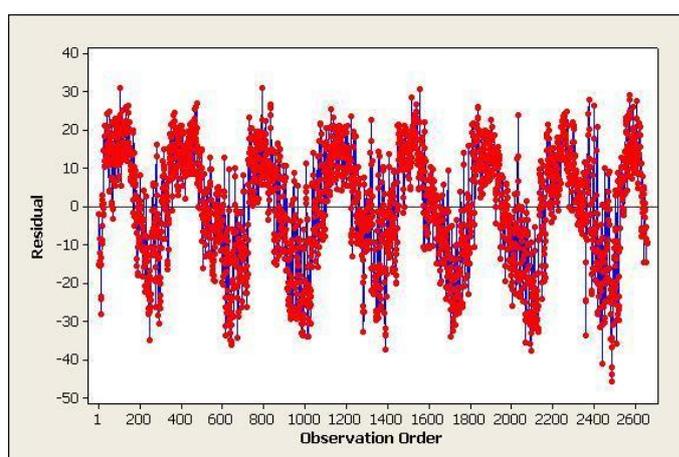


Figura 3.3: Gráfica de residuos contra tiempo para el modelo de la humedad que se obtiene usando MINITAB.

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO
3.2. REGRESIÓN DE SERIES DE TIEMPO

El valor numérico de Durbin-Watson (0.2122) es pequeño y las tablas conocidas no proporcionan valores críticos para $n = 2657$ y $k = 1$, sin embargo, empíricamente podemos afirmar que existe una alta correlación entre valores de la humedad relativa observados en periodos de tiempo cortos, esto es la suposición de independencia sufre transgresiones. Por lo tanto se considera el modelo siguiente

$$y_t = \beta_0 + \epsilon_t$$

donde

$$\epsilon_t = \phi_1 \epsilon_{t-1} + a_t.$$

En este caso, a_t es un choque aleatorio o perturbación aleatoria con media cero que satisface las suposiciones de varianza constante, independencia y normalidad.

En la Tabla 3.2 se proporcionan las partes pertinentes de la salida que ofrece MINITAB, observe que el término autorregresivo que muestra la salida de MINITAB $\phi_1 \epsilon_{t-1}$ (AR 1, en la salida) es significativo. Esto quiere decir que el término autorregresivo ayuda a modelar la estructura del error autocorrelacionado.

Tabla 3.2: Análisis con MINITAB del modelo de la humedad.

```

Final Estimates of Parameters

Type      Coef    SE Coef      T      P
AR 1      0,8941   0,0087   102,79  0,000
Constant  6,1203   0,1274    48,04  0,000
Mean      57,800   1,203
Number of observations: 2657
Residuals:  SS = 114503 (backforecasts
excluded)      MS = 43  DF = 2655
    
```

Ahora se requiere hacer un análisis profundo de la base de datos que se ha generado, para proponer un modelo adecuado que permita predecir la humedad. Se sabe que la base de datos cuenta con datos meteorológicos de registros de la humedad (H), temperatura del aire (TA), temperatura máxima (TMax), temperatura mínima (TMin), humectación de hoja (HH), barómetro (BR), velocidad del viento (VV), velocidad máxima (VM), sensación térmica (ST), recorrido del viento (RV), lluvia (LL), intensidad máxima (IM), punto rocío (PR) y evapotranspiración (ET), las cuales se decidieron tomar en consideración debido a que estas variables meteorológicas influyen sobre el comportamiento de la humedad diaria.

Algunos de los registros meteorológicos obtenidos en la zona de Texcoco en la estación Meteorológica del Departamento de Irrigación de la U.A.CH. se presentan en la siguiente tabla.

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

Fecha		Temperatura							
Fecha	Hora	Aire	Promedio diario	Máx	Promedio diario	Min	Promedio diario	ET	
09/05/2003	1:30 PM	27,9	24,88	28,4	25,53	27,5	24,31	--	
09/05/2003	2:00 PM	28,8		29,1		28,4		0,79348	
09/05/2003	2:30 PM	29,0		29,2		28,8		--	
09/05/2003	3:00 PM	25,6		28,8		23,9		0,28114	
09/05/2003	3:30 PM	26,3		27,9		24,4		--	
09/05/2003	4:00 PM	27,3		27,9		26,4		0,26573	
09/05/2003	4:30 PM	26,6		26,7		26,4		--	
09/05/2003	5:00 PM	27,1		27,2		26,7		0,29368	
09/05/2003	5:30 PM	26,1		26,8		25,4		--	
09/05/2003	6:00 PM	25,1		25,4		24,8		0,04920	
09/05/2003	6:30 PM	24,6		24,8		24,2		--	
09/05/2003	7:00 PM	23,2		24,2		22,3		0,05824	
09/05/2003	7:30 PM	22,4		22,7		22,2		--	
09/05/2003	8:00 PM	22,2		22,5		21,8		0,04273	
09/05/2003	8:30 PM	19,6		21,8		18,6		--	
09/05/2003	9:00 PM	19,2		19,4		18,7		0,06717	
09/05/2003	9:30 PM	19,6		19,7		19,4		--	
09/05/2003	10:00 PM	19,8		19,9		19,7		0,04649	
09/05/2003	10:30 PM	19,5		19,7		19,4		--	
09/05/2003	11:00 PM	19,3		19,4		19,1		0,04046	
09/05/2003	11:30 PM	18,4		19,1		18		--	
10/05/2003	12:00 AM	17,9	19,77	18	20,20	17,7	19,33	0,02581	
10/05/2003	12:30 AM	17,6		17,7		17,4		--	
10/05/2003	1:00 AM	17,3		17,4		17,2		0,01563	
10/05/2003	1:30 AM	16,9		17,2		16,6		--	
10/05/2003	2:00 AM	16,1		16,6		16,8		0,01844	

Se realizó un análisis de regresión lineal simple con respecto a cada una de las variables meteorológicas antes mencionadas una a la vez, tomando como variable respuesta o dependiente a la humedad y como variables predictoras o independientes a las demás variables meteorológicas, dos de los resultados se presentan a continuación.

The regression equation is
 $H = 68,5 - 0,669 \text{ Tmax}$

Predictor	Coef	SE Coef	T	P
Constant	68,503	1,814	37,77	0,000
Tmax	-0,6693	0,1129	-5,93	0,000

S = 14,5543 R-Sq = 1,3% R-Sq(adj) = 1,3%

En el análisis anterior obtenido con MINITAB se observa que el p -valor para la variable independiente es pequeño, con lo que sería razonable concluir que la variable independiente Tmax en el modelo es útil para pronosticar la humedad (H). Sin embargo, si observamos el coeficiente de determinación R^2 (que se representa como $R - Sq$), se observa que este porcentaje es muy bajo, por lo que se concluye que el modelo no es el adecuado para pronosticar a la humedad (H).

The regression equation is
 $H = 42,5 + 2,60 \text{ PR}$

Predictor	Coef	SE Coef	T	P
Constant	42,5312	0,2110	201,59	0,000
PR	2,59988	0,02739	94,93	0,000

S = 6,98871 R-Sq = 77,2% R-Sq(adj) = 77,2%

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

De forma que del análisis de regresión lineal simple obtenido con MINITAB el p -valor para la variable independiente es pequeño (0.000), con lo que se puede concluir que la variable independiente PR en el modelo es útil para pronosticar la humedad (H). Sin embargo, en este caso el coeficiente de determinación R^2 (que se representa como $R - Sq$) no es bajo, por lo que se concluye que este modelo podría ser el más apropiado para pronosticar a la humedad (H) [10].

De manera que al proponer un modelo de regresión lineal simple con una sola variable predictora y realizar el análisis correspondiente se encontró que el único modelo, y por lo tanto el mejor, que explica a la humedad (H) de forma conveniente es aquel que incluye como variable dependiente a la variable punto de rocío (PR).

Ahora, para proponer un modelo mejor se realiza un análisis más extenso, llevando a efecto un análisis de regresión lineal múltiple, usando todas las variables meteorológicas mencionadas anteriormente, tomando como variable respuesta a la humedad (H) y como variables predictoras el resto de ellas.

El siguiente resultado de MINITAB presenta los resultados del ajuste a un modelo de regresión lineal múltiple con las variables antes indicadas.

```
The regression equation is
H = 77,5 + 0,686 TA - 6,81 TMax + 2,99 TMin + 0,260 HH + 0,0129 BR + 1,84 VV
  - 1,05 VM + 0,563 ST - 0,0910 RV + 5,35 LL + 0,840 IM + 2,86 PR - 1,19 ET
```

Predictor	Coef	SE Coef	T	P
Constant	77,54	17,26	4,49	0,000
TA	0,6855	0,3482	1,97	0,049
TMax	-6,8068	0,4210	-16,17	0,000
TMin	2,9866	0,3999	7,47	0,000
HH	0,25970	0,01534	16,93	0,000
BR	0,01293	0,02918	0,44	0,658
VV	1,8435	0,2401	7,68	0,000
VM	-1,05470	0,09701	-10,87	0,000
ST	0,5626	0,2257	2,49	0,013
RV	-0,09095	0,02833	-3,21	0,001
LL	5,349	1,034	5,17	0,000
IM	0,8400	0,8174	1,03	0,304
PR	2,85818	0,01470	194,43	0,000
ET	-1,1884	0,9775	-1,22	0,224

```
S = 1,88120   R-Sq = 98,4%   R-Sq(adj) = 98,4%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	13	560503	43116	12183,26	0,000
Residual Error	2643	9353	4		
Total	2656	569856			

Como parte importante del análisis de regresión lineal múltiple está el modelo candidato a ser elegido que debe cumplir con los supuestos de normalidad e independencia en los residuos ya que estos son estimaciones puntuales de los términos de error. Como se puede observar, en la Figura 3.4 que la mayoría de los residuos se encuentran sobre una línea recta con excepción de algunos puntos que se encuentran alejados, pero sin embargo, ésto

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

parece indicar un comportamiento normal en los residuos. Por otro lado, en la Figura 3.5 se observa que los residuos (dada la robustez de normalidad del método) muestran un comportamiento cuadrático, esto indica que se puede estar transgrediendo la suposición de independencia de los residuos.

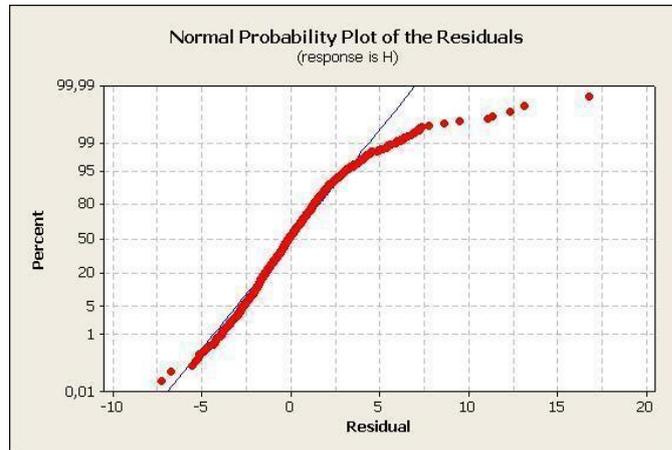


Figura 3.4: Prueba de normalidad en los residuos.

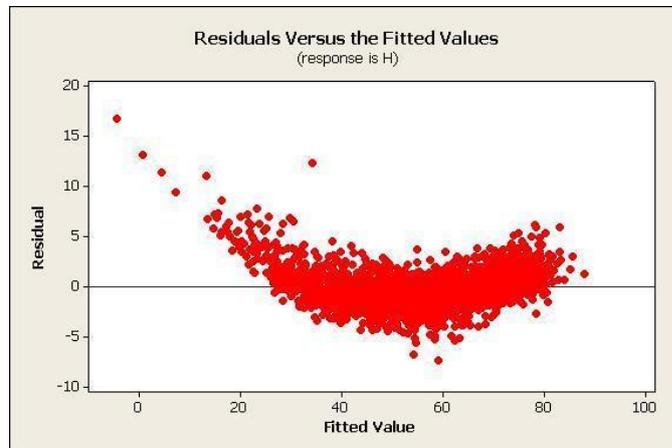


Figura 3.5: Prueba de independencia en los residuos.

Revisando los resultados obtenidos con MINITAB [11], se observa que el p -valor es muy grande para las variables meteorológicas TA, BR, ST, IM y ET, se puede decir entonces que necesitamos un modelo más sencillo para explicar a la variable dependiente humedad (H). Observando el análisis de varianza, se prueba al obtener un p -valor igual a cero que existen algunos parámetros en el modelo que no son cero.

Lo anterior indica que debe encontrarse un modelo más sencillo que se ajuste mejor a los datos, ya que el modelo completo parece tener variables meteorológicas que no ayudan a

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

la explicación del modelo. Para encontrar este modelo se usan dos métodos de selección de modelos: la regresión con los mejores subconjuntos y la regresión stepwise [6].

La selección adecuada de modelos se basa en un principio básico conocido como la navaja de Occam, el cual dice: el mejor modelo científico es el modelo mas simple que explica los hechos observados. Por otro lado, la selección de modelos, en términos de modelos lineales se basa en el principio de “parsimonia” el cual dice: un modelo debe contener el menor número de variables necesario para ajustar los datos. Los modelos que solo contienen las variables necesarias para ajustar los datos se llaman “parsimoniosos”. Aunque usar métodos para la selección de modelos resultan ser con frecuencia informales estos pueden ser sistemáticos [6].

A continuación se presenta un análisis del método de selección de los mejores conjuntos efectuado con el paquete estadístico Minitab.

```

Regresion con los mejores subconjuntos

Response is H

                T T
                M M
                T a i H B V V S R L I P E
                A x n H R V M T V L M R T
Vars  R-Sq  R-Sq(adj)  Mallows C-p      S
  1  77,2    77,2    33989,8  6,9887
  1  36,7    36,7    99241,9  11,654
  2  97,9    97,9    776,8  2,1379
  2  97,8    97,8    921,8  2,1827
  3  98,1    98,1    347,2  1,9992
  3  98,0    98,0    500,8  2,0498
  4  98,2    98,2    232,1  1,9601
  4  98,2    98,2    261,6  1,9701
  5  98,3    98,3    153,1  1,9327
  5  98,3    98,2    171,0  1,9389
  6  98,3    98,3    81,0  1,9073
  6  98,3    98,3    89,1  1,9101
  7  98,3    98,3    35,2  1,8908
  7  98,3    98,3    41,0  1,8929
  8  98,4    98,3    17,2  1,8841
  8  98,3    98,3    26,2  1,8873
  9  98,4    98,3    12,6  1,8821
  9  98,4    98,3    15,0  1,8830
 10  98,4    98,4    10,8  1,8811
 10  98,4    98,3    13,2  1,8820
 11  98,4    98,4    11,3  1,8810
 11  98,4    98,4    11,6  1,8811
 12  98,4    98,4    12,2  1,8809
 12  98,4    98,4    13,1  1,8812
 13  98,4    98,4    14,0  1,8812

```

En el análisis de selección de los mejores subconjuntos presentado anteriormente se obtienen tanto el mejor como el segundo mejor subconjunto, para el número de factores del 1 al 13, además se observa que el R^2 ajustado (que se representa como $R - Sq(adj)$) alcanza un máximo en el modelo que consta de diez variables meteorológicas que contienen las variables TA, TMax, TMin, HH, VV, VM, ST, RV, LL y PR, este resultado se corrobora con el C_p de Mallows que aunque es 10.8, éste indica el número de variables independientes que son diez [6].

En seguida se presenta el análisis de selección stepwise para determinar el mejor conjunto de variables predictoras que explican adecuadamente a la variable humedad.

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

```

Regression stepwise
  Alpha-to-Enter: 0,15  Alpha-to-Remove: 0,15

Response is H on 13 predictors, with N = 2657

Step          1          2          3          4          5          6
Constant      42,53      84,65      81,79      82,48      84,50      84,71

PR            2,5999      3,1073      2,9960      2,9812      2,8900      2,8705
T-Value       94,93      346,97      295,89      297,23      201,32      200,01
P-Value       0,000      0,000      0,000      0,000      0,000      0,000

TMax
T-Value       -2,842      -2,686      -2,661      -5,669      -5,877
P-Value       0,000      0,000      0,000      0,000      0,000

HH
T-Value       19,55      17,72      18,53      17,63
P-Value       0,000      0,000      0,000      0,000

VM
T-Value       -10,39      -12,28      -12,68
P-Value       0,000      0,000      0,000

TMin
T-Value       8,76      9,46
P-Value       0,000      0,000

LL
T-Value       8,49
P-Value       0,000

S              6,99      2,14      2,00      1,96      1,93      1,91
R-Sq          77,24      97,87      98,14      98,21      98,26      98,31
R-Sq(adj)     77,24      97,87      98,14      98,21      98,26      98,30
Mallows C-p   33989,8    776,8      347,2      232,1      153,1      81,0

Step          7          8          9          10
Constant      84,95      85,01      85,18      85,21

PR            2,864      2,860      2,862      2,862
T-Value       200,85      200,91      200,89      201,00
P-Value       0,000      0,000      0,000      0,000

TMax
T-Value       -6,04      -6,09      -6,48      -6,82
P-Value       0,000      0,000      0,000      0,000

HH
T-Value       17,66      17,19      16,93      16,95
P-Value       0,000      0,000      0,000      0,000

VM
T-Value       -10,78      -10,93      -11,04      -10,98
P-Value       0,000      0,000      0,000      0,000

TMin
T-Value       10,08      10,25      9,42      7,73
P-Value       0,000      0,000      0,000      0,000

LL
T-Value       8,56      9,66      9,45      9,47
P-Value       0,000      0,000      0,000      0,000

VV
T-Value       6,88      7,72      7,92      7,80
P-Value       0,000      0,000      0,000      0,000

RV
T-Value       -4,46      -4,39      -4,40
P-Value       0,000      0,000      0,000

ST
T-Value       2,56      2,49
P-Value       0,010      0,013

TA
T-Value       1,94
P-Value       0,052

S              1,89      1,88      1,88      1,88
R-Sq          98,34      98,35      98,35      98,36
R-Sq(adj)     98,33      98,35      98,35      98,35
Mallows C-p   35,2      17,2      12,6      10,8

```

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.2. REGRESIÓN DE SERIES DE TIEMPO

En esta regresión stepwise, al analizar los p -valores de cada una de las variables meteorológicas se obtiene de manera similar que el mejor modelo para la humedad diaria es aquel que contiene las variables meteorológicas PR, TMax, HH, VM, TMin, LL, VV, RV, ST y TA.

De los diferentes análisis de regresión lineal simple para los datos se encontró un modelo que mejor se ajustó a los datos de la humedad relativa, además se realizó un análisis de regresión lineal múltiple usando los métodos de los mejores subconjuntos y la regresión stepwise, encontrándose que el mejor modelo para pronosticar la humedad diaria es aquel que contiene las variables meteorológicas ya mencionadas en el párrafo anterior. Sin embargo según el principio de la navaja de Occan, el mejor modelo es aquel en el que sólo aparece como variable predictora la PR.

3.2.4. Aplicación de variables ficticias al análisis de la humedad relativa

El objetivo de esta sección es determinar un modelo para predecir valores futuros de la humedad usando el modelado de la variación estacional mediante variables ficticias, aquí, nuevamente se trabaja con los datos de información meteorológica. En este análisis en principio se obtienen promedios diarios de los registros de la humedad y posteriormente los promedios de cada estación del año (es decir promedios del verano, otoño, invierno y primavera)(ver Tabla 2 del Apéndice A), los datos obtenidos se grafican en la Figura 3.6.

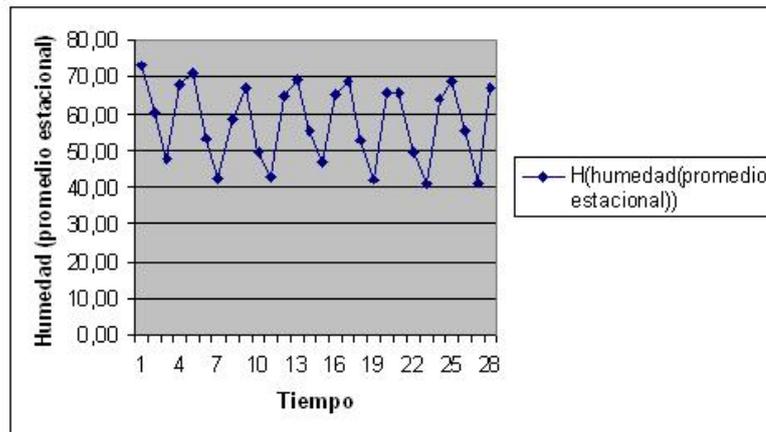


Figura 3.6: Gráfica de los promedios por estación de la variable Humedad.

En esta parte se determina un modelo de regresión con variables ficticias el cual permitirá realizar pronósticos de la humedad, tomando los datos dados en la Tabla 2 (ver Apéndice A). Se puede observar mediante la Figura 3.6 que aparte de mostrar una tendencia lineal, la serie de tiempo posee variación estacional constante. Por consiguiente, es razonable pensar que la humedad en el periodo t se describe en forma adecuada mediante el modelo de regresión

$$y_t = \beta_0 + \beta_1 t + \beta_2 Q_1 + \beta_3 Q_2 + \beta_4 Q_3 + \epsilon_t$$

donde Q_1 , Q_2 y Q_3 son variables ficticias estacionales, definidas como

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO
3.2. REGRESIÓN DE SERIES DE TIEMPO

$$Q_1 = \begin{cases} 1, & \text{si el periodo } t \text{ es el verano;} \\ 0, & \text{si no es así.} \end{cases}$$

$$Q_2 = \begin{cases} 1, & \text{si el periodo } t \text{ es el otoño;} \\ 0, & \text{si no es así.} \end{cases}$$

$$Q_3 = \begin{cases} 1, & \text{si el periodo } t \text{ es el invierno;} \\ 0, & \text{si no es así.} \end{cases}$$

Obsérvese que no se ha definido una variable ficticia para la primavera. En la Tabla 3.3 se dan las partes pertinentes de la salida que proporciona MINITAB sobre un análisis de regresión para los datos de los promedios de la humedad por estaciones del año en el que se usó el modelo de variable ficticia mencionada antes y en la cual se asumieron las suposiciones de varianza constante, independencia y normalidad.

Tabla 3.3: Análisis de regresión obtenido con MINITAB para los datos de la humedad usando el modelo $y_t = \beta_0 + \beta_1 t + \beta_2 Q_1 + \beta_3 Q_2 + \beta_4 Q_3 + \epsilon_t$.

```
The regression equation is
H(humedad(promedio estacional)) = 66,7 - 0,121 Tiempo + 4,04 Q1 - 11,1
Q2
                                - 21,4 Q3

Predictor    Coef    SE Coef    T    P
Constant    66,655    1,578    42,25 0,000
Tiempo      -0,12080    0,06972   -1,73 0,097
Q1           4,041    1,591    2,54 0,018
Q2          -11,135    1,584   -7,03 0,000
Q3          -21,397    1,579  -13,55 0,000

S = 2,95156   R-Sq = 93,3%   R-Sq(adj) = 92,2%

Analysis of Variance

Source      DF      SS      MS      F      P
Regression    4  2812,06  703,01  80,70 0,000
Residual Error 23   200,37    8,71
Total        27  3012,43

Durbin-Watson statistic = 0,893828
```

El análisis indica que debemos eliminar el término que contiene a la variable t ya que su p -valor es igual a 0.097, es decir el parámetro que representa al coeficiente de esta variable a un nivel de significancia del 5% debe ser cero.

Por consiguiente, es razonable pensar que la humedad en el periodo t se describe en forma adecuada mediante el modelo de regresión

$$y_t = \beta_0 + \beta_1 Q_1 + \beta_2 Q_2 + \beta_3 Q_3 + \epsilon_t$$

donde Q_1 , Q_2 y Q_3 son variables ficticias estacionales, definidas anteriormente.

En la Tabla 3.4 se dan las partes pertinentes de la salida que proporciona MINITAB sobre un análisis de regresión para los datos de los promedios de la humedad por estaciones

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO
3.3. MODELOS DE DESCOMPOSICIÓN

del año en el que se usó el modelo de variable ficticia mencionado antes y en el cual se hicieron las suposiciones de varianza constante, independencia y normalidad.

Tabla 3.4: Análisis de regresión obtenido con MINITAB para los datos de la humedad usando el modelo $y_t = \beta_0 + \beta_1 Q_1 + \beta_2 Q_2 + \beta_3 Q_3 + \epsilon_t$.

```
The regression equation is
H(humedad(promedio estacional)) = 64,7 + 4,40 Q1 - 10,9 Q2 - 21,3 Q3

Predictor      Coef    SE Coef      T      P
Constant      64,722   1,161    55,74  0,000
Q1             4,403    1,642     2,68  0,013
Q2            -10,893   1,642    -6,63  0,000
Q3            -21,276   1,642   -12,96  0,000

S = 3,07218   R-Sq = 92,5%   R-Sq(adj) = 91,5%

Analysis of Variance

Source         DF      SS      MS      F      P
Regression      3   2785,91  928,64  98,39  0,000
Residual Error  24    226,52    9,44
Total          27   3012,43
```

Durbin-Watson statistic = 0,791646

En la salida que ofrece MINITAB se muestra que el modelo es significativo ($F=98.39$ y el valor $p < 0.000$) y que cada una de las variables ficticias son significantes (todas las estadísticas t tienen un p -valor relacionado que es menor que .05). Además, $R^2 = 92.5$ y $s = 3.07$.

Con la ayuda de las estimaciones puntuales de mínimos cuadrados que ofrece MINITAB se determina un pronóstico puntual para y_{29} este es

$$\hat{y}_{29} = \hat{\beta}_0 + \hat{\beta}_1 Q_1 + \hat{\beta}_2 Q_2 + \hat{\beta}_3 Q_3 = 64.722 + 4.403(1) = 69.125.$$

Similarmente, un pronóstico puntual de y_{30} es

$$\hat{y}_{30} = \hat{\beta}_0 + \hat{\beta}_1 Q_1 + \hat{\beta}_2 Q_2 + \hat{\beta}_3 Q_3 = 64.722 + (-10.893)(1) = 53.829.$$

3.3. Modelos de descomposición

No hay bases teóricas para la obtención de los modelos de descomposición para pronosticar series de tiempo que manifiestan tendencia y efectos estacionales, son rigurosamente un enfoque intuitivo. No obstante, estos modelos han resultado útiles cuando los parámetros que describen a una serie de tiempo no cambian en el tiempo. La idea básica en la que se apoyan los modelos es en la descomposición de las series en varios factores: tendencia, estacionalidad, cíclica e irregular. Las estimaciones de estos factores se utilizan para describir las series de tiempo. Además, si los parámetros no cambian, las estimaciones se

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.3. MODELOS DE DESCOMPOSICIÓN

pueden usar para determinar pronósticos puntuales.

El modelo de descomposición multiplicativa es útil al modelar series de tiempo que manifiestan una variación estacional creciente o decreciente, el modelo de descomposición multiplicativa se puede plantear como sigue:

$$y_t = T_t \times S_t \times C_t \times I_t \quad (3.17)$$

donde

y_t es el valor observado de la serie de tiempo en el periodo t .

T_t es el componente (o factor) de la tendencia en el periodo t .

S_t es el componente (o factor) estacional en el periodo t .

C_t es el componente (o factor) cíclico en el periodo t .

I_t es el componente (o factor) irregular en el periodo t .

El modelo de descomposición aditiva se puede utilizar al modelar series temporales que muestran variación estacional constante.

Considere una serie de tiempo que exhibe variación estacional constante, el modelo de descomposición aditiva se puede plantear como sigue:

$$y_t = T_t + S_t + C_t + I_t \quad (3.18)$$

En este caso, T_t , S_t , C_t e I_t están definidas como la tendencia, en los factores, estacional, cíclico e irregular. Sin embargo, estos factores son aditivos en lugar de multiplicativos.

3.3.1. Aplicación de los modelos de descomposición al análisis de la humedad relativa

Lo primero que se realizó, fue obtener promedios diarios de los registros de la humedad y posteriormente se obtuvieron promedios de cada estación del año (es decir promedios del verano, otoño, invierno y primavera)(ver Tabla 2 del Apéndice A), los datos obtenidos se grafican en la Figura 3.6. Se puede observar en la Figura 3.6 que aparte de mostrar una tendencia lineal, la serie de tiempo posee variación estacional constante, por tanto es razonable pensar que la humedad en el periodo t se describe en forma adecuada mediante el modelo

$$y_t = T_t + S_t + C_t + I_t.$$

Para aplicar el método de descomposición se calculan las medias móviles y las medias móviles centradas (las cuales se denotan como CMA_t). Nuevamente el objetivo detrás del cálculo de estas medias es eliminar las variaciones estacionales y las fluctuaciones irregulares de los datos. La primera media móvil es el promedio de los primeros 4 valores de los registros de la humedad vista como promedios estacionales

$$\frac{72.99 + 60.42 + 47.62 + 67.97}{4} = 62.25321237.$$

Aquí se usó una media móvil de “4 periodos” porque los datos de la serie temporal de los registros de la humedad estacional están siendo tomados por estaciones del año ($L=4$ periodos o “estaciones” por año). La segunda media móvil se obtiene al eliminar el primer

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.3. MODELOS DE DESCOMPOSICIÓN

valor de los registros de la humedad estacional del promedio e incluir el siguiente valor de los registros de la humedad estacional. Por consiguiente se obtiene

$$\frac{60.42 + 47.62 + 67.97 + 70.96}{4} = 61.74442204.$$

Las medias móviles sucesivas se calculan en forma similar hasta que quede incluido el último valor de los registros de la humedad estacional. Obsérvese que usamos aquí el término “media móvil” porque al calcular estos promedios, avanzamos eliminando la observación más remota en la media anterior e incluyendo la observación siguiente en el promedio nuevo.

La primera media móvil corresponde a un tiempo que está en un punto intermedio entre los periodos 2 y 3, la segunda media móvil corresponde a un tiempo que está en el punto medio de los periodos 3 y 4, y así sucesivamente. Con el objeto de obtener promedios que correspondan a periodos de la serie de tiempo original de la humedad, calculamos las medias móviles centradas. Los promedios son las medias móviles de 4 periodos previamente calculadas. Por lo tanto, la primer media móvil centrada es

$$\frac{62.25321237 + 61.74442204}{2} = 61.9988172.$$

La segunda media móvil centrada es

$$\frac{61.74442204 + 59.96216661}{2} = 60.85329433.$$

Las medias móviles centradas sucesivas se determinan de manera similar. Los resultados de las medias móviles de 4 periodos y las medias móviles centradas para la serie de tiempo de los registros de la humedad estacional se dan en la Tabla 3 (ver Apéndice A). Se considera que la media móvil centrada en el periodo t , CMA_t , es igual a $t_t + c_t$, la estimación de $T_t + C_t$. La razón es que se supone que el procedimiento de obtener las medias móviles elimina

1. Las variaciones estacionales, y
2. Las fluctuaciones irregulares a corto plazo.

Se conservan los efectos de la tendencia (plazo más largo) y los efectos cíclicos, es decir, $t_t + c_t$.

Puesto que el modelo

$$y_t = T_t + S_t + C_t + I_t$$

esto quiere decir que

$$S_t + I_t = y_t - (T_t + C_t)$$

se infiere entonces que la estimación $s_t + i_t$ de $S_t + I_t$ es

$$s_t + i_t = y_t - (t_t + c_t) = y_t - CMA_t.$$

Los valores de $s_t + i_t$ son calculados y se presentan en la Tabla 4 (ver Apéndice A). Por lo que se puede encontrar s_t agrupando los valores de $s_t + i_t$ por estaciones del año

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO
3.3. MODELOS DE DESCOMPOSICIÓN

y calculando el promedio, \bar{s}_t de cada estación. Estos factores estacionales se normalizan restando la cantidad

$$\frac{\sum_{t=1}^L \bar{s}_t}{L} = 0.007174412$$

de cada uno de los valores de \bar{s}_t . Este proceso de normalización da como resultado $s_t = \bar{s}_t - (0.007174412)$ que es la estimación de S_t , donde L representa el número de estaciones del año. Estos cálculos se resumen en la Tabla 5 (ver Apéndice A). Como se ha definido, la observación compensada respecto a la variación estacional en el periodo t está dada por

$$d_t = y_t - s_t.$$

El objeto de calcular las observaciones compensadas respecto a la variación estacional es estimar mejor la componente de la tendencia T_t . Al restar el factor estacional estimado a y_t se elimina la estacionalidad de los datos y permite entender mejor la naturaleza de la tendencia. Las observaciones compensadas respecto a la variación estacional se presentan en la Tabla 6 (ver Apéndice A) y se grafican en la Figura 3.7.

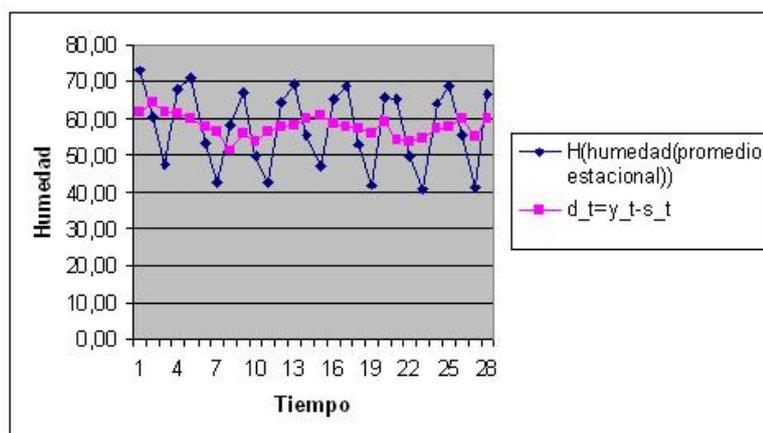


Figura 3.7: Gráfica de Excel de la humedad estacional compensada respecto a la variación estacional.

Puesto que en la Figura 3.7 las observaciones compensadas respecto a la variación estacional muestran una apariencia lineal, parece razonable suponer una tendencia lineal de la forma

$$T_t = \beta_0 + \beta_1 t.$$

Se estima T_t ajustando una recta a los datos compensados respecto a la variación estacional. Es decir, se calculan las estimaciones puntuales de mínimos cuadrados de los parámetros que están en el modelo de regresión lineal simple que relaciona la variable dependiente d_t con la variable independiente t:

$$d_t = \beta_0 + \beta_1 t + \epsilon_t.$$

Por consiguiente, se obtiene t_t , la estimación de T_t , al calcular

$$\hat{\beta}_1 = \frac{SS_{td}}{SS_{tt}} = -0.1205$$

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO
3.3. MODELOS DE DESCOMPOSICIÓN

y

$$\widehat{\beta}_0 = \bar{d} - \widehat{\beta}_1 \bar{t} = 59.5284.$$

Por lo tanto,

$$t_t = \widehat{\beta}_0 + \widehat{\beta}_1 t = 59.5284 - 0.1205t.$$

Las estimaciones puntuales de mínimos cuadrados se encuentran también, calculadas en excel, en la Tabla 7 (ver Apéndice A) y los valores de t_t , estan ya calculados en la 8 (ver Apéndice A).

Puesto que

$$y_t = T_t + S_t + C_t + I_t$$

esto quiere decir que

$$C_t + I_t = y_t - (T_t + S_t)$$

de donde se infiere que la estimación de $C_t + I_t$ esta dada por

$$c_t + i_t = y_t - (t_t + s_t).$$

Es decir, el primer valor para $c_t + i_t$ es

$$c_1 + i_1 = y_1 - (t_1 + s_1) = 2.319$$

el segundo valor para $c_t + i_t$ es

$$c_2 + i_2 = y_2 - (t_2 + s_2) = 5.4647.$$

El resto de los valores para $c_t + i_t$ se determinan de manera similar. Los resultados se resumen en la Tabla 9 (ver Apéndice A). Con objeto de promediar i_t , se calcula una media móvil de tres periodos de los valores $c_t + i_t$. Es decir la estimación de C_t es

$$c_t = \frac{(c_{t-1} + i_{t-1}) + (c_t + i_t) + (c_{t+1} + i_{t+1})}{3}.$$

Para terminar, calculamos la estimación de I_t :

$$i_t = (c_t + i_t) - c_t.$$

Los cálculos de c_t e i_t se resumen en la Tabla 10 (ver Apéndice A). Aunque no hay intervalo de predicción teóricamente correcto para y_t , un intervalo de predicción $100(1 - \alpha)\%$ aproximado para y_t es

$$[y_t \pm B_t[100(1 - \alpha)]]$$

donde $B_t[100(1 - \alpha)]$ es el límite de error en un intervalo de predicción de $100(1 - \alpha)\%$

$$[t_t \pm B_t[100(1 - \alpha)]]$$

para la observación compensada respecto a la variación estacional

$$d_t = T_t + \epsilon_t = \beta_0 + \beta_1 t + \epsilon_t.$$

CAPÍTULO 3. ANÁLISIS DE SERIES DE TIEMPO

3.3. MODELOS DE DESCOMPOSICIÓN

Por ejemplo si se usa MINITAB para predecir d_t con base en t usando la línea de tendencia anterior, tenemos que un intervalo de predicción del 95 % para d_{29} es

$$[49.875, 62.194]$$

los resultados obtenidos con MINITAB se muestran a continuación.

```

The regression equation is
d_t = 59,5 - 0,120 t

Predictor   Coef SE Coef   T   P
Constant   59,527  1,085  54,88  0,000
t          -0,12045  0,06535  -1,84  0,077

S = 2,79348  R-Sq = 11,6%  R-Sq(adj) = 8,2%

Analysis of Variance

Source      DF   SS      MS      F      P
Regression   1  26,505  26,505  3,40  0,077
Residual Error 26 202,892  7,804
Total       27 229,397

Predicted Values for New Observations

New
Obs  Fit SE Fit   95% CI      95% PI
1  56,035  1,085 (53,805; 58,264) (49,875; 62,194)

Values of Predictors for New Observations

New
Obs  t
1  29,0
    
```

Esto quiere decir que

$$B_{29}[95] = \frac{62.194 - 49.875}{2} = 6.1595.$$

Se infiere entonces que un intervalo de predicción aproximado del 95 % para y_{29} es

$$[\hat{y}_{29} - 6.1595, \hat{y}_{29} + 6.1595] = [61.1305, 73.4495].$$

Se realizó un análisis de los registros de la humedad usando un modelo de descomposición, los datos de la humedad fueron tomados cada 30 minutos, con las fechas del 9 de Mayo del 2003 al 28 de Octubre del 2010. Lo primero que se realizó, fue obtener promedios diarios de los registros de la humedad y posteriormente se obtuvieron promedios de cada estación del año (es decir promedios del verano, otoño, invierno y primavera), los datos obtenidos se grafican en la figura 3.5, la gráfica aparte de mostrar una tendencia lineal exhibe variación estacional constante, por lo que se realizó un análisis de los registros de la humedad usando un modelo de descomposición aditivo.

Capítulo 4

Suavización exponencial

La suavización exponencial es un método que muestra su mayor eficacia cuando los componentes (tendencia y variación estacional) de las series temporales manifiestan cambios en el tiempo. Es un método que no pondera de igual manera los valores observados de las series temporales. Las observaciones más recientes tienen un mayor peso que las observaciones más remotas. La ponderación desigual se consigue usando una o más constantes de suavización, lo cual determina cuanto peso se le da a cada observación.

En el método de la suavización exponencial simple se supone que la serie temporal no tiene tendencia, pero el nivel o media de la serie temporal puede modificarse con el tiempo.

Los indicios de error se usan para indicar cuando un sistema de pronóstico no está dando predicciones exactas.

El método de Holt de suavización exponencial es un método para pronosticar una serie temporal que tiene localmente una tendencia, pero una tasa de crecimiento (o pendiente) que cambia con el tiempo [12].

4.1. Suavización exponencial simple

El método de la suavización exponencial simple se usa para pronosticar una serie temporal cuando no hay tendencia o patrón estacional pero la media (o nivel) de la serie temporal y_t cambia lentamente con el tiempo.

En lugar de dar pesos iguales a cada observación, el método de suavización exponencial simple, dará a la observación más reciente, el mayor peso. Las observaciones más antiguas reciben pesos sucesivamente más pequeños. El procedimiento permite a quien elabora el pronóstico, actualizar la estimación del nivel de la serie temporal de modo que los cambios en el nivel se puedan detectar e incorporar en el sistema de pronósticos.

Estimación y pronóstico utilizando el método de suavización exponencial simple (ver Apéndice B).

1. Suponga que la serie temporal tiene un nivel (o media) que tiene la posibilidad de cambiar lentamente con el tiempo, pero no tiene tendencia ni patrón estacional. Entonces, la estimación ℓ_T del nivel (o media) de la serie temporal en el periodo T está dada por la ecuación de suavización $\ell_T = \alpha y_T + (1 - \alpha)\ell_{T-1}$ donde α es una constante de suavización entre 0 y 1, y ℓ_{T-1} es una estimación del nivel (o media)

de la serie temporal en el periodo $T - 1$.

2. Un pronóstico puntual efectuado en el periodo T para $y_{T+\tau}$ es $\hat{y}_{T+\tau}(T) = \ell_T$ ($\tau \in N$).
3. Si $\tau = 1$, entonces un intervalo de predicción del 95 por ciento calculado en el periodo T para y_{T+1} es $[\ell_T - z_{[0.025]}s, \ell_T + z_{[0.025]}s]$.
 Si $\tau = 2$, entonces un intervalo de predicción del 95 por ciento calculado en el periodo T para y_{T+2} es $[\ell_T - z_{[0.025]}s\sqrt{1 + \alpha^2}, \ell_T + z_{[0.025]}s\sqrt{1 + \alpha^2}]$.
 En general para cualquier τ , un intervalo de predicción del 95 por ciento calculado en el periodo T para $y_{T+\tau}$ es $[\ell_T - z_{[0.025]}s\sqrt{1 + (\tau - 1)\alpha^2}, \ell_T + z_{[0.025]}s\sqrt{1 + (\tau - 1)\alpha^2}]$ donde el error estándar s determinado en el tiempo T es:

$$s = \sqrt{\frac{SSE}{T-1}} = \sqrt{\frac{\sum_{t=1}^T [y_t - \hat{y}_t(t-1)]^2}{T-1}} = \sqrt{\frac{\sum_{t=1}^T [y_t - \ell_{t-1}]^2}{T-1}}.$$

4.2. Indicios de error

La primera señal que indica que algo anda mal es la llamada señal de la suma acumulada simple. Es un cociente que compara la suma acumulativa de errores con la desviación absoluta de la media suavizada.

Suponga que se tiene una historia de errores de pronóstico de un solo periodo más adelante T , $e_1(\alpha), e_2(\alpha), \dots, e_T(\alpha)$, (α) denota el valor particular de α empleado para obtener los errores de pronóstico de un solo periodo a futuro.

Definimos la **suma (Y) de los errores de pronóstico**:

$$Y(\alpha, T) = \sum_{t=1}^T e_t(\alpha),$$

luego

$$Y(\alpha, T - 1) = \sum_{t=1}^{T-1} e_t(\alpha)$$

y por tanto

$$Y(\alpha, T) = Y(\alpha, T - 1) + e_T(\alpha)$$

y la desviación absoluta de la media suavizada (DAMS), se define como:

$$DAMS(\alpha, T) = \alpha |e_T(\alpha)| + (1 - \alpha)DAMS(\alpha, T - 1). \quad (4.1)$$

Luego, el **indicio de error de la suma acumulada simple** $C(\alpha, T)$ se define como:

$$C(\alpha, T) = \left| \frac{Y(\alpha, T)}{DAMS(\alpha, T)} \right|. \quad (4.2)$$

Si $C(\alpha, T)$ es grande quiere decir que el sistema de pronóstico está generando errores consecuentemente positivos, o consecuentemente negativos.

Otro indicio de error que es muy usado en la práctica, es el **indicio de error suavizado**.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.3. MÉTODO DE HOLT, DE LA SUAVIZACIÓN EXPONENCIAL CORREGIDA DE LA TENDENCIA

Se define el error suavizado (E) de los errores de pronóstico de un solo periodo posterior como:

$$E(\alpha, T) = e_T(\alpha) + \alpha E(\alpha, T - 1). \quad (4.3)$$

Luego el indicio de error suavizado se define como:

$$S(\alpha, T) = \left| \frac{E(\alpha, T)}{DAMS(\alpha, T)} \right|. \quad (4.4)$$

4.3. Método de Holt, de la suavización exponencial corregida de la tendencia

1. Suponga que la serie de tiempo y_1, y_2, \dots, y_n muestra una tendencia lineal, para la cual el nivel y la tasa de crecimiento podrían estar cambiando sin ningún patrón estacional. Entonces la estimación ℓ_T del nivel de la serie temporal y la estimación b_T de tasa de crecimiento de la serie temporal en el periodo T están dadas por las ecuaciones de suavización:

$$\begin{aligned} \ell_T &= \alpha y_T + (1 - \alpha) [\ell_{T-1} + b_{T-1}] \\ b_T &= \gamma [\ell_T - \ell_{T-1}] + (1 - \gamma) b_{T-1} \end{aligned}$$

donde α y γ son contantes de suavización entre 0 y 1, y ℓ_{T-1} y b_{T-1} son estimaciones en el tiempo $T - 1$ para el nivel, y la tasa de crecimiento, respectivamente [14].

2. Un pronóstico puntual efectuado en el periodo T para $y_{T+\tau}$ es $\hat{y}_{T+\tau}(T) = \ell_T + \tau b_T$ ($\tau \in N$).
3. Si $\tau = 1$, entonces un intervalo de predicción del 95 por ciento calculado en el periodo T para y_{T+1} es $[(\ell_T + b_T) \pm z_{[0.025]} s]$.
Si $\tau = 2$, entonces un intervalo de predicción del 95 por ciento calculado en el periodo T para y_{T+2} es $[(\ell_T + 2b_T) \pm z_{[0.025]} s \sqrt{1 + \alpha^2(1 + \gamma)^2}]$.
Si $\tau = 3$, entonces un intervalo de predicción del 95 por ciento calculado en el periodo T para y_{T+3} es $[(\ell_T + 3b_T) \pm z_{[0.025]} s \sqrt{1 + \alpha^2(1 + \gamma)^2 + \alpha^2(1 + 2\gamma)^2}]$.
En general para $\tau \geq 2$, un intervalo de predicción del 95 por ciento calculado en el periodo T para $y_{T+\tau}$ es:

$[(\ell_T + \tau b_T) \pm z_{[0.025]} s \sqrt{1 + \alpha^2(1 + \gamma)^2 + \alpha^2(1 + 2\gamma)^2 + \dots + \alpha^2(1 + (\tau - 1)\gamma)^2}]$, es decir,
 $[(\ell_T + \tau b_T) \pm z_{[0.025]} s \sqrt{1 + \sum_{i=1}^{\tau-1} \alpha^2(1 + i\gamma)^2}]$ donde el error estándar s determinado en el tiempo T es:

$$s = \sqrt{\frac{SSE}{T-2}} = \sqrt{\frac{\sum_{t=1}^T [y_t - \hat{y}_t(t-1)]^2}{T-2}} = \sqrt{\frac{\sum_{t=1}^T [y_t - (\ell_{t-1} + b_{t-1})]^2}{T-2}}.$$

4.4. Método aditivo de Holt-Winters

1. Suponga que la serie de tiempo y_1, y_2, \dots, y_n manifiesta tendencia lineal y sigue un patrón con variación estacional constante (aditiva), y que el nivel, la tasa de crecimiento y el patrón estacional podrían estar cambiando. Entonces las ecuaciones de suavización que se dan enseguida, proporcionan la estimación ℓ_T del nivel, la estimación b_T de tasa de crecimiento, y la estimación sn_T del factor estacional de la serie temporal en el periodo T están dadas por las ecuaciones de suavización:

$$\begin{aligned}\ell_T &= \alpha(y_T - sn_{T-L}) + (1 - \alpha)[\ell_{T-1} + b_{T-1}] \\ b_T &= \gamma[\ell_T - \ell_{T-1}] + (1 - \gamma)b_{T-1} \\ sn_T &= \delta(y_T - \ell_T) + (1 - \delta)sn_{T-L}\end{aligned}$$

donde α, γ y δ son contantes de suavización entre 0 y 1, ℓ_{T-1} y b_{T-1} son estimaciones en el tiempo $T - 1$ para el nivel, y la tasa de crecimiento, y sn_{T-L} es la estimación en el periodo $T - L$ del factor estacional [14].

2. Un pronóstico puntual efectuado en el periodo T para $y_{T+\tau}$ es $\hat{y}_{T+\tau}(T) = \ell_T + \tau b_T + sn_{T+\tau-L}$ ($\tau \in N$).
3. Un intervalo de predicción del 95 por ciento calculado en el periodo T para $y_{T+\tau}$ es :

$$[\hat{y}_{T+\tau}(T) \pm z_{[0.025]} s \sqrt{c_\tau}]$$

Si $\tau = 1$, entonces $c_1 = 1$.

Si $2 \leq \tau \leq L$, entonces $c_\tau = \left[1 + \sum_{j=1}^{\tau-1} \alpha^2 (1 + j\gamma)^2\right]$.

Si $L \leq \tau$, entonces $c_\tau = 1 + \sum_{j=1}^{\tau-1} [\alpha(1 + j\gamma) + d_{j,L}(1 - \alpha)\gamma]^2$, donde $d_{j,L} = 1$ si $j = L$ y 0 si no es así, donde el error estándar s determinado en el periodo T es:

$$s = \sqrt{\frac{SSE}{T-3}} = \sqrt{\frac{\sum_{t=1}^T [y_t - \hat{y}_t(t-1)]^2}{T-3}} = \sqrt{\frac{\sum_{t=1}^T [y_t - (\ell_{t-1} + b_{t-1} + sn_{t-L})]^2}{T-3}}$$

4.5. Método multiplicativo de Holt-Winters

1. Suponga que la serie de tiempo y_1, y_2, \dots, y_n manifiesta tendencia lineal local, y sigue un patrón con variación estacional creciente (multiplicativa), y que el nivel, la tasa de crecimiento y el patrón estacional podrían estar cambiando. Entonces las ecuaciones de suavización que se dan enseguida, proporcionan la estimación ℓ_T del nivel, la estimación b_T de tasa de crecimiento, y la estimación sn_T del factor estacional de la serie de tiempo en el periodo T :

$$\begin{aligned}\ell_T &= \alpha\left(\frac{y_T}{sn_{T-L}}\right) + (1 - \alpha)[\ell_{T-1} + b_{T-1}] \\ b_T &= \gamma[\ell_T - \ell_{T-1}] + (1 - \gamma)b_{T-1} \\ sn_T &= \delta\left(\frac{y_T}{\ell_T}\right) + (1 - \delta)sn_{T-L}\end{aligned}$$

donde α, γ y δ son contantes de suavización entre 0 y 1, ℓ_{T-1} y b_{T-1} son estimaciones en el tiempo $T - 1$ para el nivel, y la tasa de crecimiento, y sn_{T-L} es la estimación en el periodo $T - L$ del factor estacional [14].

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.6. TENDENCIA AMORTIGUADA, Y OTROS MÉTODOS DE SUAVIZACIÓN EXPONENCIAL

2. Un pronóstico puntual hecho en el periodo T para $y_{T+\tau}$ es $\hat{y}_{T+\tau}(T) = (\ell_T + \tau b_T)sn_{T+\tau-L}$ ($\tau \in N$), donde $sn_{T+\tau-L}$ es la estimación más reciente del factor estacional para la estación correspondiente al periodo $T + \tau$.
3. Un intervalo de predicción del 95 por ciento calculado en el periodo T para $y_{T+\tau}$ es :

$$[\hat{y}_{T+\tau}(T) \pm z_{[0.025]} s_r (\sqrt{c_\tau}) (sn_{T+\tau-L})]$$

Si $\tau = 1$, entonces $c_1 = (\ell_T + b_T)^2$.

Si $\tau = 2$, entonces $c_2 = \alpha^2(1 + \gamma)^2(\ell_T + b_T)^2 + (\ell_T + 2b_T)^2$.

Si $\tau = 3$, entonces $c_3 = \alpha^2(1 + 2\gamma)^2(\ell_T + b_T)^2 + \alpha^2(1 + \gamma)^2(\ell_T + 2b_T)^2 + (\ell_T + 3b_T)^2$.

Si $2 \leq \tau \leq L$, entonces $c_\tau = \sum_{j=1}^{\tau-1} \alpha^2(1 + [\tau - j]\gamma)^2(\ell_T + jb_T)^2 + (\ell_T + \tau b_T)^2 = \alpha^2(1 + [\tau - 1]\gamma)^2(\ell_T + b_T)^2 + \dots + \alpha^2(1 + \gamma)^2(\ell_T + [\tau - 1]b_T)^2 + (\ell_T + \tau b_T)^2$.

El error estándar relativo s_r determinado en el periodo T es:

$$s_r = \sqrt{\frac{\sum_{t=1}^T \left[\frac{y_t - \hat{y}_t(t-1)}{\hat{y}_t(t-1)} \right]^2}{T-3}} = \sqrt{\frac{\sum_{t=1}^T \left[\frac{y_t - (\ell_{t-1} + b_{t-1})sn_{t-L}}{(\ell_{t-1} + b_{t-1})sn_{t-L}} \right]^2}{T-3}}.$$

4.6. Tendencia amortiguada, y otros métodos de suavización exponencial

Hasta este momento, entre los métodos presentados se encuentran los métodos exponenciales que se aplican con más frecuencia. Otro método de suavización exponencial común es la suavización exponencial de la tendencia amortiguada de Gardner y McKenzie. Este método es apropiado para pronosticar una serie de tiempo cuya tasa de crecimiento no se mantendrá en el futuro, cuyos efectos tienen que ser amortiguados. El amortiguamiento de la tasa de crecimiento, significa reducirla de tamaño de modo que disminuya el incremento o decremento de los pronósticos.

4.6.1. Método de la tendencia amortiguada

1. Suponga que la serie de tiempo y_1, y_2, \dots, y_n manifiesta tendencia lineal en la cual el nivel y la tasa de crecimiento cambian sin ningún patrón estacional. Además suponga que la pregunta es si la tasa de crecimiento al final de la serie temporal continuará en el futuro. Entonces, la estimación ℓ_T del nivel y la estimación b_T de tasa de crecimiento de la serie de tiempo en el periodo T están dadas por las ecuaciones de suavización:

$$\begin{aligned} \ell_T &= \alpha y_T + (1 - \alpha)[\ell_{T-1} + \phi b_{T-1}] \\ b_T &= \gamma[\ell_T - \ell_{T-1}] + (1 - \gamma)\phi b_{T-1} \end{aligned}$$

donde α y γ son constantes de suavización entre 0 y 1, y ϕ es un factor de amortiguamiento entre 0 y 1.

2. Un pronóstico puntual hecho en el periodo T para $y_{T+\tau}$ es $\hat{y}_{T+\tau}(T) = \ell_T + (\phi b_T + \phi^2 b_T + \dots + \phi^\tau b_T)$ ($\tau \in N$).

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.6. TENDENCIA AMORTIGUADA, Y OTROS MÉTODOS DE SUAVIZACIÓN EXPONENCIAL

3. Si $\tau = 1$, entonces un intervalo de predicción del 95 por ciento calculado en el periodo T para y_{T+1} es :

$$[\hat{y}_{T+1}(T) \pm z_{[0.025]}s]$$

Si $\tau = 2$, entonces un intervalo de predicción del 95 por ciento calculado en el periodo T para y_{T+2} es :

$$[\hat{y}_{T+2}(T) \pm z_{[0.025]}s\sqrt{1 + \alpha^2(1 + \phi\gamma)^2}]$$

Si $\tau = 3$, entonces un intervalo de predicción del 95 por ciento calculado en el periodo T para y_{T+3} es :

$$[\hat{y}_{T+3}(T) \pm z_{[0.025]}s\sqrt{1 + \alpha^2(1 + \phi\gamma)^2 + \alpha^2(1 + \phi\gamma + \phi^2\gamma)^2}]$$

En general si $\tau \geq 2$, entonces un intervalo de predicción del 95 por ciento calculado en el periodo T para $y_{T+\tau}$ es :

$$[\hat{y}_{T+\tau}(T) \pm z_{[0.025]}s\sqrt{1 + \sum_{j=1}^{\tau-1} \alpha^2(1 + \phi_j\gamma)^2}]$$

donde $\phi_j = \phi + \phi^2 + \dots + \phi^j$ [12] y [14].

4.6.2. Método aditivo de Holt-Winters con tendencia amortiguada

1. La estimación ℓ_T del nivel, la estimación b_T de la tasa de crecimiento, y la estimación sn_T del factor estacional de la serie de tiempo en el periodo T se determinan, mediante las ecuaciones de suavización

$$\begin{aligned} \ell_T &= \alpha(y_T - sn_{T-L}) + (1 - \alpha)(\ell_{T-1} + \phi(b_{T-1})) \\ b_T &= \gamma(\ell_T - \ell_{T-1}) + (1 - \gamma)\phi(b_{T-1}) \\ sn_T &= \delta(y_T - \ell_T) + (1 - \delta)sn_{T-L} \end{aligned}$$

donde α , γ y δ son constantes de suavización entre 0 y 1, y ϕ es un factor de amortiguamiento entre 0 y 1. La forma de corrección del error de las ecuaciones de suavización es

$$\begin{aligned} \ell_T &= \ell_{T-1} + \phi(b_{T-1}) + \alpha[y_T - (\ell_{T-1} + \phi(b_{T-1}) + sn_{T-L})] \\ b_T &= \phi(b_{T-1}) + \alpha\gamma[y_T - (\ell_{T-1} + \phi(b_{T-1}) + sn_{T-L})] \\ sn_T &= sn_{T-L} + (1 - \alpha)\delta[y_T - (\ell_{T-1} + \phi(b_{T-1}) + sn_{T-L})] \end{aligned}$$

2. Un pronóstico puntual realizado en el periodo T para $y_{T+\tau}$ es

$$\hat{y}_{T+\tau}(T) = \ell_T + (\phi(b_T) + \phi^2(b_T) + \phi^3(b_T) + \dots + \phi^\tau(b_T)) + sn_{T+\tau-L}$$

donde $sn_{T+\tau-L}$ es la estimación mas reciente del factor estacional de la estación correspondiente al periodo $T + \tau$.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.6. TENDENCIA AMORTIGUADA, Y OTROS MÉTODOS DE SUAVIZACIÓN EXPONENCIAL

3. Un intervalo de predicción del 95 por ciento calculado en el periodo T para $y_{T+\tau}$ es

$$[\hat{y}_{T+\tau}(T) \pm z_{[0.025]} s \sqrt{c_\tau}]$$

Si $\tau = 1$ entonces $c_1 = 1$

Si $\tau \geq 2$ entonces $c_\tau = 1 + \sum_{i=1}^{\tau-1} [\alpha(1 + \phi_j \gamma) + d_{j,L}(1 - \alpha)\delta]^2$

donde $d_{j,L} = 1$ si j es un entero múltiplo de L y 0 si no es así y $\phi_j = \phi + \phi^2 + \dots + \phi^j$ [12] y [14].

4.6.3. Método multiplicativo de Holt-Winters con tendencia amortiguada

1. La estimación ℓ_T del nivel, la estimación b_T de la tasa de crecimiento, y la estimación sn_T del factor estacional de la serie temporal en el periodo T se determinan, mediante las ecuaciones de suavización

$$\begin{aligned} \ell_T &= \alpha \left(\frac{y_T}{sn_{T-L}} \right) + (1 - \alpha)(\ell_{T-1} + \phi(b_{T-1})) \\ b_T &= \gamma(\ell_T - \ell_{T-1}) + (1 - \gamma)\phi(b_{T-1}) \\ sn_T &= \delta \left(\frac{y_T}{\ell_T} \right) + (1 - \delta)sn_{T-L} \end{aligned}$$

donde α , γ y δ son constantes de suavización entre 0 y 1, y ϕ es un factor de amortiguamiento entre 0 y 1. La forma de corrección del error de las ecuaciones de suavización es

$$\begin{aligned} \ell_T &= \ell_{T-1} + \phi(b_{T-1}) + \alpha \frac{[y_T - (\ell_{T-1} + \phi(b_{T-1}))sn_{T-L}]}{sn_{T-L}} \\ b_T &= \phi(b_{T-1}) + \alpha \gamma \frac{[y_T - (\ell_{T-1} + \phi(b_{T-1}))sn_{T-L}]}{sn_{T-L}} \\ sn_T &= sn_{T-L} + (1 - \alpha) \delta \frac{[y_T - (\ell_{T-1} + \phi(b_{T-1}))sn_{T-L}]}{\ell_T} \end{aligned}$$

2. Un pronóstico puntual realizado en el periodo T para $y_{T+\tau}$ es

$$\hat{y}_{T+\tau}(T) = (\ell_T + (\phi(b_T) + \phi^2(b_T) + \phi^3(b_T) + \dots + \phi^\tau(b_T)))sn_{T+\tau-L}$$

donde $sn_{T+\tau-L}$ es la estimación mas reciente del factor estacional de la estación correspondiente al periodo $T + \tau$.

3. Un intervalo de predicción aproximado del 95 por ciento calculado en el periodo T para $y_{T+\tau}$ es

$$[\hat{y}_{T+\tau}(T) \pm z_{[0.025]} s_r(\sqrt{c_\tau})(sn_{T+\tau-L})]$$

Si $\tau = 1$ entonces $c_1 = (\ell_T + \phi(b_T))^2$

Si $2 \leq \tau \leq L$ entonces

$c_\tau = [\sum_{j=1}^{\tau-1} \alpha^2(1 + [\tau - j]\gamma)^2(\ell_T + \phi_j(b_T))^2] + (\ell_T + \phi_\tau(b_T))^2 = \alpha^2(1 + [\tau - 1]\gamma)^2(\ell_T + \phi_1(b_T))^2 + \alpha^2(1 + [\tau - 2]\gamma)^2(\ell_T + \phi_2(b_T))^2 + \dots + \alpha^2(1 + \gamma)^2(\ell_T + \phi_{\tau-1}(b_T))^2 + (\ell_T + \phi_\tau(b_T))^2$

donde $\phi_j = \phi + \phi^2 + \dots + \phi^j$ [12] y [14].

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.7. MODELOS PARA SUAVIZACIÓN EXPONENCIAL, E INTERVALOS DE PREDICCIÓN

4.6.4. Método multiplicativo de Holt-Winters sin tendencia

1. La estimación ℓ_T del nivel, la estimación b_T de la tasa de crecimiento, y la estimación sn_T del factor estacional de la serie temporal en el periodo T se determinan, mediante las ecuaciones de suavización

$$\begin{aligned}\ell_T &= \alpha\left(\frac{y_T}{sn_{T-L}}\right) + (1 - \alpha)(\ell_{T-1}) \\ sn_T &= \delta\left(\frac{y_T}{\ell_T}\right) + (1 - \delta)sn_{T-L}\end{aligned}$$

donde α y δ son constantes de suavización entre 0 y 1.

2. Un pronóstico puntual realizado en el periodo T para $y_{T+\tau}$ es

$$\hat{y}_{T+\tau}(T) = (\ell_T)sn_{T+\tau-L}$$

donde $sn_{T+\tau-L}$ es la estimación mas reciente del factor estacional de la estación correspondiente al periodo $T + \tau$.

3. Un intervalo de predicción aproximado del 95 por ciento calculado en el periodo T para $y_{T+\tau}$ donde $1 \leq \tau \leq L$ es

$$\left[\hat{y}_{T+\tau}(T) \pm z_{[0.025]} s_r (\sqrt{1 + (\tau - 1)\alpha^2}) (\ell_T sn_{T+\tau-L}) \right].$$

4.7. Modelos para suavización exponencial, e intervalos de predicción

Todos los métodos de suavización exponencial tienen un modelo estadístico. Los modelos estadísticos son necesarios para deducir fórmulas para los intervalos de predicción. Usaremos los modelos del espacio del estado con una sola fuente de error para los modelos de suavización exponencial. Los componentes de la suavización exponencial se llaman estados en los modelos del espacio del estado. La notación para los componentes (estados) es

Componente (estado)	Modelo	Estimación
Nivel en el periodo t	L_t	ℓ_t
Tasa de crecimiento en el periodo t	B_t	b_t
Factor estacional en el periodo t	SN_t	sn_t

La notación para las constantes de suavización y del factor de amortiguamiento son

	Parametro	Estimación
Constante de suavización para el nivel	α	$\hat{\alpha}$
Constante de suavización para la tasa de crecimiento	γ	$\hat{\gamma}$
Constante de suavización para la estación	δ	$\hat{\delta}$
Factor de amortiguamiento	ϕ	$\hat{\phi}$

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

Todas las ecuaciones del espacio del estado tienen una ecuación de observación, y una o mas ecuaciones de estado. La ecuación de observación es una ecuación para el valor y_t , el cual puede ser observado. Las ecuaciones de estado muestran como los componentes no observados (estados), que son el nivel, la tasa de crecimiento, y el factor estacional, cambian de un periodo al siguiente. Los modelos tienen una fuente aleatoria de error ϵ_t . El término de error ϵ_t tiene el valor de una distribución normal cuya media es cero, es decir $E(\epsilon_t) = 0$, y una desviación estándar s , es decir $Var(\epsilon_t) = \sigma^2$, que es la misma para cada uno de los periodos. Además, se supone que los términos de error $\epsilon_1, \epsilon_2, \epsilon_3, \dots$, en diferentes periodos son estadísticamente independientes entre si.

Los modelos del espacio del estado para los modelos con una sola fuente de error se presentan en la Tabla 4.0.

Tabla 4.0: Modelos del espacio del estado para los métodos de suavización exponencial

Método	Modelo	Modelo
	Ecuación de observación	Ecuaciones de estado
SES	$y_t = L_{t-1} + \epsilon_t$	$L_t = L_{t-1} + \alpha\epsilon_t$
MHSECT	$y_t = L_{t-1} + B_{t-1} + \epsilon_t$	$L_t = L_{t-1} + B_{t-1} + \alpha\epsilon_t$ $B_t = B_{t-1} + \alpha\gamma\epsilon_t$
MA H-W	$y_t = L_{t-1} + B_{t-1} + SN_{t-L} + \epsilon_t$	$L_t = L_{t-1} + B_{t-1} + \alpha\epsilon_t$ $B_t = B_{t-1} + \alpha\gamma\epsilon_t$ $SN_t = SN_{t-L} + (1 - \alpha)\delta\epsilon_t$
MM H-W	$y_t = (L_{t-1} + B_{t-1})SN_{t-L}(1 + \epsilon_t)$	$L_t = L_{t-1} + B_{t-1} + \alpha(L_{t-1} + B_{t-1})\epsilon_t$ $B_t = B_{t-1} + \alpha\gamma(L_{t-1} + B_{t-1})\epsilon_t$ $SN_t = SN_{t-L} + (1 - \alpha)\delta(SN_{t-L})\epsilon_t$
Método de la TA	$y_t = L_{t-1} + \phi(B_{t-1}) + \epsilon_t$	$L_t = L_{t-1} + \phi(B_{t-1}) + \alpha\epsilon_t$ $B_t = \phi(B_{t-1}) + \alpha\gamma\epsilon_t$
MA H-W con TA	$y_t = L_{t-1} + \phi(B_{t-1}) + SN_{t-L} + \epsilon_t$	$L_t = L_{t-1} + \phi(B_{t-1}) + \alpha\epsilon_t$ $B_t = \phi(B_{t-1}) + \alpha\gamma\epsilon_t$ $SN_t = SN_{t-L} + (1 - \alpha)\delta\epsilon_t$
MM H-W con TA	$y_t = (L_{t-1} + \phi(B_{t-1}))SN_{t-L}(1 + \epsilon_t)$	$L_t = L_{t-1} + \phi(B_{t-1}) + \alpha(L_{t-1} + \phi(B_{t-1}))\epsilon_t$ $B_t = \phi(B_{t-1}) + \alpha\gamma(L_{t-1} + \phi(B_{t-1}))\epsilon_t$ $SN_t = SN_{t-L} + (1 - \alpha)\delta(SN_{t-L})\epsilon_t$
MM H-W sin tendencia	$y_t = (L_{t-1})(SN_{t-L})(1 + \epsilon_t)$	$L_t = L_{t-1} + \alpha(L_{t-1})\epsilon_t$ $SN_t = SN_{t-L} + (1 - \alpha)\delta(SN_{t-L})\epsilon_t$

4.8. Aplicación de los métodos de suavización

Los datos a analizar provienen de la estación Meteorológica del Departamento de Irrigación de la U.A.CH. (Universidad Autónoma de Chapingo) por medio del método de la suavización exponencial, los datos de la humedad fueron tomados cada 30 minutos, con las fechas del 9 de Mayo del 2003 al 28 de Octubre del 2010. Lo primero que se realizó, fue obtener promedios diarios de los registros de la humedad, los datos de la humedad se presentan en la Tabla 3.1 (ver Apéndice A), la gráfica de los datos (ver Figura 3.1) sugiere que no hay tendencia o patrón estacional muy bien definido. Es posible también que el nivel (o media) pudiera estar cambiando lentamente a lo largo del tiempo.

4.8.1. Aplicación del método de la suavización exponencial simple

El procedimiento de suavización exponencial simple, inicia con el cálculo de una estimación inicial ℓ_0 del nivel (o media) de la serie en el periodo $t = 0$. Se calcula esta estimación

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

ℓ_0 promediando los primeros 1328 valores de la serie (casi la mitad de los datos), obteniéndose:

$$\ell_0 = \frac{\sum_{t=1}^{1328} y_t}{1328} = 58.819.$$

Luego se supone que al final del periodo $T - 1$ se tiene una estimación ℓ_{T-1} para el nivel (o media) de la serie de tiempo. Después se asume que en el periodo T se obtiene una nueva observación y_T , por lo que se puede actualizar ℓ_{T-1} a ℓ_T , la cual es la nueva actualización del nivel (o media) en el periodo T . Se calcula la estimación actualizada mediante la ecuación de suavización

$$\ell_T = \alpha(y_T) + (1 - \alpha)\ell_{T-1}.$$

En este caso, α es una constante de suavización entre 0 y 1. La ecuación de suavización plantea que ℓ_T , la estimación del nivel efectuada en el periodo T , es igual a una fracción α (por ejemplo, 0.1) de la observación y_T apenas observada de la serie de tiempo, más una fracción $(1 - \alpha)$ (por ejemplo, 0.9) de ℓ_{T-1} , la estimación del nivel efectuada en el periodo $T - 1$. A medida que cambie más el nivel del proceso, más influye en la estimación un valor de la serie de tiempo recientemente observado y, por consiguiente, más grande debe ser la constante de suavización α .

Se empieza con la estimación inicial $\ell_0 = 58.819$ misma que se actualiza aplicando la ecuación de suavización a los 2657 datos observados de la humedad. Para hacer esto, fijamos arbitrariamente α en 0.1, y con el fin de juzgar la propiedad de esta elección de α , se calculan pronósticos para “solo un periodo más adelante” de los registros de la humedad antiguos, a medida que efectuamos el procedimiento de suavización. Un pronóstico de y_T un periodo más adelante es la estimación ℓ_{T-1} del nivel (o media) realizado en el periodo $T - 1$. Entonces, puesto que la estimación inicial del nivel es $\ell_0 = 58.819$, se infiere que 58.819 es el pronóstico de y_1 hecho en el tiempo 0. Como se puede ver en la Tabla 4.1 que $y_1 = 42.65$, se tiene un error de pronóstico de $42.65 - 58.819 = -16.17$.

Al utilizar $y_1 = 42.65$, se puede actualizar ℓ_0 a ℓ_1 , una estimación efectuada en el periodo 1 del nivel (o media) de la serie de tiempo, de esta manera

$$\ell_1 = \alpha(y_1) + (1 - \alpha)\ell_0 = (0.1)(42.65) + (0.9)(58.819) = 57.202.$$

Como $\ell_1 = 57.202$ es un pronóstico hecho en el periodo 1 para y_2 , y como se puede ver en la Tabla 4.1 que $y_2 = 55.88$, se tiene un error de pronóstico de $55.88 - 57.202 = -1.32$. Si se usa $y_2 = 55.88$, se puede actualizar ℓ_1 a ℓ_2 , una estimación del nivel (o media) hecha en el periodo 2, usando la ecuación

$$\ell_2 = \alpha(y_2) + (1 - \alpha)\ell_1 = (0.1)(55.88) + (0.9)(57.202) = 57.070.$$

Por referencia lógica, esto lleva a que, al ser 57.070 el pronóstico hecho en el periodo 2 para y_3 , y según la Tabla 4.1 se puede ver que $y_3 = 49.38$, se tiene un pronóstico de error de $49.38 - 57.070 = -7.69$.

Se continúa con este procedimiento considerando los 2657 periodos de los datos anteriores. En la Tabla 4.1 se presenta una porción de los resultados de los 2657 periodos de los datos, de la humedad diaria, en una hoja de cálculo de Excel.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

Tabla 4.1: Hoja de cálculo para la suavización exponencial simple en el caso de la humedad diaria con $\alpha = 0.1$.

n	alfa	SSE	ssquare	s	
2657	0,1	210579,4529	79,28443258	8,90418062	
		Estimacion	Pronostico		Error
Periodo	Humedad diaria	suavizada por nivel	hecho en el ultimo periodo	Error del pronostico	cuadrado del pronostico
0		58,81976657			
1	42,65	57,20278991	58,81976657	-16,17	261,461351
2	55,88	57,07051092	57,20278991	-1,32	1,74977315
3	49,38	56,30145983	57,07051092	-7,69	59,1439582
4	49,44	55,61531384	56,30145983	-6,86	47,079631
5	45,94	54,64778246	55,61531384	-9,68	93,611698
6	42,52	53,43500421	54,64778246	-12,13	147,083107
7	43,73	52,46450379	53,43500421	-9,71	94,1871068
8	43,73	51,59105341	52,46450379	-8,73	76,2915565
9	42,71	50,70294807	51,59105341	-8,88	78,8731097
10	33,90	49,02265326	50,70294807	-16,80	282,339064
11	29,81	47,10138794	49,02265326	-19,21	369,126045
12	34,42	45,83324914	47,10138794	-12,68	160,8176
13	44,46	45,69592423	45,83324914	-1,37	1,88581321
14	48,25	45,95133181	45,69592423	2,55	6,52330304
15	49,08	46,26419863	45,95133181	3,13	9,78856466
16	52,88	46,92577876	46,26419863	6,62	43,7688278
17	60,77	48,31020089	46,92577876	13,84	191,662462
18	59,69	49,4481808	48,31020089	11,38	129,499828
19	58,48	50,35136272	49,4481808	9,03	81,5737581
20	58,75	51,19122645	50,35136272	8,40	70,5371082
.
.
.
2648	57,31	58,88888567	59,06431741	-1,75	3,07762958
2649	51,81	58,1809971	58,88888567	-7,08	50,1106223
2650	48,96	57,25889739	58,1809971	-9,22	85,0267876
2651	49,58	56,49100765	57,25889739	-7,68	58,9654652
2652	52,48	56,08990689	56,49100765	-4,01	16,0881824
2653	43,23	54,8039162	56,08990689	-12,86	165,377205
2654	43,27	53,65052458	54,8039162	-11,53	133,031223
2655	49,04	53,18947212	53,65052458	-4,61	21,2569369
2656	48,40	52,71052491	53,18947212	-4,79	22,9390432
2657	48,67	52,30647242	52,71052491	-4,04	16,3258415

Puesto que se ejecutan los mismos cálculos para cada uno de los 2657 periodos, la suavización exponencial simple se adecúa con facilidad a una hoja de cálculo, ya que las fórmulas para el periodo 1 se copian para los otros 2656 periodos. Con los resultados de la Tabla 4.1, tenemos que para $\alpha = 0.1$, la suma de cuadrados del error del pronóstico es 210579.4529. Para determinar un buen valor de α , recurrimos al Solver de Excel (ver Apéndice C) para especificar un valor de α que proporcione un valor mínimo para la suma

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

de cuadrados del error (SSE), cuando los valores posibles de α varían entre 0 y 1. En la Tabla 4.2 se presenta la hoja de cálculo de Excel con la solución que proporciona el Solver.

Tabla 4.2: Hoja de cálculo de Excel con valores de α , s , y ℓ_{2657} cuando SSE tiene un valor mínimo.

n	alfa	SSE	ssquare	s	
2657	0,97592247	121144,8083	45,6117501	6,75364717	
		Estimacion	Pronostico		Error
	Humedad	suavizada	hecho en el	Error del	cuadrado del
Periodo	diaria	por nivel	ultimo periodo	pronostico	pronostico
0		58,81976657			
1	42,65	43,03932796	58,81976657	-16,17	261,461351
2	55,88	55,57082839	43,03932796	12,84	164,882858
3	49,38	49,52905983	55,57082839	-6,19	38,3263562
4	49,44	49,44214434	49,52905983	-0,09	0,00793165
5	45,94	46,02432297	49,44214434	-3,50	12,265015
6	42,52	42,60437542	46,02432297	-3,50	12,2802795
7	43,73	43,70289775	42,60437542	1,13	1,26703068
8	43,73	43,72934744	43,70289775	0,03	0,00073453
9	42,71	42,73454336	43,72934744	-1,02	1,03906921
10	33,90	34,11271394	42,73454336	-8,83	78,0491564
11	29,81	29,9135987	34,11271394	-4,30	18,5133473
12	34,42	34,31149701	29,9135987	4,51	20,3076526
13	44,46	44,21564916	34,31149701	10,15	102,992113
14	48,25	48,15286282	44,21564916	4,03	16,2759867
15	49,08	49,05767683	48,15286282	0,93	0,85958336
16	52,88	52,78796792	49,05767683	3,82	14,6101544
17	60,77	60,57781242	52,78796792	7,98	63,7128362
18	59,69	59,71137633	60,57781242	-0,89	0,78821089
19	58,48	58,50964849	59,71137633	-1,23	1,51628766
20	58,75	58,74421293	58,50964849	0,24	0,05776885
.
.
.
2648	57,31	57,25860757	55,17554372	2,13	4,55590363
2649	51,81	51,94118899	57,25860757	-5,45	29,6873245
2650	48,96	49,03177965	51,94118899	-2,98	8,88748777
2651	49,58	49,56680021	49,03177965	0,55	0,30054555
2652	52,48	52,40985736	49,56680021	2,91	8,48673301
2653	43,23	43,45102825	52,40985736	-9,18	84,2697811
2654	43,27	43,27435871	43,45102825	-0,18	0,03277123
2655	49,04	48,90117763	43,27435871	5,77	33,2426195
2656	48,40	48,41206712	48,90117763	-0,50	0,25117901
2657	48,67	48,66378961	48,41206712	0,26	0,06652937

Se puede ver, que cuando la SSE está en su valor mínimo de 121144.8083, el valor de la constante de suavización es $\alpha = 0.9759$, y la estimación final del nivel (o media) es $\ell_{2657} = 48.66$. Este valor de α indica que el nivel (o media) de la serie de tiempo no se

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

modifica con el tiempo.

En el método de la suavización exponencial simple, un pronóstico puntual en el tiempo T de cualquier valor futuro $y_{T+\tau}$ de una serie de tiempo es la última estimación ℓ_T para la media de la serie de tiempo porque no hay tendencia o patrón estacional que explotar. Sin embargo, se espera, que será menos exacto cuando pronosticamos muy en el futuro. Por tanto, los intervalos de predicción serán menos exactos (o sea más amplios).

Como $\alpha = 0.9759$ es un “buen valor” para la constante de suavización al pronosticar las 2657 observaciones de la humedad, que se proporcionan en la Tabla 4.2. Por tanto, al aplicar la suavización exponencial simple con $\alpha = 0.9759$ para pronosticar la humedad futura diaria. En la Tabla 4.2 se observa que $\ell = 48.66$ es la estimación hecha en el día 2657 del nivel (o media) para los datos de la humedad diaria.

Se infiere que el pronóstico puntual elaborado en el día 2657 para la humedad diaria del día 2658, y para cualquier otro día futuro de la humedad es

$$\widehat{y}_{2657+\tau}(2657) = \ell_{2657} = 48.66.$$

Para el paso siguiente, se necesita primero determinar el error estándar:

$$s = \sqrt{\frac{(y_1 - \ell_0)^2 + (y_2 - \ell_1)^2 + \dots + (y_{2657} - \ell_{2656})^2}{2656}} = 6.753.$$

Luego, se pueden calcular los intervalos de predicción como se indica a continuación:

- Un intervalo de predicción de 95 % en el día 2657 para y_{2658} es

$$[48.66 \pm (1.96)(6.75364717)] = [35.42, 61.89].$$

- Un intervalo de predicción de 95 % en el día 2657 para y_{2659} es

$$[48.66 \pm (1.96)(6.75364717)\sqrt{1 + (0.9759)^2}] = [30.16, 67.15].$$

- Un intervalo de predicción de 95 % en el día 2657 para y_{2660} es

$$[48.66 \pm z_{0.025}s\sqrt{1 + 2\alpha^2}] = [26.10, 71.22].$$

Como la constante de suavización α es grande, el incremento en la amplitud del intervalo de predicción también es grande.

En la Figura 4.1 y en la Tabla 4.3 se presentan la salida de MINITAB luego de aplicar la suavización exponencial simple, para pronosticar en el día 2657 la humedad para los siguientes tres días.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

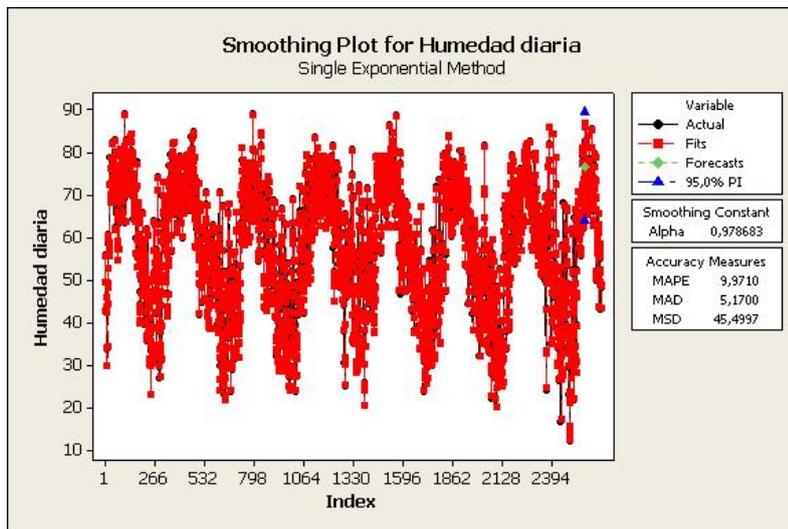


Figura 4.1: Salida de MINITAB de los resultados de optimización para los datos de la humedad (constante de suavización, y gráfica de pronósticos).

Tabla 4.3: Salida de MINITAB de los resultados de optimización para los datos de la humedad (pronósticos puntuales e intervalos de predicción).

```

Data      Humedad diaria
Length   2657

Smoothing Constant

Alpha    0,978683

Accuracy Measures

MAPE     9,9710
MAD      5,1700
MSD      45,4997

Forecasts

Period   Forecast   Lower   Upper
2658    48,6645   35,9981 61,3308
2659    48,6645   35,9981 61,3308
2660    48,6645   35,9981 61,3308
    
```

MINITAB usa un pronóstico hacia atrás o retropronóstico para encontrar la estimación inicial del nivel. En este procedimiento se empieza en el final de los datos y se efectúan

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

pronósticos para predecir un nivel inicial. Se utiliza el promedio de casi la mitad de los datos, (1328 observaciones) como la estimación inicial ℓ_0 del nivel.

La constante de suavización de 0.9786 que proporciona MINITAB, y la de 0.9759 que da Excel son distintas. No obstante, las dos indican que el nivel de los datos de la humedad no se modifica mucho. Los pronósticos puntuales de 48.6645 de MINITAB y 48.66 de Excel también difieren. Dado que hay muy poca variación en los datos y que el nivel (o la media) no cambia mucho, estos pronósticos no son muy diferentes.

4.8.2. Aplicación del método de Holt, de la suavización exponencial corregida de la tendencia

A continuación se realiza un análisis de los datos de información meteorológica los cuales fueron obtenidos en la estación Meteorológica del Departamento de Irrigación de la U.A.CH. (Universidad Autónoma de Chapingo) usando ahora el método de Holt, de la suavización corregida de la tendencia, los datos de la humedad fueron tomados en principio cada 30 minutos, con las fechas del 9 de Mayo del 2003 al 28 de Octubre del 2010.

Lo primero que se realizó, fue obtener promedios diarios de los registros de la humedad y posteriormente se obtuvieron promedios semanales, los datos de la humedad se presentan en la Tabla 4.4, la gráfica de los datos (ver Figura 4.2) señala una tendencia hacia abajo en las últimas semanas, la tasa de crecimiento de la humedad ha estado cambiando claramente en el periodo de estas semanas. Además se sugiere que no hay patrón estacional. Por tanto, el método de Holt, de la suavización exponencial corregida de la tendencia es un procedimiento apropiado que se puede aplicar en esta serie de tiempo.

Tabla 4.4: Promedios semanales de la humedad.

H (humedad(promedio semanal))	H (humedad(promedio semanal))	H (humedad(promedio semanal))	H (humedad(promedio semanal))
70,06	46,19	61,19	73,20
74,59	54,26	49,24	77,07
72,12	58,53	66,26	66,86
62,11	50,37	70,89	65,42
70,63	49,40	68,30	67,14
75,92	45,92	75,96	54,29
72,39	46,54	75,22	.
72,10	39,37	70,29	.
77,31	33,12	67,83	.
75,87	34,85	73,06	76,20
73,76	50,49	70,64	72,10
75,07	55,35	75,25	75,85
76,72	49,90	68,95	68,14
74,31	54,14	68,05	74,06
77,71	50,68	70,44	72,20
71,98	35,93	74,32	54,51
69,60	39,50	73,44	58,35
67,39	57,14	72,73	57,68
61,89	65,57	72,14	48,34
62,24	57,90	78,65	48,54

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

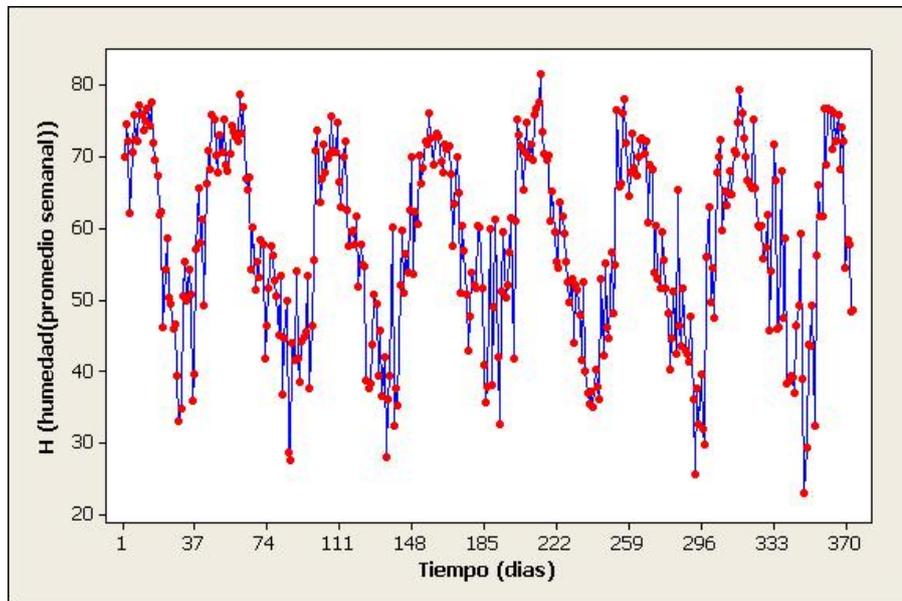


Figura 4.2: Gráfica de MINITAB para los promedios semanales de la humedad.

Para iniciar el procedimiento por medio de las dos ecuaciones de suavización, primero se obtiene una estimación inicial ℓ_0 para el nivel, y una estimación inicial b_0 para la tasa de crecimiento en el periodo 0.

Una manera de determinarlas es ajustar una recta de tendencia de mínimos cuadrados a una parte (por ejemplo, la mitad) de los datos anteriores o históricos de los promedios semanales de la humedad y hacer que la ordenada al origen sea ℓ_0 y que la pendiente sea b_0 .

Se consideró a las 374 observaciones de los valores de la humedad de la Tabla 4.4. Se ajustó una recta de tendencia de mínimos cuadrados a los primeros 187 de estos valores, esto es

$$\hat{y}_t = 62.46 - 0.043t.$$

Esto lleva a que $\ell_0 = 62.46$ y $b_0 = -0.043$. Se han encontrado estos dos valores efectuando un análisis de regresión lineal simple mediante Excel.

En la tabla 4.5 se presenta una hoja de cálculo con los resultados de la regresión en las últimas dos columnas de la derecha. Se han copiado los valores de la columna “Nivel”, y “Tasa de crecimiento” en el renglón “periodo 0”.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

Tabla 4.5: Hoja de cálculo de Excel para el método Holt de la suavización exponencial corregida de la tendencia, aplicado a los datos semanales de la humedad, $\alpha = 0.20$, $\gamma = 0.10$.

n	alfa	gamma	SSE		ssquare	s	
374	0,2	0,1	37620,46		101,130279	10,06	
				Pronóstico			
				hecho en		Error	
	Humedad real		Tasa de	el último	Error de	cuadrado del	
Periodo	semanal	Nivel	crecimiento	periodo	pronóstico	pronóstico	
0		62,46	-0,04				Resumen
1	70,06	63,95	0,11	62,42	7,64	58,30	
2	74,59	66,16	0,32	64,06	10,53	110,97	<i>Estadísticas de la regresión</i>
3	72,12	67,61	0,43	66,48	5,64	31,81	Coefficiente de corr
4	62,11	66,86	0,31	68,05	-5,93	35,21	Coefficiente de dete
5	70,63	67,87	0,38	67,17	3,46	11,97	R^2 ajustado
6	75,92	69,78	0,54	68,25	7,67	58,80	Error típico
7	72,39	70,73	0,58	70,32	2,07	4,29	Observaciones
8	72,10	71,47	0,59	71,31	0,79	0,63	
9	77,31	73,11	0,70	72,07	5,24	27,46	ANÁLISIS DE VARIANZA
10	75,87	74,22	0,74	73,81	2,06	4,23	<i>Grados de libert</i>
11	73,76	74,72	0,72	74,96	-1,21	1,45	Regresión
12	75,07	75,37	0,71	75,44	-0,36	0,13	Residuos
13	76,72	76,20	0,72	76,07	0,64	0,41	Total
14	74,31	76,40	0,67	76,92	-2,61	6,83	
15	77,71	77,20	0,68	77,07	0,64	0,41	<i>Coefficientes</i>
16	71,98	76,70	0,56	77,88	-5,91	34,87	Intercepción
17	69,60	75,73	0,41	77,26	-7,66	58,74	Variable X 1
18	67,39	74,39	0,24	76,14	-8,75	76,60	
19	61,89	72,08	-0,02	74,63	-12,73	162,17	
20	62,24	70,10	-0,22	72,06	-9,82	96,52	
.	
.	
.	
370	54,51	74,24	1,04	79,17	-24,66	608,23	
371	58,35	71,89	0,70	75,27	-16,93	286,47	
372	57,68	69,61	0,40	72,59	-14,91	222,23	
373	48,34	65,67	-0,03	70,01	-21,67	469,60	
374	48,54	62,22	-0,37	65,64	-17,11	292,69	

Se empieza con ℓ_0 y b_0 para calcular un pronóstico puntual de y_1 , a partir del origen 0

$$\hat{y}_1(0) = \ell_0 + b_0 = 62.46 + (-0.043) = 62.42.$$

Este pronóstico puntual, se muestra en la hoja de cálculo de la Tabla 4.5 en la columna con el encabezado “Pronóstico hecho en el último periodo”. También en la hoja de cálculo se muestran, la humedad real semanal $y_1 = 70.06$, y el error de pronóstico, que es

$$y_1 - \hat{y}_1(0) = 70.06 - 62.42 = 7.64.$$

Luego, se escogen valores de las constantes de suavización α y γ . Una elección razonable es $\alpha = 0.2$ y $\gamma = 0.1$.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

Entonces, si se usa $y_1 = 70.06$, y la ecuación para ℓ_T , se infiere que la estimación para la serie de tiempo en el periodo 1 es

$$\ell_1 = \alpha(y_1) + (1 - \alpha)[\ell_0 + b_0] = 0.2(70.06) + (1 - 0.2)[62.42] = 63.95.$$

Además, al aplicar la ecuación para b_T , la estimación para la tasa de crecimiento de la serie de tiempo en el periodo 1 es

$$b_1 = \gamma[\ell_1 - \ell_0] + (1 - \gamma)b_0 = 0.1[63.95 - 62.46] + (1 - 0.1)(-0.043) = 0.11.$$

De dónde se tiene un pronóstico puntual de $y_2 = 74.59$, realizado en el periodo 1 el cual es,

$$\hat{y}_2(1) = \ell_1 + b_1 = 64.06.$$

Como el valor real de la humedad semanal en el periodo 2 es $y_2 = 74.59$, el error de pronóstico es

$$y_2 - \hat{y}_2(1) = 10.53.$$

En la hoja de cálculo de Excel de la Tabla 4.5, se ilustra el proceso completo usando las dos ecuaciones de suavización con el método Holt de la suavización exponencial corregida de la tendencia, para encontrar nuevas estimaciones periodo por periodo del nivel, y de la tasa de crecimiento de toda la serie de tiempo. En la hoja de cálculo también se encuentran los pronósticos para el siguiente periodo, los errores de pronóstico, y los errores cuadráticos del pronóstico. En dicha hoja se observa también, que la suma de cuadrados del error de pronóstico (**sum of squared forecast errors, SSE**) es 37620.46 cuando $\alpha = 0.2$ y $\gamma = 0.1$.

Con el fin de determinar cuáles son los “mejores” valores para α y γ , se puede recurrir al Solver de Excel (ver Apéndice C), con objeto de encontrar α y γ que generen un valor mínimo para SSE.

En la Tabla 4.6 se muestra la hoja de cálculo de todos los valores cuando SSE alcanza su valor mínimo de 23255.35. Cuando SSE es mínimo, $\alpha = 0.56$, $\gamma = 0$, $s = 7.91$, $\ell_{374} = 50.41$, y $b_{374} = -0.04$. Con estos valores se pueden encontrar pronósticos e intervalos de predicción de 95 %, para la humedad futura.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

Tabla 4.6: Hoja de cálculo de Excel con valores de α , γ , s , ℓ_{374} , y b_{374} cuando la SSE es mínima.

n	alfa	gamma	SSE		ssquare	s
374	0,561839794	0	23255,35		62,5143835	7,91
				Pronóstico		
				hecho en		Error
	Humedad real		Tasa de	el último	Error de	cuadrado del
Periodo	semanal	Nivel	crecimiento	periodo	pronóstico	pronóstico
0		62,46	-0,04			
1	70,06	66,71	-0,04	62,42	7,64	58,30
2	74,59	71,12	-0,04	66,67	7,92	62,79
3	72,12	71,67	-0,04	71,08	1,05	1,10
4	62,11	66,28	-0,04	71,62	-9,51	90,45
5	70,63	68,71	-0,04	66,24	4,40	19,34
6	75,92	72,74	-0,04	68,66	7,25	52,62
7	72,39	72,52	-0,04	72,70	-0,30	0,09
8	72,10	72,27	-0,04	72,48	-0,38	0,14
9	77,31	75,08	-0,04	72,23	5,08	25,80
10	75,87	75,50	-0,04	75,04	0,83	0,69
11	73,76	74,50	-0,04	75,46	-1,70	2,90
12	75,07	74,81	-0,04	74,46	0,61	0,38
13	76,72	75,86	-0,04	74,76	1,95	3,81
14	74,31	74,97	-0,04	75,82	-1,51	2,27
15	77,71	76,49	-0,04	74,93	2,78	7,73
16	71,98	73,93	-0,04	76,45	-4,47	20,00
17	69,60	71,48	-0,04	73,89	-4,29	18,42
18	67,39	69,16	-0,04	71,44	-4,05	16,38
19	61,89	65,06	-0,04	69,12	-7,23	52,24
20	62,24	63,46	-0,04	65,02	-2,78	7,73
.
.
.
370	54,51	62,33	-0,04	72,35	-17,85	318,52
371	58,35	60,07	-0,04	62,28	-3,93	15,48
372	57,68	58,71	-0,04	60,03	-2,35	5,51
373	48,34	52,86	-0,04	58,67	-10,33	106,68
374	48,54	50,41	-0,04	52,82	-4,29	18,37

Para ilustrar el pronóstico de la humedad semanal, se usa $\ell_{374} = 50.41$ y $b_{374} = -0.04$ para determinar los pronósticos puntuales de y_{375} , y_{376} y y_{377} como se indica a continuación

$$\hat{y}_{375}(374) = \ell_{375} + b_{375} = 50.37$$

$$\hat{y}_{376}(374) = \ell_{375} + 2b_{375} = 50.33$$

$$\hat{y}_{377}(374) = \ell_{375} + 3b_{375} = 50.29.$$

Con objeto de calcular los intervalos de predicción de 95 %, se usan los errores de pronóstico de la Tabla 4.6 para determinar el error estándar s . En la Tabla 4.5 se proporcionan los errores de pronóstico para el siguiente periodo cuando se usa $\alpha = 0.56$ y $\gamma = 0$, las

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

constantes de suavización que minimizan la SSE.

Se tiene

$$s = \sqrt{\frac{\sum_{t=1}^T [y_t - \hat{y}_t(t-1)]^2}{T-2}} = 7.91.$$

Entonces, un intervalo de predicción de 95 % para y_{375} es

$$[\hat{y}_{375}(374) \pm z_{[0.025]}s] = [34.87, 65.87].$$

Entonces, un intervalo de predicción de 95 % para y_{376} es

$$\begin{aligned} [\hat{y}_{376}(374) \pm z_{[0.025]}s\sqrt{1 + \alpha^2(1 + \gamma)^2}] &= [(\ell_{375} + 2b_{375}) \pm z_{[0.025]}s\sqrt{1 + \alpha^2(1 + \gamma)^2}] = \\ &= [32.56, 68.10]. \end{aligned}$$

Entonces, un intervalo de predicción de 95 % para y_{377} es

$$\begin{aligned} [\hat{y}_{377}(374) \pm z_{[0.025]}s\sqrt{1 + \alpha^2(1 + \gamma)^2 + \alpha^2(1 + 2\gamma)^2}] &= \\ = [(\ell_{375} + 3b_{375}) \pm z_{[0.025]}s\sqrt{1 + \alpha^2(1 + \gamma)^2 + \alpha^2(1 + 2\gamma)^2}] &= [30.51, 70.07]. \end{aligned}$$

En la realidad, es muy común para una serie de tiempo tener un nivel que cambia sin que haya modificaciones en la tasa de crecimiento. Por consiguiente, puede parecer que una serie de tiempo tiene una tendencia cambiante, pero lo único que cambia es el nivel. En la salida de MINITAB de la Figura 4.3 se puede ver que los valores iniciales no se ajustan muy bien a los datos. Esto ocasiona la elección de una α grande ($=1.2733$), y una SSE también grande ($=88.6020(374)=33137.15$). Como resultado, los intervalos de predicción son más amplios de lo necesario (ver Tabla 4.7), y el valor grande para y_{374} podría haber causado que los pronósticos sean muy altos.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

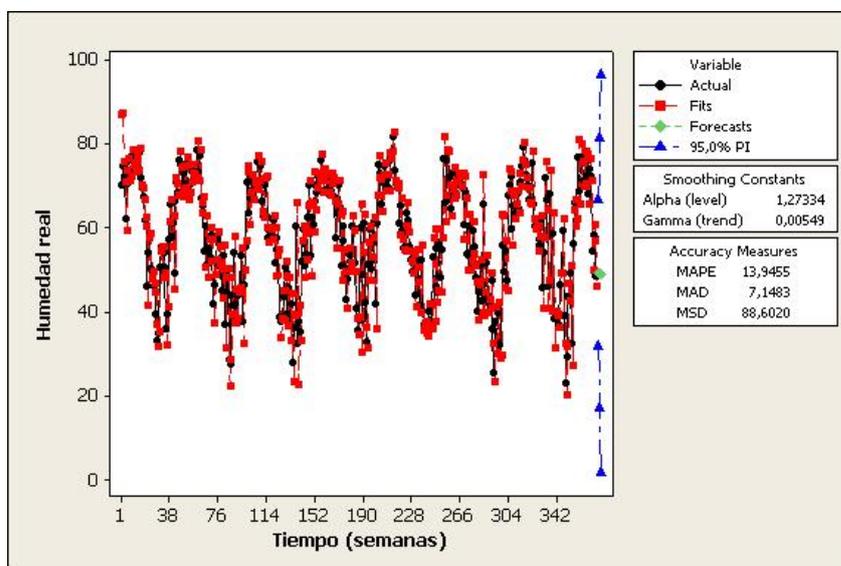


Figura 4.3: Salida de MINITAB del Método de Holt de la suavización exponencial corregida de la tendencia, aplicado a la humedad semanal (Constantes de suavización y gráfica de pronósticos).

Tabla 4.7: Salida de MINITAB del Método de Holt de la suavización exponencial corregida de la tendencia, aplicado a la humedad semanal (Pronósticos puntuales e intervalos de predicción).

```

Data Humedad real
Length 374

Smoothing Constants

Alpha (level) 1,27334
Gamma (trend) 0,00549

Accuracy Measures

MAPE 13,9455
MAD 7,1483
MSD 88,6020

Forecasts

Period Forecast Lower Upper
375 49,1771 31,6641 66,6902
376 49,1058 16,9153 81,2964
377 49,0346 1,8094 96,2597
    
```

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

4.8.3. Aplicación del método multiplicativo de Holt-Winters

A continuación se realiza un análisis de los datos de información meteorológica que fueron obtenidos en la estación Meteorológica del Departamento de Irrigación de la U.A.CH. usando ahora el método multiplicativo de Holt-Winters, los datos de la humedad fueron tomados en principio cada 30 minutos, con las fechas del 9 de Mayo del 2003 al 28 de Octubre del 2010. Nuevamente, en principio se tomaron los promedios diarios de los registros de la humedad y posteriormente se obtuvieron promedios de cada estación del año (es decir, promedios del verano, otoño, invierno y primavera)(ver Tabla 3.4), los datos obtenidos se grafican en la Figura 3.5. En ella se puede observar un decremento lineal de la humedad en un periodo de siete años, y que el patrón estacional es creciente cuando aumenta el nivel de la serie de tiempo. Este patrón sugiere que el método multiplicativo de Holt-Winters se puede aplicar para pronosticar la humedad futura.

Para iniciar el procedimiento por medio de las tres ecuaciones de suavización, primero se obtiene una estimación inicial del nivel ℓ_0 , la tasa de crecimiento b_0 y los factores estacionales, sn_{-3} , sn_{-2} , sn_{-1} y sn_0 . Como se requieren datos de por lo menos cuatro o cinco años para determinar los valores de los factores estacionales, por tanto se usó casi la mitad de los datos es decir, 16 valores, para determinar los valores iniciales. Al igual que en el método de suavización exponencial corregida para la tendencia de Holt, se utiliza la regresión, para ajustar una recta de tendencia a los primeros 16 valores de los promedios estacionales de la humedad. En la Tabla 4.8, se da una salida de regresión obtenida con Excel que proporciona la ecuación siguiente de regresión de mínimos cuadrados:

$$\hat{y}_t = 61.67011077 + (-0.376297811)t.$$

Por consiguiente, para los valores iniciales del nivel y la tasa de crecimiento, se usa $\ell_0 = 61.6701$ y $b_0 = -0.3762$. Estos valores se copian en las celdas que corresponden al nivel y la tasa de crecimiento en el tiempo o periodo 0 de la Tabla 4.9.

Tabla 4.8: Aplicación de Excel para determinar las estimaciones iniciales del nivel ℓ_0 , la tasa de crecimiento b_0 y los factores estacionales sn_{-3} , sn_{-2} , sn_{-1} y sn_0 .

Humedad real	t	Estación	Estimaciones de regresión	Sin tendencia	verano	otoño	invierno	primavera	Índice
72,99	1	1 (verano)	61,2939	1,1909	1,1909	0,9919	0,7866	1,1297	1,1872
60,42	2	2 (otoño)	60,9177	0,9919	1,1868	0,8970	0,7212	0,9935	0,9322
47,62	3	3 (invierno)	60,5415	0,7866	1,1513	0,8556	0,7441	1,1331	0,7729
67,97	4	4 (primavera)	60,1653	1,1297	1,2199	0,9845	0,8397	1,1746	1,1077
70,96	5	1 (verano)	59,7891	1,1868					
53,30	6	2 (otoño)	59,4129	0,8970	Promedios por estación				Suma
42,58	7	3 (invierno)	59,0367	0,7212	1,1872	0,9322	0,7729	1,1077	4,0001
58,28	8	4 (primavera)	58,6605	0,9935					
67,10	9	1 (verano)	58,2843	1,1513	Grados de libertad de cuadrado de los cua				F
49,54	10	2 (otoño)	57,9081	0,8556	Regresión	1	48,1440145	48,1440145	0,43426582
42,81	11	3 (invierno)	57,5319	0,7441	Residuos	14	1552,08208	110,863006	
64,76	12	4 (primavera)	57,1557	1,1331	Total	15	1600,2261		
69,26	13	1 (verano)	56,7795	1,2199					
55,53	14	2 (otoño)	56,4033	0,9845		Coeficientes	Error típico	Estadístico t	Probabilidad
47,04	15	3 (invierno)	56,0271	0,8397	Intercepción	61,6701108	5,521533	11,1690197	2,3364E-08
65,37	16	4 (primavera)	55,6509	1,1746	Variable X 1	-0,37629781	0,57102335	-0,65898849	0,52059536

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

Los factores estacionales iniciales se determinan mediante el siguiente procedimiento de cuatro pasos:

1. Se aplica la ecuación de regresión de mínimos cuadrados para calcular \hat{y}_t de cada periodo que se usa en la determinación de la ecuación de regresión de mínimos cuadrados. En este caso, se calcula \hat{y}_t de los cuatro años de datos, $t = 1, 2, \dots, 16$.

Estos valores se muestran en la columna de “estimaciones de regresión” de la Tabla 4.8. Por ejemplo,

$$\hat{y}_1 = 61.6701 + (-0.3762)(1) = 61.29$$

$$\hat{y}_2 = 61.6701 + (-0.3762)(2) = 60.92$$

$$\hat{y}_3 = 61.6701 + (-0.3762)(3) = 60.54.$$

2. Enseguida se elimina la tendencia de los datos calculando $S_t = \frac{y_t}{\hat{y}_t}$ de cada periodo que se uso para estimar la recta de regresión de mínimos cuadrados.

Estos valores se muestran en la columna “Sin tendencia” de la Tabla 4.8. Por ejemplo,

$$S_1 = \frac{y_1}{\hat{y}_1} = 1.19$$

$$S_2 = \frac{y_2}{\hat{y}_2} = 0.99$$

$$S_3 = \frac{y_3}{\hat{y}_3} = 0.79.$$

3. Luego se calculan los valores estacionales promedio para cada una de las estaciones L . Estos L promedios, $\bar{S}_{[1]}, \bar{S}_{[2]}, \bar{S}_{[3]}, \dots, \bar{S}_{[L]}$, se encuentran determinando el promedio de los valores sin tendencia para la estación correspondiente. Por lo que toca a los datos de la humedad, se encontro $L = 4$ factores estacionales, $\bar{S}_{[1]}, \bar{S}_{[2]}, \bar{S}_{[3]}$ y $\bar{S}_{[4]}$. Para el verano, hay cuatro primeros veranos de datos sin tendencia. Por tanto,

$$\bar{S}_{[1]} = \frac{S_1 + S_5 + S_9 + S_{13}}{4}$$

de donde

$$\bar{S}_{[1]} = 1.1872.$$

De igual manera, se encontró que

$$\bar{S}_{[2]} = 0.9322$$

es el promedio estacional del otoño,

$$\bar{S}_{[3]} = 0.7729$$

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

es el promedio estacional del invierno,

$$\bar{S}_{[4]} = 1.1077$$

es el promedio estacional de la primavera.

4. Por último, se quiere que el promedio de los factores estacionales sea 1. Esto se consigue al multiplicar los valores estacionales promedio por el factor de corrección

$$CF = \frac{L}{\sum_{i=1}^L \bar{S}_{[i]}}.$$

Se observa que si $\sum_{i=1}^L \bar{S}_{[i]} = L$, no hay que hacer correcciones porque $CF = 1$. Con el objeto de usar las ecuaciones de suavización en la hoja de cálculo de la Tabla 4.9, los factores estacionales iniciales tienen subíndices para los cuales la primera estación es en el periodo $1 - L$, la segunda estación es en el periodo $2 - L$, y la última estación es en el periodo $L - L = 0$. Por consiguiente, los factores estacionales iniciales son

$$sn_{i-L} = \bar{S}_{[i]}(CF) \quad (i = 1, 2, \dots, L).$$

En cuanto a los datos estacionales de la humedad, se encontraron cuatro factores estacionales iniciales. En este caso, $CF = 4/4.0001 = 1.0000$ y

$$sn_{-3} = sn_{1-4} = \bar{S}_{[1]}(CF) = 1.1872$$

$$sn_{-2} = sn_{2-4} = \bar{S}_{[2]}(CF) = 0.9322$$

$$sn_{-1} = sn_{3-4} = \bar{S}_{[3]}(CF) = 0.7729$$

$$sn_0 = sn_{4-4} = \bar{S}_{[4]}(CF) = 1.1077$$

Estos valores se listan bajo el encabezado “índice” en la Tabla 4.8, y se copian en las celdas de los periodos -3,-2,-1 y 0 bajo “Factor estacional” en la Tabla 4.9.

Después de determinar los valores iniciales del nivel, tendencia, y cuatro factores estacionales, se está listo para aplicar las ecuaciones de suavización.

Para empezar con los valores iniciales, se calcula un pronóstico puntual de y_1 en el origen 0:

$$\hat{y}_1(0) = (\ell_0 + b_0)sn_{1-4} = (\ell_0 + b_0)sn_{-3} = 72.77$$

. Este pronóstico puntual se proporciona en la Tabla 4.9 bajo el encabezado “Pronóstico hecho en el último periodo”. En la hoja de cálculo también se puede ver el valor de la humedad real $y_1 = 72.99$, y el error de pronóstico, que es

$$y_1 - \hat{y}_1(0) = 72.99 - 72.77 = 0.22$$

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

. La hoja de cálculo de la Tabla 4.9 se prepara con $\alpha = 0.2$, $\gamma = 0.1$ y $\delta = 0.1$. Al usar $y_1 = 72.99$, y la ecuación de suavización para ℓ_T la estimación del nivel de la serie de tiempo en el periodo 1 es

$$\ell_1 = \alpha(y_1/sn_{-3}) + (1 - \alpha)(\ell_0 + b_0) = 61.34.$$

Al aplicar la ecuación de suavización para b_T , la estimación de la tasa de crecimiento en el periodo 1 es

$$b_1 = \gamma(\ell_1 - \ell_0) + (1 - \gamma)b_0 = -0.3.$$

Al aplicar la ecuación de suavización para sn_T , la estimación del factor estacional en el periodo 1 es

$$sn_1 = \delta(y_1/\ell_1) + (1 - \delta)sn_{-3} = 1.181.$$

Se infiere entonces que un pronóstico puntual de y_2 en el periodo 1 es

$$\hat{y}_2(1) = (\ell_1 + b_1)sn_{2-4} = (\ell_1 + b_1)sn_{-2} = 53.83.$$

Puesto que el valor real de la humedad estacional en el periodo 2 es $y_2 = 60.42$, el error de pronóstico es

$$y_2 - \hat{y}_2(1) = 6.59.$$

Ahora se pueden calcular las estimaciones actualizadas ℓ_2 , b_2 y sn_2 como se indica a continuación:

$$\ell_2 = \alpha(y_2/sn_{-2}) + (1 - \alpha)(\ell_1 + b_1) = 61.76$$

$$b_2 = \gamma(\ell_2 - \ell_1) + (1 - \gamma)(b_1) = -0.2957$$

$$sn_2 = \delta(y_2/\ell_2) + (1 - \delta)(sn_{-2}) = 0.9368.$$

El proceso continúa hasta abarcar los 28 periodos. También se calculan los errores cuadrados del pronóstico y la suma de cuadrados del error (SSE).

Los resultados de este proceso se muestran en la Tabla 4.9, donde se ve que $SSE = 271.8779$.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

Tabla 4.9: Hoja de cálculo de Excel para el método multiplicativo de Holt-Winters, aplicado a la humedad estacional con $\alpha = 0.2$, $\gamma = 0.1$ y $\delta = 0.1$.

n	alfa	gamma	delta	SSE	ssquare	s	SSRE	sr
28	0,2	0,1	0,1	271,8779	10,875117	3,29774422	0,087369	0,05911654
					Pronóstico		Error	
					del		cuadrado	Error
	Humedad		Tasa de	Factor	último	Error de	del	relativo
Periodo	real	Nivel	crecimiento	estacional	periodo	pronóstico	pronóstico	cuadrado
-3				1,1872				
-2				0,9322				
-1				0,7729				
0		61,6701	-0,3762	1,1077				
1	72,99	61,3317	-0,3724	1,1875	72,7689	0,2245	0,0504	0,000010
2	60,42	61,7307	-0,2953	0,9369	56,8286	3,5955	12,9275	0,004003
3	47,62	61,4717	-0,2916	0,7731	47,4841	0,1404	0,0197	0,000009
4	67,97	61,2162	-0,2880	1,1080	67,7709	0,2000	0,0400	0,000009
5	70,96	60,6933	-0,3115	1,1857	72,3524	-1,3943	1,9440	0,000371
6	53,30	59,6823	-0,3815	0,9325	56,5716	-3,2765	10,7355	0,003354
7	42,58	58,4562	-0,4659	0,7686	45,8452	-3,2652	10,6613	0,005073
8	58,28	56,9122	-0,5737	1,0996	64,2526	-5,9721	35,6656	0,008639
9	67,10	56,3899	-0,5686	1,1861	66,7987	0,3048	0,0929	0,000021
10	49,54	55,2830	-0,6224	0,9289	52,0538	-2,5102	6,3011	0,002325
11	42,81	54,8674	-0,6017	0,7698	42,0135	0,7950	0,6320	0,000358
12	64,76	55,1920	-0,5091	1,1070	59,6703	5,0928	25,9365	0,007284
13	69,26	55,4253	-0,4349	1,1925	64,8593	4,4033	19,3889	0,004609
14	55,53	55,9483	-0,3391	0,9352	51,0792	4,4484	19,7881	0,007584
15	47,04	56,7101	-0,2290	0,7758	42,8071	4,2374	17,9557	0,009799
16	65,37	56,9946	-0,1776	1,1110	62,5233	2,8419	8,0764	0,002066
17	68,97	57,0212	-0,1572	1,1942	67,7516	1,2179	1,4834	0,000323
18	53,10	56,8465	-0,1590	0,9351	53,1812	-0,0818	0,0067	0,000002
19	42,05	56,1901	-0,2087	0,7730	43,9760	-1,9295	3,7229	0,001925
20	65,74	56,6197	-0,1449	1,1160	62,1934	3,5461	12,5746	0,003251
21	65,66	56,1760	-0,1748	1,1916	67,4402	-1,7845	3,1846	0,000700
22	49,42	55,3712	-0,2378	0,9309	52,3678	-2,9458	8,6776	0,003164
23	40,78	54,6579	-0,2853	0,7703	42,6190	-1,8379	3,3779	0,001860
24	64,03	54,9737	-0,2252	1,1209	60,6785	3,3542	11,2506	0,003056
25	68,94	55,3692	-0,1631	1,1970	65,2395	3,6982	13,6767	0,003213
26	55,49	56,0880	-0,0749	0,9367	51,3893	4,1050	16,8512	0,006381
27	41,24	55,5174	-0,1245	0,7676	43,1483	-1,9094	3,6458	0,001958
28	66,91	56,2525	-0,0385	1,1277	62,0875	4,8177	23,2103	0,006021

Para encontrar “buenos” valores para α , γ , y δ , se seleccionan los valores que minimizan la suma de cuadrados del error del pronóstico (SSE).

En la Tabla 4.10 se muestran los resultados para SSE mínima que se obtienen con el Solver. Se observa que la SSE mínima es igual a 218.770 cuando $\alpha = 0.567$, $\gamma = 0$ y $\delta = 0.534$. Además, también se puede ver que las estimaciones finales para el nivel, tasa de crecimiento y factores estacionales son $\ell_{28} = 57.0475$, $b_{28} = -0.3700$, $sn_{25} = 1.2011$, $sn_{26} = 0.9360$, $sn_{27} = 0.7529$ y $sn_{28} = 1.1604$.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

Tabla 4.10: Hoja de cálculo de Excel en la que se muestra la SSE mínima con los valores para α , γ , δ , ℓ_{28} , b_{28} , sn_{25} , sn_{26} , sn_{27} y sn_{28} .

n	alfa	gamma	delta	SSE	ssquare	s	SSRE	sr
28	0,5674595	0	0,53437565	218,7704	8,75081482	2,95817762	0,068131	0,05220382
					Pronóstico		Error	
					del		cuadrado	Error
	Humedad		Tasa de	Factor	último	Error de	del	relativo
Periodo	real	Nivel	crecimiento	estacional	periodo	pronóstico	pronóstico	cuadrado
-3				1,1872				
-2				0,9322				
-1				0,7729				
0		61,6700	-0,3700	1,1077				
1	72,99	61,4038	-0,3700	1,1880	72,7761	0,2172	0,0472	0,000009
2	60,42	63,1801	-0,3700	0,9451	56,8981	3,5260	12,4327	0,003840
3	47,62	62,1331	-0,3700	0,7695	48,5467	-0,9222	0,8504	0,000361
4	67,97	61,5347	-0,3700	1,1061	68,4168	-0,4459	0,1988	0,000042
5	70,96	60,3492	-0,3700	1,1815	72,6655	-1,7073	2,9148	0,000552
6	53,30	57,9417	-0,3700	0,9316	56,6887	-3,3936	11,5162	0,003584
7	42,58	56,3030	-0,3700	0,7624	44,3003	-1,7203	2,9596	0,001508
8	58,28	54,0941	-0,3700	1,0907	61,8650	-3,5844	12,8480	0,003357
9	67,10	55,4671	-0,3700	1,1966	63,4745	3,6291	13,1701	0,003269
10	49,54	54,0098	-0,3700	0,9240	51,3285	-1,7849	3,1859	0,001209
11	42,81	55,0633	-0,3700	0,7704	40,8960	1,9125	3,6576	0,002187
12	64,76	57,3502	-0,3700	1,1113	59,6560	5,1070	26,0815	0,007329
13	69,26	57,4920	-0,3700	1,2010	68,1833	1,0793	1,1648	0,000251
14	55,53	58,8103	-0,3700	0,9348	52,7786	2,7490	7,5569	0,002713
15	47,04	59,9277	-0,3700	0,7782	45,0252	2,0194	4,0778	0,002011
16	65,37	59,1377	-0,3700	1,1081	66,1877	-0,8225	0,6764	0,000154
17	68,97	58,0081	-0,3700	1,1945	70,5772	-1,6077	2,5846	0,000519
18	53,10	57,1653	-0,3700	0,9316	53,8782	-0,7788	0,6065	0,000209
19	42,05	55,2250	-0,3700	0,7692	44,2001	-2,1535	4,6377	0,002374
20	65,74	57,3921	-0,3700	1,1281	60,7852	4,9543	24,5448	0,006643
21	65,66	55,8536	-0,3700	1,1844	68,1156	-2,4599	6,0509	0,001304
22	49,42	54,1025	-0,3700	0,9219	51,6895	-2,2674	5,1411	0,001924
23	40,78	53,3259	-0,3700	0,7668	41,3322	-0,5511	0,3037	0,000178
24	64,03	55,1166	-0,3700	1,1461	59,7374	4,2953	18,4499	0,005170
25	68,94	56,7099	-0,3700	1,2011	64,8401	4,0976	16,7905	0,003994
26	55,49	58,5267	-0,3700	0,9360	51,9415	3,5528	12,6227	0,004679
27	41,24	55,6721	-0,3700	0,7529	44,5965	-3,3576	11,2733	0,005668
28	66,91	57,0475	-0,3700	1,1604	63,3802	3,5250	12,4258	0,003093

Usando las estimaciones de la Tabla 4.10 que minimizan SSE, los pronósticos puntuales de y_{29} , y_{30} , y_{31} y y_{32} son

$$\hat{y}_{29}(28) = (\ell_{28} + b_{28})sn_{29-4} = (\ell_{28} + b_{28})sn_{25} = (57.0475 + (-0.3700))(1.2011) = 68.0753$$

$$\hat{y}_{30}(28) = (\ell_{28} + 2b_{28})sn_{30-4} = (\ell_{28} + 2b_{28})sn_{26} = (57.0475 + 2(-0.3700))(0.9360) = 52.7038$$

$$\hat{y}_{31}(28) = (\ell_{28} + 3b_{28})sn_{31-4} = (\ell_{28} + 3b_{28})sn_{27} = (57.0475 + 3(-0.3700))(0.7529) = 42.1153$$

$$\hat{y}_{32}(28) = (\ell_{28} + 4b_{28})sn_{32-4} = (\ell_{28} + 4b_{28})sn_{28} = (57.0475 + 4(-0.3700))(1.1604) = 64.4805.$$

Antes de proceder a calcular los intervalos de predicción de 95%, se observa que en las fórmulas de estos intervalos se usa el error estándar relativo s_r en el tiempo T , y no el error estándar s en el tiempo T . La razón de este cambio es el resultado de usar un modelo multiplicativo donde la tendencia está multiplicada tanto por el factor estacional como por el factor irregular. Para determinar s_r , es necesario determinar la suma de los cuadrados

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

de los errores relativos $[y_t - \hat{y}_t(t-1)]/\hat{y}_t(t-1)$, $t = 1, 2, \dots, T$, y no la suma de los cuadrados de los errores, $[y_t - \hat{y}_t(t-1)]$, como sigue:

$$s_r = \sqrt{\frac{\sum_{t=1}^{28} \left[\frac{y_t - \hat{y}_t(t-1)}{\hat{y}_t(t-1)} \right]^2}{28 - 3}} = 0.052.$$

Entonces, un intervalo de predicción de 95 % para y_{29} es

$$[\hat{y}_{29}(28) \pm z_{[0.025]} s_r (\sqrt{c_1})(sn_{29-4})] = [61.1104, 75.0402].$$

Para calcular un intervalo de predicción de 95 % para y_{30} primero se determina

$$c_2 = \alpha^2(1 + \gamma)^2(\ell_{28} + b_{28})^2 + (\ell_{28} + 2b_{28})^2 = 4203.2662$$

entonces,

$$[\hat{y}_{30}(28) \pm z_{[0.025]} s_r (\sqrt{c_2})(sn_{30-4})] = [46.5190, 58.8886].$$

Ahora para calcular un intervalo de predicción de 95 % para y_{31} primero se determina

$$c_3 = \alpha^2(1 + 2\gamma)^2(\ell_{28} + b_{28})^2 + \alpha^2(1 + \gamma)^2(\ell_{28} + 2b_{28})^2 + (\ell_{28} + 3b_{28})^2$$

de donde

$$c_3 = 5181.0275$$

entonces,

$$[\hat{y}_{31}(28) \pm z_{[0.025]} s_r (\sqrt{c_3})(sn_{31-4})] = [36.5919, 47.6387].$$

Si se dispusiera de nuevos valores estacionales de la humedad, entonces se pueden revisar los pronósticos y los intervalos de predicción de 95 %, por medio de las ecuaciones de predicción para actualizar las estimaciones del nivel, tasa de crecimiento y factor estacional sin volver a estimar los parámetros de suavización.

Tanto las versiones aditivas como las multiplicativas de los métodos de Holt-Winters se encuentran en el software MINITAB para la suavización exponencial.

En la Figura 4.4 se presenta una salida de MINITAB en la que se usaron los valores que asigna automáticamente el programa a las constantes de suavización ($\alpha = \gamma = \delta = 0.2$). Los métodos de Holt-Winters en MINITAB, carecen de la opción de búsqueda automática de constantes óptimas de suavización.

Una SSE grande siempre se refleja en la longitud de los intervalos de predicción, de aquí que los intervalos son muchos más amplios en la Tabla 4.11 que proporciona MINITAB.

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL
4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

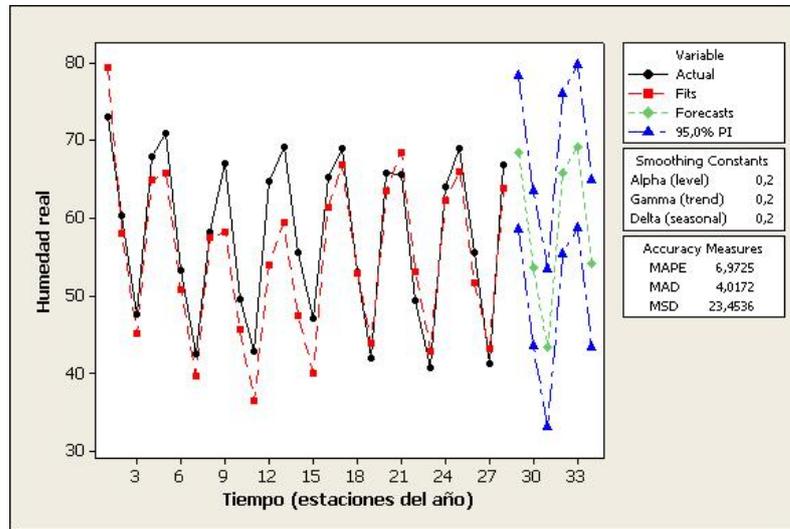


Figura 4.4: Salida de MINITAB del Método multiplicativo de Holt-Winters, aplicado a la humedad estacional (constantes de suavización y gráfica de pronósticos).

Tabla 4.11: Salida de MINITAB del Método multiplicativo de Holt-Winters, aplicado a la humedad estacional (Pronósticos puntuales e intervalos de predicción).

```

Multiplicative Method

Data      Humedad real
Length   28

Smoothing Constants

Alpha (level)    0,2
Gamma (trend)    0,2
Delta (seasonal) 0,2

Accuracy Measures

MAPE    6,9725
MAD     4,0172
MSD     23,4536

Forecasts

Period  Forecast    Lower    Upper
29      68,4839    58,6419  78,3260
30      53,5797    43,5834  63,5760
31      43,2938    33,1256  53,4620
32      65,7457    55,3887  76,1027
33      69,2460    58,6843  79,8077
34      54,1742    43,3928  64,9557
    
```

CAPÍTULO 4. SUAVIZACIÓN EXPONENCIAL

4.8. APLICACIÓN DE LOS MÉTODOS DE SUAVIZACIÓN

Para finalizar esta sección se hace una comparación de los tres métodos de suavización expuestos anteriormente, con el objeto de destacar cual es el mejor método para realizar pronósticos.

Se llevaron a cabo tres análisis usando los métodos de la suavización exponencial, el primero en aplicarse fue el método de la suavización exponencial simple, para los datos de la humedad tomados como promedios diarios, luego se aplicó el método de Holt, de la suavización exponencial corregida de la tendencia, para los datos de la humedad ahora tomados como promedios semanales, por último se aplicó el método multiplicativo de Holt-Winters, para los datos de la humedad tomados como promedios de cada estación, es decir, promedios del verano, otoño, invierno y primavera.

Como el objetivo de cada método es encontrar buenos valores de pronóstico, se seleccionan los valores que minimizan la suma de cuadrados del error de pronóstico (SSE), en las Tablas 4.2, 4.6 y 4.10 se observa que el método multiplicativo de Holt-Winters arroja una SEE menor que en los demás métodos y en consecuencia se obtienen intervalos de predicción más cortos, es decir más exactos, por lo tanto, el método multiplicativo de Holt-Winters es el mejor para realizar pronóstico de la serie de tiempo observada.

Capítulo 5

Metodología de Box-Jenkins

En 1970, Box y Jenkins desarrollaron un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series de tiempo en los que la variable tiempo juega un papel fundamental. Una parte importante de esta metodología está pensada para liberar al investigador de la tarea de especificación de los modelos dejando que los propios datos temporales de la variable a estudiar indiquen las características de la estructura probabilística subyacente.

La metodología de Box-Jenkins consta de los siguientes pasos:

Paso 1: Identificación tentativa del modelo. Se propone un modelo tentativo utilizando la función de autocorrelación muestral (**FACM**) y la función de autocorrelación parcial muestral (**FACPM**).

Paso 2: Estimación de los parámetros. Se estiman los parámetros del modelo identificado en forma tentativa.

Paso 3: Comprobación y diagnóstico. Comprobar el modelo, si el modelo propuesto en forma tentativa resulta ser inadecuado, el diagnóstico sugiere un nuevo modelo tentativo, que se somete a estos tres pasos y así sucesivamente hasta obtener el modelo adecuado.

Paso 4: Predicción. Pronosticar o calcular valores futuros usando el modelo final establecido en los pasos anteriores.

A continuación se definirán algunos conceptos teóricos fundamentales para el análisis de series de tiempo.

5.1. Modelos estacionarios y la función de autocorrelación

Definición 5.1. *Un proceso es estacionario de orden r si todos sus momentos hasta el orden r existen y son estables.*

Definición 5.2. *Una serie de tiempo $\{X_t; t \in T\}$, es estacionaria en sentido estricto si y sólo si:*

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS

5.1. MODELOS ESTACIONARIOS Y LA FUNCIÓN DE AUTOCORRELACIÓN

$$F[x_{t_1}, x_{t_2}, \dots, x_{t_n}] = F[x_{t_1+k}, x_{t_2+k}, \dots, x_{t_n+k}] \quad \forall t_1, t_2, \dots, t_n \in T \text{ y } k \in \mathbb{N}$$

es decir, si la función de distribución en cualquier conjunto finito de n variables aleatorias de la serie no se altera si se desplaza k periodos en el tiempo.

Definición 5.3. Una serie de tiempo $\{X_t; t \in T\}$, es estacionaria de segundo orden o débilmente estacionaria si y sólo si:

1. Es estacionaria en media, es decir, todas las variables aleatorias de la serie tienen la misma media y ésta es finita:

$$E(X_t) = \mu < \infty, \quad \forall t \in T.$$

2. Todas las variables tienen la misma varianza y es finita, es decir, la dispersión en torno a la media constante a lo largo del tiempo es la misma para todas las variables de la serie de tiempo:

$$V(X_t) = E[X_t - \mu]^2 = \sigma_x^2 < \infty, \quad \forall t \in T.$$

3. Las autocovarianzas sólo dependen del número de periodos de separación entre las variables y no del tiempo, es decir, la covarianza lineal entre dos variables aleatorias de la serie de tiempo que disten k periodos de tiempo es la misma que existe entre cualesquiera otras dos variables que estén también separadas k periodos, independientemente del momento concreto de tiempo al que estén referidas

$$\text{cov}(X_{t+k}, X_t) = E[X_{t+k} - \mu][X_t - \mu] = \gamma_{|(t+k)-t|} = \gamma_k < \infty, \quad \forall k \in \mathbf{Z}^+. [1]$$

Observación 5.1. Si $\{X_t; t \in T\}$ es estrictamente estacionaria y $E(X_t^2) < \infty$ para toda t , entonces $\{X_t; t \in T\}$ es estacionaria de segundo orden.

Observación 5.2. Si una serie de tiempo es estacionaria de segundo orden y su distribución es normal, es estacionaria en sentido estricto [1].

A partir de ahora se usa el término serie de tiempo estacionaria para hacer referencia a una serie débilmente estacionaria, a menos que se indique otra cosa.

Definición 5.4. Sea $\{X_t; t \in T\}$ una serie de tiempo estacionaria, el coeficiente de autocovarianza de la serie en el lapso k (número de periodos de separación entre las variables) es

$$\gamma_k = \text{cov}(X_{t+k}, X_t). \quad (5.1)$$

Definición 5.5. El conjunto $\{\gamma_k; k \in \mathbf{Z}\}$ se denomina función de autocovarianza (**FACV**).

Propiedades 5.1. (FACV)

1. Incluye la varianza del proceso para $k = 0$: $\gamma_0 = E[X_t - \mu][X_t - \mu] = \text{Var}(X_t)$.
2. $\gamma_0 \geq 0$.
3. $|\gamma_k| \leq \gamma_0 \quad \forall k \in \mathbf{Z}^+$.

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS

5.1. MODELOS ESTACIONARIOS Y LA FUNCIÓN DE AUTOCORRELACIÓN

4. Es una función simétrica: $\gamma_k = E[X_t - \mu][X_{t+k} - \mu] = E[X_t - \mu][X_{t-k} - \mu] = \gamma_{-k}$.

La función de autocovarianzas de una serie de tiempo recoge toda la información sobre la estructura dinámica lineal de la misma. Pero depende de las unidades de medida de la variable, por lo que, en general, se suele utilizar la función de autocorrelación.

Definición 5.6. Sea $\{X_t; t \in T\}$ una serie de tiempo estacionaria, el coeficiente de autocorrelación de la serie en el lapso k (número de periodos de separación entre las variables) es

$$\rho_k = \frac{\gamma_k}{\sqrt{\gamma_0 \gamma_0}} = \frac{\gamma_k}{\gamma_0}. \quad (5.2)$$

Definición 5.7. El conjunto $\{\rho_k; k \in \mathbb{Z}\}$ se denomina función de autocorrelación (FAC).

La función de autocorrelación se suele representar gráficamente por medio de un gráfico de barras denominado correlograma.

Propiedades 5.2. (FAC)

1. El coeficiente de autocorrelación de orden 0 es, por definición, 1. Por eso, a menudo, no se le incluye explícitamente en la función de autocorrelación.

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1.$$

2. $|\rho_k| \leq 1 \forall k \in \mathbb{Z}^+$.

3. Es una función simétrica: $\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_{-k}}{\gamma_0} = \rho_{-k}$. Por ello, en el correlograma se representa la función de autocorrelación solamente para los valores positivos del retardo k .

4. La función de autocorrelación de una serie de tiempo estacionaria tiende a cero rápidamente cuando k tiende a ∞ .

Ejemplos de modelos estacionarios.

Ejemplo 5.1. (Ruido iid).

Una sucesión $\{X_t; t \in T\}$ de variables aleatorias que son independientes e idénticamente distribuidas (v. a. iid) con media 0 y varianza σ^2 , es denominada un ruido iid y es denotado de la siguiente forma

$$X_t \sim IID(0, \sigma^2).$$

Ya que satisface que $E(X_t) = 0$ y

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

entonces el ruido iid es estacionario.

Ejemplo 5.2. (Ruido blanco).

Si $\{X_t; t \in T\}$ es una sucesión de variables aleatorias no correlacionadas, cada una con media cero y varianza σ^2 , entonces se dice que la sucesión es un ruido blanco. Para indicar que la sucesión $\{X_t; t \in T\}$ es un ruido blanco se usará la siguiente notación:

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS

5.1. MODELOS ESTACIONARIOS Y LA FUNCIÓN DE AUTOCORRELACIÓN

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

Dado que $E(X_t) = 0$ y

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

no depende de t , entonces el ruido blanco es estacionario.

Observación 5.3. Claramente se puede observar que toda sucesión $IID(0, \sigma^2)$ es un $RB(0, \sigma^2)$, sin embargo, el recíproco no se cumple [2].

Ya se vio como calcular la **FAC** para una serie de tiempo. Sin embargo, al trabajar con fenómenos reales, se cuenta con datos observados $\{x_1, x_2, \dots, x_n\}$ y generalmente no se conocen los parámetros de las variables aleatorias del proceso. Para calcular el grado de independencia de los datos así como para seleccionar un modelo para estos, es necesario usar la función de autocorrelación muestral (**FACM**) que se obtiene en base a los datos observados. Si se cree que los datos son valores realizados de una serie de tiempo estacionaria $\{X_t; t \in T\}$, entonces la función de autocorrelación muestral dará un estimador de la función de autocorrelación de $\{X_t; t \in T\}$.

Definición 5.8. Sean $\{x_1, x_2, \dots, x_n\}$ observaciones de una serie de tiempo.

La media muestral de estas observaciones es

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (5.3)$$

La varianza muestral de estas observaciones es

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5.4)$$

La autocovarianza muestral de estas observaciones es

$$\hat{\gamma}_k := \frac{1}{n} \sum_{t=1}^{n-|k|} (x_{t+|k|} - \bar{x})(x_t - \bar{x}) \quad (5.5)$$

para $-n < k < n$.

La autocorrelación muestral de estas observaciones es

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (5.6)$$

para $-n < k < n$.

La **FACM** es la representación gráfica de las autocorrelaciones muestrales en los lapsos k y en base a ésta diremos que:

1. Si el comportamiento de la **FACM** de la serie de tiempo se corta bastante rápido o tiende a cero rápidamente, entonces la serie de tiempo será considerada estacionaria.

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.2. MODELO ESTACIONARIO LINEAL GENERAL

2. Si la **FACM** de la serie de tiempo tiende a cero de forma extremadamente lenta, entonces la serie será considerada no estacionaria.

Otro elemento importante en la identificación del modelo es la función de autocorrelación parcial muestral (**FACPM**), la cual se define de la siguiente manera:

Definición 5.9. Sean $\{x_1, x_2, \dots, x_n\}$ observaciones de una serie de tiempo estacionaria. El coeficiente de autocorrelación parcial de la muestra en el lapso k es

$$\hat{\rho}_{kk} = \begin{cases} \hat{\rho}_1, & \text{si } k=1 \\ \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} (\hat{\rho}_{k-1,j} \hat{\rho}_{k-j})}{1 - \sum_{j=1}^{k-1} (\hat{\rho}_{k-1,j} \hat{\rho}_j)}, & \text{si } k=2,3,\dots \end{cases} \quad (5.7)$$

donde $\hat{\rho}_{kj} = \hat{\rho}_{k-1,j} - \hat{\rho}_{kk} \hat{\rho}_{k-1,k-j}$ para $j = 1, 2, \dots, k-1$.

Definición 5.10. El conjunto $\{\hat{\rho}_{kk}; k \in \mathbb{Z}\}$ se denomina función de autocorrelación parcial muestral (**FACPM**).

5.2. Modelo estacionario lineal general

Definición 5.11. Un operador de retardo en un periodo (denotado por B) se define como

$$BX_t = X_{t-1} \quad (5.8)$$

para toda t , el cual retarda a la variable X_t un periodo.

De manera general se tiene

$$B^k X_t = X_{t-k} \quad (5.9)$$

para toda t y $k = 0, 1, 2, \dots$, la cual retarda k periodos a la variable X_t .

Observación 5.4. El operador adelanto o cambio hacia adelante (en ingles, forward shift operator) se denota por F y se define como $FX_t = X_{t+1}$, cumpliéndose que $F = B^{-1}$.

Definición 5.12. Un operador diferencia (denotado por ∇) se define de la siguiente manera

$$\nabla X_t = X_t - X_{t-1} \quad (5.10)$$

para toda t .

De manera general se tiene

$$\nabla_k X_t = \sum_{i=0}^k \frac{k!}{i!(k-i)!} (-1)^i X_{t-i} \quad (5.11)$$

para toda t y $k = 0, 1, 2, \dots$

Definición 5.13. El modelo (o proceso) lineal general expresa a cada variable aleatoria \tilde{X}_t como una combinación lineal de todos sus retardos pasados más un término de error puramente aleatorio o un ruido blanco

$$\tilde{X}_t = \pi_1 \tilde{X}_{t-1} + \pi_2 \tilde{X}_{t-2} + \dots + \epsilon_t = \sum_{j=1}^{\infty} \pi_j \tilde{X}_{t-j} + \epsilon_t \quad (5.12)$$

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.2. MODELO ESTACIONARIO LINEAL GENERAL

en donde $\tilde{X}_{t-j} = X_{t-j} - \mu$ ($j = 1, 2, \dots$) son los retardos de la variable aleatoria $\tilde{X}_t = X_t - \mu$, μ es la media del proceso X_t , π_j ($j = 1, 2, \dots$) son los coeficientes asociados a las variables explicativas y ϵ_t es un proceso de ruido blanco.

Las condiciones generales que ha de cumplir el proceso son:

1. Que el proceso sea no anticipante, es decir, que el presente no venga determinado por el futuro, luego el valor de \tilde{X} en el momento t no puede depender de valores futuros de \tilde{X} o de ϵ_t .
2. Que el proceso sea invertible, es decir, que el presente dependa de forma convergente de su propio pasado lo que implica que la influencia de \tilde{X}_{t-k} en \tilde{X}_t ha de ir disminuyendo conforme nos alejemos en el pasado. Esta condición se cumple si los parámetros del modelo general (5.12) cumplen con la siguiente restricción

$$\sum_{i=1}^{\infty} \pi_i^2 < \infty. \quad (5.13)$$

Observación 5.5. *El proceso lineal general (5.12) puede contemplarse como un modelo de regresión lineal con infinitas variables explicativas, que son los retardos de la propia variable dependiente.*

El modelo (5.12) se puede escribir de forma más compacta en términos del operador de retardo:

$$\tilde{X}_t = (\pi_1 B + \pi_2 B^2 + \dots) \tilde{X}_t + \epsilon_t \Rightarrow (1 - \pi_1 B - \pi_2 B^2 - \dots) \tilde{X}_t = \epsilon_t \Rightarrow \pi_{\infty}(B) \tilde{X}_t = \epsilon_t.$$

Teorema 5.1. (Representación de Wold (1938)). *El proceso lineal general puede expresarse como una combinación lineal de los retardos de un proceso puramente aleatorio*

$$\tilde{X}_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots = \epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i} = \psi_{\infty}(B) \epsilon_t \quad (5.14)$$

en donde $\psi_{\infty}(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$ es un polinomio en B de orden infinito con $\psi_0 = 1$.

Demostración. Partiendo de que $\pi_{\infty}(B) \tilde{X}_t = \epsilon_t$, podemos escribir $\tilde{X}_t = (\frac{1}{\pi_{\infty}(B)}) \epsilon_t = \psi_{\infty}(B) \epsilon_t$. Vemos que los polinomios $\psi_{\infty}(B)$ y $\pi_{\infty}(B)$ cumplen con la relación $\psi_{\infty}(B) \pi_{\infty}(B) = 1$, de modo que podemos obtener los pesos ψ a partir de los pesos de π , y viceversa. ■

Proposición 5.1. *El proceso lineal general es estacionario si ψ_0, ψ_1, \dots es una serie convergente, esto es, si la suma de los valores absolutos de los pesos ψ_i es finita, $\sum_{i=0}^{\infty} |\psi_i| < \infty$.*

Demostración. Puesto que $E(\tilde{X}_t) = 0 \Rightarrow E(X_t - \mu) = 0 \Rightarrow E(X_t) = \mu$, además

$$E(\tilde{X}_t)^2 = E(\epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots)^2 = E(\epsilon_t^2 + \psi_1^2 \epsilon_{t-1}^2 + \psi_2^2 \epsilon_{t-2}^2 + \dots + 2\psi_1 \epsilon_t \epsilon_{t-1} + \dots) = \sigma_{\epsilon}^2 (1 + \psi_1^2 + \dots)$$

y por último se tiene

$$E(\tilde{X}_t \tilde{X}_{t-k})^2 = E[(\epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots + \psi_k \epsilon_{t-k} + \psi_{k+1} \epsilon_{t-k-1} + \psi_{k+2} \epsilon_{t-k-2} + \dots)(\epsilon_{t-k} + \psi_1 \epsilon_{t-k-1} + \psi_2 \epsilon_{t-k-2} + \dots)] = \sigma_\epsilon^2 \psi_k + \sigma_\epsilon^2 \psi_{k+1} \psi_1 + \dots = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_{k+i} \psi_i. \blacksquare$$

Definición 5.14. *Un proceso es invertible si π_0, π_1, \dots es una serie convergente, esto es, $\sum_{i=1}^{\infty} |\pi_i| < \infty$. En este caso, el pasado muy remoto es irrelevante en la explicación del presente.*

Como en este trabajo se va a estudiar una serie de tiempo finita, los modelos no van a poder expresar dependencias infinitas sin restricciones sino que tendrán que especificar una dependencia en el tiempo acotada y con restricciones. Será necesario simplificar la formulación del modelo (5.12) de forma que tenga un número finito de parámetros. Acudiendo a la teoría de polinomios, bajo condiciones muy generales, se puede aproximar un polinomio de orden infinito mediante un cociente de polinomios finitos, es decir:

$$\pi_\infty(B) \simeq \frac{\phi_p(B)}{\theta_q(B)} \quad (5.15)$$

donde $\phi_p(B)$ y $\theta_q(B)$ son polinomios en el operador de retardos finitos de orden p y q respectivamente, es decir:

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p. \\ \theta_q(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q. \end{aligned}$$

5.3. Procesos ARMA

Definición 5.15. *Sea $\{X_t; t \in T\}$ una serie de tiempo. Se dice que es un proceso autorregresivo de promedios móviles de orden \mathbf{p}, \mathbf{q} , denotado por **ARMA**(\mathbf{p}, \mathbf{q}), si $\{X_t; t \in T\}$ es estacionario y para toda t se tiene que:*

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (5.16)$$

donde $\epsilon_t \sim RB(0, \sigma^2)$ y los polinomios $(1 - \phi_1 x - \dots - \phi_p x^p)$ y $(1 + \theta_1 x + \dots + \theta_q x^q)$ no tienen factores comunes.

El proceso $\{X_t; t \in T\}$ es un proceso **ARMA**(\mathbf{p}, \mathbf{q}) con media μ si $\{X_t - \mu; t \in T\}$ es un modelo **ARMA**(\mathbf{p}, \mathbf{q}). Una forma abreviada de expresar (5.16) es de la siguiente manera:

$$\phi_p(B)X_t = \theta_q(B)\epsilon_t. \quad (5.17)$$

Una condición para la existencia de una solución para el proceso estacionario es que el polinomio $\phi(x) = 1 - \phi_1 x - \dots - \phi_p x^p$ sea distinto de cero para todo número complejo x con $|x| = 1$.

Proposición 5.2. (Existencia y unicidad). *Una solución de (5.16) existe y además es única para la serie $\{X_t; t \in T\}$ si y sólo si, $\phi(x) = 1 - \phi_1 x - \dots - \phi_p x^p \neq 0$, para todo $|x| = 1$.*

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.3. PROCESOS ARMA

Proposición 5.3. (Causal). *Un proceso ARMA(p,q), $\{X_t; t \in T\}$ es causal o una función causal de $\{\epsilon_t\}$ si existen constantes $\{\psi_i\}$ tal que $\sum_{i=0}^{\infty} |\psi_i| < \infty$ y además*

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \quad (5.18)$$

para toda $t \in T$.

Una condición equivalente para que la serie sea causal es que $1 - \phi_1 x - \dots - \phi_p x^p \neq 0$, para todo $|x| \leq 1$.

La sucesión $\{\psi_i\}$ en (5.18) se determina por la relación $\psi(x) = \sum_{i=0}^{\infty} \psi_i x^i = \frac{\theta_q(x)}{\phi_p(x)}$ o equivalentemente,

$$(1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p)(\psi_0 + \psi_1 x + \dots) = 1 + \theta_1 x + \theta_2 x^2 + \dots + \theta_q x^q.$$

Igualando los coeficientes correspondientes a x^i con $i = 0, 1, \dots$, se encuentra que

$$\begin{aligned} 1 &= \psi_0 \\ \theta_1 &= \psi_1 - \psi_0 \phi_1 \\ \theta_2 &= \psi_2 - \psi_1 \phi_1 - \psi_0 \phi_2 \\ &\vdots \\ &\vdots \\ &\vdots \end{aligned}$$

o equivalentemente,

$$\psi_i - \sum_{k=1}^p \phi_k \psi_{i-k} = \theta_i$$

con $i = 0, 1, 2, \dots$ donde $\theta_0 := 1, \theta_i := 0$ para $i > q$, además $\psi_i := 0$ para $i < 0$.

Proposición 5.4. (Invertibilidad). *Un proceso ARMA(p,q), $\{X_t; t \in T\}$ es invertible si existen constantes $\{\pi_i\}$ tales que $\sum_{i=0}^{\infty} |\pi_i| < \infty$ y*

$$\epsilon_t = \sum_{i=0}^{\infty} \pi_i X_{t-i} \quad (5.19)$$

para toda $t \in T$.

La condición de invertibilidad es equivalente a que $\theta_q(x) = 1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_q x^q \neq 0$ para todo, $|x| \leq 1$.

Para poder obtener los valores de la sucesión $\{\pi_i\}$ se realiza el mismo razonamiento que fue utilizado para hallar los coeficientes de $\{\psi_i\}$. Así los valores serán determinados mediante la ecuación,

$$\pi_i + \sum_{k=1}^q \theta_k \pi_{i-k} = -\phi_i$$

para $i = 0, 1, \dots$ donde $\phi_0 := -1, \phi_j = 0$ para $j > p$, además $\pi_j := 0$ para $j < 0$.

5.3.1. Función de autocovarianza de un proceso ARMA(p,q)

A continuación se presentan diferentes métodos para calcular el coeficiente de autocovarianza de un proceso causal ARMA.

Primer Método. El proceso causal ARMA(p,q) definido por $\phi_p(B)X_t = \theta_q(B)\epsilon_t$ y en el que la causalidad implica que $X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ donde $\sum_{i=0}^{\infty} \psi_i x^i = \frac{\theta_q(x)}{\phi_p(x)}$, con $|x| \leq 1$, se tiene como coeficiente de autocovarianza γ' a $\gamma'_k = E(X_{t+k}X_t) = E[(\sum_{i=-\infty}^{\infty} \psi_i \epsilon_{t+k-i})(\sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j})] = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j E(\epsilon_{t+k-i} \epsilon_{t-j}) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j \gamma'_{k-i+j} = \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+k} \sigma^2 = \sum_{i=0}^{\infty} \psi_i \psi_{i+|k|} \sigma^2$.

$$\therefore \gamma'_k = \sum_{i=0}^{\infty} \psi_i \psi_{i+|k|} \sigma^2. \quad (5.20)$$

Segundo Método. Si se multiplica por la derecha cada lado de la ecuación

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

por X_{t-h} , $h = 0, 1, 2, \dots$, y calculamos la esperanza de cada lado, esto es, para el miembro izquierdo se tiene que,

$$A = X_t X_{t-h} - \phi_1 X_{t-1} X_{t-h} - \dots - \phi_p X_{t-p} X_{t-h}$$

entonces

$$E(A) = E(X_t X_{t-h} - \phi_1 X_{t-1} X_{t-h} - \dots - \phi_p X_{t-p} X_{t-h}) = E(X_t X_{t-h}) - \phi_1 E(X_{t-1} X_{t-h}) - \dots - \phi_p E(X_{t-p} X_{t-h})$$

$$\therefore E(A) = \gamma'_h - \phi_1 \gamma'_{h-1} - \dots - \phi_p \gamma'_{h-p}$$

y para el miembro derecho, sea

$$B = \epsilon_t X_{t-h} + \theta_1 \epsilon_{t-1} X_{t-h} + \dots + \theta_q \epsilon_{t-q} X_{t-h}$$

entonces

$$E(B) = E(\epsilon_t X_{t-h} + \theta_1 \epsilon_{t-1} X_{t-h} + \dots + \theta_q \epsilon_{t-q} X_{t-h})$$

$$E(B) = E(\epsilon_t X_{t-h}) + \theta_1 E(\epsilon_{t-1} X_{t-h}) + \dots + \theta_q E(\epsilon_{t-q} X_{t-h})$$

$$E(B) = E(\epsilon_t \sum_{i=0}^{\infty} \psi_i \epsilon_{t-(i+h)}) + \theta_1 E(\epsilon_{t-1} \sum_{i=0}^{\infty} \psi_i \epsilon_{t-(i+h)}) + \dots + \theta_q E(\epsilon_{t-q} \sum_{i=0}^{\infty} \psi_i \epsilon_{t-(i+h)})$$

$$E(B) = \sum_{i=0}^{\infty} \psi_i E(\epsilon_t \epsilon_{t-(i+h)}) + \theta_1 \sum_{i=0}^{\infty} \psi_i E(\epsilon_{t-1} \epsilon_{t-(i+h)}) + \dots + \theta_q \sum_{i=0}^{\infty} \psi_i E(\epsilon_{t-q} \epsilon_{t-(i+h)}).$$

Ahora, si se toma un h que se encuentre entre $[0, m)$ donde $m = \max(p, q + 1)$, como primera parte se toma $h = 0$, entonces la ecuación anterior queda

$$\sigma^2(\theta_0 \psi_0 + \theta_1 \psi_1 + \dots + \theta_q \psi_q)$$

si $h = 1$

$$0 + \sigma^2(\theta_1 \psi_0 + \theta_2 \psi_1 + \dots + \theta_q \psi_{q-1})$$

·
·

si $h = q$

$$0 + 0 + \dots + \theta_q \psi_0$$

si $h = m$, en este caso si $m = q + 1$

$$0 + 0 + \dots + 0$$

pero si $m = q$ entonces $p > q + 1$, por lo tanto, la ecuación es cero, al igual que si se toma $h > m$. Por lo tanto

$$\gamma'_h - \phi_1 \gamma'_{h-1} - \dots - \phi_p \gamma'_{h-p} = \sigma^2 \sum_{i=0}^{\infty} \theta_{h+i} \psi_i \quad (5.21)$$

para $0 \leq h < m$ y para $h \geq m$ se tiene

$$\gamma'_h - \phi_1 \gamma'_{h-1} - \dots - \phi_p \gamma'_{h-p} = 0 \quad (5.22)$$

donde $\psi_i := 0$ para $i < 0$, $\theta_0 := 1$ y $\theta_i := 0$, para algún i que no pertenece al conjunto $\{0, \dots, q\}$. Las ecuaciones (5.22) son un conjunto de ecuaciones de diferencia lineales no homogéneas con coeficientes constantes cuya solución esta dada por

$$\gamma'_k = \alpha_1 \xi_1^{-k} + \alpha_2 \xi_2^{-k} + \dots + \alpha_p \xi_p^{-k} \quad (5.23)$$

para $k \geq m - p$, donde las ξ_1, \dots, ξ_p son las raíces de la ecuación $\phi(x) = 0$ y $\alpha_1, \alpha_2, \dots, \alpha_p$ son constantes arbitrarias. Después se sustituye la ecuación (5.23) en (5.21) para obtener un conjunto de m ecuaciones que determinaran a las constantes $\alpha_1, \dots, \alpha_p$ y las $m - p$ autocovarianzas γ'_k , $0 \leq k < m - p$.

Tercer Método. Se puede encontrar el coeficiente de autocovarianza γ' usando las ecuaciones (5.21) y (5.22) para resolver las primeras $p + 1$ ecuaciones para $\gamma'_0, \dots, \gamma'_p$ y usando las ecuaciones subsecuentes para determinar $\gamma'_{p+1}, \gamma'_{p+2}, \dots$ [1]

Definición 5.16. El conjunto $\{\gamma'_k; k \in \mathbb{Z}\}$ se denomina función de autocovarianza (**FACV**).

5.3.2. Función de autocorrelación de un proceso ARMA(p,q)

El coeficiente de autocorrelación de un proceso **ARMA(p,q)** $\{X_t; t \in T\}$ es el coeficiente ρ' encontrado inmediatamente desde el coeficiente de autocovarianza γ' y se escribe como

$$\rho'_k = \frac{\gamma'_k}{\gamma'_0}.$$

Definición 5.17. El conjunto $\{\rho'_k; k \in \mathbb{Z}\}$ se denomina función de autocorrelación (**FAC**).

5.3.3. Función de autocorrelación parcial de un proceso ARMA(p,q)

El coeficiente de autocorrelación parcial de un proceso ARMA(p,q) $\{X_t; t \in T\}$ está dado por el coeficiente α el cual esta definido mediante las siguientes ecuaciones $\alpha_0 = 1$ y $\alpha_k = \phi_{kk}$, para $k \geq 1$, donde ϕ_{kk} es la última componente de

$$\phi_k = \Gamma_k^{-1} \gamma_k,$$

donde $\Gamma_k = [\gamma'_{(i-j)}]_{i,j=1}^k$ y $\gamma_k = [\gamma'_{(1)}, \gamma'_{(2)}, \dots, \gamma'_{(k)}]^t$.

Definición 5.18. *El conjunto $\{\alpha_k; k \in \mathbb{Z}\}$ se denomina función de autocorrelación parcial (FACP).*

La función de autocorrelación parcial mide la relación que existe en las observaciones hechas en la serie de tiempo separadas por un retraso de k unidades de tiempo eliminando el efecto de las observaciones intermedias.

Función de autocorrelación parcial muestral de un proceso ARMA(p,q)

Para un conjunto de observaciones $\{x_1, x_2, \dots, x_n\}$ con $x_i \neq x_j$ para algún i y j , el coeficiente de autocorrelación muestral parcial, $\hat{\alpha}_k$ esta definido mediante:

$$\hat{\alpha}_0 = 1 \text{ y } \hat{\alpha}_k = \hat{\phi}_{kk} \text{ con } k \leq 1,$$

donde $\hat{\phi}_{kk}$ es la última componente correspondiente de

$$\hat{\phi}_k = \hat{\Gamma}_k^{-1} \hat{\gamma}_k.$$

Definición 5.19. *El conjunto $\{\hat{\alpha}_k; k \in \mathbb{Z}\}$ se denomina función de autocorrelación parcial muestral (FACPM).*

5.4. Modelos autorregresivos AR

Definición 5.20. *El proceso autorregresivo de primer orden, denotado por AR(1), es*

$$X_t = \delta + \phi X_{t-1} + \epsilon_t \Leftrightarrow (1 - \phi B)X_t = \delta + \epsilon_t \quad (5.24)$$

en donde δ , ϕ y $\sigma_\epsilon^2 = E(\epsilon_t^2)$ son los parámetros del modelo y ϵ_t es un proceso de ruido blanco.

Observación 5.6. *El proceso AR(1) se obtiene como un caso especial del proceso lineal general cuando $\pi_1 = \phi_1$ y $\pi_j = 0$ para $j > 1$.*

Observación 5.7. *El proceso AR(1) puede contemplarse como un modelo de regresión simple*

$$y_t = \beta_1 + \beta_2 x_t + u_t \quad (5.25)$$

en donde la variable explicativa es la propia variable dependiente retardada un periodo. El término autorregresivo significa que la variable dependiente se explica por si misma, sin auxilio de otras variables.

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.4. MODELOS AUTORREGRESIVOS AR

Proposición 5.5. *El proceso AR(1) tendrá una media estable si $\phi \neq 1$.*

Demostración. El proceso lineal general tiene media estable si $\sum_{i=1}^{\infty} \pi_i \neq 1$. En nuestro caso, $\sum_{i=1}^{\infty} \pi_i = \phi$. ■

Proposición 5.6. *Bajo estacionariedad de primer orden podemos escribir, el proceso AR(1) en desviaciones respecto a la media*

$$\tilde{X}_t = \phi \tilde{X}_{t-1} + \epsilon_t \Leftrightarrow (1 - \phi B) \tilde{X}_t = \epsilon_t. \quad (5.26)$$

Proposición 5.7. *La representación de Wold del proceso AR(1) es*

$$\tilde{X}_t = \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots = \epsilon_t + \sum_{j=1}^{\infty} \phi^j \epsilon_{t-j}. \quad (5.27)$$

Demostración. La forma de ψ puede obtenerse siguiendo dos aproximaciones alternativas.

1. Sustitución reiterada de retardos:

$$\begin{aligned} \tilde{X}_t &= \phi \tilde{X}_{t-1} + \epsilon_t \\ \tilde{X}_t &= \phi[\phi \tilde{X}_{t-2} + \epsilon_{t-1}] + \epsilon_t = \phi^2 \tilde{X}_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \\ \tilde{X}_t &= \phi^2[\phi \tilde{X}_{t-3} + \epsilon_{t-2}] + \phi \epsilon_{t-1} + \epsilon_t = \phi^3 \tilde{X}_{t-3} + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \\ &\vdots \\ \tilde{X}_t &= \phi^k \tilde{X}_{t-k} + \phi^{k-1} \epsilon_{t-(k-1)} + \dots + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \end{aligned}$$

Si $|\phi| < 1$, entonces el término $\phi^k \tilde{X}_{t-k}$ es despreciable y tenemos el resultado buscado.

2. Inversión del polinomio autorregresivo:

$$(1 - \phi B) \tilde{X}_t = \epsilon_t \Rightarrow \tilde{X}_t = \frac{1}{(1 - \phi B)} \epsilon_t = \psi(B) \epsilon_t$$

en donde $\psi(B)$ es un polinomio en B de orden infinito

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

cuyos coeficientes pueden encontrarse de la relación $\psi(B)(1 - \phi B) = 1$. El polinomio producto es

$$\begin{aligned} \psi(B)(1 - \phi B) &= (1 + \psi_1 B + \psi_2 B^2 + \dots)(1 - \phi B) \\ \psi(B)(1 - \phi B) &= 1 + \psi_1 B + \psi_2 B^2 + \dots - \phi B - \psi_1 \phi B^2 - \psi_2 \phi B^3 - \dots \\ \psi(B)(1 - \phi B) &= 1 + (\psi_1 - \phi) B + (\psi_2 - \psi_1 \phi) B^2 + \dots + (\psi_j - \phi \psi_{j-1}) B^j + \dots \end{aligned}$$

será igual a uno si sus coeficientes $\psi_j - \phi \psi_{j-1}$ son nulos. De esto se encuentra que $\psi_j = \phi^j$, lo cual implica que

$$\tilde{X}_t = \psi(B) \epsilon_t = \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots$$

Esta segunda aproximación a veces se obtiene directamente como una aplicación de la suma de una serie geométrica

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.4. MODELOS AUTORREGRESIVOS AR

$$1 + x + x^2 + x^3 + \dots = \frac{1}{1-x} \text{ cuando } -1 < x < 1$$

de manera que

$$(1 - \phi B)^{-1} = 1 + \phi B + \phi^2 B^2 + \dots \blacksquare$$

Proposición 5.8. *El proceso AR(1) es estacionario si $-1 < \phi < 1$.*

Demostración. El proceso lineal general es estacionario si $\sum_{j=1}^{\infty} \psi_j < \infty$. En nuestro caso, $\psi_j = \phi^j$, la condición de estacionariedad requiere que $|\phi| < 1$. ■

Proposición 5.9. *El proceso AR(1) siempre es invertible.*

Demostración. Es claro que el proceso AR(1) no depende del pasado remoto $\sum_{j=1}^{\infty} \pi_j = \phi < \infty$. ■

Proposición 5.10. *La función de autocovarianzas $\{\gamma_k; k \in \mathbb{Z}\}$ de un proceso AR(1) es*

$$\gamma_k = \phi^k \gamma_0$$

en donde $\gamma_0 = \frac{\sigma_\epsilon^2}{(1-\phi^2)}$.

Demostración. Para obtener la covarianza en el retardo k , $\gamma_k = E[(X_t - \mu)(X_{t-k} - \mu)]$, multiplicamos por la derecha la ecuación (5.26) por \tilde{X}_{t-k} y tomamos la esperanza matemática

$$E(\tilde{X}_t \tilde{X}_{t-k}) = E[(\phi \tilde{X}_{t-1} + \epsilon_t) \tilde{X}_{t-k}] = \phi E[\tilde{X}_{t-1} \tilde{X}_{t-k}] + E[\epsilon_t \tilde{X}_{t-k}]$$

en donde $E[\epsilon_t \tilde{X}_{t-k}] = \sigma_\epsilon^2$ si $k = 0$ y $E[\epsilon_t \tilde{X}_{t-k}] = 0$ si $k > 0$, Vemos que para $k = 0$, $\gamma_0 = \frac{\sigma_\epsilon^2}{(1-\phi^2)}$, y para $k > 0$, γ_k cumple una ecuación en diferencias de primer orden $\gamma_k = \phi \gamma_{k-1}$. ■

Observación 5.8. *La varianza γ_0 del proceso puede obtenerse directamente de la forma ψ . En efecto,*

$$\gamma_0 = E(\tilde{X}_t^2) = E\left[\left(\sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}\right)^2\right] = E\left[\sum_{j=0}^{\infty} \phi^{2j} \epsilon_{t-j}^2 + \sum_{j=0}^{\infty} \sum_{h \neq j}^{\infty} \phi^j \phi^h \epsilon_{t-j} \epsilon_{t-h}\right] = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \phi^{2j}$$

que es la suma de una progresión geométrica de razón ϕ^2 . De aquí, $\gamma_0 = \frac{\sigma_\epsilon^2}{(1-\phi^2)}$.

Sabemos que la función de autocovarianza es una función simétrica, es decir, $\gamma_k = \gamma_{-k}$, para toda $k \in \mathbb{Z}$, entonces se tiene que

$$\gamma_k = \gamma_0 \phi^{|k|} \text{ para toda } k \in \mathbb{Z}.$$

Proposición 5.11. *La función de autocorrelación $\{\rho_k; k \in \mathbb{Z}\}$ de un proceso AR(1) es*

$$\rho_k = \phi^{|k|}.$$

Observación 5.9. *Se dice que el proceso AR(1) tiene una memoria infinita para indicar que X_t está correlacionado con cualquier retardo X_{t-k} .*

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.4. MODELOS AUTORREGRESIVOS AR

Definición 5.21. Sea $\{X_t; t \in T\}$ una serie de tiempo tal que

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t \quad (5.28)$$

donde $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$, $|\phi_i| < 1$ para $i = 1, \dots, p$. Al proceso (5.28) se le llama proceso autorregresivo de orden p , y se denota por $\mathbf{AR}(p)$.

Este modelo se podría tomar como un modelo de regresión múltiple donde $\{X_t; t \in T\}$ es regresado por valores pasados de $\{X_t; t \in T\}$ mas que ser regresado en variables predictoras separadas.

El proceso $\mathbf{AR}(p)$ puede ser representado en términos del operador de retardo B de la siguiente forma:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = \epsilon_t \Rightarrow \phi_p(B) X_t = \epsilon_t. \quad (5.29)$$

Teorema 5.2. Un proceso autorregresivo finito $\mathbf{AR}(p)$ es estacionario si y sólo si el módulo de las raíces del polinomio autorregresivo $\phi_p(B)$ está fuera del círculo unitario.

Una manera de verificar si el proceso es estacionario es usando el Teorema de Schur, el cual se usa para verificar la estacionaridad en términos de los parámetros $\phi_1, \phi_2, \dots, \phi_p$ y en el cual solo se requiere que los p determinantes que se presentan a continuación sean positivos.

Teorema 5.3. (De Schur) Los módulos de las raíces de la ecuación $g^p - a_1 g^{p-1} - a_2 g^{p-2} - \dots - a_{p-1} g - a_p = 0$ serán todos menores que la unidad, si y sólo si los p determinantes (D_1, \dots, D_p) que se muestran a continuación son positivos [4].

$$D_1 = \begin{vmatrix} -1 & a_p \\ a_p & -1 \end{vmatrix}.$$

$$D_2 = \begin{vmatrix} -1 & 0 & a_p & a_{p-1} \\ a_1 & -1 & 0 & a_p \\ a_p & 0 & -1 & a_1 \\ a_{p-1} & a_p & 0 & -1 \end{vmatrix}.$$

$$\vdots$$

$$D_p = \begin{vmatrix} -1 & 0 & \dots & 0 & a_p & a_{p-1} & \dots & a_1 \\ a_1 & -1 & \dots & 0 & 0 & a_p & \dots & a_2 \\ \dots & \dots \\ a_{p-1} & a_{p-2} & \dots & -1 & 0 & 0 & \dots & a_p \\ a_p & 0 & \dots & 0 & -1 & a_1 & \dots & a_{p-1} \\ a_{p-1} & a_p & \dots & 0 & 0 & -1 & \dots & a_{p-2} \\ \dots & \dots \\ a_1 & a_2 & \dots & a_p & 0 & 0 & \dots & -1 \end{vmatrix}.$$

Por lo que en los modelos autorregresivos se debe verificar que los siguientes determinantes sean positivos, para los ϕ_i , $i = 1, 2, \dots, p$.

$$D_1 = \begin{vmatrix} -1 & \phi_p \\ \phi_p & -1 \end{vmatrix}.$$

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.5. MODELOS DE PROMEDIOS MÓVILES MA

$$D_2 = \begin{vmatrix} -1 & 0 & \phi_p & \phi_{p-1} \\ \phi_1 & -1 & 0 & \phi_p \\ \phi_p & 0 & -1 & \phi_1 \\ \phi_{p-1} & \phi_p & 0 & -1 \end{vmatrix}.$$

$$\vdots$$

$$D_p = \begin{vmatrix} -1 & 0 & \dots & 0 & \phi_p & \phi_{p-1} & \dots & \phi_1 \\ \phi_1 & -1 & \dots & 0 & 0 & \phi_p & \dots & \phi_2 \\ \dots & \dots \\ \phi_{p-1} & \phi_{p-2} & \dots & -1 & 0 & 0 & \dots & \phi_p \\ \phi_p & 0 & \dots & 0 & -1 & \phi_1 & \dots & \phi_{p-1} \\ \phi_{p-1} & \phi_p & \dots & 0 & 0 & -1 & \dots & \phi_{p-2} \\ \dots & \dots \\ \phi_1 & \phi_2 & \dots & \phi_p & 0 & 0 & \dots & -1 \end{vmatrix}.$$

Por otro lado, si multiplicamos cada lado de (5.28) por X_{t-k} , para $k = 1, 2, \dots, p$, posteriormente calculamos la esperanza y por último dividimos entre γ_0 , obtenemos

$$\frac{E(X_t X_{t-k})}{\gamma_0} = \frac{E[(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t) X_{t-k}]}{\gamma_0}$$

$$\frac{\gamma_k}{\gamma_0} = \frac{\phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}}{\gamma_0}$$

$$\therefore \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (5.30)$$

para toda $k > 0$. Las ecuaciones en (5.30) se denominan las ecuaciones de Yule-Walker, y pueden también escribirse de la siguiente forma

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \rho_k = 0 \quad (5.31)$$

para $k = 1, 2, \dots, p$, de forma que la estacionaridad se da si cada una de las raíces caen fuera del círculo unitario.

5.5. Modelos de promedios móviles MA

Definición 5.22. La ecuación de un proceso de promedios móviles de primer orden, denotado por **MA(1)**, es

$$X_t = \mu + \epsilon_t - \theta \epsilon_{t-1} \Leftrightarrow X_t = \mu + (1 - \theta B) \epsilon_t \quad (5.32)$$

en donde μ , θ , y $\sigma_\epsilon^2 = E(\epsilon_t^2)$ son los parámetros del modelo y ϵ_t es un proceso de ruido blanco.

Observación 5.10. El proceso **MA(1)** es un caso especial de la forma ψ del proceso lineal general que se obtiene fijando $\psi_1 = \theta$ y $\psi_j = 0$ para toda $j > 1$.

Proposición 5.12. El proceso **MA(1)** en desviaciones respecto a la media es

$$\tilde{X}_t = \epsilon_t - \theta \epsilon_{t-1} \text{ o } \tilde{X}_t = (1 - \theta B) \epsilon_t$$

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.5. MODELOS DE PROMEDIOS MÓVILES MA

en donde $\tilde{X}_t = X_t - \mu$.

Demostración. Es inmediato comprobar que el término constante μ es la media del proceso

$$E(X_t) = \mu + E(\epsilon_t) - \theta(E(\epsilon_{t-1})) = \mu. \blacksquare$$

Proposición 5.13. *La forma π del proceso MA(1) es*

$$\tilde{X}_t = -\theta\tilde{X}_{t-1} - \theta^2\tilde{X}_{t-2} - \dots + \epsilon_t = -\sum_{j=1}^{\infty} \theta^j \tilde{X}_{t-j} + \epsilon_t.$$

Demostración. Seguimos las dos aproximaciones descritas para el proceso AR(1).

1. Sustitución reiterada de errores retardados:

$$\begin{aligned} \tilde{X}_t &= \epsilon_t - \theta\epsilon_{t-1} \\ \tilde{X}_t &= \epsilon_t - \theta(\tilde{X}_{t-1} + \theta\epsilon_{t-2}) = \epsilon_t - \theta\tilde{X}_{t-1} - \theta^2\epsilon_{t-2} \\ \tilde{X}_t &= \epsilon_t - \theta\tilde{X}_{t-1} - \theta^2(\tilde{X}_{t-2} + \theta\epsilon_{t-3}) = \epsilon_t - \theta\tilde{X}_{t-1} - \theta^2\tilde{X}_{t-2} - \theta^3\epsilon_{t-3} \\ &\vdots \\ \tilde{X}_t &= \epsilon_t - \theta\tilde{X}_{t-1} - \theta^2\tilde{X}_{t-2} - \dots - \theta^{k-1}\tilde{X}_{t-k+1} - \theta^k\epsilon_{t-k} \end{aligned}$$

en donde el término $\theta^k\epsilon_{t-k}$ tenderá a cero cuando $k \rightarrow \infty$ si $|\theta| < 1$.

2. Inversión del polinomio de medias móviles:

$$\tilde{X}_t = (1 - \theta B)\epsilon_t \implies \left(\frac{1}{(1 - \theta B)}\right)\tilde{X}_t = \epsilon_t \implies \pi(B)\tilde{X}_t = \epsilon_t$$

en donde $\pi(B)$ es un polinomio en B de orden infinito

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$$

cuyos coeficientes pueden encontrarse de la relación $\pi(B)(1 - \theta B) = 1$. El polinomio producto

$$\begin{aligned} \pi(B)(1 - \theta B) &= 1 - \pi_1 B - \pi_2 B^2 - \dots - \theta B + \theta\pi_1 B^2 + \theta\pi_2 B^3 + \dots \\ \therefore \pi(B)(1 - \theta B) &= 1 - (\pi_1 + \theta)B - (\pi_2 + \theta\pi_1)B^2 - \dots - (\pi_j + \theta\pi_{j-1})B^j + \dots \end{aligned}$$

será igual a 1 si sus coeficientes $\pi_j + \theta\pi_{j-1}$ son nulos. De aquí encontramos que los pesos $\pi_j = -\theta^j$. \blacksquare

Proposición 5.14. *Un proceso MA(1) siempre es estacionario.*

Demostración. Como se cumple que $\sum_{j=1}^{\infty} |\psi_j| = |\theta| < \infty$. \blacksquare

Proposición 5.15. *Un proceso MA(1) es invertible si $-1 < \theta < 1$.*

Demostración. Se cumplirá que $\sum_{j=1}^{\infty} |\pi_j| = \sum_{j=1}^{\infty} |\theta^j| < \infty$ cuando $|\theta| < 1$. \blacksquare

Proposición 5.16. *La función de autocovarianzas $\{\gamma_k; k \in \mathbb{Z}\}$ de un proceso MA(1) es*

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.5. MODELOS DE PROMEDIOS MÓVILES MA

$$\gamma_k = \begin{cases} (1 + \theta^2)\sigma_\epsilon^2, & k = 0; \\ -\theta\sigma_\epsilon^2, & k = 1; \\ 0, & k > 1. \end{cases}$$

Demostración. Obsérvese que

$$E(\tilde{X}_t) = E(\epsilon_t - \theta\epsilon_{t-1}) = 0.$$

Por tanto

$$\begin{aligned} \gamma_0 &= E[(\tilde{X}_t - E(\tilde{X}_t))^2] = E[(\tilde{X}_t)^2] = E[(\epsilon_t - \theta\epsilon_{t-1})^2] = E[\epsilon_t^2] - (2\theta)E[\epsilon_t\epsilon_{t-1}] + \theta^2E[\epsilon_{t-1}^2] \\ &\therefore \gamma_0 = \sigma_\epsilon^2 - 0 + \theta^2\sigma_\epsilon^2 = (1 + \theta^2)\sigma_\epsilon^2 < \infty. \end{aligned}$$

Las autocovarianzas, $\gamma_k, k > 0$:

$$\begin{aligned} \gamma_1 &= E[(\tilde{X}_t - E(\tilde{X}_t))(\tilde{X}_{t-1} - E(\tilde{X}_{t-1}))] = E[(\epsilon_t - \theta\epsilon_{t-1})(\epsilon_{t-1} - \theta\epsilon_{t-2})] \\ &\therefore \gamma_1 = E[\epsilon_t\epsilon_{t-1}] - \theta E[\epsilon_{t-1}^2] - (\theta)E[\epsilon_t\epsilon_{t-2}] + \theta^2 E[\epsilon_{t-1}\epsilon_{t-2}] = -\theta\sigma_\epsilon^2 \\ \gamma_2 &= E[(\tilde{X}_t - E(\tilde{X}_t))(\tilde{X}_{t-2} - E(\tilde{X}_{t-2}))] = E[(\epsilon_t - \theta\epsilon_{t-1})(\epsilon_{t-2} - \theta\epsilon_{t-3})] \\ &\therefore \gamma_2 = E[\epsilon_t\epsilon_{t-2}] - \theta E[\epsilon_{t-1}\epsilon_{t-2}] - \theta E[\epsilon_t\epsilon_{t-3}] + \theta^2 E[\epsilon_{t-1}\epsilon_{t-3}] = 0. \end{aligned}$$

Por lo tanto, es claro que para $k > 1$

$$\gamma_k = E[(\tilde{X}_t\tilde{X}_{t-k})] = E[(\epsilon_t - \theta\epsilon_{t-1})(\epsilon_{t-k} - \theta\epsilon_{t-k-1})] = 0. \blacksquare$$

Proposición 5.17. La función de autocorrelación simple $\{\rho_k; k \in \mathbb{Z}\}$ de un proceso MA(1) es

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

Observación 5.11. Se dice que la memoria del proceso MA(1) es de un periodo porque $\rho_k = 0$ para $k > 1$.

Definición 5.23. Un proceso $\{X_t; t \in T\}$ se dice ser un proceso de promedios móviles de orden q , que se denotará por MA(q), si

$$X_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \dots - \theta_q\epsilon_{t-q} \quad (5.33)$$

donde $\epsilon_{t-j} \sim RB(0, \sigma_\epsilon^2)$ para $j = 0, 1, 2, \dots, q$ y $\theta_1, \theta_2, \dots, \theta_q$ son constantes.

El modelo de promedios móviles finito de orden q definido anteriormente, es la aproximación natural del modelo lineal general (5.14), ya que supone su truncamiento bajo el supuesto de que

$$\psi_i = 0 \text{ para toda } i > q.$$

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.5. MODELOS DE PROMEDIOS MÓVILES MA

Se dice que todo proceso de promedios móviles es estacionario ya que éste puede expresarse como un proceso lineal y $\{\theta_i; i = 1, 2, \dots, q\}$ es una sucesión de constantes absolutamente sumable. Notemos que la ecuación (5.33) se puede escribir en términos del operador de retardos de la siguiente manera

$$X_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \dots - \theta_q\epsilon_{t-q} = \theta_q(B)\epsilon_t \quad (5.34)$$

donde $\theta_q(B)$ es un polinomio de orden q llamado polinomio de medias móviles y $(\theta_1, \theta_2, \dots, \theta_q)$ es el vector de parámetros de medias móviles. Además se tiene que

$$E(X_t) = E(\epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \dots - \theta_q\epsilon_{t-q}) = 0 \quad (5.35)$$

y

$$\gamma_0 = Var(X_t) = Var(\epsilon_t - \theta_1\epsilon_{t-1} - \dots - \theta_q\epsilon_{t-q}) = (1 - \theta_1^2 - \dots - \theta_q^2)\sigma_\epsilon^2 \quad (5.36)$$

y como coeficiente de autocovarianza

$$\begin{aligned} \gamma_k = Cov(X_t, X_{t+k}) &= Cov(\epsilon_t + \theta_1\epsilon_{t-1} + \dots + \theta_q\epsilon_{t-q}, \epsilon_{t+k} + \theta_1\epsilon_{t+k-1} + \dots + \theta_q\epsilon_{t+k-q}) = \\ &Cov(\epsilon_t, \epsilon_{t+k}) + \theta_1Cov(\epsilon_t, \epsilon_{t+k-1}) + \dots + \theta_qCov(\epsilon_t, \epsilon_{t+k-q}) + \theta_1Cov(\epsilon_{t-1}, \epsilon_{t+k}) + \\ &\theta_1^2Cov(\epsilon_{t-1}, \epsilon_{t+k-1}) + \dots + \theta_qCov(\epsilon_{t-q}, \epsilon_{t+k}) + \theta_q\theta_1Cov(\epsilon_{t-q}, \epsilon_{t+k-1}) + \dots + \\ &\theta_q^2Cov(\epsilon_{t-q}, \epsilon_{t+k-q}) \end{aligned}$$

así que el coeficiente de autocovarianza queda de la siguiente manera

$$\gamma_k = \begin{cases} 0, & \text{si } |k| > q; \\ \sigma_\epsilon^2 \sum_{i=0}^{q-k} \theta_i\theta_{i+k}, & \text{si } k = 0, \pm 1, \dots, \pm q. \end{cases} \quad (5.37)$$

y por lo tanto el coeficiente de autocorrelación está dado por

$$\rho_k = \begin{cases} 0, & \text{si } k > 0; \\ \frac{\sum_{i=0}^{q-k} \theta_i\theta_{i+k}}{\sum_{i=0}^q \theta_i^2}, & \text{si } k = 1, \dots, q. \end{cases} \quad (5.38)$$

Debido a que la suma $\sum_{i=0}^q |\theta_i|$, cuando se considera un orden finito, es una constante finita, se tiene que todo proceso de promedios móviles es estacionario. Además para los procesos de promedios móviles se requiere que los coeficientes $\theta_1, \theta_2, \dots, \theta_q$ cumplan con la condición de invertibilidad.

Teorema 5.4. *Un proceso de promedios móviles finito $MA(q)$ es invertible si y sólo si el módulo de las raíces del polinomio de promedios móviles $\theta_q(B)$ está fuera del círculo unidad.*

Nuevamente, usando el teorema de Schur, se puede verificar que un proceso de promedios móviles es invertible si los q determinantes (D_1, D_2, \dots, D_q) son todos positivos.

$$\begin{aligned}
 D_1 &= \begin{vmatrix} -1 & \theta_q \\ \theta_q & -1 \end{vmatrix} \\
 D_2 &= \begin{vmatrix} -1 & 0 & \theta_q & \theta_{q-1} \\ \theta_1 & -1 & 0 & \theta_q \\ \theta_q & 0 & -1 & \theta_1 \\ \theta_{q-1} & \theta_q & 0 & -1 \end{vmatrix} \\
 &\quad \cdot \\
 &\quad \cdot \\
 D_p &= \begin{vmatrix} -1 & 0 & \dots & 0 & \theta_q & \theta_{q-1} & \dots & \theta_1 \\ \theta_1 & -1 & \dots & 0 & 0 & \theta_q & \dots & \theta_2 \\ \dots & \dots \\ \theta_{q-1} & \theta_{q-2} & \dots & -1 & 0 & 0 & \dots & \theta_q \\ \theta_q & 0 & \dots & 0 & -1 & \theta_1 & \dots & \theta_{q-1} \\ \theta_{q-1} & \theta_q & \dots & 0 & 0 & -1 & \dots & \theta_{q-2} \\ \dots & \dots \\ \theta_1 & \theta_2 & \dots & \theta_q & 0 & 0 & \dots & -1 \end{vmatrix} \cdot
 \end{aligned}$$

5.6. Modelos ARIMA

Los modelos **ARIMA** fueron popularizados en los años 70 por George Box y Gwilym Jenkins, y sus nombres se utilizan, frecuentemente, como sinónimos de la metodología **ARIMA** aplicada al análisis y predicción de series. Esta familia de modelos ha sido utilizada ampliamente a partir de los años 80, debido a los avances de recursos de cálculo y optimización.

El acrónimo **ARIMA** significa modelo autorregresivo integrado de media móvil (AutoRegresive Integrated Moving Average). Cada una de las tres partes del acrónimo se le denomina componente y modela un comportamiento distinto de la serie.

Los modelos **ARIMA** se construyen a partir de los modelos **ARMA**, pero considerando que la serie en estudio para que sea estacionaria en media tendrá que diferenciarse una serie de veces.

Un modelo **ARIMA(p,d,q)** es un modelo **ARMA(p,q)** sobre la serie diferenciada d veces. Es decir, su expresión algebraica, sera

$$W_t = \nabla^d X_t = (1 - B)^d X_t \quad (5.39)$$

para toda t .

Para esta nueva serie podría ya ser posible obtener un modelo **ARMA**:

$$\phi_p(B)W_t = \theta_q(B)\epsilon_t, \quad (5.40)$$

lo cual es equivalente al modelo **ARIMA**

$$\phi_p(B)\nabla^d X_t = \theta_q(B)\epsilon_t \quad (5.41)$$

en donde $d \geq 1$, para $\{W_t; t \in T\}$, en donde $\{\epsilon_t; t \in T\}$ es un proceso ruido blanco. El término *integrado* se refiere a que W_t se obtiene de la relación (5.39) por inversión del operador ∇^d , dando como resultado una suma infinita (o una integración) de términos W_t . De (5.41) tenemos

$$W_t - \phi_1 W_{t-1} - \dots - \phi_p W_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (5.42)$$

El orden del polinomio de retraso $\phi_p(B)$, del exponente en el operador diferencia y el orden del polinomio de retraso $\theta_q(B)$, se acostumbra mencionarlo en esa secuencia, de manera que el modelo **ARIMA(p,d,q)** indica que consta de un polinomio autorregresivo de orden p , de una diferencia de orden d y de un polinomio de promedios móviles de orden q .

Algunas Propiedades de un modelo ARIMA óptimo

A continuación se verán las características generales que ha de tener un buen modelo **ARIMA**. Estas características nos servirán para posteriormente poder identificar, estimar y verificar el comportamiento de un modelo **ARIMA** que ha sido optimizado.

Parsimonia (Parquedad). Box y Jenkins ponen énfasis en que la clave de un buen modelo **ARIMA** es que cumpla el principio de parsimonia, que significa sencillez. Así, un modelo se dice que es parsimonioso si se ajusta a la serie de forma adecuada sin usar coeficientes innecesarios.

Estacionariedad. Otra condición de gran importancia para lograr un buen modelo **ARIMA** es que la serie sea estacionaria.

Asumir que una serie sea estacionaria nos permite desarrollar un marco de trabajo bastante simple y usar herramientas estadísticas de muestreo de gran potencia. Así, si la media de un proceso es constante, podremos usar N observaciones para estimarla, mientras que sería mucho más complicado si la media no fuese estacionaria. Si una serie no es estacionaria podemos modificar dicha serie para convertirla en estacionaria.

Buenos coeficientes estimados. Que un modelo tenga unos buenos coeficientes estimados, está relacionado con dos vertientes distintas.

La primera es que los coeficientes, tanto los de la componente autorregresiva (ϕ 's) como los de la componente de media móvil (θ 's) sean significativamente distintos de cero. Esto se realiza mediante contrastes de hipótesis.

La segunda es que las estimaciones de los coeficientes ϕ 's y θ 's no deben estar altamente correlacionadas entre sí. Si están muy correlacionados, tienden a ser inestables, incluso siendo estadísticamente significativos.

Los residuos son ruido blanco. Esta proposición es muy importante a la hora de verificar un modelo **ARIMA**, una vez se han realizado las etapas de identificación y ajuste.

Debe ajustarse bien a los datos. Que un modelo se ajuste todo lo bien posible a los datos de los que es generado, es una hipótesis asumible y lógica. Esta bondad del ajuste se mide en términos de error.

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.6. MODELOS ARIMA

Debe dar buenas predicciones. Aunque el modelo haya sido ajustado y prediga el pasado de una forma suficientemente correcta, lo que realmente se requiere de cualquier modelo de predicción es que realice predicciones satisfactorias. La evaluación de un modelo según este criterio se debe realizar mediante el uso durante un periodo de prueba.

En la Tabla 5.1 se resume el comportamiento de la función de autocorrelación (FAC) y de la función de autocorrelación parcial (FACP) para cada uno de los modelos generales no estacionales que se trataron.

Tabla 5.1.

Modelo	FAC	FACP
Promedio móvil de primer orden $X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1}$	Se corta después del desfase 1; específicamente: $\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$ $\rho_k = 0$ para $k > 1$	Se corta de manera que domina el decremento exponencial amortiguado.
Promedio móvil de segundo orden $X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$	Se corta después del desfase 2; específicamente: $\rho_1 = \frac{-\theta_1(1-\theta_1)}{1+\theta_1^2+\theta_2^2}$ $\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$ $\rho_k = 0$ para $k > 2$	Se corta según una combinación de exponenciales amortiguados, ondas seno amortiguadas, o ambos.
Promedio móvil de orden q $X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$	Se trunca después del desfase q .	Se extingue o muere.
Autorregresivo de primer orden $X_t = \delta + \phi_1 X_{t-1} + \epsilon_t$	Se corta de modo exponencial amortiguado; específicamente: $\rho_k = (\phi_1)^k$ para $k \geq 1$	Se corta después del desfase 1.
Autorregresivo de segundo orden $X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$	Se corta según una combinación exponenciales amortiguados, de senoidales amortiguadas, o ambos; específicamente: $\rho_1 = \frac{\phi_1}{1-\phi_2}$ $\rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2$ $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$ para $k \geq 3$	Se corta después del desfase 2.
Autorregresivo de orden p $X_t = \delta + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t$	Se extingue o muere.	Se trunca después del desfase p .
Combinado de promedio móvil-autorregresivo de orden (1,1)	Se corta de manera exponencial amortiguada; específicamente: $\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\theta_1\phi_1}$ $\rho_k = \phi_1 \rho_{k-1}$ para $k \geq 2$	Se corta de manera que domina el decremento exponencial amortiguado.
Combinado de promedio móvil-autorregresivo de orden (p, q) $\phi_p(B)X_t = \theta_q(B)\epsilon_t$	Se extingue	Se extingue

Para finalizar esta sección en la Tabla 5.2 se resume la forma de proponer un modelo de serie de tiempo para un conjunto de datos analizando las gráficas de la función de autocorrelación muestral (FACM) y de la función de autocorrelación parcial muestral (FACPM) para los modelos generales no estacionales de series de tiempo.

Tabla 5.2.

Modelo	FACM	FACPM
Promedio móvil de primer orden $X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1}$	Se corta después del desfaseamiento 1.	Se corta de manera que domina el decremento exponencial amortiguado.
Promedio móvil de segundo orden $X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$	Se corta después del desfaseamiento 2.	Se corta según una combinación de exponenciales amortiguados, ondas seno amortiguadas, o ambos.
Promedio móvil de orden q $X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$	Se trunca después del desfaseamiento q .	Se extingue o muere.
Autorregresivo de primer orden $X_t = \delta + \phi_1 X_{t-1} + \epsilon_t$	Se corta de modo exponencial amortiguado.	Se corta después del desfaseamiento 1.
Autorregresivo de segundo orden $X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$	Se corta según una combinación de exponenciales amortiguados, de senoidales amortiguadas, o ambos.	Se corta después del desfaseamiento 2.
Autorregresivo de orden p $X_t = \delta + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t$	Se extingue o muere.	Se trunca después del desfaseamiento p .
Combinado de promedio móvil-autorregresivo de orden (1,1)	Se corta de manera exponencial amortiguada.	Se corta de manera que domina el decremento exponencial amortiguado.
Combinado de promedio móvil-autorregresivo de orden (p, q) $\phi_p(B)X_t = \theta_q(B)\epsilon_t$	Se extingue	Se extingue

El Modelo General Estacional

El modelo general de Box-Jenkins de orden (p, P, q, Q) es

$$\phi_p(B)\phi_P(B^L)X_t = \delta + \theta_q(B)\theta_Q(B^L)\epsilon_t$$

en donde B es el operador de retraso y además:

1. $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$
se denomina operador no estacional autorregresivo de orden p .
2. $\phi_P(B^L) = (1 - \phi_{1,L} B^L - \phi_{2,L} B^{2L} - \dots - \phi_{P,L} B^{PL})$
se llama operador estacional autorregresivo de orden P .
3. $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$
es el operador no estacional de media móvil de orden q .
4. $\theta_Q(B^L) = (1 - \theta_{1,L} B^L - \theta_{2,L} B^{2L} - \dots - \theta_{Q,L} B^{QL})$
se denomina como operador estacional de media móvil de orden Q .
5. $\delta = \phi_p(B)\phi_P(B^L)\mu$
es un término constante, donde μ es la media verdadera de la serie de tiempo estacional que está siendo modelada, y $B\mu = \mu$.
6. $\phi_1, \phi_2, \dots, \phi_p; \phi_{1,L}, \phi_{2,L}, \dots, \phi_{P,L}; \theta_1, \theta_2, \dots, \theta_q; \theta_{1,L}, \theta_{2,L}, \dots, \theta_{Q,L};$ y δ son parámetros desconocidos que es necesario estimar partir de los datos de la muestra.

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS

5.7. APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS

7. $\epsilon_t, \epsilon_{t-1}, \dots$ son perturbaciones aleatorias que, se supone, son estadísticamente independientes entre sí; se supone también que cada una fue seleccionada al azar de una distribución normal cuya media es cero, y su varianza es la misma para todos los periodos t .

Mediante el uso de la función de autocorrelación muestral (**FACM**) y la función de autocorrelación parcial muestral (**FACPM**) se puede identificar un modelo de este tipo.

1. Se usan los comportamientos de la función de autocorrelación muestral (**FACM**) y la función de autocorrelación parcial muestral (**FACPM**) de los valores X_1, X_2, \dots, X_n en el nivel no estacional para determinar cuál (si acaso) de los operadores no estacionales se deben utilizar: el operador no estacional **MA**(q) o el operador no estacional **AR**(p).
2. Se usan los comportamientos de la función de autocorrelación y la función de autocorrelación parcial de los valores X_1, X_2, \dots, X_n en el nivel estacional para determinar cuál (si acaso) de los operadores estacionales se deben utilizar: el operador estacional **MA**(Q) o el operador estacional **AR**(P).

5.7. Aplicación de la metodología de Box-Jenkins

Lo que se pretende realizar ahora, dado que se tiene una serie de tiempo para la humedad relativa, es encontrar un modelo para predecir valores futuros de la humedad relativa semanal usando la metodología de Box-Jenkins. Lo que se procede a realizar a continuación, es encontrar algún modelo **AR**, **MA** o **ARIMA** (**ARMA**), que se ajuste a los datos y que haga una buena predicción de los mismos.

Los modelos clásicos de Box-Jenkins describen series de tiempo estacionarias. Por consiguiente, para identificar en forma tentativa un modelo de Box-Jenkins, es necesario determinar primero si la serie de tiempo que se quiere pronosticar es estacionaria. En forma intuitiva, una serie de tiempo es estacionaria si las propiedades estadísticas (por ejemplo, la media y la varianza) de la serie de tiempo son esencialmente constantes a través del tiempo.

Como se puede observar en la Figura 4.2 que los valores originales de la serie de tiempo parecen fluctuar con respecto de una media constante y, por tanto, parece en forma tentativa que estos valores son estacionarios.

Los modelos de pronóstico de Box-Jenkins se identifican en forma tentativa examinando el comportamiento de la función de autocorrelación muestral (**FACM**) y la función de autocorrelación parcial muestral (**FACPM**) para los valores de una serie de tiempo estacionaria.

Por tanto, lo primero que se tiene que revisar es que la serie de tiempo sea estacionaria, esto se hace calculando la media y la función de autocovarianza, y determinando que éstas no dependen del tiempo, es decir, son constantes. Dado que los datos de la humedad relativa semanal con los que se está trabajando es una muestra de observaciones, entonces se requiere usar las ecuaciones (5.3), (5.4), (5.5) y (5.6) del capítulo 5 sección 1, para encontrar la media muestral, la varianza muestral, la autocovarianza muestral y la autocorrelación

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS

5.7. APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS

muestral para determinar si es estacionaria. Este procedimiento es muy complejo, ya que se realizan estos cálculos para diferentes muestras del total de los datos y para cada una se tiene que observar si la media muestral y la función de autocovarianza muestral es constante en el tiempo. Para facilitar esto lo que se propone es observar la FACM del modelo que se requiere usar para los datos de la humedad relativa semanal, la cual se muestra en la Figura 5.1:

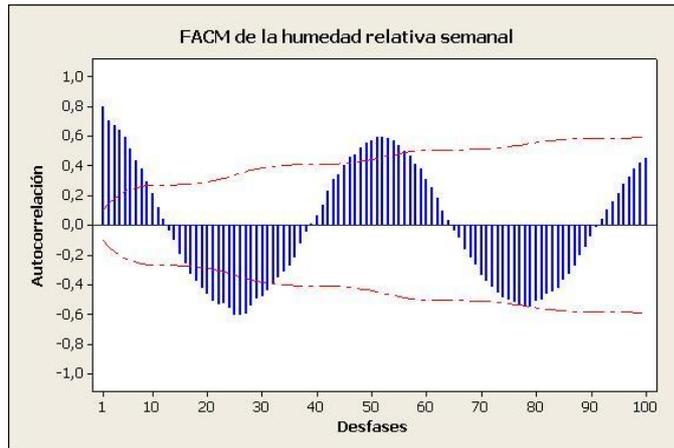


Figura 5.1: Función de autocorrelación muestral de la serie de tiempo.

En ésta se aprecia que la serie no es estacionaria. Para tratar de transformarla en una serie estacionaria, se le calcula una primera diferencia. La función de autocorrelación de la primera diferencia se puede apreciar en la Figura 5.2:

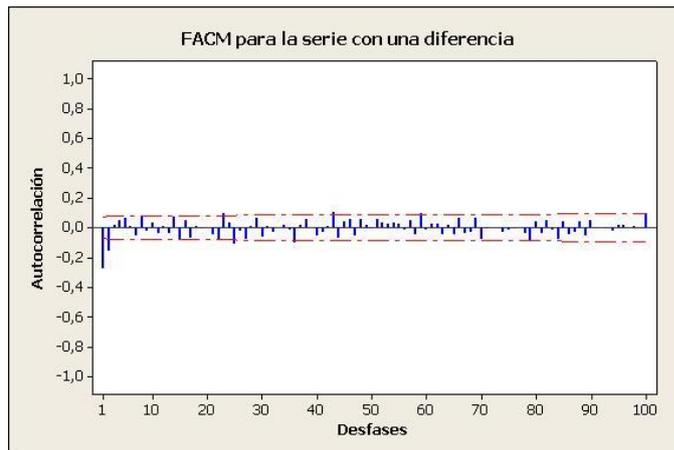


Figura 5.2: Función de autocorrelación muestral de la serie de tiempo con una diferencia.

En esta se observa que con una sola diferencia fue suficiente para lograr la estacionariedad.

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS

5.7. APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS

De la forma que tiene la función de autocorrelación de la primera diferencia, se podría proponer un modelo **MA**(2) para la serie, pues sólo tiene las dos primeras autocorrelaciones significativamente distintas de cero, sin embargo, recordando que es preferible un modelo **AR** sobre un modelo **MA**, ya que con un modelo **AR** nos da mayor certeza de que los valores obtenidos son realmente un óptimo, se observa la gráfica de la Función de autocorrelación parcial muestral de la primera diferencia (ver Figura 5.3) y ésta sugiere un modelo **AR**(3) o **AR**(2).

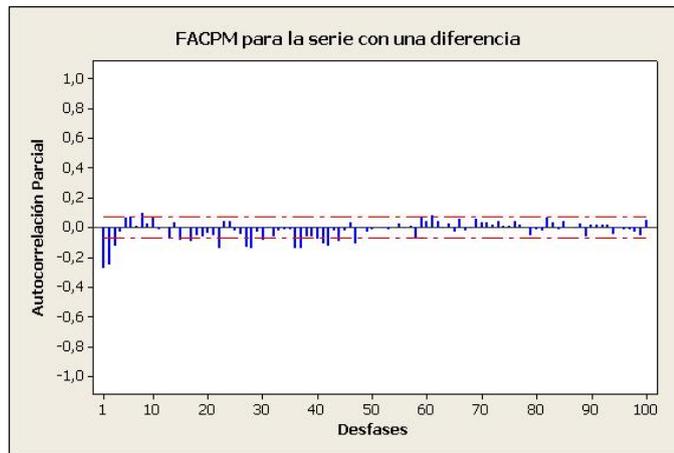


Figura 5.3: Función de autocorrelación parcial muestral de la serie de tiempo con una diferencia.

A continuación, en la Figura 5.4 se muestra la gráfica de la primera diferencia para la humedad relativa:

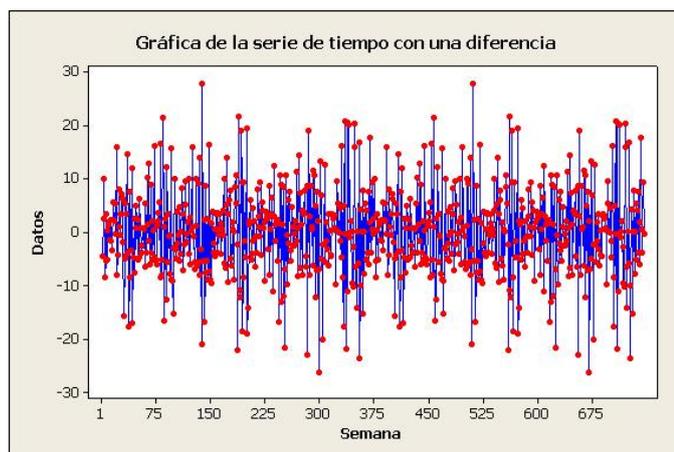


Figura 5.4: Serie de tiempo de la humedad con una diferencia.

El problema que se tiene ahora es encontrar qué modelo es el que se ajusta a los

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS

5.7. APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS

datos de la serie de tiempo con una diferencia. Siguiendo los pasos del método para series estacionales se tratará de encontrar ese modelo.

Primero se identificará un modelo no estacional que describa los valores de la serie de tiempo.

Al observar la función de autocorrelación muestral se tiene que hay dos espigas en los desfases 1 y 2 (ver Figura 5.2). La función de autocorrelación parcial muestral sugiere un modelo $AR(3)$ o $AR(2)$ (ver Figura 5.3).

Como segundo paso se debe identificar un modelo estacional que describa los valores de la serie de tiempo.

Para identificar qué modelo se puede determinar, se realizó una diferencia de un periodo, es decir, $L=52$ (ver Figura 5.5). La función de autocorrelación muestral se corta después del segundo desfase en el nivel estacional (ver Figura 5.6). Al observar la función de autocorrelación parcial en el nivel estacional se tiene que esta se corta después del quinto desfase (ver Figura 5.7).

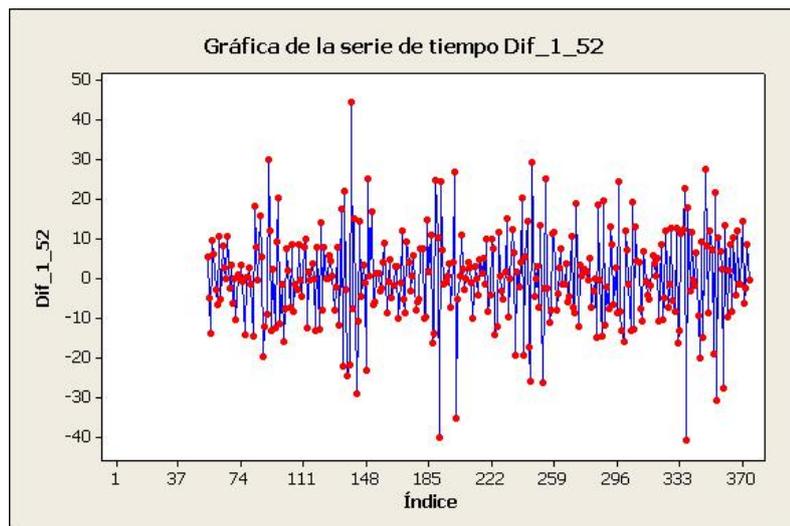


Figura 5.5: Gráfica de la serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional.

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.7. APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS

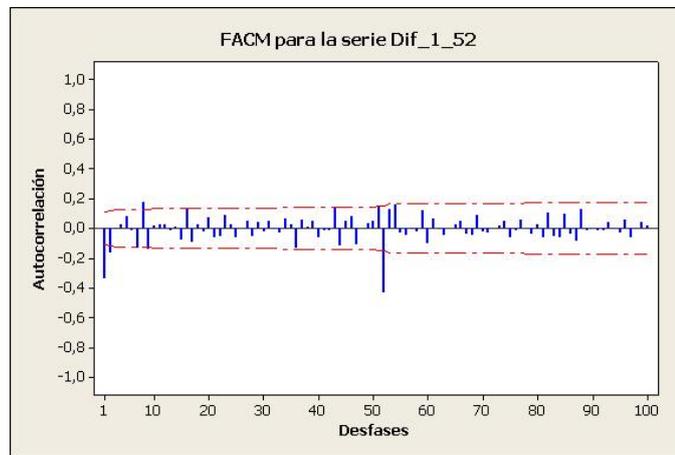


Figura 5.6: Función de autocorrelación muestral de la serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional.

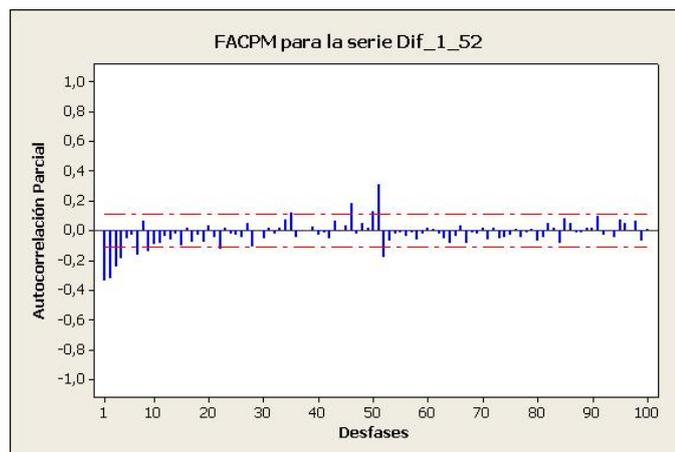


Figura 5.7: Función de autocorrelación parcial muestral de la serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional.

Como tercer paso se tienen que combinar todos los modelos anteriores, se determinaron algunos modelos **ARIMA**(2,1,1)(1,1,1), **ARIMA**(2,1,1)(1,1,0), **ARIMA**(2,1,0)(1,1,0), **ARIMA**(2,1,2)(0,1,1), **ARIMA**(1,1,2)(0,1,1) y **ARIMA**(0,1,2)(0,1,1) encontrándose que el mejor modelo fue el modelo **ARIMA**(0,1,2)(0,1,1).

Entonces tenemos:

$$X_t = y_t - y_{t-1}.$$

Por el primer paso tenemos:

$$X_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2}$$

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.7. APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS

y por el segundo paso tenemos:

$$X_t^* = X_t - X_{t-52}$$

$$X_t^* = \theta_{1,52} X_{t-52}^* + \epsilon_t$$

entonces la combinación es

$$X_t = X_{t-52} + \theta_{1,52}(X_t - X_{t-52}) + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$

Con la ayuda del software estadístico MINITAB se obtuvieron los valores de los parámetros del modelo anterior. Para verificar qué modelo **ARIMA** fue el que mejor se ajusta a los datos, fue necesario observar en cada una de las pruebas lo siguiente: el Criterio de información de Akaike (ver Apéndice E), los p -valores de la estadística de Ljung-Box (ver Apéndice D) además de los p -valores de la estimación de cada parámetro.

Se revisarán primero las estimaciones de los parámetros del modelo **ARIMA(0,1,2)(0,1,1)**, éstos se muestran en la Tabla 5.3:

Tabla 5.3: Análisis de la humedad relativa usando el modelo estacional **ARIMA(0,1,2)(0,1,1)**.

Final Estimates of Parameters					
Type		Coef	SE Coef	T	P
MA	1	0,6384	0,0544	11,74	0,000
MA	2	0,2593	0,0555	4,67	0,000
SMA	52	0,7951	0,0558	14,26	0,000

Differencing: 1 regular, 1 seasonal of order 52					
Number of observations: Original series 374, after differencing 321					
Residuals: SS = 16897,9 (backforecasts excluded)					
MS = 53,1 DF = 318					

Para comparar la significancia estadística de los coeficientes se puede ver la última columna de esta tabla (valor p) y se concluye que los tres coeficientes son significativamente diferentes de cero con un nivel de significancia de $\alpha = 0.05$.

A continuación se muestra en la Tabla 5.4, los resultados del estadístico de Ljung-Box para el modelo **ARIMA(0,1,2)(0,1,1)**, en donde se observa, que no hay evidencia en contra del supuesto del modelo sobre las autocorrelaciones del error, que por ser ruido blanco, deberían ser todas cero para cualquier retraso. Se puede apreciar que (ver último renglón de la Tabla 5.4) con este estadístico, aún en retrasos grandes para el modelo **ARIMA(0,1,2)(0,1,1)**, no hay evidencia en contra de las hipótesis de que las autocorrelaciones son todas cero.

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS
5.7. APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS

Tabla 5.4: Análisis de la humedad relativa usando el modelo estacional $ARIMA(0,1,2)(0,1,1)$.

Modified Box-Pierce (Ljung-Box) Chi-Square statistic				
Lag	12	24	36	48
Chi-Square	13,0	30,6	42,0	58,8
DF	9	21	33	45
P-Value	0,162	0,081	0,136	0,081

Una vez que ya se obtuvo un modelo que se ajusta a los datos de la humedad relativa, mediante este modelo se procede a realizar pronósticos de datos de aproximadamente cuatro semanas en la que se pueda observar el comportamiento de cada uno de ellos.

Los resultados presentados en la Tabla 5.5 son referentes a los valores pronosticados mediante el uso del modelo estacional $ARIMA(0,1,2)(0,1,1)$. En la Figura 5.8, se presentan estos pronósticos con su banda de confianza.

Tabla 5.5: Pronósticos del modelo de la humedad.

Datos originales	Pronósticos hechos en MINITAB
58.3486	63.7195
57.6814	67.8229
48.3386	66.1412
48.5350	68.4621

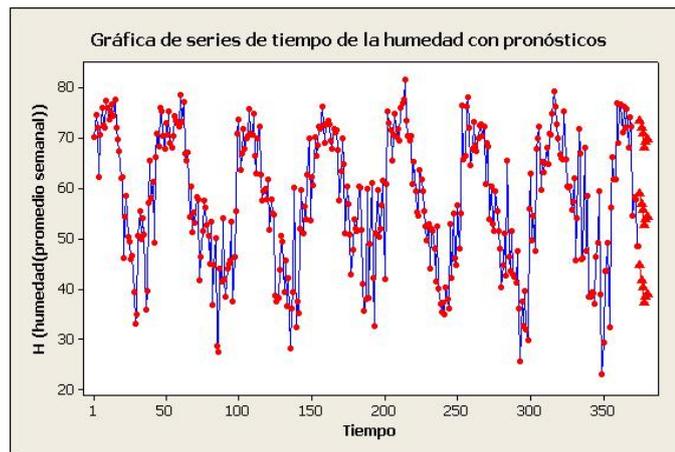


Figura 5.8: Pronósticos del modelo $ARIMA(0,1,2)(0,1,1)$.

Se puede decir, que el modelo opera adecuadamente, pues cubre, dentro de la banda de 95 % de confianza los valores observados las últimas cuatro semanas. Además no hay

CAPÍTULO 5. METODOLOGÍA DE BOX-JENKINS

5.7. APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS

diferencias considerables entre el valor observado y el pronóstico.

Anteriormente, la prueba de Ljung-Box mostró que no hay evidencia contra el supuesto de que las autocorrelaciones del término de error son cero (son variables aleatorias independientes). Se revisarán ahora aspectos adicionales sobre los residuales del modelo elegido **ARIMA**(0,1,2)(0,1,1).

Para la validación del modelo se observa la gráfica de autocorrelación de los residuales del modelo estimado (**ARIMA**(0,1,2)(0,1,1)). (ver Figura 5.9)

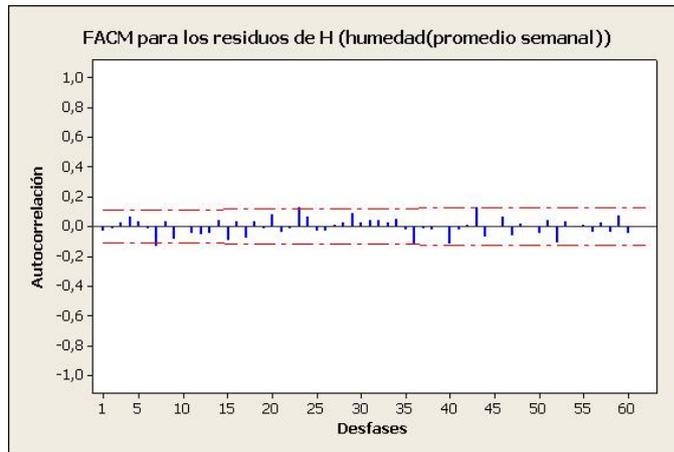


Figura 5.9: Función de autocorrelación muestral de los residuos para el modelo **ARIMA**(0,1,2)(0,1,1).

Los residuales son valores aleatorios independientes. Se observan fluctuando sobre el cero, lo cual es consistente con la hipótesis de los modelos **ARIMA** de que el término aleatorio tiene media cero.

Por lo tanto no hay evidencias en contra de los supuestos del modelo. El modelo **ARIMA**(0,1,2)(0,1,1) es una buena alternativa para hacer pronósticos en la serie de tiempo de la humedad relativa.

Capítulo 6

Discusión y conclusiones

El trabajo realizado en esta tesis propone utilizar bases de datos de fenómenos meteorológicos que influyen día a día en nuestro medio ambiente para introducir conocimientos de estadística que puedan ser utilizados en estos y tratar de determinar que pasa con estos fenómenos, en este caso se estudió principalmente a la humedad relativa.

Lo primero que se realizó fue tratar de encontrar mediante el análisis de regresión lineal simple un modelo que tomara como variable respuesta a la humedad relativa y como variables predictoras a las demás variables meteorológicas, de los diferentes análisis de regresión lineal simple para los datos de la humedad se encontró que el modelo que mejor se ajustó a los datos de humedad relativa fue aquel que tiene como variable predictora a la variable punto rocío (PR). Además se realizó un análisis de regresión lineal múltiple usando los métodos de los mejores subconjuntos y la regresión stepwise, encontrándose que el mejor modelo para pronosticar la humedad diaria es aquel que contiene las variables meteorológicas PR, TMax, HH, VM, TMin, LL, VV, RV, ST, y TA. Sin embargo según el principio de la navaja de Occan, el mejor modelo es aquel en el que sólo aparece como variable predictora la PR.

Se realizó un ajuste tomando como método el modelo con variables ficticias y se encontró que un buen modelo es aquel que elimina el término del tiempo para pronosticar a la humedad relativa.

Debido a que se requería encontrar un modelo que se ajustara a los datos y posteriormente hacer predicciones sobre éste, se descartó el uso del análisis de regresión. El análisis de regresión no funcionaría en el sentido de que para saber el valor de un dato futuro de humedad relativa es necesario saber cual es el valor de ese día para cada variable meteorológica, lo cual no facilita el manejo de este modelo.

Se mencionó que la serie de tiempo de la humedad relativa se pudo descomponer en los diferentes tipos de variación (tendencia, estacionalidad, etc.). Se realizó un análisis de los registros de la humedad relativa usando un modelo de descomposición, los datos de la humedad fueron tomados cada 30 minutos, con las fechas del 9 de Mayo del 2003 al 28 de Octubre del 2010. Lo primero que se realizó, fue obtener

CAPÍTULO 6. DISCUSIÓN Y CONCLUSIONES

promedios diarios de los registros de la humedad y posteriormente se obtuvieron promedios de cada estación del año (es decir promedios del verano, otoño, invierno y primavera), los datos obtenidos se graficaron y la gráfica aparte de mostrar una tendencia lineal exhibe variación estacional constante, por lo que se realizó un análisis de los registros de la humedad usando un modelo de descomposición aditivo.

Una vez que se conoce la parte teórica acerca del tema de la suavización exponencial, se procedió a aplicar este procedimiento a la base de datos. Se llevaron a cabo tres análisis usando los métodos de la suavización exponencial, el primero en aplicarse fue el método de la suavización exponencial simple, para los datos de la humedad tomados como promedios diarios, luego se aplicó el método de Holt, de la suavización exponencial corregida de la tendencia, para los datos de la humedad ahora tomados como promedios semanales, luego por último se aplicó el método multiplicativo de Holt-Winters, para los datos de la humedad tomados como promedios de cada estación, es decir, promedios del verano, otoño, invierno y primavera.

Como el objetivo de cada método es encontrar buenos valores de pronóstico, se seleccionan los valores que minimizan la suma de cuadrados del error de pronóstico (SSE), en las Tablas 4.2, 4.6 y 4.10 se observa que el método multiplicativo de Holt-Winters arroja una SEE menor que en los demás métodos y en consecuencia se obtienen intervalos de predicción más cortos, es decir más exactos, por lo tanto, el método multiplicativo de Holt-Winters es el mejor para realizar pronóstico de la serie de tiempo observada.

Finalmente después de presentar la parte teórica acerca del tema de los modelos autorregresivos de promedios móviles e integrados (**ARIMA**), se aplica esta metodología para determinar un modelo adecuado a los datos de la humedad relativa que permita obtener buenas predicciones. Como primera parte fué importante analizar si esta serie era estacionaria, ya que esta propiedad permite aplicar el método de Box-Jenkins para obtener el modelo que mejor explique los datos (esto se hace a través de la media y la **FAC**). Dado que la serie en principio no es estacionaria se procedió a realizar una primera diferencia para encontrar un modelo **ARIMA** que se ajustará a los datos.

El modelo que se obtuvo para ajustar a los datos de la humedad relativa fue un **ARIMA**(0,1,2)(0,1,1) representado por la siguiente ecuación:

$$X_t = X_{t-52} + \theta_{1,52}(X_t - X_{t-52}) + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2}$$

donde $\{\epsilon_t\} \sim RB(0, \sigma^2)$ y los parámetros son $\theta_1 = 0.63$, $\theta_2 = 0.25$ y $\theta_{1,52} = 0.79$. Este modelo se obtuvo a través de diferentes tipos de análisis, como lo son: la observación de la **FACM** de la serie para determinar si para ésta se podría utilizar un modelo **ARIMA**; mediante la observación del valor del Criterio de información de Akaike y el valor del error medio cuadrático se encontró este modelo de entre algunos que probablemente se ajustarían a los datos, una vez encontrado el modelo se determinó que la serie es estacionaria; los p -valores del estadístico de Ljung Box resultaron ser grandes, tanto para no rechazar la hipótesis nula (los coeficientes de la **FAC** del modelo están no correlacionados) lo cual indica que los residuales son independientes y aleatorios.

CAPÍTULO 6. DISCUSIÓN Y CONCLUSIONES

Se obtuvo así un modelo que se ajusta a los datos y que ayudó a pronosticar datos futuros de una semana.

Se presentan a continuación las conclusiones obtenidas como parte de los objetivos presentados al inicio de este trabajo.

- Se realizaron varios análisis de regresión lineal simple para los datos y se encontró un modelo que mejor se ajustó a los datos de la humedad relativa y fue aquel que tiene como variable predictora al punto rocío (PR).
- Se realizó un análisis de los registros de la humedad determinándose un modelo de descomposición aditivo.
- De los tres análisis aplicados a los métodos de la suavización exponencial: el método de la suavización exponencial simple, el método de Holt de la suavización exponencial corregida de la tendencia y el método multiplicativo de Holt-Winters. Se observa que el método multiplicativo de Holt-Winters es el mejor para realizar pronóstico de la serie de tiempo observada.
- Se estableció un modelo estadístico paramétrico de series de tiempo que describe a la humedad relativa. Determinándose que el modelo que mejor se ajusta a los datos es el modelo **ARIMA**(0,1,2)(0,1,1) y con dicho modelo se realizaron predicciones de aproximadamente una semana obteniéndose buenas predicciones al compararlas con los datos reales.
- Mediante los resultados presentados en este trabajo se pretende contribuir al análisis del estudio climático así como determinar las consecuencias en la variación del clima.
- Para este trabajo se requirió el uso de los siguientes software: MINITAB y Excel. El uso de estos programas fué de gran importancia para la obtención de los resultados en el análisis de los datos.

Un posible trabajo futuro puede consistir en efectuar un análisis para determinar predicciones usando métodos de redes neuronales y comparar estas predicciones con las obtenidas al usar el método de Box-Jenkins para series de tiempo.

Apéndice A

Anexo de datos

Tabla 1: Promedios diarios de la humedad.

Tiempo (días)	H (humedad)						
1	42,65	26	78,92	51	71,94	76	70,54
2	55,88	27	76,15	52	76,33	77	69,96
3	49,38	28	69,88	53	74,94	78	63,42
4	49,44	29	68,15	54	82,85	79	75,92
5	45,94	30	68,67	55	72,27	80	80,81
6	42,52	31	72,48	56	70,04	81	72,15
7	43,73	32	78,90	57	73,77	82	79,15
8	43,73	33	72,67	58	76,38	83	73,35
9	42,71	34	72,52	59	75,98	84	70,21
10	33,90	35	70,52	60	74,50	85	79,83
11	29,81	36	69,06	61	80,13	86	77,90
12	34,42	37	69,00	62	73,06	87	68,90
13	44,46	38	72,46	63	65,65	.	.
14	48,25	39	74,17	64	59,17	.	.
15	49,08	40	82,27	65	57,98	.	.
16	52,88	41	78,73	66	54,77	2648	57,31
17	60,77	42	71,29	67	57,19	2649	51,81
18	59,69	43	76,77	68	69,67	2650	48,96
19	58,48	44	77,50	69	73,31	2651	49,58
20	58,75	45	78,83	70	64,63	2652	52,48
21	56,04	46	72,54	71	57,23	2653	43,23
22	55,98	47	68,56	72	68,90	2654	43,27
23	55,63	48	68,94	73	73,42	2655	49,04
24	57,35	49	61,77	74	76,88	2656	48,40
25	72,75	50	62,25	75	71,31	2657	48,67

APÉNDICE A. ANEXO DE DATOS

Tabla 2: Promedio por estación del año de la Humedad.

Año	t	Estacion	H(humedad(promedio estacional))
1	1	1 (Verano)	72,99
	2	2 (Otoño)	60,42
	3	3 (Invierno)	47,62
	4	4 (Primavera)	67,97
2	5	1 (Verano)	70,96
	6	2 (Otoño)	53,30
	7	3 (Invierno)	42,58
	8	4 (Primavera)	58,28
3	9	1 (Verano)	67,10
	10	2 (Otoño)	49,54
	11	3 (Invierno)	42,81
	12	4 (Primavera)	64,76
4	13	1 (Verano)	69,26
	14	2 (Otoño)	55,53
	15	3 (Invierno)	47,04
	16	4 (Primavera)	65,37
5	17	1 (Verano)	68,97
	18	2 (Otoño)	53,10
	19	3 (Invierno)	42,05
	20	4 (Primavera)	65,74
6	21	1 (Verano)	65,66
	22	2 (Otoño)	49,42
	23	3 (Invierno)	40,78
	24	4 (Primavera)	64,03
7	25	1 (Verano)	68,94
	26	2 (Otoño)	55,49
	27	3 (Invierno)	41,24
	28	4 (Primavera)	66,91

Tabla 3: Medias móviles y medias móviles centradas para la Humedad(promedio estacional).

H(humedad(promedio estacional))	Media movil de 4 periodos	CMA _{t=t-t+c}
72,99		
60,42		
47,62	62,25321237	61,9988172
67,97	61,74442204	60,85329433
70,96	59,96216661	59,33160209
53,30	58,70103758	57,48974682
42,58	56,27845605	55,7966281
58,28	55,31480014	54,84585992
67,10	54,37691971	54,40548153
49,54	54,43404336	55,24435586
42,81	56,05466836	56,32454739
64,76	56,59442643	57,34242914
69,26	58,09043186	58,61993455
55,53	59,14943724	59,22470898
47,04	59,29998072	59,26335437
65,37	59,22672803	58,92319542
68,97	58,61966281	57,99491819
53,10	57,37017356	57,41695345
42,05	57,46373335	57,04949947
65,74	56,6352656	56,17560528
65,66	55,71594495	55,55775947
49,42	55,39957398	55,18623159
40,78	54,9728892	55,38314458
64,03	55,79339995	56,55243528
68,94	57,31147061	57,36870179
55,49	57,42593297	57,78499547
41,24	58,14405797	
66,91		

APÉNDICE A. ANEXO DE DATOS

Tabla 4: Valores de $s_t + i_t$ para la Humedad(promedio estacional).

H(humedad(promedio estacional))	Media movil de 4 periodos	CMA_t=t_t+c_t	s_t+i_t=y_t-(t_t+c_t)
72,99			
60,42			
47,62	62,25321237	61,9988172	-14,37
67,97	61,74442204	60,85329433	7,12
70,96	59,96216661	59,33160209	11,63
53,30	58,70103758	57,48974682	-4,19
42,58	56,27845605	55,7966281	-13,22
58,28	55,31480014	54,84585992	3,43
67,10	54,37691971	54,40548153	12,70
49,54	54,43404336	55,24435586	-5,70
42,81	56,05466836	56,32454739	-13,52
64,76	56,59442643	57,34242914	7,42
69,26	58,09043186	58,61993455	10,64
55,53	59,14943724	59,22470898	-3,70
47,04	59,29998072	59,26335437	-12,22
65,37	59,22672803	58,92319542	6,44
68,97	58,61966281	57,99491819	10,97
53,10	57,37017356	57,41695345	-4,32
42,05	57,46373335	57,04949947	-15,00
65,74	56,6352656	56,17560528	9,56
65,66	55,71594495	55,55775947	10,10
49,42	55,39957398	55,18623159	-5,76
40,78	54,9728892	55,38314458	-14,60
64,03	55,79339995	56,55243528	7,48
68,94	57,31147061	57,36870179	11,57
55,49	57,42593297	57,78499547	-2,29
41,24	58,14405797		
66,91			

Tabla 5: Valores de s_t para la Humedad(promedio estacional).

Estaciones del año	Año 1	Año 2	Año 3	Año 4	Año 5	Año 6	Promedio de s_t
1 Primavera	7,11757524	3,434683555	7,42061433	6,442021973	9,563851245	7,48028211	6,909838077
2 Verano	11,62656995	12,69806685	10,6426461	10,97465171	10,09793946	11,5690401	11,26815237
3 Otoño	-4,19463812	-5,7007689	-3,69710028	-4,31760563	-5,76416637	-2,2906476	-4,327487825
4 Invierno	-14,3743011	-13,2166281	-13,5160528	-12,2188382	-15,0029403	-14,602069	-13,82180497
						SUMA	0,028697649
L							
4						SUMA/L	0,007174412
						s_t=Promedio de s_t(SUMA/L)	6,902663664
							11,26097796
							-4,334662238
							-13,82897938

APÉNDICE A. ANEXO DE DATOS

Tabla 6: Observaciones compensadas respecto a la variación estacional calculadas por medio de Excel.

H(humedad(promedio estacional))	Media movil de 4 periodos	CMA_t=t_t+c_t	s_t+i_t=y_t-(t_t+c_t)	s_t	d_t=y_t-s_t
72,99				11,260978	61,73
60,42				-4,33466224	64,76
47,62	62,25321237	61,9988172	-14,37	-13,8289794	61,45
67,97	61,74442204	60,85329433	7,12	6,90266366	61,07
70,96	59,96216661	59,33160209	11,63	11,260978	59,70
53,30	58,70103758	57,48974682	-4,19	-4,33466224	57,63
42,58	56,27845605	55,7966281	-13,22	-13,8289794	56,41
58,28	55,31480014	54,84585992	3,43	6,90266366	51,38
67,10	54,37691971	54,40548153	12,70	11,260978	55,84
49,54	54,43404336	55,24435586	-5,70	-4,33466224	53,88
42,81	56,05466836	56,32454739	-13,52	-13,8289794	56,64
64,76	56,59442643	57,34242914	7,42	6,90266366	57,86
69,26	58,09043186	58,61993455	10,64	11,260978	58,00
55,53	59,14943724	59,22470898	-3,70	-4,33466224	59,86
47,04	59,29998072	59,26335437	-12,22	-13,8289794	60,87
65,37	59,22672803	58,92319542	6,44	6,90266366	58,46
68,97	58,61966281	57,99491819	10,97	11,260978	57,71
53,10	57,37017356	57,41695345	-4,32	-4,33466224	57,43
42,05	57,46373335	57,04949947	-15,00	-13,8289794	55,88
65,74	56,6352656	56,17560528	9,56	6,90266366	58,84
65,66	55,71594495	55,55775947	10,10	11,260978	54,39
49,42	55,39957398	55,18623159	-5,76	-4,33466224	53,76
40,78	54,9728892	55,38314458	-14,60	-13,8289794	54,61
64,03	55,79339995	56,55243528	7,48	6,90266366	57,13
68,94	57,31147061	57,36870179	11,57	11,260978	57,68
55,49	57,42593297	57,78499547	-2,29	-4,33466224	59,83
41,24	58,14405797			-13,8289794	55,07
66,91				6,90266366	60,00

Tabla 7: Analisis de regresión lineal simple obtenido con Excel.

Resumen								
<i>Estadísticas de la regresión</i>								
Coefficiente de correlación múltiple	0,33991566							
Coefficiente de determinación R ²	0,11554265							
R ² ajustado	0,08152506							
Error típico	2,79348009							
Observaciones	28							
ANÁLISIS DE VARIANZA								
	<i>Grados de libertad</i>		<i>de cuadrado de los cua</i>		<i>F</i>	<i>valor crítico de F</i>		
Regresión	1	26,505131	26,505131	3,39655612	0,07676272			
Residuos	26	202,891806	7,803531					
Total	27	229,396937						
	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95,0%</i>	<i>superior 95,0%</i>
Intercepción	59,5274801	1,08476861	54,8757401	2,0618E-28	57,2977063	61,7572539	57,2977063	61,7572539
Variable X 1	-0,12044692	0,06535462	-1,8429748	0,07676272	-0,25478527	0,01389143	-0,25478527	0,01389143

APÉNDICE A. ANEXO DE DATOS

Tabla 8: Valores de t_t y \hat{y}_t calculados con Excel.

Año	Estacion	t	H(humedad(promedio estacional))	s_t+i_t-y_t-(t-t+c_t)	s_t	d_t=y_t-s_t	t_t=59,5284-0,1205t	y^_t=t_t+s_t
1	1 (Verano)	1	72,99		11,260978	61,73	59,41	70,67
	2 (Otoño)	2	60,42		-4,33466224	64,76	59,29	54,95
	3 (Invierno)	3	47,62	-14,37	-13,8289794	61,45	59,17	45,34
	4 (Primavera)	4	67,97	7,12	6,90266366	61,07	59,05	65,95
2	1 (Verano)	5	70,96	11,63	11,260978	59,70	58,93	70,19
	2 (Otoño)	6	53,30	-4,19	-4,33466224	57,63	58,81	54,47
	3 (Invierno)	7	42,58	-13,22	-13,8289794	56,41	58,68	44,86
	4 (Primavera)	8	58,28	3,43	6,90266366	51,38	58,56	65,47
3	1 (Verano)	9	67,10	12,70	11,260978	55,84	58,44	69,70
	2 (Otoño)	10	49,54	-5,70	-4,33466224	53,88	58,32	53,99
	3 (Invierno)	11	42,81	-13,52	-13,8289794	56,64	58,20	44,37
	4 (Primavera)	12	64,76	7,42	6,90266366	57,86	58,08	64,99
4	1 (Verano)	13	69,26	10,64	11,260978	58,00	57,96	69,22
	2 (Otoño)	14	55,53	-3,70	-4,33466224	59,86	57,84	53,51
	3 (Invierno)	15	47,04	-12,22	-13,8289794	60,87	57,72	43,89
	4 (Primavera)	16	65,37	6,44	6,90266366	58,46	57,60	64,50
5	1 (Verano)	17	68,97	10,97	11,260978	57,71	57,48	68,74
	2 (Otoño)	18	53,10	-4,32	-4,33466224	57,43	57,36	53,02
	3 (Invierno)	19	42,05	-15,00	-13,8289794	55,88	57,24	43,41
	4 (Primavera)	20	65,74	9,56	6,90266366	58,84	57,12	64,02
6	1 (Verano)	21	65,66	10,10	11,260978	54,39	57,00	68,26
	2 (Otoño)	22	49,42	-5,76	-4,33466224	53,76	56,88	52,54
	3 (Invierno)	23	40,78	-14,60	-13,8289794	54,61	56,76	42,93
	4 (Primavera)	24	64,03	7,48	6,90266366	57,13	56,64	63,54
7	1 (Verano)	25	68,94	11,57	11,260978	57,68	56,52	67,78
	2 (Otoño)	26	55,49	-2,29	-4,33466224	59,83	56,40	52,06
	3 (Invierno)	27	41,24		-13,8289794	55,07	56,27	42,45
	4 (Primavera)	28	66,91		6,90266366	60,00	56,15	63,06

Tabla 9: Valores de $c_t + i_t$ calculados con Excel.

Año	t	Estacion	H(humedad(promedio estacional))	s_t	t_t=59,5284-0,1205t	y^_t=t_t+s_t	c_t+i_t
1	1	1 (Verano)	72,99	11,26097796	59,41	70,67	2,32
		2 2 (Otoño)	60,42	-4,334662238	59,29	54,95	5,47
		3 3 (Invierno)	47,62	-13,82897938	59,17	45,34	2,29
		4 4 (Primavera)	67,97	6,902663664	59,05	65,95	2,02
2		5 1 (Verano)	70,96	11,26097796	58,93	70,19	0,77
		6 2 (Otoño)	53,30	-4,334662238	58,81	54,47	-1,18
		7 3 (Invierno)	42,58	-13,82897938	58,68	44,86	-2,28
		8 4 (Primavera)	58,28	6,902663664	58,56	65,47	-7,19
3		9 1 (Verano)	67,10	11,26097796	58,44	69,70	-2,60
		10 2 (Otoño)	49,54	-4,334662238	58,32	53,99	-4,45
		11 3 (Invierno)	42,81	-13,82897938	58,20	44,37	-1,57
		12 4 (Primavera)	64,76	6,902663664	58,08	64,99	-0,22
4		13 1 (Verano)	69,26	11,26097796	57,96	69,22	0,04
		14 2 (Otoño)	55,53	-4,334662238	57,84	53,51	2,02
		15 3 (Invierno)	47,04	-13,82897938	57,72	43,89	3,15
		16 4 (Primavera)	65,37	6,902663664	57,60	64,50	0,86
5		17 1 (Verano)	68,97	11,26097796	57,48	68,74	0,23
		18 2 (Otoño)	53,10	-4,334662238	57,36	53,02	0,07
		19 3 (Invierno)	42,05	-13,82897938	57,24	43,41	-1,36
		20 4 (Primavera)	65,74	6,902663664	57,12	64,02	1,72
6		21 1 (Verano)	65,66	11,26097796	57,00	68,26	-2,60
		22 2 (Otoño)	49,42	-4,334662238	56,88	52,54	-3,12
		23 3 (Invierno)	40,78	-13,82897938	56,76	42,93	-2,15
		24 4 (Primavera)	64,03	6,902663664	56,64	63,54	0,49
7		25 1 (Verano)	68,94	11,26097796	56,52	67,78	1,16
		26 2 (Otoño)	55,49	-4,334662238	56,40	52,06	3,43
		27 3 (Invierno)	41,24	-13,82897938	56,27	42,45	-1,21
		28 4 (Primavera)	66,91	6,902663664	56,15	63,06	3,85

APÉNDICE A. ANEXO DE DATOS

Tabla 10: Valores de c_t e i_t calculados con Excel.

c_t+i_t	c_t	i_t
2,32		
5,47	1,82379756	3,65
2,29	0,7621985	1,52
2,02	0,6739353	1,35
0,77	0,25709803	0,51
-1,18	-0,39187636	-0,78
-2,28	-0,75864021	-1,52
-7,19	-2,39550673	-4,79
-2,60	-0,86710986	-1,73
-4,45	-1,48171694	-2,96
-1,57	-0,52180866	-1,04
-0,22	-0,07400673	-0,15
0,04	0,01323423	0,03
2,02	0,67362364	1,35
3,15	1,05086517	2,10
0,86	0,28738458	0,57
0,23	0,07623065	0,15
0,07	0,02487002	0,05
-1,36	-0,45445383	-0,91
1,72	0,57279762	1,15
-2,60	-0,86772634	-1,74
-3,12	-1,04022418	-2,08
-2,15	-0,71561512	-1,43
0,49	0,16455124	0,33
1,16	0,38695466	0,77
3,43	1,14453669	2,29
-1,21	-0,40233196	-0,80
3,85		

Apéndice B

Suavización exponencial

Sea y_1, y_2, \dots, y_n una serie de tiempo, en la cual se supone que no tiene tendencia, es decir, la serie es localmente constante, más un comportamiento irregular. Se utiliza como predicción de los valores futuros, a partir del origen n el promedio de la serie en el instante $T = n$.

Note que si la serie de tiempo y_1, y_2, \dots, y_n consiste de una muestra aleatoria de tamaño n , entonces es natural pensar en un promedio muestral de las observaciones que da igual peso a cada observación. Sin embargo, en el contexto de series de tiempo, es razonable dar un mayor peso a las observaciones más recientes que las observaciones del pasado más remoto. Una manera de lograr esto es dando ponderaciones distintas a las observaciones a través de un promedio ponderado, cuyos pesos decaen geoméricamente. De este modo el nivel medio de la serie ($\bar{\ell}_T$) en el instante T es estimado por:

$$\bar{\ell}_T = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots \quad (\text{B.1})$$

con $0 < \alpha < 1$ y $1 \leq T \leq n$.

Luego la predicción de ℓ_{T+k} es la estimación del nivel en el instante $T+k$. Por ejemplo, si se denota por $\hat{y}_{T+1}(T)$ el valor estimado de la serie en el instante T , entonces

$$\ell_T = \hat{y}_{T+1}(T) = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots \quad (\text{B.2})$$

con $0 < \alpha < 1$ y $1 \leq T \leq n$.

Note que la ecuación (B.2), también puede ser aplicada para aquellos puntos en el segmento observado de la serie, es decir, para $1 \leq T \leq n$. También note que la ecuación (B.2) contiene una cantidad infinita de términos, pero en la práctica sólo se dispone de un número finito de ellos. Una manera de representar (B.2) en una forma más compacta es la siguiente

$$\ell_T = \alpha y_T + (1 - \alpha)\ell_{T-1} \quad (\text{B.3})$$

con $0 < \alpha < 1$ y $1 \leq T \leq n$.

APÉNDICE B. SUAVIZACIÓN EXPONENCIAL

Observación B.1. *Los pesos $\{\alpha_i\}$ decaen en forma geométrica. En efecto*

$$\sum_{i=0}^{\infty} \alpha(1-\alpha)^i = 1.$$

Además, como $\alpha(1-\alpha)^i \leq \alpha(1-\alpha)^j$, si $i \geq j$, entonces los pesos asignan más importancia al pasado reciente de la serie y menos importancia al pasado remoto de la serie satisfaciendo nuestro requerimiento inicial.

Observación B.2. *Si $\hat{y}_1(0) = \ell_0$, entonces es posible usar la ecuación (B.3) de manera recursiva para generar una nueva serie que contendrá los valores predichos asociados a la serie original. La forma de inicializar la serie de valores predichos no es única. Otra manera simple de inicializar la serie es considerando $\hat{y}_1(0)$ como el promedio de las primeras observaciones de la serie (por ejemplo, considerando el primer periodo o el 10% de los valores observados).*

El procedimiento definido por la ecuación (B.3) se denomina **suavización exponencial simple**. La constante α se llama constante de suavización, la cual debe elegirse usando algún criterio de optimalidad.

Considere nuevamente la ecuación (B.3),

$$\ell_T = \alpha y_T + (1-\alpha)\ell_{T-1}$$

$$\ell_T = \alpha y_T + \ell_{T-1} - \alpha \ell_{T-1}$$

$$\ell_T = \alpha(y_T - \ell_{T-1}) + \ell_{T-1}$$

$$\ell_T = \alpha(e_T) + \ell_{T-1},$$

donde $e_T = y_T - \ell_{T-1}$, e_T se denomina error de predicción a un paso en el tiempo T y mide la distancia entre el valor predicho y el valor original de la serie en el instante T , e_T es similar al concepto de residuo en el contexto de regresión.

Aún está pendiente el problema de cómo estimar α . Un método muy sencillo consiste en calcular los errores de predicción y luego considerar aquel valor de α que minimice la suma de cuadrados de dichos errores, es decir,

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \sum_{T=1}^n e_T^2.$$

En la práctica es posible calcular la cantidad $\sum_{T=1}^n e_T^2$ para varios valores de α en el intervalo $(0, 1)$ y luego encontrar el valor de α que minimice el error cuadrático medio (EMC) de predicción $\sum_{T=1}^n e_T^2$.

Observación B.3. *Considere nuevamente la ecuación (B.3) pero escribimos la predicción de la variable y_T , τ pasos adelante tomando como punto de partida $T = n$,*

$$\ell_n = \hat{y}_{n+\tau}(n) = \alpha y_n + (1-\alpha)\ell_{n-1} \tag{B.4}$$

para todo $\tau \geq 1$. Entonces la predicción es constante cuando se considera más de un paso adelante ya que el lado derecho de la ecuación (B.4) no depende de τ . Una modificación

APÉNDICE B. SUAVIZACIÓN EXPONENCIAL

que parece razonable es considerar una ecuación que sea recursiva en τ a partir de $T = n$, es decir,

$$\ell_n = \widehat{y}_{n+\tau}(n) = \alpha y_n + (1 - \alpha)\ell_{n-1} \quad (\text{B.5})$$

para todo $\tau \geq 1$. Esto permite adaptar la predicción considerando una combinación lineal convexa entre el último valor de la serie y las predicciones para $\tau \geq 1$.

Suponga ahora que $y_0, y_1, y_2, \dots, y_n$ es una colección de variables aleatorias y $\alpha \in (0, 1)$ es una constante. Entonces podemos reescribir la ecuación (B.3) considerando la condición inicial $\widehat{y}_1(0) = \ell_0$, y entendiendo que la predicción del suavizamiento exponencial simple es a un paso, entonces

$$\ell_T = \alpha y_T + (1 - \alpha)\ell_{T-1} \quad (\text{B.6})$$

para todo $T \geq 1$.

Resultado B.1. Considere un suavizador ℓ_T como en la ecuación (B.6) y $\alpha \in (0, 1)$. Entonces

$$\ell_T = \alpha \sum_{j=0}^{T-1} (1 - \alpha)^j y_{T-j} + (1 - \alpha)^T y_0 \quad (\text{B.7})$$

para toda $T \geq 1$.

Demostración. La demostración se realiza por inducción sobre T . Para $T = 1$ se tiene que

$$\ell_1 = \alpha y_1 + (1 - \alpha)y_0.$$

Supongamos que la ecuación (B.7) se satisface para T , entonces

$$\ell_{T+1} = \alpha y_{T+1} + (1 - \alpha)\ell_T = \alpha y_{T+1} + (1 - \alpha) \left[\alpha \sum_{j=0}^{T-1} (1 - \alpha)^j y_{T-j} + (1 - \alpha)^T y_0 \right]$$

$$\ell_{T+1} = \alpha y_{T+1} + \alpha \sum_{j=0}^{T-1} (1 - \alpha)^{j+1} y_{T-j} + (1 - \alpha)^{T+1} y_0$$

$$\ell_{T+1} = \alpha y_{T+1} + \alpha \sum_{k=1}^T (1 - \alpha)^k y_{T-(k-1)} + (1 - \alpha)^{T+1} y_0$$

$$\ell_{T+1} = \alpha \sum_{j=0}^T (1 - \alpha)^j y_{T+1-j} + (1 - \alpha)^{T+1} y_0. \blacksquare$$

Resultado B.2. Supongamos que $y_0, y_1, y_2, \dots, y_n$ es una colección de variables aleatorias con media constante μ y varianza constante σ^2 . Entonces para ℓ_T como en (B.7) se tiene $E(\ell_T) = \mu$.

Demostración. Usando el resultado anterior tenemos que

$$E(\ell_T) = \alpha \sum_{j=0}^{T-1} (1 - \alpha)^j \mu + (1 - \alpha)^T \mu = \alpha \mu \sum_{j=0}^{T-1} (1 - \alpha)^j + (1 - \alpha)^T \mu$$

APÉNDICE B. SUAVIZACIÓN EXPONENCIAL

$$E(\ell_T) = \alpha\mu \left(\frac{1 - (1 - \alpha)^T}{1 - (1 - \alpha)} \right) + (1 - \alpha)^T \mu = \mu(1 - (1 - \alpha)^T) + (1 - \alpha)^T \mu$$

$$E(\ell_T) = \mu - \mu(1 - \alpha)^T + (1 - \alpha)^T \mu = \mu. \blacksquare$$

Resultado B.3. Supongamos que $y_0, y_1, y_2, \dots, y_n$ es una colección de variables aleatorias no correlacionadas con media constante μ y varianza constante σ^2 , entonces para ℓ_T como en (B.7) se tiene $V(\ell_T) < \sigma^2$, cuando $T \rightarrow \infty$.

Demostración.

$$V(\ell_T) = \alpha^2 \sum_{j=0}^{T-1} (1 - \alpha)^{2j} V(y_{T-j}) + (1 - \alpha)^{2T} V(y_0) = \alpha^2 \sum_{j=0}^{T-1} (1 - \alpha)^{2j} \sigma^2 + (1 - \alpha)^{2T} \sigma^2$$

$$V(\ell_T) = \alpha^2 \sigma^2 \left(\frac{1 - (1 - \alpha)^{2T}}{1 - (1 - \alpha)^2} \right) + (1 - \alpha)^{2T} \sigma^2 = \frac{\sigma^2 \alpha}{2 - \alpha} < \sigma^2,$$

cuando $T \rightarrow \infty$. \blacksquare

Resultado B.4. Supongamos que $y_0, y_1, y_2, \dots, y_n$ es una colección de variables aleatorias que satisfacen $E(y_T) = \mu$, $0 < T \leq N - 1$ y $E(y_T) = \lambda$, $\forall T \geq N$ entonces

$$E(\ell_T) = \lambda,$$

cuando $T \rightarrow \infty$.

Demostración.

$$E(\ell_T) = \alpha \sum_{j=0}^{T-N} (1 - \alpha)^j \lambda + \alpha \sum_{j=T-(N-1)}^{T-1} (1 - \alpha)^j \mu + (1 - \alpha)^T \mu$$

$$E(\ell_T) = \alpha \lambda \sum_{j=0}^{T-N} (1 - \alpha)^j + \alpha \mu \sum_{j=T-(N-1)}^{T-1} (1 - \alpha)^j + (1 - \alpha)^T \mu$$

$$E(\ell_T) = \alpha \lambda \left(\frac{1 - (1 - \alpha)^{T-N+1}}{1 - (1 - \alpha)} \right) + \alpha \mu (1 - \alpha)^{T-(N-1)} \left(\frac{1 - (1 - \alpha)^{N-1}}{1 - (1 - \alpha)} \right) + (1 - \alpha)^T \mu$$

$$E(\ell_T) = \lambda(1 - (1 - \alpha)^{T-N+1}) + \mu(1 - \alpha)^{T-(N-1)}(1 - (1 - \alpha)^{N-1}) + (1 - \alpha)^T \mu$$

$$E(\ell_T) = \lambda - \lambda(1 - \alpha)^{T-N+1} + \mu \left[(1 - \alpha)^{T-(N-1)}(1 - (1 - \alpha)^{N-1}) + (1 - \alpha)^T \right]$$

$\therefore E(\ell_T) \rightarrow \lambda$, cuando $T \rightarrow \infty$. \blacksquare

El resultado anterior establece que el valor medio del suavizador es sesgado para valores de T mayores a N , sin embargo, a medida que la serie se aleja del punto $T = N$ el efecto de las observaciones iniciales en el valor medio cada vez es menos influyente.

Apéndice C

Herramienta Solver

Solver es una herramienta para resolver y optimizar ecuaciones mediante el uso de métodos numéricos.

Con Solver, se puede buscar el valor óptimo para una celda, denominada celda objetivo, en donde se escribe la fórmula de la función objetivo.

Solver cambia los valores de un grupo de celdas, denominadas celdas cambiantes, y que estén relacionadas, directa o indirectamente, con la fórmula de la celda objetivo. En estas celdas se encuentran los valores de las variables controlables y_1, y_2, \dots, y_n .

Se puede agregar restricciones a Solver, escribiendo una fórmula en una celda, y especificando que la celda deberá ser mayor o igual, igual, o menor o igual que otra celda que contiene la constante.

Solver ajustará los valores de las celdas cambiantes, para generar el resultado especificado en la fórmula de la celda objetivo.

Instalacion del Solver

La herramienta Solver no se instala por defecto:

- En primer lugar se debe tener instalada la versión profesional del Office, la versión estándar no la considera, si no se tiene instalada la versión profesional consulte con su proveedor habitual de software para que se la instale.
- Para saber si la tiene instalada se mira el menú de Herramientas si tiene una opción de nombre Solver.
- Si no la tiene instalada, debe instalarla y para eso debe hacer lo siguiente:
 - Menú Herramientas
 - Sub-Menú Complementos
 - Activar la opción Solver y Aceptar.

APÉNDICE C. HERRAMIENTA SOLVER

Ahora se entra a usar el Solver:

Entrar a Menu: Herramientas → Solver

Aparece el siguiente cuadro de diálogo:



- Celda objetivo: Seleccionar o escribir por ejemplo: \$D\$13
- Valor de la celda objetivo por ejemplo: Máximo
- Cambiando las celdas: Seleccionar o escribir por ejemplo: \$D\$1:\$D\$2
- Se va a introducir las restricciones: Hacer clic en Agregar

Aparece el siguiente cuadro de diálogo:



Usando este cuadro de diálogo se van colocando una a una todas las restricciones que tiene el problema planteado.

En el cuadro de diálogo siguiente se muestra que ya están ingresados todos los parámetros que necesita Solver, entonces una vez revisados debe darse un clic en el botón Resolver a fin de que Solver busque una solución que cumpla con todas las condiciones:

APÉNDICE C. HERRAMIENTA SOLVER



Solver se ejecuta en forma iterativa hasta encontrar una solución óptima para el problema propuesto y entonces entrega el siguiente cuadro de diálogo:



Los diversos botones que aparecen en este cuadro de diálogo tiene el siguiente significado:

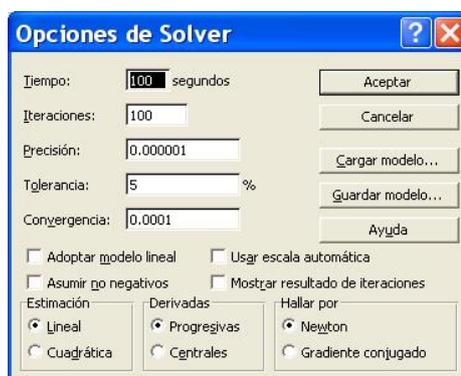
Utilizar solución de Solver: cambia los valores de las variables en la hoja de cálculo.

Restaurar valores originales: deja los valores iniciales de las variables.

Guardar escenario: guarda los valores de las variables como escenario.

Opciones de Solver Se debe hacer clic en el botón Opciones, para que aparezca el cuadro de diálogo Opciones de Solver, que se muestra a continuación:

APÉNDICE C. HERRAMIENTA SOLVER



Descripción de las opciones

Este cuadro de diálogo indica las opciones del Solver en forma general, pero que se pueden cambiar para la resolución de un problema real específico.

Pueden controlarse las características avanzadas del proceso de solución, cargarse o guardarse definiciones de problemas y definirse parámetros para los problemas lineales y no lineales. Cada opción tiene una configuración predeterminada adecuada a la mayoría de los problemas.

Tiempo máximo

Limita el tiempo que tarda el proceso de solución en segundos. Puede introducirse un valor de hasta 32.367, pero el valor predeterminado 100 (segundos) es adecuado para la mayor parte de los problemas.

Iteraciones

Limita el tiempo que tarda el proceso de solución, limitando el número de cálculos provisionales. Aunque puede introducirse un valor de hasta 32 767, el valor predeterminado 100 es adecuado para la mayor parte de los problemas pequeños.

Precisión

Controla la precisión de las soluciones utilizando el número que se introduce para averiguar si el valor de una restricción cumple un objetivo o satisface un límite inferior o superior. Debe indicarse la precisión mediante una fracción entre 0 (cero) y 1. Cuantos más decimales tenga el número que se introduzca, mayor será la precisión; por ejemplo, 0.0001 indica una precisión mayor que 0.01. Sin embargo, cuanto mayor sea la precisión, más tiempo se tardará en encontrar una solución.

Tolerancia

El porcentaje mediante el cual la celda objetivo de una solución satisface las restricciones externas puede diferir del valor óptimo verdadero y todavía considerarse aceptable. Esta opción sólo se aplica a los problemas que tengan restricciones enteras. Una tolerancia mayor tiende a acelerar el proceso de solución.

Convergencia

Si el valor del cambio relativo en la celda objetivo es menor que el número introducido en el cuadro Convergencia para las últimas cinco iteraciones, Solver se detendrá. La convergencia se aplica únicamente a los problemas no lineales y debe indicarse mediante una fracción entre 0 (cero) y 1. Cuantos más decimales tenga el número que se introduzca, menor será la convergencia; por ejemplo, 0.0001 indica un cambio relativo menor que 0.01. Cuanto menor sea el valor de convergencia, más tiempo se tardará en encontrar una solución.

Adoptar modelo lineal

Selecciónelo cuando todas las relaciones en el modelo sean lineales y desee resolver un problema de optimización o una aproximación lineal a un problema no lineal.

Mostrar resultado de iteraciones

Selecciónelo para que Solver muestre temporalmente los resultados de cada iteración. Esta opción es válida sólo en modelos no lineales.

Usar escala automática

Selecciónelo para utilizar la escala automática cuando haya grandes diferencias de magnitud entre las entradas y los resultados.

Adoptar no-negativo

Hace que Solver suponga un límite inferior de 0 (cero) para todas las celdas ajustables en las que no se haya definido un límite inferior en el cuadro Restricción del cuadro de diálogo Agregar restricción.

Cargar modelo

Muestra el cuadro de diálogo Cargar modelo, donde puede especificarse la referencia (nombre) del modelo que desee cargar.

Guardar modelo

Muestra el cuadro de diálogo Guardar modelo, donde puede especificar la ubicación en que desee guardar el modelo. Úselo únicamente cuando desee guardar más de un modelo con una hoja de cálculo; el primer modelo se guardará de forma automática, al guardar el

libro de Excel.

Opciones para modelos no-lineales

Estimación

Especifica el enfoque que se utiliza para obtener las estimaciones iniciales de las variables básicas en cada una de las búsquedas dimensionales.

- **Lineal:** Utiliza la extrapolación lineal de un vector tangente.
- **Cuadrática:** Utiliza la extrapolación cuadrática, que puede mejorar en gran medida los resultados de problemas no lineales.

Derivadas

Especifica la diferencia que se utiliza para estimar las derivadas parciales del objetivo y las funciones de la restricción.

- **Progresivas:** Se utilizan para la mayor parte de los problemas, en que los valores de restricción cambien relativamente poco.
- **Centrales:** Se utiliza en los problemas en que las restricciones cambian rápidamente, especialmente cerca de los límites. Aunque esta opción necesita más cálculos, puede ser útil cuando Solver devuelve un mensaje diciendo que no puede mejorarse la solución.

Buscar

Especifica el algoritmo que se utiliza en cada iteración para determinar la dirección en que se hace la búsqueda.

- **Newton:** Utiliza un método cuasi Newton que normalmente necesita más memoria pero menos iteraciones que el método de gradiente conjugado.
- **Gradiente Conjugado:** Necesita menos memoria que el método Newton, pero normalmente necesita más iteraciones para alcanzar un determinado nivel de precisión. Se usa esta opción cuando se trata de un problema grande o cuando al hacer un recorrido a través de iteraciones se descubre un progreso lento.

Apéndice D

Estadístico de Box-Pierce y Ljung-Box

Una buena manera de comprobar la suficiencia del modelo de Box-Jenkins, es analizar los residuos obtenidos a partir del modelo, de la manera siguiente. Del mismo modo en que se calcularón las funciones de autocorrelación muestral, y de la autocorrelación parcial muestral de la serie de tiempo, se pueden determinar dichas funciones para los residuos.

Una forma de aprovechar los residuos para comprobar la suficiencia del modelo, es examinar una estadística o dato numérico que determine si las primeras K autocorrelaciones muestrales de los residuos, considerados juntos, indican suficiencia del modelo.

Box y Pierce desarrollaron un estadístico que, basado en los cuadrados de los primeros coeficientes de autocorrelación de los residuos, permite analizar si existe o no autocorrelación. Estos datos numéricos se resumen a continuación:

1. La estadística de Box-Pierce es

$$Q = n' \sum_{l=1}^K \hat{\rho}_l^2$$

siendo

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{n'} e_t e_{t-l}}{\sum_{t=1}^{n'} e_t^2}$$

donde $e_1, e_2, \dots, e_{n'}$ son los residuos ordenados en el tiempo.

2. La estadística de Ljung-Box es

$$Q^* = n'(n' + 2) \sum_{l=1}^K \frac{\hat{\rho}_l^2}{n-l}$$

Aquí, $n' = n - d$, donde n es el número de observaciones en la serie de tiempo original, y d es el grado de diferenciación no estacional usado para transformar los valores originales

APÉNDICE D. ESTADÍSTICO DE BOX-PIERCE Y LJUNG-BOX

de la serie de tiempo en valores estacionarios de la serie de tiempo. Además, $\hat{\rho}_l^2$ es el cuadrado de $\hat{\rho}_l$, la autocorrelación muestral de los residuos en el desfaseamiento l , es decir, la autocorrelación muestral de los residuos separada por un desfaseamiento de unidades de tiempo l .

Apéndice E

Criterio de información de Akaike

Cuando se requiere que una serie $\{X_t\}$ sea ajustada por un modelo **ARMA** con media cero, se encuentra con el problema de encontrar el orden de p y q así como el encontrar su valor. Una aproximación alternativa para la selección de este modelo es llevar a cabo una serie de test de hipótesis o usar un criterio de selección de modelos.

No es prudente simplemente elegir un modelo para dar un mejor ajuste minimizando la suma de los cuadrados de los residuales, como el último generalmente decrecerá cuando el número de parámetros sea incrementado a pesar de si la complejidad adicional es realmente valiosa. Pero hay un estadístico de ajuste que puede ser de más ayuda, el R^2 ajustado el cual intenta tomar en cuenta el número de parámetros ajustados, sin embargo, los modelos de selección estadística más sofisticados usualmente son preferidos. Un modelo que es muy usado, es el criterio de información de Akaike (CIA), esta dado por:

$$\text{CIA} = -2\ln(\text{máxima verosimilitud}) + 2r$$

donde r denota el número de parámetros independientes que son ajustados por el modelo que está siendo evaluado. El CIA esencialmente escoge el modelo con el mejor ajuste, moderado por la función de verosimilitud, sometido a un término penalty para prevenir el sobreajuste, que incrementa con el número de parámetros en el modelo. Para un modelo **ARMA(p,q)**, note que $r = p + q + 1$ es la varianza residual que es incluida como un parámetro. Ignorando constantes arbitrarias, el primer término de verosimilitud es usualmente aproximado por $N\ln(\frac{S}{N})$, donde S denota la suma de cuadrados de residuales, y N es el número de observaciones.

Bibliografía

- [1] Brockwell, P. J. & Davis, R. A. (1987). *Time Series: Theory and Methods*. Primera edicion. Editorial Springer-Verlag.
- [2] Brockwell, P. J. & Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Segunda edicion. Editorial Springer-Verlag.
- [3] Chatfield, C. (2004). *The Analisis of Time Series an Introduction*. Sexta edicion. Editorial Chapman & and Hall.
- [4] Guerrero, V. (1987). *Modelos estadísticos para series de tiempo univariadas*. V Coloquio del Departamento de Matematicas. Mexico, D.F.
- [5] Morris, W. (2007). *Elementary Meteorology*. Editorial Ginn & Company.
- [6] Navidi, W. C. (2006). *Estadística para Ingenieros y científicos*. Segunda edicion. McGrall-Hill.
- [7] Oliver, J. E. (2005). *Encyclopedia of world climatology*. Primera edicion. Editorial Springer. Gran Bretaña.
- [8] Reyes, S. (2002). *Introducción a la Meteorología*. Primera edicion. Editorial UABC.
- [9] Taylor, G. F. (1954). *Elementary Meteorology*. Editorial Prentice-Hall. New York.
- [10] Wackerly, D., Mendenhall III, W. & Scheaffer, R. (2002). *Estadística matemática con aplicaciones*. Sexta edicion. Editorial Thomson.
- [11] Minitab. (2008). (Version 15 para Windows). Manual de Usuario.
- [12] Bowerman, Bruce L. & O'Connell, Richard T. & Koehler, Anne B. (2007). *Pronósticos, series de tiempo y regresión. Un enfoque aplicado*. Cuarta edición. Editorial Cengage Learning Editores S.A. de C.V.
- [13] Excel. (Version 2003 para Windows). Manual de Usuario.
- [14] Ronny Vallejos. (2012). *Introducción a las series Cronológicas*. Universidad Técnica Federico Santa María.