



---

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**FACULTAD DE QUÍMICA**

**PRONÓSTICOS DE VENTAS POR MEDIO DE SERIES DE TIEMPO  
CON APLICACIONES EN EL *SOFTWARE R***

**TESINA**

**QUE PARA OBTENER EL TÍTULO DE  
INGENIERO QUÍMICO**

**PRESENTA**

**CARLOS MANUEL DURÁN HERNÁNDEZ**



**MÉXICO, D.F. 2015**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

## **JURADO ASIGNADO:**

**PRESIDENTE:** Profesor: ALEJANDRO LEON IÑIGUEZ HERNANDEZ

**VOCAL:** Profesor: ALEJANDRO ZANELLI TREJO

**SECRETARIO:** Profesor: ANGEL MANUEL GODOY AGUILAR

**1er. SUPLENTE:** Profesor: NAYELI MANZANO SANCHEZ

**2° SUPLENTE:** Profesor: JORGE RAFAEL MARTINEZ PENICHE

## **ASESOR DEL TEMA:**

ANGEL MANUEL GODOY AGUILAR

## **SUSTENTANTE:**

CARLOS MANUEL DURAN HERNANDEZ

# Índice

<b>1. Introducción</b>	<b>3</b>
Objetivo . . . . .	3
Justificación . . . . .	3
Conceptos de probabilidad . . . . .	6
Series de tiempo . . . . .	13
<b>2. Procedimiento</b>	<b>19</b>
Construcción de modelos para series de tiempo . . . . .	19
Identificación . . . . .	19
Pronóstico . . . . .	27
El <i>software</i> R . . . . .	31
<b>3. Resultados</b>	<b>33</b>
Análisis de series de tiempo con el <i>software</i> R aplicada a un caso de estudio . . . . .	33
Resultados del modelo ARIMA con R . . . . .	33
Verificación del modelo . . . . .	39
<b>4. Discusión, Conclusiones y Recomendaciones</b>	<b>43</b>
Discusión . . . . .	43
Conclusiones . . . . .	45
Recomendaciones . . . . .	47
<b>Bibliografía</b>	<b>49</b>
<b>A. Códigos en R</b>	<b>51</b>

---

Librerías, vectores, gráficas y estacionariedad . . . . .	51
Transformada de Box-Cox . . . . .	52
Transformación de datos y diferenciación . . . . .	53
ACF, PACF y Ajuste de modelo ARIMA . . . . .	53
Pronóstico, verificación y actualización . . . . .	54

# Capítulo 1

## Introducción

### Objetivo

Mostrar cómo construir una serie de tiempo por la metodología Box-Jenkins, para realizar pronósticos de ventas a través del *software* libre R.

### Justificación

Las previsiones de ventas constituyen una parte fundamental de los sistemas de planeación y por ende de la economía en general. Los pronósticos de la demanda ejercen una gran influencia en la determinación de factores claves de los procesos, factores como lo son la capacidad instalada (equipos, almacenes, plantas), requerimientos financieros (inventarios, flujo de caja), estructura organizativa (personas, sistemas, servicios), contratos con terceros (compras, operadores), etc. A causa de la extensa influencia de la previsión de ventas en cualquier sistema productivo, se considera que la gestión de la demanda constituye un factor fundamental para el éxito de cualquier organización. Las organizaciones hacen grandes inversiones monetarias basadas en el pronóstico de la demanda de un nuevo producto o en inversión para incrementar la producción de un producto existente del que se espera se incrementen las ventas. Las agencias gubernamentales necesitan pronósticos económicos, de impacto ambiental, espacios deportivos y los efectos de posibles programas sociales.<sup>1</sup>

Si clasificamos a los pronósticos en cualitativos y cuantitativos, y a su vez los métodos cuantitativos dentro de 2 categorías según el tipo de modelo que siguen: Los métodos de serie de tiempo que son los que utilizan datos históricos como base para estimar resultados futuros, es decir asume que la variable es función del tiempo, el método de serie de tiempo considera que los datos históricos están restringidos a valores pasados de la variable que se

---

<sup>1</sup>Principles of Forecasting: A Handbook for Researchers and Practitioners, J.S. Armstrong, Springer Science Business Media, 2001

**Descripción y aplicaciones de los modelos más comunes de pronóstico**

Modelo	Método	Descripción	Aplicaciones
<b>Modelos Causales de pronósticos</b>	Análisis de regresión	Pronóstico explicativo; supone una relación de causa - efecto entre la información de entrada y la información resultante	Pronóstico de corto y mediano alcance de productos y servicios existentes; estrategias de marketing , producción, contratación de personal y planeación de instalaciones
	Regresión Múltiple	Pronóstico explicativo; supone una relación de causa - efecto entre la información introducida en el sistema y la información que resulta de este	Pronóstico de corto y mediano alcance de productos y servicios existentes; estrategias de marketing , producción, contratación de personal y planeación de instalaciones
<b>Modelos de pronósticos de series de tiempo</b>	Método de descomposición	Pronóstico explicativo; supone una relación de causa - efecto entre el tiempo y la información resultante del sistema; el sistema se descompone en sus elementos	Pronóstico de medio alcance para una planta nueva y planeación del equipo, financiamiento, desarrollo del nuevo producto y nuevos métodos de ensamblaje; pronóstico de corto alcance para personal, publicidad, inventario, financiamiento y planeación de la producción
	Promedios Móviles	Se usan para eliminar la aleatoriedad en una serie de tiempo; el pronóstico se basa en los datos de la serie de tiempo suavizados por un promedio móvil	Pronóstico de corto alcance para operaciones como inventario, programación, control, fijación de precios y calendarización de promociones especiales; se emplea para calcular los componentes cíclico y estacional del método de descomposición de corto plazo
	Suavización exponencial	Similar a los promedios móviles, pero los valores son ponderados exponencialmente, otorgando mayor peso a los datos más recientes	Pronóstico de corto alcance para operaciones como inventario, programación, control, fijación de precios y calendarización de promociones especiales
	Modelos autorregresivos	Se emplean con variables económicas para explicar las relaciones entre observaciones adyacentes en una serie de tiempo	Pronósticos de corto y mediano plazos para datos económicos - incluyendo precio,- inventario, producción, acciones y ventas ordenados en una serie de tiempo
	Técnicas Box - Jenkins	No supone ningún patrón particular en los datos históricos de la serie que va a ser pronosticada; usa un método iterativo de identificación y ajuste de un modelo posiblemente útil tomada de un grupo general de modelos	pronósticos de corto y mediano plazos para datos económicos - incluyendo precio,- inventario, producción acciones y ventas ordenados en una serie de tiempo
	Redes Neurales	Por medio de un <i>software</i> se asimilan datos relevantes y se reconocen patrones mediante "Aprendizaje" como lo hacen los humanos	Uso creciente de una gran variedad de aplicaciones de pronóstico.

**Figura 1.1:** Descripción y aplicaciones de los modelos más comunes de pronóstico (modelos causales y de series de tiempo)

está pronosticando. Los métodos de pronóstico causales se basan en la suposición de que la variable que se está pronosticando exhibe una relación de causa y efecto con otra u otras variables, éste método relaciona la variable que se está pronosticando con otras variables que se piensa la influyen o la explican. Un factor principal que influye en la selección de una técnica de pronóstico consiste en la identificación y comprensión de patrones históricos en los datos. Si se pueden reconocer patrones de tendencia, cíclicos o estacionales, entonces se pueden seleccionar las técnicas con la capacidad de utilizar eficazmente estos patrones. El horizonte en el tiempo para un pronóstico tiene una relación directa con la selección de una técnica de pronóstico. Para los pronósticos de corto y mediano plazos, se pueden aplicar diversas técnicas cuantitativas. Sin embargo, al aumentar el horizonte del pronóstico, algunas de estas técnicas se hacen menos aplicables. Por ejemplo, los promedios móviles, la atenuación exponencial y los modelos de Box-Jenkins no son muy buenos pronósticos de cambios económicos radicales, mientras que los modelos econométricos son más útiles para este fin. Los modelos de regresión, son apropiados para los periodos corto, mediano y largo. Las proyecciones de medias, promedios móviles, descomposición clásica y tendencia son técnicas cuantitativas apropiadas para horizontes de corto y mediano plazos. Las técnicas más complejas de Box-Jenkins y los modelos econométricos resultan también apropiados para pronósticos de corto y mediano plazos. Para horizontes mayores en el tiempo, se usan con frecuencia los métodos cualitativos. Una breve descripción de los distintos métodos así como de sus aplicaciones lo encontramos en la figura 1.1

## Patrones de datos para los diferentes métodos de pronóstico

Método	Patrón de Datos	Horizonte de tiempo	Tipo de modelo	Datos mínimos requeridos	
				No estacionales	Estacionales
Simple	ST,T,S	S	TS	1	
Promedios simples	ST	S	TS	30	
Promedios móviles	ST	S	TS	4-20	
Suavizamiento exponencial	ST	S	TS	2	
Suavizamiento exponencial lineal	T	S	TS	3	
Suavizamiento exponencial cuadrático	T	S	TS	4	
Suavizamiento exponencial estacional	S	S	TS		2xs
Filtración adaptativa	S	S	TS		
Regresión simple	T	I	C	10	
Regresión múltiple	C,S	I	C	10xV	
Descomposición clásica	S	S	TS		5xs
Modelos de tendencia exponencial	T	I,L	TS	10	
Ajuste de la curva S	T	I,L	TS	10	
Modelo de Gompertz	T	I,L	TS	10	
Curvas de crecimiento	T	I,L	TS	10	
Census X-12	S	S	TS		6xs
Box-Jenkins	ST,T,C,S	S	TS	24	3xs
Indicadores principales	C	S	C	24	
Modelos econométricos	C	S	C	30	
Regresión múltiple de series de tiempo	T,S	I,L	C		6xs

Patrón de datos: ST, estacional; T, de tendencia; S, estacional; C, cíclico

Horizonte de tiempo: S, corto plazo (menos de tres meses); I, mediano plazo; L, largo plazo

Tipo de modelo: TS, serie de tiempo; C, causal

Estacional: s, longitud de la estacionalidad

Variable: V, número de variables

**Figura 1.2:** Patrones de datos, horizonte de tiempo y tipo de modelo para los diferentes métodos de pronóstico

En la figura 1.2 encontraremos los patrones de datos, horizontes de tiempo, tipo de modelo y los datos mínimos requeridos para los más comunes modelos de pronóstico.

En este documento se revisará un caso de estudio con datos reales de una compañía dedicada a la manufactura, la cual por seguridad llamaremos "CDcorp". La compañía "CDcorp" cuenta con los datos mensuales de ventas de los últimos 24 meses, los datos anteriores son inexactos ya que no se medían con la metodología correcta. En este documento buscamos desarrollar un método, que nos permitirá obtener pronósticos de ventas soportadas por una firme base teórica. El primer paso es identificar el modelo de pronóstico que mejor se adapte a nuestros datos, considerando que estamos buscando pronósticos de mediano plazos para datos económicos ordenados en una serie de tiempo, la recomendación de Hanke<sup>2</sup> para una serie con esas características, la que encontramos en la figura 1.1, donde vemos que se nos recomienda los modelos autoregresivos y las técnicas Box - Jenkins. Por otra parte de acuerdo a la cantidad de datos disponibles (24) y al horizonte de tiempo previsto, si consideramos la figura 1.2 vemos que se nos recomienda el modelo causal de Indicadores principales y el modelo de series de tiempo de Box-Jenkins, sin embargo el patrón de datos que cubre el modelo de Box-Jenkins es más amplio. Por lo que el método estadístico que desarrollaremos en este documento, es el de "Series de Tiempo.<sup>a</sup> a través de la metodología de Box-Jenkins,

<sup>2</sup>Pronósticos en los negocios, John E. Hanke y Dean W Wichern, novena edición, Pearson, 2010

---

donde se mostrará cómo se puede construir un pronóstico que sirva como base para la sólida toma de decisiones.

En este documento encontraremos las bases teóricas necesarias para poder modelar series de tiempo. En el capítulo 1, encontraremos los conceptos básicos de probabilidad y estadística, partiendo de la explicación del modelo matemático usado para estudiar los fenómenos aleatorios. En este capítulo se abordarán los conceptos de: variable aleatoria, función de probabilidad, función de distribución y procesos estocásticos. Con lo que podremos entender a las series de tiempo como la realización de un proceso estocástico. En el capítulo 1 estudiaremos los componentes de las series de tiempo, las propiedades estadísticas, los diferentes tipos y sus clasificaciones. Una vez que tenemos los conocimientos acerca de las series de tiempo, los podremos aplicar en el pronóstico a través de la metodología de Box-Jenkins que se explicará en el capítulo 2, el cual presenta la estrategia de construcción de modelos para series de tiempo estacionarias. En este capítulo se explicará cada una de las 4 etapas en la construcción de un modelo para series de tiempo. Estas etapas son:

1. Identificación. Se identifica un posible modelo dentro de las distintas clases y tipos.
2. Estimación. Se realizan las estimaciones de los parámetros involucrados en el modelo.
3. Verificación. Se verifica que el modelo proporcione un ajuste adecuado.
4. Uso del modelo. En esta etapa es cuando podemos ocupar nuestra serie de tiempo para poder hacer un pronóstico de la variable estudiada.

En el capítulo 3 se analizará un caso de estudio que mostrará cómo se aplican las series de tiempo de manera práctica, a través del *software* "R". Dicho *software* se distribuye de manera gratuita y puede ser extrapolado fácilmente al *software* comercial que regularmente es usado en las empresas como SAS<sup>3</sup> y SAP<sup>4</sup>. Este documento presentará una metodología para realizar pronósticos, que puede complementar y orientar a los directores comerciales en los ajustes finales de los objetivos de ventas.

## Conceptos de probabilidad

### Fenómenos deterministas y aleatorios

Dentro de las posibles maneras que encontramos de modelar la realidad, podemos distinguir dos tipos de fenómenos: El determinista y el aleatorio. El fenómeno determinista es aquel que produce el mismo resultado cuando se le repite bajo las mismas condiciones. Por ejemplo, medir el volumen de un gas cuando la presión y la temperatura son constantes, esto

---

<sup>3</sup><http://www.sas.com/>

<sup>4</sup><http://www.sap.com>

---

produce teóricamente siempre el mismo resultado. El fenómeno aleatorio es aquel que cuando se le repite bajo las mismas condiciones, el resultado que se observa no siempre es el mismo y tampoco es predecible, un ejemplo sería el resultado lanzar una moneda al aire, la cara de la moneda que se mostraría no siempre será la misma y tampoco es predecible. La teoría de la probabilidad es la parte de las matemáticas que se encarga del estudio de los fenómenos o experimentos aleatorios.

## Espacio muestra y la $\sigma$ -álgebra

En Probabilidad el conjunto de todos los resultados posibles de un fenómeno aleatorio o experimento se llama espacio muestra, y se le denota generalmente por la letra griega  $\Omega$  (omega mayúscula). A un resultado particular se le denota por  $\omega$  (omega minúscula). Como ya hemos mencionado  $\Omega$  es el conjunto de todos los resultados posibles, es decir, todos los posibles resultados de todas las posibles variables de un experimento o fenómeno, por ello necesitamos definir un conjunto que agrupe a todos los eventos de un mismo fenómeno aleatorio para los cuales se pueda definir o calcular sus probabilidades. A esta colección de subconjuntos se llama álgebra.

Una colección  $\mathcal{F}$  de subconjuntos de un espacio muestra  $\Omega$  es un álgebra si cumple las siguientes 3 condiciones:

1.  $\Omega \in \mathcal{F}$
2. Si  $A \in \mathcal{F}$  entonces  $A^c \in \mathcal{F}$
3. Si  $A_1, A_2, \dots, A_n \in \mathcal{F}$  entonces  $\bigcup_{k=1}^n A_k \in \mathcal{F}$

La primera condición establece que todo  $\Omega$  debe pertenecer a la colección  $\mathcal{F}$ , la segunda condición asegura que si algún subconjunto  $A$  es de interés y por lo tanto se le considera un evento, entonces el complemento de dicho conjunto también debe ser un evento. Es decir, la probabilidad del evento  $A^c$  está siempre dada por  $1 - P(A)$ . El tercer requisito establece que si se tiene una sucesión finita de eventos entonces la unión de ellos también es un evento, lo cual nos asegura la ocurrencia de por lo menos uno de los eventos de la sucesión. Cuando el tercer punto es válido para sucesiones infinitas, esta colección de subconjuntos de  $\Omega$  se llama  $\sigma$ -álgebra, el prefijo  $\sigma$  se refiere a la operación infinita involucrada. A cada uno de los elementos de  $\mathcal{F}$  se les llama eventos.

Una colección  $\mathcal{F}$  de subconjuntos de un espacio muestra  $\Omega$  es un  $\sigma$ -álgebra si cumple las siguientes 3 condiciones:

1.  $\Omega \in \mathcal{F}$
2. Si  $A \in \mathcal{F}$  entonces  $A^c \in \mathcal{F}$

---

3. Si  $A_1, A_2, \dots, A_n \in \mathcal{F}$  entonces  $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$

En otras palabras la  $\sigma$ -álgebra es una clase o colección no vacía  $\mathcal{F}$  de subconjuntos de  $\Omega$  si es cerrada bajo las operaciones de tomar complementos y uniones numerables. De esta forma en la  $\sigma$ -álgebra  $\mathcal{F}$  agrupa a todos los subconjuntos de  $\Omega$  para los que estamos interesados en calcular su probabilidad y tal colección constituye al dominio sobre el cual se define una medida de probabilidad. Así a cada fenómeno aleatorio se le puede asociar una pareja compuesta por el espacio muestra y una  $\sigma$ -álgebra de eventos.

## Medida de Probabilidad

En 1933 el matemático ruso Andrey Nikolaevich Kolmogorov, propuso las reglas que el cálculo de probabilidades debe satisfacer. Estas reglas las estableció mediante los siguientes tres axiomas

1.

$$P(A) \geq 0$$

2.

$$P(\Omega) = 1$$

3.

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \quad \text{cuando } A_1, A_2, \dots \text{ son sucesos mutuamente excluyentes}$$

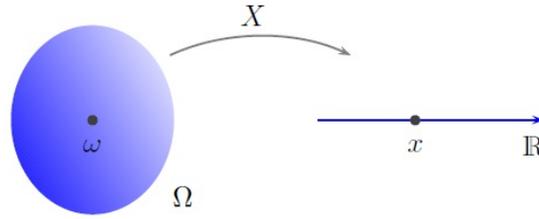
A cualquier función  $P$  definida sobre una colección de eventos que satisfaga los tres axiomas de Kolmogorov se le llama medida de probabilidad, o simplemente probabilidad.

## Espacios de probabilidad

El modelo matemático creado para estudiar los experimentos aleatorios, es el así llamado espacio de probabilidad.

Un espacio de probabilidad es una terna  $(\Omega, \mathcal{F}, P)$  en donde  $\Omega$  es un conjunto arbitrario,  $\mathcal{F}$  es una  $\sigma$ -álgebra de subconjuntos de  $\Omega$  y  $P$  es una medida de probabilidad definida sobre una  $\mathcal{F}$ .

El conjunto arbitrario  $\Omega$  representa usualmente el espacio muestra de un fenómeno aleatorio e inicialmente tal conjunto no tiene ninguna estructura matemática asociada pues sus elementos pueden ser de muy distinta naturaleza como mediciones, personas, objetos, categorías, etc. La  $\sigma$ -álgebra tiene el objetivo de agrupar en una sola colección a todos los



**Figura 1.3:** Representación de una variable aleatoria

subconjuntos de  $\Omega$ , llamados eventos, para los cuales uno está interesado en definir o calcular su probabilidad. Finalmente la medida de probabilidad  $P$  es una función definida sobre la  $\sigma$ -álgebra. Tal función indica la probabilidad de ocurrencia de cada uno de los eventos contenidos en la  $\sigma$ -álgebra, sin especificar la forma concreta en que estas probabilidades son calculadas. Para cada fenómeno aleatorio supondremos que existe un espacio de probabilidad asociado  $(\Omega, \mathcal{F}, P)$ , esta terna es un modelo matemático cuyo objetivo es capturar los elementos esenciales para el estudio científico del experimento aleatorio.

## Variabes aleatorias

Una variable aleatoria es una función con imagen inversa o dominio en la  $\sigma$ -álgebra, lo que asigna a cada uno de sus elementos un número real. En otras palabras una variable aleatoria es una función del espacio muestra en el conjunto de número reales, lo que nos permitirá considerar que el resultado del experimento aleatorio es un número real tomado por la variable aleatoria. Es decir una variable aleatoria se define como una transformación  $X$  del espacio de resultados  $\Omega$  al conjunto de los números reales, esto es,

$$X: \Omega \rightarrow \mathfrak{R}$$

Tal que para cualquier número real  $x$ ,

$$\{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{F} \quad (1.1)$$

Supongamos entonces que se efectúa un experimento aleatorio una vez y se obtiene el resultado  $\omega$  en  $\Omega$ . Al transformar este resultado con la variable aleatoria  $X$  se obtiene un número real  $X(\omega) = x$ . La figura 1.3 ilustra el concepto de variable aleatoria.

La expresión (1.1) se le conoce como la condición de medibilidad de la función  $X$  respecto de la  $\sigma$ -álgebra  $\mathcal{F}$  del espacio de probabilidad y la  $\sigma$ -álgebra de  $\mathfrak{R}$ , esto debido a que nos permite trasladar la medida de probabilidad del espacio de probabilidad a la  $\sigma$ -álgebra.

Para entender la expresión (1.1) es necesario que definamos la imagen inversa de un conjunto determinado. Si  $A$  es un conjunto de  $\mathfrak{R}$ , tenemos que la imagen inversa de  $A$  definida

---

por la expresión  $(X \in A)$  y denotada por  $X^{-1}A$ , expresa lo siguiente:

$$(X \in A) = \{\omega \in \Omega : X(\omega) \in A\}$$

En palabras, la expresión  $(X \in A)$  denota aquel conjunto de elementos de  $\omega$  del espacio muestra  $\Omega$ , tales que bajo la aplicación de la función  $X$  toman un valor dentro del conjunto  $A$ . Consideremos que el conjunto  $A$  es el intervalo  $(a, b)$  entonces el evento  $(X \in A)$  ó  $(X \in (a, b))$  también puede escribirse como  $(a < X < b)$  y esta será una abreviación del siguiente evento:

$$\{\omega \in \Omega : a < X(\omega) < b\}.$$

Consideremos que el conjunto  $A$  es el intervalo infinito  $(-\infty, x]$  para algún valor real  $x$  fijo. Entonces el evento  $(X \in (-\infty, x])$  también puede escribirse como  $(X \leq x)$  y significa

$$\{\omega \in \Omega : -\infty < X(\omega) \leq x\},$$

que es justamente el conjunto al que se hace referencia en la expresión (1.1), y que indica la condición de medibilidad.

## **Función de probabilidad**

La medida de probabilidad de  $X$  se define como aquella función que toma los valores de la variable aleatoria y les asigna valores entre  $[0, 1]$  con ciertas reglas. Lo anterior lo podemos definir formalmente de la siguiente manera:

Sea  $X$  una variable aleatoria discreta con valores  $x_0, x_1, \dots$ . La función de probabilidad de  $X$ , denotada por  $P(x) : \mathfrak{R} \rightarrow \mathfrak{R}$ , se define como sigue

$$P(x) = \begin{cases} P(X = x) & \text{si } x = x_0, x_1, \dots \\ 0 & \text{otro caso.} \end{cases} \quad (1.2)$$

En otras palabras la función de probabilidad es aquella función que indica la probabilidad en los distintos valores que toma la variable aleatoria y puede escribirse mediante una tabla como se muestra en 1.1. Dicho distintamente, sea  $X$  una variable aleatoria discreta que toma los valores  $x_0, x_1, \dots$  con probabilidades

$$\begin{aligned} p_0 &= P(X = x_0) \\ p_1 &= P(X = x_1) \\ p_2 &= P(X = x_2) \\ \vdots & \quad \quad \quad \vdots \end{aligned}$$

Ahora consideremos el caso de una variable aleatoria continua.

---

$x$	$x_0$	$x_1$	$\dots$
$P(x)$	$p_0$	$p_1$	$\dots$

**Cuadro 1.1:** Tabla de función de probabilidad

Sea  $X$  una variable aleatoria continua. Decimos que la función integrable y no negativa  $f(x): \mathfrak{R} \rightarrow \mathfrak{R}$  es la función de densidad de  $X$  si para cualquier intervalo  $[a, b]$  de  $\mathfrak{R}$  se cumple que la igualdad mostrada en 1.3

$$P(X \in [a, b]) = \int_b^a f(x) dx \tag{1.3}$$

Expresado de otra manera, la probabilidad de que la variable tome un valor dentro del intervalo  $[a, b]$  se puede calcular o expresar como el área bajo la función  $f(x)$  en dicho intervalo.

### Función de distribución (función de probabilidad acumulada)

Otra función importante que puede asociarse a una variable aleatoria es la siguiente:

Sea  $X$  una variable aleatoria cualquiera. La función de distribución de  $X$ , denotada por  $F(x): \mathfrak{R} \rightarrow [0, 1]$ , se define como la probabilidad

$$F(x) = P(X \leq x) \tag{1.4}$$

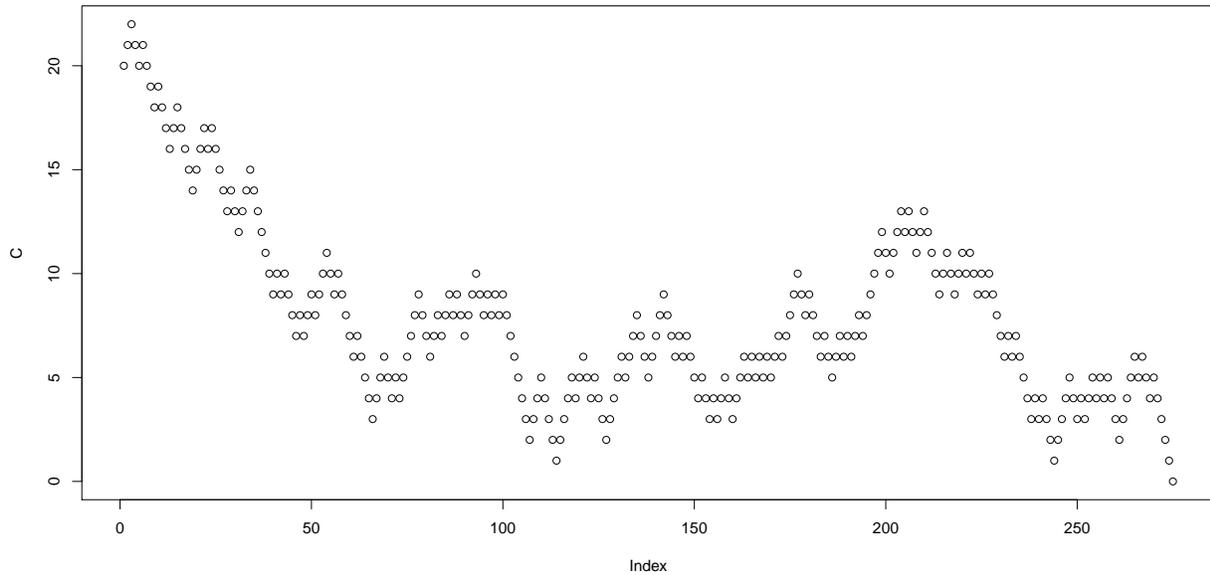
y se puede leer de la siguiente manera: La función de distribución evaluada en un número  $x$  cualquiera es la probabilidad de que la variable aleatoria tome un valor menor o igual a  $x$ , o en otras palabras, que tome un valor en el intervalo  $(-\infty, x]$ . De esta manera siendo  $F(x)$  una probabilidad, sus valores están siempre entre cero y uno. En el caso discreto, suponiendo que  $P(x)$  es la función de probabilidad de  $X$ , la función de distribución (1.4) se calcula de la siguiente manera

$$F(x) = \sum_{u \leq x} P(u) \tag{1.5}$$

Y corresponde a sumar todos los valores positivos de la función de probabilidad evaluada en aquellos números menores o iguales a  $x$ . En el caso continuo, si  $f(x)$  es la función de densidad de  $X$ , por (1.4) se tiene que

$$F(x) = \int_{-\infty}^x P(u) du \tag{1.6}$$

La función de distribución resulta ser importante desde el punto de vista matemático, pues siempre puede definirse dicha función para cualquier variable aleatoria y a través de ella quedan representadas todas las propiedades de la variable aleatoria.



**Figura 1.4:** Proceso estocástico a tiempo discreto (Ruina en volados)

## Procesos estocásticos

Un proceso estocástico es una colección de variables aleatorias  $\{X_t, t \in T\}$  indexadas por un conjunto  $T$  y definidas en algún espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ . Al conjunto de índices  $T$  se le puede interpretar como un parámetro temporal.  $T$  será  $\{0, \dots, n\}$ , algún intervalo  $[0, t]$  ó  $[0, \infty)$ . Interpretamos a un proceso estocástico como la evolución en el tiempo de algún fenómeno cuya dinámica se rige por el azar. Un ejemplo sencillo de esto, es si tengo un capital de 20 pesos al tiempo cero y cada instante de tiempo apuesto un peso en un volado, ganando si cae águila y paro hasta llegar a la ruina.

Otro ejemplo es la evolución en el tiempo de la reserva de una compañía de seguros. En el primer ejemplo, se puede indexar al proceso por algún intervalo de naturales, en cuyo caso hablaremos de un proceso estocástico a tiempo discreto (figura 1.4). En el segundo caso, se puede pensar en un modelo indexado por un subintervalo de  $[0, \infty)$  y hablaremos de un proceso estocástico a tiempo continuo.

---

# Series de tiempo

## Definición de una serie de tiempo y sus componentes

Tal como se reviso en el capitulo anterior, un proceso estocástico es una familia de variables aleatorias  $\{X_t, t \in T\}$  definido en un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ . Las funciones  $\{X(\omega), \omega \in \Omega\}$  en  $T$  son conocidas como las realizaciones del proceso  $\{X_t, t \in T\}$ . Supongamos que observamos una muestra de tamaño  $T$  de alguna variable aleatoria  $Y_t: \{y_1, y_2, \dots, y_T\}$  Dicha muestra observada representa  $T$  números particulares, pero en realidad este conjunto de  $T$  números, es solamente una realización del proceso estocástico subyacente que genera los datos.

Por lo tanto, podemos definir una serie de tiempo  $\{X_t\}_{t=1}^T$  de datos de longitud  $T$ , como una realización de un proceso estocástico en el cual  $t$  representa al tiempo. Nótese que, mientras un proceso estocástico es la colección de las variables aleatorias, una serie de tiempo es una realización finita de un proceso estocástico. Es decir, la serie de tiempo es el resultado de observar la colección de las variables aleatorias. De esta manera, existe un número infinito de realizaciones, entendiendo por realización a una serie de observaciones. Habiendo así un número infinito de observaciones podemos tener un infinito de realizaciones (series de tiempo) de un mismo proceso estocástico.

El primer paso en el análisis de cualquier serie de tiempo es graficar los datos, ya que esto permite sugerir la posibilidad de representar los datos como una realización del proceso. El modelo clásico de descomposición de una serie de tiempo se puede presentar de la siguiente manera

$$X_t = m_t + s_t + Y_t,$$

donde  $m_t$  es una función conocida como el componente de tendencia,  $s_t$  es una función de período conocido  $d$  y se refiere al componente estacional y  $Y_t$  es el componente de ruido aleatorio. En este trabajo haremos uso de un caso particular de ruido aleatorio conocido como ruido blanco (WN), el es una secuencia de variables aleatorias de media cero y varianza constante  $\sigma^2$ , en cual puede ser denotado con de la siguiente manera  $\{Y_t\} \sim WN(0, \sigma^2)$ . nuestro objetivo es estimar y extraer los componentes determinísticos  $m_t$  y  $s_t$ , esperando que el ruido aleatorio restante sea una serie de tiempo estacionaria. En la siguiente sección hablaremos sobre ¿qué es una serie estacionaria y cuáles son sus características?.

## Series de tiempo estacionarias

En términos generales, para una serie de tiempo  $\{X_t, t = 0, 1, \dots\}$  podemos decir que es estacionaria si sus propiedades estadísticas son similares en todos los valores de la serie de tiempo desplazada, es decir sus componentes son semejantes para cualquier secuencia del tipo  $\{X_{t+h}, t = 0, 1, \dots\}$  para toda  $h$  entera. las variables estadísticas que nos interesan para definir esta estacionalidad son las asociadas al primero y segundo momento correspondientes

---

a la media y a la covarianza respectivamente.

La media de la función de  $\{X_t\}$  es

$$\mu_X(t) = E(X_t).$$

La covarianza de la función de  $\{X_t\}$  es

$$\gamma_X(r, s) = Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

para todos los enteros  $r$  y  $s$ .

Una vez revisados el primer y segundo momento podemos decir que una serie de tiempo debe cumplir con las siguientes propiedades para poder ser estacionaria:

1.  $\mu_X(t)$  es independiente del tiempo, esto significa que la media  $\mu_X(t) = \mu_X(t+h)$  para todo  $t, h$
2.  $\gamma_X(t, t+h)$  es independiente de  $t$  para cada  $h$ , esto significa que la covarianza  $\gamma_X(t, t+h) = \gamma_X(t+h, t+h+r)$  para todo  $t, h, r$ .

Como podemos ver en la segunda condición, para una serie estacionaria  $\{X_t\}$  cuando nos referimos al termino covarianza, nos referimos a la función  $\gamma_x$  de una variable definida por

$$\gamma_X(h) := \gamma_X(h, 0) = \gamma_X(t+h, t).$$

La función  $\gamma(\cdot)$  se conoce como la función de autocovarianza y  $\gamma(h)$  como su valor en el intervalo  $h$ .

Siendo  $\{X_t\}$  una serie de tiempo estacionaria. La función de autocovarianza de  $\{X_t\}$  en el intervalo  $h$  es

$$\gamma_X(h) = Cov(X_{t+h}, X_t)$$

la función de autocorrelación para  $\{X_t\}$  en el intervalo  $h$  es

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_{t+h}, X_t)$$

La función de autocovarianza (ACVF) y la función de autocorrelación (ACF) nos dan una útil medida del grado de dependencia entre los valores de la serie a diferentes tiempos. Por esta razón juegan un papel importante cuando consideramos la predicción de valores futuros de una serie en función de los valores pasados y presentes.

Podemos clasificar las series de tiempo estacionarias en tres categorías

1.  $X_t$  es una serie de tiempo estrictamente estacionaria si  $(X_1, \dots, X_n)' \stackrel{d}{=} (X_{1+h}, \dots, X_{n+h})'$  para cualquier  $t_1 < t_2, \dots, < t_n$  la distribución de probabilidad es la misma para cualquier  $n$  y  $h \geq 1$
2.  $X_t$  es una serie de tiempo débilmente estacionaria de primer orden si  $E(X_t)$  es constante para todo  $t$

- 
3.  $X_t$  es una serie de tiempo débilmente estacionaria de segundo orden si  $E(X_t)$  y  $Var(X_t)$  son constantes para todo  $t$  y la  $Cov(X_t, X_{t+h})$  depende a lo más de  $h$ , pero no de  $t$  para todo  $t$ .

## Modelos para series de tiempo estacionarias AR y MA

En el análisis de series de tiempo, existen dos representaciones muy útiles para expresar una serie de tiempo. Los modelos autoregresivos (AR) que expresa la serie de tiempo como una función de valores previos y los modelos de medias móviles (MA) que representa al proceso estocástico como una suma finita ponderada de errores aleatorios.

### Modelos Autoregresivos (AR)

Este modelo puede ser representado básicamente como una ecuación de regresión lineal en el que la variable dependiente  $X$  en el periodo  $t$  depende de sus propios valores observados en periodos anteriores a  $t$  y ponderados con coeficientes para cada periodo, los cuales son llamados coeficientes autoregresivos. En general un proceso autoregresivo de orden  $p$ ,  $AR(p)$  se puede escribir mediante la siguiente ecuación

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \dots + \phi_p \tilde{X}_{t-p} + Y_t \quad (1.7)$$

en donde  $\tilde{X}_t = X_t - \mu$ ,  $\phi_1, \dots, \phi_p$  son los coeficientes autoregresivos y  $Y_t$  el componente de ruido aleatorio.

Otra manera de escribir la ecuación (1.7), es a través de polinomios de retraso. Un polinomio de retraso puede expresarse de la siguiente manera

$$\phi(B)X_t$$

en donde

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_k B^k = 1 - \sum_{j=1}^k \phi_j B^j$$

$\phi_1, \dots, \phi_k$  se refiere a las constantes que ponderan la importancia de los retrasos con los cuales están asociados y  $B^j$  es el operador de retraso, el cual aplicado sobre una variable aleatoria  $X_t$  produce la siguiente transformación:

$$B^k X_t = X_{t-k} \text{ para } k = 0, 1, 2, \dots \text{ y toda } t$$

---

A continuación ocupando los polinomios de retraso podremos encontrar la expresión (1.7)

$$\begin{aligned}
\phi(B)\tilde{X}_t &= Y_t \\
(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_k B^k)\tilde{X}_t &= Y_t \\
\tilde{X}_t - \phi_1 B\tilde{X}_t - \phi_2 B^2\tilde{X}_t - \dots - \phi_k B^k\tilde{X}_t &= Y_t \\
\tilde{X}_t - \phi_1\tilde{X}_{t-1} - \phi_2\tilde{X}_{t-2} - \dots - \phi_k\tilde{X}_{t-k} &= Y_t \\
\tilde{X}_t &= Y_t + \phi_1\tilde{X}_{t-1} + \phi_2\tilde{X}_{t-2} + \dots + \phi_k\tilde{X}_{t-k}
\end{aligned} \tag{1.8}$$

De esta manera podemos decir que ambas expresiones son análogas y son conocidas como el modelo AR de orden  $p$ ,  $AR(p)$ .

## Modelos de promedios móviles (MA)

La idea básica de los modelos de promedios móviles consiste en representar un proceso estocástico  $\{X_t\}$  (cuyos valores pueden ser dependientes unos de otros), como una suma finita ponderada de choques aleatorios independientes  $\{Y_t\}$

$$\tilde{X}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)Y_t = \theta(B)Y_t \tag{1.9}$$

en donde  $\{X_t\}$  representa a las desviaciones de  $\{X_t\}$  respecto a sus nivel medio.  $\mu$  y  $\theta_1, \theta_2, \dots, \theta_q$  son las ponderaciones (parámetros de promedios móviles) asociados con los choques aleatorios en los periodos  $t - 1, t - 2, \dots, t - q$  respectivamente. La ecuación (1.9) es conocida como el modelo MA de orden  $q$ ,  $MA(q)$ .

Las ecuaciones (1.7) y (1.9), son similares excepto en que la variable dependiente  $\{X_t\}$  en los modelos de promedios móviles depende del ruido aleatorio para cada  $t$  del pasado y no de los valores de la variable misma en el pasado como en los modelos Autoregresivos.

Los modelos de promedios móviles (MA) permiten estimar  $X_t$  como una combinación lineal de un número finito de ruido aleatorio pasado  $Y_t$ , mientras que los modelos autoregresivos (AR) estiman  $X_t$  como una función lineal de un número finito de valores pasados de  $X_t$ .

## Modelos ARMA y ARIMA

Una generalización de los modelos AR y MA consiste en combinar ambas clases de modelos para obtener lo que se conoce como modelos autoregresivos y de promedios móviles (ARMA). Los Modelos autoregresivos e integrados de promedios móviles (ARIMA) pueden ser vistos como una generalización de los modelos ARMA

---

## Modelos Autoregresivos y de promedios móviles (ARMA)

El proceso ARMA(p,q) se representa mediante

$$\phi(B)\tilde{X}_t = \theta(B)Y_t$$

en donde  $\phi(B)$  y  $\theta(B)$  son polinomios de retraso de orden p y k respectivamente,  $\{Y_t\}$  es un proceso de ruido blanco y  $\tilde{X}_t$  es la serie de desviaciones de la variable  $X_t$  respecto a su nivel  $\mu$ .

Esta generalización de modelos de series de tiempo estacionarios, surge del hecho de que muchos procesos en la práctica presentan características tanto como de procesos AR como procesos MA.

## Modelos Autoregresivos Integrados de promedios móvil (ARIMA)

En la practica la gran mayoría de los de las series de tiempo con las que se trabaja son no estacionarias y en muchos casos esta no-estacionariedad es debido a que existe alguna tendencia polinomial adaptativa, esto quiere decir que la tendencia puede ser descompuesta en varios polinomios. En estos casos es posible eliminar dicha tendencia , y por lo tanto volver la serie estacionaria, mediante la aplicación de un operador diferencia  $\nabla$  un numero apropiado de veces. Si se define a  $\nabla$  mediante

$$\nabla X_t = X_t - X_{t-1} \text{ para toda } t$$

el operador diferencia  $\nabla$  puede ligarse con el operador retraso  $B$  retraso de la siguiente manera

$$\nabla = 1 - B \text{ por tanto; } \nabla Z_t = (1 - B)Z_t$$

Por lo tanto la forma general para un retraso  $k$  en el operador diferencia puede ser expresado de la siguiente manera

$$\nabla^k Z_t = (1 - B)^k Z_t$$

como se puede ver  $\nabla^k$  es en realidad un binomio elevado a la  $k$ -ésima potencia. Dicho polinomio puede ser generalizado de la siguiente manera

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

Los Modelos autoregresivos e integrados de promedios móviles (ARIMA) pueden ser vistos como una generalización de los modelos ARMA, ya que lo que en realidad hace es aplicar al

---

operador diferencia  $\nabla^d$  es eliminar una posible tendencia polinomial. Es posible construir el proceso estacionario  $\{W_t\}$ , en donde

$$W_t = \nabla^d \tilde{X}_t \quad \text{para toda } t$$

para esta nueva serie podría ser posible ya aplicar un modelo ARMA  $\phi(B)W_t = \theta(B)Y_t$ , lo cual seria equivalente a considerar el modelo ARIMA

$$\phi(B)\nabla^d \tilde{X}_t = \theta(B)Y_t, \quad d \geq 1 \tag{1.10}$$

para  $\{\tilde{X}_t\}$ , en donde  $Y_t$  es un proceso de ruido blanco.

El orden del polinomio de retraso  $\phi(B)$ , el exponente del operador diferencia y el orden del polinomio de retraso  $\theta(B)$ , se acostumbra mencionarlos en la siguiente secuencia,  $p, d, q$ . De manera que un modelo ARIMA( $p, d, q$ ) indica que consta de un polinomio autoregresivo de orden  $p$ , de una diferencia de orden  $d$  y de un polinomio de promedios móviles de orden  $q$ .

# Capítulo 2

## Procedimiento

### Construcción de modelos para series de tiempo

En este capítulo se presenta la estrategia de construcción de modelos para series de tiempo estacionarias por Box y Jenkins (1970)<sup>1</sup>.

Dicha estrategia consta de las siguientes cuatro etapas fundamentales:

1. Identificación. En esta etapa se identifica un posible modelo dentro de la clase ARIMA; es decir, determinación de los valores  $p$ ,  $d$  y  $q$  que especifiquen el modelo apropiado para la serie de estudio,
2. Estimación. En esta etapa se realizan las estimaciones de los parámetros involucrados en el modelo.
3. Verificación. Para esta etapa se verifica que el modelo proporcione un ajuste adecuado y de que los supuesto básicos, implícitos en el modelo, se satisfacen. En caso de no ser adecuado volver al paso 1
4. Uso del modelo. En esta etapa es cuando podemos ocupar nuestra serie de tiempo para poder hacer proyecciones de la variable estudiada.

### Identificación

Esta etapa tiene como objetivo principal determinar los órdenes de los polinomios auto-regresivo y de promedios móviles, es decir, que reproduce las características de la serie. La identificación del modelo se lleva a cabo en dos fases principales.

---

<sup>1</sup>Análisis estadístico de series de tiempo económicas, Victor M. Guerrero, segunda edición, Thomson, 2004

---

## 1. Análisis de estacionariedad

- Estacionariedad en varianza
- Estacionariedad en media o de nivel.

## 2. Elección de los órdenes $p$ y $q$

De manera más general, podría decirse que la etapa de la identificación consiste en determinar, primero, una serie estacionaria en función de la serie original, para la cual se pueda tener una representación ARMA  $(p,q)$  y, posteriormente, en fijar los valores  $p$  y  $q$ .

## Verificación de la estacionariedad de la serie de tiempo

Para convertir una serie de tiempo no estacionaria en otra estacionaria, suelen emplearse en la práctica dos tipos de transformaciones. Un primer tipo para estabilizar su dispersión; esto significa inducir estacionariedad en varianza. Un segundo tipo de estabilización es necesaria y es para estabilizar el nivel, es decir, para eliminar su tendencia o para inducir estacionariedad en media.

### Estabilización de la varianza

En muchas ocasiones la no estacionariedad se debe a que la varianza no es constante, esto quizá debido a que cada punto de observación  $t$ , la variable  $X_t$  tiene varianza  $\sigma_t^2$  la cual es función de su media  $\mu_t$ , esto quiere decir que la varianza y la media existen y dependen de  $t$ , por lo que no son constantes. Un argumento derivado del trabajo de Box-Cox, conduce a determinar una transformación potencia del tipo

$$T(X_t) = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(X_t) & \lambda = 0 \end{cases} \quad (2.1)$$

La cual puede ser útil para estabilizar la varianza de la serie, antes de estabilizar el nivel o media.

### Estabilización del nivel

Una vez determinada la transformación apropiada para estabilizar la varianza de una serie, se procede a estabilizar el nivel de la serie mediante la aplicación del operador diferencia un número apropiado de veces. La principal herramienta para determinar el grado de diferenciación apropiado es la función de autorrelación (ACF). De esta manera lo que comúnmente se hace en la práctica es graficar la ACF correspondiente a cada una de las series  $\{T(X_t)\}$ ,  $\{\nabla T(X_t)\}$ ,  $\{\nabla^2 T(X_t)\}$ , ya que la experiencia ha demostrado que solo en raras ocasiones se

---

requieren diferencias de grado más alto, y recuérdese que debe evitarse la sobre diferenciación de la serie porque esto podría causar problemas al tratar de identificar un posible proceso generador de la serie observada al analizar las ACF y las PACF. Se dice que la serie está sobre diferenciada si se toman más diferencias de las necesarias, por ejemplo, si se elige un orden de integración  $d$  cuando la serie  $\nabla^{d-1}X_t$  ya es estacionaria.

En este punto conviene recordar que si se diferencia un proceso estacionario sigue siendo estacionario. Por lo tanto, en principio, el hecho de que  $\nabla^d X_t$  sea estacionario, no significa necesariamente que  $\nabla^{d-1}X_t$  no lo sea, por lo que hay tener cuidado ya que el objetivo en esta fase de la modelización ARIMA es determinar el menor número de diferencias  $d$  capaz de convertir a una serie en estacionaria.

Como ya hemos mencionado para las series económicas los valores de  $d$  más habituales son  $d = 0, 1, 2$  y para decidir cuál es el más apropiado para la serie bajo estudio utilizaremos los siguientes instrumentos:

- Gráfico de la serie original y las transformaciones correspondientes, para observar si se cumple o no la condición de estacionariedad de oscilar en torno a un nivel constante.
- función de autocorrelación de la serie original y de las transformaciones correspondientes, para comprobar si decrece rápidamente hacia cero o no.
- Contrastes de raíces unitarias, que proporcionan contrastes estadísticos que permiten, a partir del conjunto de información, hacer inferencia sobre la existencia o no de una raíz unitaria en una serie, es decir, sobre la no estacionariedad de la serie. Si se rechaza la hipótesis nula de existencia de raíz unitaria en  $\nabla^{d-1}X_t$ , no se diferenciará más la serie. En caso contrario, si no se rechaza la hipótesis nula se tomará una diferencia más de orden 1.

## Identificación del modelo estacionario ARMA ( $p,q$ )

Una vez determinado el orden de diferenciación  $d$ , se tiene la transformación estacionaria de la serie  $\{\nabla^d T(X_t)\}$  que puede representarse mediante un proceso ARMA( $p,q$ ) estacionario. En esta fase se trata de identificar los órdenes  $p$  y  $q$  del proceso que puede reproducir las características de la serie estacionaria.

Las características dinámicas del proceso estacionario están recogidas en la función de autocorrelación, ACF, por lo que ésta será el instrumento básico para identificar los ordenes  $p$  y  $q$  del modelo ARMA adecuado para representar las características de la serie estacionaria  $X_t$ . Para identificar los ordenes  $p$  y  $q$ , se compararán las funciones de autocorrelación muestrales con las ACF teóricas de los modelos ARMA cuyas características conocemos.

Si el ACF de la serie  $X_t$  presenta un corte a partir de un retardo finito  $j$ , la identificación del proceso adecuado para la misma es sencilla, ya que se correspondería con la FAC teórica

de un MA(j). Pero si el ACF no presenta ningún corte sino que parece decrecer rápidamente siguiendo una estructura exponencial o de onda seno-coseno, la identificación no es tan clara, ya que basándose únicamente en la FAC podría corresponder a un modelo teórico AR o ARMA de cualquier orden. Para ayudarnos en la identificación de modelos ARMA( $p,q$ ) en estos casos acudimos a la función de autocorrelación parcial (PACF). El coeficiente de autocorrelación parcial mide el grado de asociación lineal existente entre las variables  $X_t$  y  $X_{t-i}$ . Para denotar las autocorrelaciones parciales es costumbre escribir  $\phi_{pi}$  para denotar al  $i$ -ésimo coeficiente de un modelo AR( $p$ ), de manera que la PACF viene a ser la sucesión de valores  $\phi_{ii}$ . Los Parámetros  $\phi_{11}, \phi_{22}, \dots, \phi_{pp}$  pueden calcularse a partir de las ecuaciones de Yule-Walker como:

$$\phi_{11} = \rho_1, \quad \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}, \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \dots,$$

$$\phi_{pp} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-3} & \rho_2 \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{p-4} & \rho_3 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-2} & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-3} & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{p-4} & \rho_{p-3} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & \rho_1 & 1 \end{vmatrix}} \quad (2.2)$$

mientras que  $\phi_{ii} = 0$  en caso de que  $i > p$ . De esta manera, el número de autocorrelaciones parciales distintas de cero indica el orden del proceso AR. En la práctica se requiere estimar a las autocorrelaciones parciales con los datos observados, para este fin se sustituye en (2.2) a las autocorrelaciones teóricas,  $\{\rho_k\}$ , por sus correspondientes estimaciones,  $\{r_k\}$ . No debe perderse de vista que la PACF estimada está sujeta a errores muestrales, así que para decidir si  $\phi_{ii}$  puede ser o no considerada como cero, se requiere cuantificar la variación muestral de su estimador  $\Phi_{ii}$ . Una aproximación sugerida por Quenouille (1949) establece que  $\phi_{ii}$  es distinto de cero (al nivel de significación de aproximadamente el 5%) si el valor calculado de  $\Phi_{ii}$  se encuentra fuera del intervalo definido por

$$\pm \sqrt{Var(\phi_{ii})} = \frac{\pm 2}{\sqrt{N-d}}, \quad \text{para } i > p$$

Mismo intervalo de confianza que se aplica las autocorrelaciones  $\rho_k$ .

Proceso	ACF	PACF
AR( $p$ )	Convergencia a cero, con comportamiento dictado por la ecuación $\phi(B)_{\rho_k} = 0$ , para $k \geq p$	Solamente las primeras $p$ autocorrelaciones parciales son distintas de cero.
MA( $q$ )	Sólo las primeras $q$ autocorrelaciones son distintas de cero.	Sucesión infinita convergente a cero.
ARMA( $p, q$ )	Comportamiento irregular de las primeras $q$ autocorrelaciones y después convergencia a cero de acuerdo con $\phi(B)_{\rho_k} = 0$ , para $k \geq p$ .	Sucesión infinita convergente a cero.

**Cuadro 2.1:** Comportamiento de la ACF y la PACF para procesos AR, MA y ARMA

## Comportamientos típicos de la ACF y PACF

En conclusión, para llevar a cabo la etapa de identificación es conveniente recordar las características de la ACF y de la PACF; dichas características se resumen en el cuadro 2.1 para una serie estacionaria  $W_t = \nabla^d T(Z_t)$

Para facilitar la identificación de modelos a partir de los valores muestrales, en la figura X se muestran los comportamientos típicos de la ACF y PACF, para algunos modelos que ocurren con frecuencia en la práctica.

## Estimación

La etapa de estimación presupone que se ha identificado ya un modelo y que, de ser éste adecuado, lo único que resta es encontrar los mejores valores de los parámetros para que dicho modelo represente apropiadamente a la serie en consideración. Es decir, una vez conocidos los órdenes de los polinomios autoregresivo y de promedios móviles  $p$  y  $q$ , así como el grado de diferenciación  $d$ , se postula que el modelo

$$\phi(B)\nabla^d T(X_t) = \theta_0 + \theta(B)Y_t$$

resulta adecuado para representar la serie  $\{T(X_t)\}$  y se requiere entonces asignar valores a  $\phi_1, \dots, \phi_p, \theta_0, \theta_1, \dots, \theta_q$ , lo cual podría hacerse de manera arbitraria, pero evidentemente es preferible utilizar un método objetivo y estadísticamente apropiado. Dicho método será en este caso el de máxima verosimilitud.

La función de verosimilitud nos indica qué tan verosímil es que una muestra observada sea función de los posibles valores de parámetro. Llevar al máximo la verosimilitud dará los valores del parámetro para los cuales es más creíble que la muestra observada se haya generado, es decir, los valores de parámetro que concuerdan de manera más cercana con los datos observados.

Es decir, sean  $X_1, X_2, \dots, X_n$  con una función de distribución

$$f(x_1, x_2, \dots, x_n; \theta_1, \dots, \theta_m) \quad (2.3)$$

donde los parámetros  $\theta_1, \dots, \theta_m$  tienen valores desconocidos. Cuando  $x_1, \dots, x_n$  son los valores muestrales observados y la ecuación (2.3) es considerada como una función de  $\theta_1, \dots, \theta_m$ , recibe el nombre de función de verosimilitud. Los estimadores de máxima verosimilitud  $\hat{\theta}_1, \dots, \hat{\theta}_m$  son los valores de las  $\theta_i$  que maximizan la función de verosimilitud, de modo que

$$f(x_1, \dots, x_n; \hat{\theta}_1, \dots, \hat{\theta}_m) \geq f(x_1, \dots, x_n; \theta_1, \dots, \theta_m) \quad \text{para toda } \theta_1, \dots, \theta_m$$

## Verificación

La etapa de verificación de la metodología de Box-Jenkins tiene su origen en la idea de que todo modelo es erróneo, puesto que los modelos son meras representaciones simplificadas de la realidad. Lógicamente, si hay que elegir entre varios modelos, habrá que elegir aquél que presente menos fallas, o bien, fallas menos importantes; por este motivo habrá que poner todos los posibles modelos en tela de juicio para detectar sus fallas. Una de las formas más claras y simples para detectar violaciones a los supuestos de los modelos es a través del análisis de residuales, en donde, como residual se considera aquella parte de las observaciones que no es explicada por el modelo. Es decir, los residuales miden la discrepancia entre los valores observados y los valores estimados por el modelo.

$$\hat{Y} = W_t - \hat{W}_t$$

además, cuando el tamaño de la muestra es grande, los errores aleatorios y los residuales (que también son variables aleatorias) son esencialmente iguales; por esta razón, al analizar los residuales observados  $\{\hat{Y}\}$  se analiza básicamente lo que debería ser una realización del proceso de ruido blanco  $\{Y_t\}$ . Por lo anterior, los supuestos acerca del proceso  $\{Y_t\}$  pueden verificarse y posiblemente corregirse de la siguiente manera:

1.  $\{Y_t\}$  tiene media cero. Calcúlese la media aritmética y la desviación estándar muestral de los residuales

$$m(\hat{Y}) = \sum_{t=t'}^N \hat{Y}_t / (N - d - p) \quad \text{y} \quad \hat{\sigma}_Y = \sqrt{\sum_{t=t'}^N [\hat{Y}_t - m(\hat{Y})]^2 / (N - d - p - q)}$$

con  $t' = d + p + 1$ , para construir el cociente

$$t = \sqrt{N - d - p} \frac{m(\hat{Y})}{\hat{\sigma}_Y} \quad N(0, 1)$$

Si el valor absoluto de dicho cociente es menor que dos, se dirá que no hay evidencia de que la media del proceso de ruido blanco sea distinta de cero y por lo mismo no se rechaza el supuesto. Esto debido a que el estadístico de prueba a un nivel de significancia del 5% para una distribución normal es de  $|t| \leq 1,96$ .

- 
2.  $\{Y_t\}$  tiene varianza constante. Si en el gráfico de los residuos la dispersión es constante, concluiremos que la varianza de  $Y_t$  permanece constante. Esta verificación visual podría pensarse que es muy burda, pero la idea es que solamente las violaciones muy notorias a este supuesto son las que realmente llegan a causar problemas. es decir si notamos algún patrón de crecimiento o decrecimiento diremos que la varianza no es constante.
  3. Las variables aleatorias  $\{Y_t\}$  son mutuamente independientes. Debido a que independencia implica no-autocorrelación, se debe requerir que  $\rho_k(Y) = 0$  para todo  $k \neq 0$ . Esto último puede verificarse calculando, primero, la ACF muestral de los residuales  $\{r_k(\hat{Y})\}$ , que en el supuesto de que su media es cero, está dada por

$$r_k(\hat{Y}) = \frac{\sum_{t=t'}^{N-k} \hat{Y}_t \hat{Y}_{t+k}}{\sum_{t=t'}^N \hat{Y}_t^2}, \quad k = 1, 2, \dots$$

con  $t' = d + p + 1$ , posteriormente estímesese la desviación estándar de  $r_k(\hat{Y})$  como

$$\sqrt{Var[r_k(\hat{Y})]} = 1/\sqrt{N - d - p}$$

para determinar el nivel de significancia estadística individual de las autocorrelaciones de los residuales: o sea, si  $|r_k(\hat{Y})| \geq 2\sqrt{N - d - p}$  se dirá que la autocorrelación  $k$ -ésima es significativamente distinta de cero. Es decir se dirá que  $\{\hat{Y}_t\}$  es ruido blanco si los coeficientes de autocorrelación estimados están dentro del intervalo de no significación, es decir, si

$$|\hat{\rho}_k| \leq \frac{2}{\sqrt{N - d - p}}$$

para todo  $k$  o al menos para el 95% de los estimados. No obstante esta prueba de significación no es del todo válida para autocorrelaciones correspondientes a retrasos pequeños ( $k \leq 3$ ) y conviene por ello realizar una prueba simultánea de la significación de las primeras  $K$  autocorrelaciones. Esto significa realizar un contraste de significatividad sobre un conjunto de coeficientes de autocorrelación.

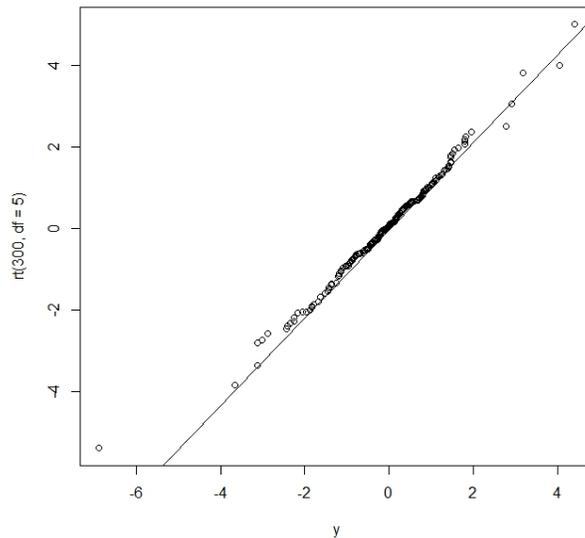
$$\rho_1(Y) = \rho_2(Y) = \dots = \rho_K(Y) = 0$$

El estadístico más utilizado para contrastar esta hipótesis es el propuesto por Ljung-Box(1978)

$$Q' = (N - d - p)(N - d - p + 2) \sum_{k=1}^K r_k^2(\hat{Y}) / (N - d - p - k)$$

que resulta ser el más adecuado para realizar la prueba. Si  $K$  es grande ( $K > 20$ ), sigue aproximadamente una distribución Ji-cuadrada con  $K-p-q$  grados de libertad, por lo tanto para un nivel de significación del 5%, podemos decir que si

$$Q' < \chi_{0,05}^2(K - p - k)$$



**Figura 2.1:** Ajuste de normalidad en una grafica QQ

Si esto se cumple, podríamos decir que las variables aleatorias son mutuamente independientes.

4.  $\hat{Y}_t$  tienen una distribución normal. Se sabe que para una distribución normal, aproximadamente del 95 % de las observaciones deben localizarse dentro de un intervalo que se extienda dos desviaciones estándar por abajo y por arriba de la media; entonces, sí se cumple que la media de los residuales sea cero, se esperaría que a lo más un total de  $(N - d - p)/20$  observaciones se localizan fuera del intervalo  $(-2\hat{\sigma}_Y, 2\hat{\sigma}_Y)$ . También se pueden ocupar algunas pruebas de normalidad como la prueba de Jarque-Bera o la prueba de Shapiro, otra opción sería graficar los residuales en una grafica de cuantiles teóricos vs cuantiles muestrales, también conocidos como gráfico QQ, como se muestra en la figura 2.1
  
5. Implícitamente se ha supuesto que no existen observaciones raras. La grafica de residuales contra el tiempo permitirá visualizar si existe este tipo de observaciones anómalas. Por ejemplo, un residual que se encuentra fuera de  $(-3\hat{\sigma}_Y, 3\hat{\sigma}_Y)$  implicará que, o bien sucedió un evento cuya probabilidad de ocurrencia era aproximadamente 0.2 % o el residual en cuestión corresponde a una observación que no fue generada por el mismo proceso generador del resto de la serie. De esta manera, como una regla empírica de trabajo, podría considerarse como sospechosas a las observaciones cuyos residuales estén fuera del intervalo  $(-3\hat{\sigma}_Y, 3\hat{\sigma}_Y)$

---

## Pronóstico

Uno de los objetivos más importantes en el análisis de series de tiempo es el pronóstico de los valores futuros. El termino *pronóstico* es usado más frecuentemente en la literatura reciente que el termino *predicción*. Mucho acerca de la teoría de pronósticos de serie de tiempo se deriva de la teoría general de la predicción lineal desarrollada por Kolmogorov (1939,1941), Wiener (1949), Kalman(1960), Yaglom(1962) y Whittle (1983), entre otros.

### Pronóstico con error cuadrático medio mínimo para modelos AR-MA

En pronósticos, nuestro objetivo es producir un pronóstico óptimo, es decir que no contenga error o que este sea lo menor posible; lo que nos conduce al concepto del error cuadrático medio mínimo. Esta predicción, producirá un valor futuro óptimo en términos del criterio del error medio mínimo cuadrado. Para desarrollar el pronóstico con error cuadrático medio mínimo, consideraremos el caso cuando  $d = 0$  y  $\mu = 0$ , es decir, el modelo estacionario ARMA

$$\phi(B)X_t = \theta(B)Y_t \quad (2.4)$$

Debido a que el modelo es estacionario, podemos reescribir (2.4) en representación de medias móviles.

$$\begin{aligned} X_t &= \psi(B)Y_t \\ &= \psi_0 Y_t + \psi_1 Y_{t-1} + \psi_2 Y_{t-2} + \dots, \end{aligned}$$

donde

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = \frac{\theta(B)}{\phi(B)}$$

y  $\psi_0 = 1$ . Para  $t = t + h$ , tenemos

$$X_{t+h} = \sum_{j=0}^{\infty} \psi_j Y_{t+h-j} \quad (2.5)$$

En particular si a partir de  $t$ , se desea pronosticar a la observación  $X_{t+h}$ , un pronóstico cualquiera de esta observación (es decir cualquier valor para la observación), que se obtenga como combinación lineal de los valores de la serie  $\{X_t\}$ , y en consecuencia de los errores  $\{Y_t\}$ , será denotado por  $\tilde{X}_t(h)$ , mientras que el pronóstico óptimo se escribirá como  $\hat{X}_t(h)$ . Como ya se ha mencionado el criterio que se usará para determinar la optimalidad será la del error cuadrático medio mínimo, es decir,  $\hat{X}_t(h)$  deberá satisfacer la condición siguiente

$$E_t[X_{t+h} - \hat{X}_t(h)]^2 = \min_{\tilde{X}_t(h)} E_t[X_{t+h} - \tilde{X}_t(h)]^2 \quad (2.6)$$

en el cual  $E$  denota a la esperanza condicional, dada toda la información hasta el momento  $t$ , o sea

$$E_t[X_{t+h} - \tilde{X}_t(h)]^2 = E\{[X_{t+h} - \tilde{X}_t(h)]^2 | Z_t, Z_{t-1}, \dots\}$$

Ya que  $\tilde{X}_t(h)$ , y por consiguiente también  $\hat{X}_t(h)$  pueden expresarse de la misma manera que (2.5) el problema de obtener  $\hat{X}_t(h)$  se traduce en la especificación de los valores de  $\psi_1, \psi_2, \dots$ , de manera tal que se cumpla (2.6).  $X_{t+h}$  como se muestra en (2.5), puede ser separado en 2 sumas, en donde la primera suma corresponda a la información desconocida al tiempo  $t$  (ya que abarcará observaciones desde  $t + 1$  hasta  $t + h$ )

$$\sum_{j=0}^{h-1} \psi_j Y_{t+h-j}$$

mientras que la segunda suma consta de información conocida al tiempo  $t$  (desde  $-\infty$  hasta  $t$ )

$$\sum_{j=h}^{\infty} \psi_j Y_{t+h-j}$$

es decir

$$X_{t+h} = \sum_{j=0}^{h-1} \psi_j Y_{t+h-j} + \sum_{j=h}^{\infty} \psi_j Y_{t+h-j} \quad (2.7)$$

Recordemos que  $\tilde{X}_t(h)$  es una combinación lineal de los valores de  $\{X_t\}$  y los errores  $\{Y_t\}$ , por lo que consta de la información conocida al tiempo  $t$

$$\tilde{X}_t(h) = C_h Y_t + C_{h+1} Y_{t-1} + \dots = \sum_{j=h}^{\infty} C_j Y_{t+h-j} \quad (2.8)$$

Ahora ocupando (2.7) y (2.8) podemos expresar  $X_{t+h} - \tilde{X}_t(h)$  como sigue

$$X_{t+h} - \tilde{X}_t(h) = \sum_{j=0}^{h-1} \psi_j Y_{t+h-j} + \sum_{j=h}^{\infty} \psi_j Y_{t+h-j} - \sum_{j=h}^{\infty} C_j Y_{t+h-j}$$

Agrupando las sumas con valores conocidos al tiempo  $t$

$$X_{t+h} - \tilde{X}_t(h) = \sum_{j=0}^{h-1} \psi_j Y_{t+h-j} + \sum_{j=h}^{\infty} (\psi_j - C_j) Y_{t+h-j}$$

y, debido a que  $E(Y_{t+i} Y_{t+j}) = 0$  para  $i \neq j$ , se sigue que

$$E_t[X_{t+h} - \tilde{X}_t(h)]^2 = \sum_{j=0}^{h-1} \psi_j^2 \sigma_Y^2 + \sum_{j=h}^{\infty} (\psi_j - C_j)^2 \sigma_Y^2 \quad (2.9)$$

el mínimo de esta expresión se obtiene haciendo  $C_j = -\psi_j$ , para  $j = h, h + 1, \dots$ , lo que nos lleva a la siguiente expresión

$$E_t[X_{t+h} - \hat{X}_t(h)]^2 = \sum_{j=0}^{h-1} \psi_j \sigma_Y^2 \quad (2.10)$$

En otras palabras,  $\hat{X}_t(h)$  será el pronóstico óptimo por el criterio de del error cuadrático mínimo siempre y cuando cumpla la condición (2.10).

$\hat{X}_t(h)$  al igual que  $\tilde{X}_t(h)$  en (2.8), es una combinación lineal de los valores de  $\{X_t\}$  y los errores  $\{Y_t\}$ . Por lo tanto solo depende de la información conocida al tiempo  $t$ .

$$\hat{X}_t(h) = \psi_h Y_t + \psi_{h+1} Y_{t-1} + \dots = \sum_{j=h}^{\infty} \psi_j Y_{t+h-j} \quad (2.11)$$

Ahora bien, tomando (2.7) y por definición de la esperanza de la esperanza condicional

$$E_t(Y_{t+h-j}) = \begin{cases} Y_{t+h-j} & \text{si } j \leq h \\ 0 & \text{si } j < h \end{cases} \quad (2.12)$$

se tiene

$$E_t(X_{t+h}) = E_t\left(\sum_{j=0}^{h-1} \psi_j Y_{t+h-j}\right) + E_t\left(\sum_{j=h}^{\infty} \psi_j Y_{t+h-j}\right)$$

y debido a (2.12) toda la primera suma que contiene los valores desconocidos desde  $t + 1$  hasta  $t + h$  es decir para  $j < h$  llegamos a que

$$E_t(X_{t+h}) = \sum_{j=h}^{\infty} \psi_j Y_{t+h-j} \quad (2.13)$$

La expresión anterior (2.13) es la misma que tenemos en (2.11), por lo tanto podemos decir que la esperanza condicional de  $X_{t+h}$  nos proporciona el pronóstico con error cuadrático medio mínimo (2.14).

$$\hat{X}_t(h) = E_t(X_{t+h}) \quad (2.14)$$

El error del pronóstico (2.14), usualmente es leído como el pronóstico  $h$  periodos adelante a partir de  $t$ , y viene dado por

$$\begin{aligned} e_t(h) &= X_{t+h} - \hat{X}_t(h) \\ &= -\sum_{j=h}^{h-1} \psi_j Y_{t+h-j} \end{aligned} \quad (2.15)$$

por lo tanto tenemos que

$$E_t[e_t(h)] = 0 \quad \text{y} \quad Var_t[e_t(h)] = \sum_{j=0}^{h-1} \psi_j^2 \sigma_Y^2 \quad (2.16)$$

con  $Var_t[e_t(h)] = Var_t[e_t(h)|X_t, X_{t-1}, \dots]$ . Además, considerando que

$$Var_t[e_t(h)] - Var_t[e_t(h-1)] = \psi_{h-1}^2 \sigma_Y^2 \geq 0 \quad \text{para } h \geq 1$$

podemos decir que al emplear los pronósticos óptimos, mientras más alejado se desee el pronóstico (mayor sea  $h$ ) mayor será la varianza del mismo, lo cual indica que habrá menor precisión. Derivado de que tenemos en mente una aplicación práctica, es importante obtener límites de confianza para los pronósticos que se realicen; dichos límites del  $100(1 - \alpha)\%$  de confianza cuando se ocupa (2.16) y el supuesto de que  $Y_t \sim N(0, \sigma_Y^2)$  son

$$\hat{X}_t(h) \pm N_{\alpha/2} \left[ 1 + \sum_{j=1}^{h-1} \psi_j^2 \right] \sigma_Y \quad (2.17)$$

donde  $N_{\alpha/2}$  es la desviación normal estándar de tal forma que  $P(N > N_{\alpha/2}) = \alpha/2$ .

Tomando  $h = 1$  en (2.15) obtenemos

$$e_t(1) = X_{t+1} - \hat{X}_t(1) = Y_{t+1} \quad (2.18)$$

Entonces podemos ver que la predicción un paso adelante ( $h = 1$ ) es independiente, lo cual implica que  $\hat{X}_t(1)$  es la mejor predicción de  $X_{t+1}$ ; por otra parte los errores de  $h \geq 2$  en general se encuentran correlacionados.

Ya que se desea obtener pronósticos de la serie  $\{X_t\}$ , considere el modelo  $\phi(B)X_t = \theta(B)Y_t$ , entonces

$$\begin{aligned} \hat{X}_t(h) &= E_t(X_{t+h}) \\ &= E_t(\phi_1 X_{t+h-1} + \dots + \phi_p X_{t+h-p} + Y_{t+h} - \theta_1 Y_{t+h-1} - \dots - \theta_q Y_{t+h-q}) \\ &= \phi_1 E_t(X_{t+h-1}) + \dots + \phi_p E_t(X_{t+h-p}) \\ &\quad + E_t(Y_{t+h}) - \theta_1 E_t(Y_{t+h-1}) - \dots - \theta_q E_t(Y_{t+h-q}) \end{aligned} \quad (2.19)$$

donde la esperanza condicional para  $\{X_{t+h-j}\}$  está dada por

$$E_t(X_{t+h-j}) = \begin{cases} X_{t+h-j} & \text{si } j \geq h \\ \hat{X}_t(h-j) & \text{si } j < h \end{cases} \quad (2.20)$$

A diferencia de la esperanza condicional previa (2.20), la de  $\{X_{t+h-j}\}$  (2.12) depende de los errores aleatorios  $\{Y_{t+h,j}\}$  que no son observables. Por lo tanto haciendo uso de (2.18)

y de (2.12) podemos obtener una expresión que dependa de los valores de la serie y de sus pronósticos

$$E_t(Y_{t+h-j}) = \begin{cases} X_{t+h-j} - \hat{X}_{t+h-j}(1) & \text{si } j \geq h \\ 0 & \text{si } j < h \end{cases} \quad (2.21)$$

Como podemos ver (2.21) solo depende de valores que si son observables.

## Pronóstico con error cuadrático medio mínimo para modelos ARI-MA

Debido a que la mayoría de las series que se observan en la práctica son no estacionarias, necesitamos una generalización de lo visto en la sección anterior. Dicha generalización se obtiene de manera inmediata si se supone que la no estacionariedad de la serie original  $\{X_t\}$  se cancela aproximadamente al determinar alguna transformación  $T(X_t)$  y aplicando un número apropiado de veces el operador diferencia sobre  $T(X_t)$ . Considerando el hecho de que el pronóstico óptimo de  $T(X_t)$  sea su esperanza condicional al tiempo  $t$ ; es decir, si el modelo para  $T(X_t)$  puede escribirse como

$$\varphi(B)T(X_t) = \theta(B)Y_t, \quad \text{con } \varphi(B) = \nabla^d \phi(B)$$

entonces

$$\begin{aligned} \hat{T}(X_t)(h) &= E_t[T(X_{t+h})] \\ &= E_t[\varphi_1 T(X_{t+h-1}) + \dots + \varphi_{p+d} T(X_{t+h-p-d}) \\ &\quad + Y_{t+h} - \theta_1 Y_{t+h-1} - \dots - \theta_q Y_{t+h-q}] \\ &= \varphi_1 E_t[T(X_{t+h-1})] + \dots + \varphi_p E_t[T(X_{t+h-p})] \\ &\quad + E_t[Y_{t+h}] - \theta_1 E_t[Y_{t+h-1}] - \dots - \theta_q E_t[Y_{t+h-q}] \end{aligned} \quad (2.22)$$

donde

$$E_t[T(X_{t+h-j})] = \begin{cases} T(X_{t+h-j}) & \text{si } j \geq h \\ \hat{T}(X_t)(h-j) & \text{si } j < h \end{cases} \quad (2.23)$$

que es la expresión equivalente a la (2.20), mientras que la equivalente a (2.21) vendría ser

$$E_t(Y_{t+h-j}) = \begin{cases} T(X_{t+h-j}) - \hat{T}(X_{t+h-j-1})(1) & \text{si } j \geq h \\ 0 & \text{si } j < h \end{cases} \quad (2.24)$$

## El software R

R es un lenguaje y entorno de programación para análisis estadístico y gráfico, el cual se distribuye de manera gratuita en <http://www.r-project.org/>. R proporciona un amplio

---

abanico de herramientas estadísticas (modelos lineales y no lineales, pruebas estadísticas, análisis de series de tiempo, algoritmos de clasificación y agrupamiento, etc.) y gráficas. R es probablemente uno de los dos lenguajes más utilizados en investigación por la comunidad estadística<sup>2</sup>, siendo además muy populares en el campo de la investigación biomédica, la bioinformática y las matemáticas financieras; a este fin contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con finalidades específicas de cálculo o gráfico. Fue desarrollado inicialmente por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993. Su desarrollo actual es responsabilidad del R *Development Core Team*<sup>3</sup>.

Dentro del *software* ampliamente usado en la industria, son los desarrollos de las compañías SAS<sup>4</sup> y SAP<sup>5</sup> el cual presenta módulos estadísticos específicos para realizar pronósticos, estas compañías dan soporte sobre la plataforma que ofrecen, ya que no son *software* libre, es decir se debe pagar una licencia para su uso. Los desarrollos de R puede ser fácilmente extrapolados al *software* comercial, además de contar con una gran variedad de nuevos paquetes que constantemente lanzados por la comunidad estadística mundial.

El uso de herramientas estadísticas nos permite robustecer las áreas de desarrollo de negocios, aportando proyecciones reales para una efectiva planeación del curso del negocio. En este documento mostraremos como desarrollar un pronóstico ocupando algunos de los paquetes disponibles para R; este software fue elegido debido a su gran uso en la comunidad estadística y debido a que se distribuye de manera gratuita, por lo que cualquier empresa puede descargar el *software* y utilizar este documento para realizar mejoras en sus métodos de pronóstico.

---

<sup>2</sup>[http://en.wikipedia.org/wiki/R\\_\(programming\\_language\)](http://en.wikipedia.org/wiki/R_(programming_language))

<sup>3</sup><http://www.r-project.org/contributors>

<sup>4</sup><http://www.sas.com/>

<sup>5</sup><http://www.sap.com>

# Capítulo 3

## Resultados

### Análisis de series de tiempo con el *software* R aplicada a un caso de estudio

En este capítulo se analizará un caso de estudio con datos reales de una compañía dedicada a la manufactura, la cual por seguridad llamaremos "CDcorp". "CDcorp" se encuentra a unos meses de su cierre fiscal, por lo que el equipo de liderazgo pronto llevará a cabo la junta de planeación estratégica de cierre de año. Para esta reunión, se necesita conocer cuál sería el pronóstico de las ventas de los próximos 12 meses para los clientes existentes, considerando que no hay eventos relevantes en la empresa (despido de personal, fusión con otra compañía, etc.), ni en el país (devaluaciones, cambio de reformas fiscales, guerras, etc.). Esta compañía mantiene una distribución periódica de productos a sus diferentes clientes, por lo que la labor de los vendedores del área de expansión de negocios, es incrementar las ventas, ya sea con nuevos productos o en expansión de ventas de los productos ya existentes. El área de Business Intelligence de la compañía (BI), es la encargada de entregar el pronóstico al grupo de liderazgo. Para llevar a cabo la proyección, los analistas de BI cuentan con la información de las ventas mensuales de los últimos 2 años de la empresa (cuadro 3.1).

### Resultados del modelo ARIMA con R

El equipo de BI, ocupa el modelo ARIMA de series de tiempo para realizar el pronóstico solicitado por el equipo de liderazgo. Para la construcción de modelos ARIMA con R debemos seguir la estrategia analizada en el capítulo 2. Lo que desarrollaremos a continuación es el uso de las herramientas estadísticas de R, para cumplir satisfactoriamente la construcción de la serie de tiempo y su posterior aplicación en el pronóstico de nuestro caso de estudio. El primer paso es graficar el histórico de ventas para tener una idea general del comportamiento del negocio en los 2 años correspondientes. Para tal efecto construiremos un vector que

Mes	2012	2013
Ene	8,705	7,232
Feb	8,471	8,384
Mar	9,380	9,418
Abr	10,067	8,261
May	10,838	8,363
Jun	10,188	8,672
Jul	10,488	8,361
Ago	9,824	9,155
Sep	9,496	9,528
Oct	11,566	9,404
Nov	10,155	10,291
Dic	8,236	8,532

**Cuadro 3.1:** Ventas mensuales 2012 y 2013 (M\$MXN)

contenga todos los datos de las ventas y después graficaremos los mismos, el código para esta construcción lo encontramos en el apéndice A el cual nos genera la gráfica 3.1 correspondiente al histórico de las ventas.

Como se menciona en la sección 2 debemos garantizar la estacionariedad de la serie de tiempo. Por lo tanto necesitamos una prueba que nos indique si la serie de tiempo de nuestro caso de estudio es estacionaria. Dentro de las pruebas de estacionariedad más utilizadas se encuentra la prueba de Dickey-Fuller aumentada, la cual a base de un análisis en las raíces del polinomio determina el grado de estacionariedad de una serie de tiempo. De antemano sabemos que los datos históricos no son estacionarios, esto debido a que presentan estacionalidad y visibles desviaciones en la varianza según se aprecia en la gráfica 3.1. Sin embargo para comprobar nuestra hipótesis aplicaremos la prueba de Dickey-Fuller al os datos históricos de ventas; el código para esta prueba lo encontramos en el apéndice A. El resultado de la prueba de Dickey-Fuller para las ventas 2012-2013 es el siguiente:

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 1

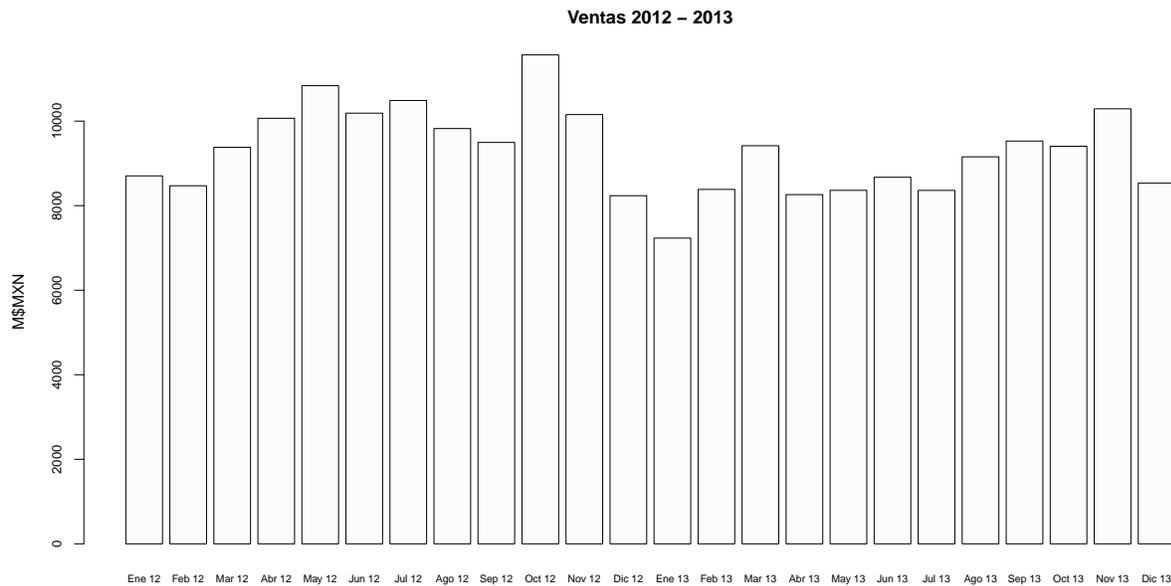
STATISTIC:

Dickey-Fuller: -0.1852

P VALUE:

0.5491

Calibraciones previas para el modelo de Dickey-Fuller con series altamente estacionarias y altamente no estacionarias, muestran que un p-value alto oscila entre 0.5 y 0.6. Un p-value



**Figura 3.1:** Ventas 2012-2013 (M\$MXN)

alto denota alta no estacionariedad ya que nuestra prueba de hipótesis alternativa es que la serie es estacionaria. Por lo tanto el que p-value de la prueba a nuestro caso de estudio sea de 0.5491, nos dice que efectivamente debemos hacer estabilizaciones en la media y varianza. La estabilización en la varianza la conseguiremos aplicando la transformada de Box-Cox 2.1, el valor de  $\lambda$  lo determináramos a través del método de máxima verosimilitud. Los resultados del parámetro calculado vs el estimador máximo verosímil los encontramos en la gráfica 3.2, de donde podemos notar que el valor de  $\lambda$  con el estimador máximo verosímil es  $\lambda=0.25$ . Una vez aplicada la transformada de Box-Cox con  $\lambda=0.25$  nuestros datos se comportan como se muestra en la gráfica 3.3. En el apéndice A encontramos el código para generar la gráfica 3.2 y obtener el valor de  $\lambda$ .

Antes de alcanzar la estacionariedad debemos estabilizar el nivel. Lo cual lograremos por medio de un número apropiado de diferenciaciones. El número apropiado de diferenciaciones lo determinaremos a través de una prueba Dickey-Fuller en cada diferenciación. Resultado de la prueba de estacionariedad a la primera diferenciación (código apéndice A).

Title:

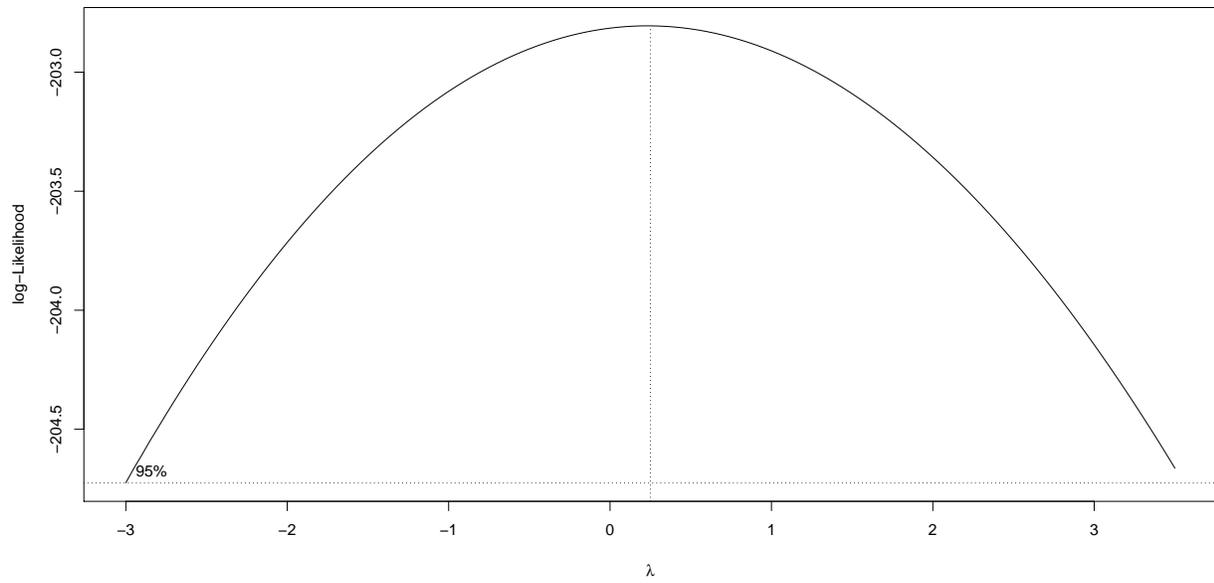
Augmented Dickey-Fuller Test

Test Results:

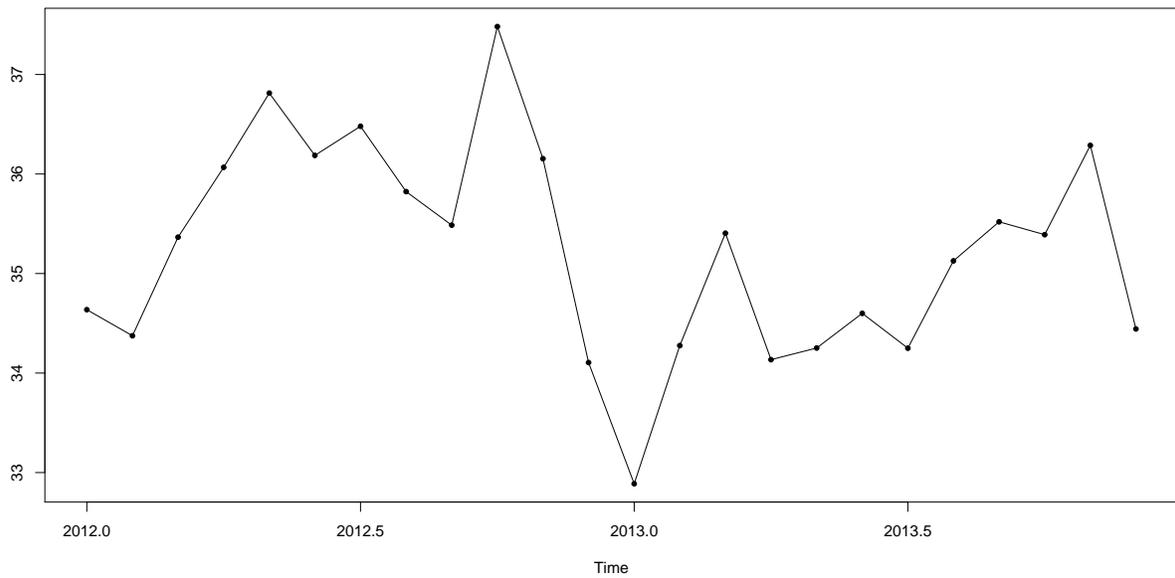
PARAMETER:

Lag Order: 1

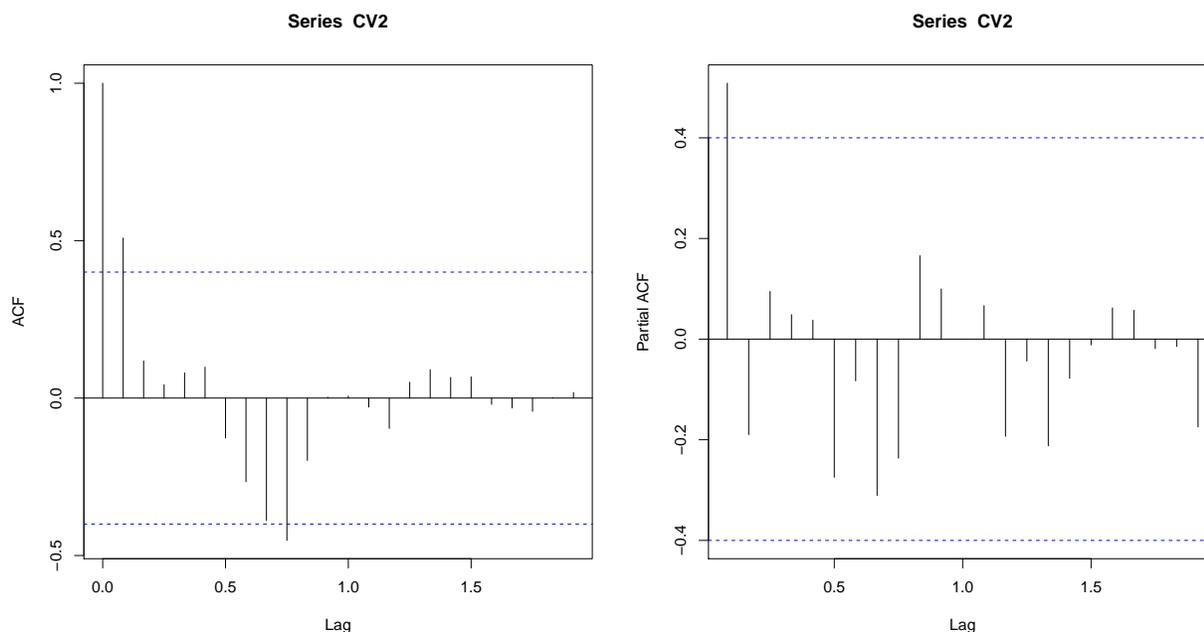
STATISTIC:



**Figura 3.2:** Parámetro calculado vs Estimador máximo verosímil



**Figura 3.3:** Datos resultantes de la Transformada de Box-Cox



**Figura 3.4:** Gráfica de autocorrelación (ACF) y Autocorrelación parcial (PACF)

Dickey-Fuller:  $-4.5803$   
P VALUE:  
0.01

Como se ve en la respuesta para la primera diferenciación. Tenemos un p-value suficientemente pequeño para asegurar que la serie de tiempo de nuestro caso de estudio con una transformación de Box-Cox con el valor adecuado de  $\lambda$  y una diferenciación, alcanza la estacionariedad.

Una vez que hemos alcanzado la estacionariedad y hemos determinado el orden de diferenciación "d", debemos identificar los órdenes p y q del proceso ARMA(p,q) que puede reproducir las características de la serie estacionaria. Para ello haremos la inspección de las gráficas de correlación (ACF) y correlación parcial(PACF). El código que encontramos en el apéndice A nos genera la gráfica que se muestra en la figura 3.4 que contiene al ACF y PACF correspondiente a la serie de tiempo de nuestro caso de estudio. Como se reviso en la sección 2, para el proceso AR(p) al examinar el PACF sólo las primera p autocorrelaciones son distintas de cero, y para el proceso MA(q) al examinar el ACF sólo las primera q autocorrelaciones son distintas de cero. Por lo que observamos en el PACF podemos decir que solo 2 autocorrelaciones son distintas de 0, por lo tanto  $p=2$ . De la observación del ACF podemos ver que ninguna autocorrelacion parcial es distinta de 0, por lo que  $q=0$ . Sin embargo al observar el ACF y PACF de la serie transformada, solo podemos hacernos una idea del modelo que describe nuestra serie, o al menos de cuáles son los primeros candidatos que debemos probar. Para comprobar analíticamente (no visualmente), ajustaremos varios mo-

---

delos ARIMA que pudieran modelar nuestra información, suponiendo que alguno de nuestros datos es extraordinario y esto provoco una alteración en el ACF y PACF provocando un  $p$  y  $0$   $q$  distinto a  $0$ . En base a esta suposición otros modelos que pudieran modelar nuestra serie de tiempo serían el ARIMA(0,1,1), ARIMA(1,1,1) y ARIMA(0,1,2). Escogeremos como un buen modelo aquel que tenga los residuales semejantes al de un ruido blanco, además que tenga los valores del AIC (Criterio de Información de Akaike) y BIC (Criterio de Información Bayesiana) menores con relación al resto de los modelos candidatos. El código para ajustar los distintos valores de  $(p,q)$  en un modelo ARIMA, se encuentra en el apéndice A. Mostraremos la respuesta que R nos genera para los 3 distintas ternas de valores  $(p,d,q)$ . Es decir, los coeficientes del modelo seleccionado, así como los valores del AIC y BIC.

- Arima(0,1,1)

```
Series: CV2
ARIMA(0,1,1)
```

```
Coefficients:
          ma1
      -0.2625
s.e.    0.4211
```

```
sigma^2 estimated as 1.058:  log likelihood=-33.32
AIC=70.64  AICc=71.24  BIC=72.92
```

- Arima(1,1,1)

```
Series: CV2
ARIMA(1,1,1)
```

```
Coefficients:
          ar1      ma1
      0.5720  -1.0000
s.e.  0.1875   0.1505
```

```
sigma^2 estimated as 0.834:  log likelihood=-31.54
AIC=69.08  AICc=70.34  BIC=72.48
```

- Arima(0,1,2)

```
ARIMA(0,1,2)
```

```
Coefficients:
          ma1      ma2
```

---

```
      0.0709  -0.9291
s.e.  0.2361   0.2294
```

```
sigma^2 estimated as 0.6657:  log likelihood=-29.89
AIC=65.78  AICc=67.04  BIC=69.18
```

Una vez comparadas las respuestas para los distintos modelos ARIMA, observamos que el menor AIC y BIC lo encontramos en el ARIMA(0,1,2) mismo que será el que ocupemos para la predicción de los valores futuros. Una vez identificado el modelo realizaremos el pronóstico de los valores futuros con el código que encontramos en el apéndice A, el cual nos arroja la respuesta para los próximos 12 meses así como las bandas de confianza al 80 % y 95 %

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2014	7932.079	7071.249	8869.212	6644.981	9397.678
Feb 2014	9084.236	7736.121	10601.311	7086.398	11477.405
Mar 2014	9084.236	7730.237	10608.769	7077.972	11489.511
Apr 2014	9084.236	7724.383	10616.195	7069.593	11501.571
May 2014	9084.236	7718.559	10623.592	7061.260	11513.586
Jun 2014	9084.236	7712.764	10630.960	7052.972	11525.556
Jul 2014	9084.236	7706.999	10638.298	7044.727	11537.481
Aug 2014	9084.236	7701.262	10645.607	7036.527	11549.362
Sep 2014	9084.236	7695.554	10652.888	7028.370	11561.201
Oct 2014	9084.236	7689.874	10660.141	7020.255	11572.997
Nov 2014	9084.236	7684.221	10667.366	7012.183	11584.751
Dec 2014	9084.236	7678.596	10674.564	7004.152	11596.463

En la gráfica 3.5 podemos observar el comportamiento de los valores futuros de las ventas.

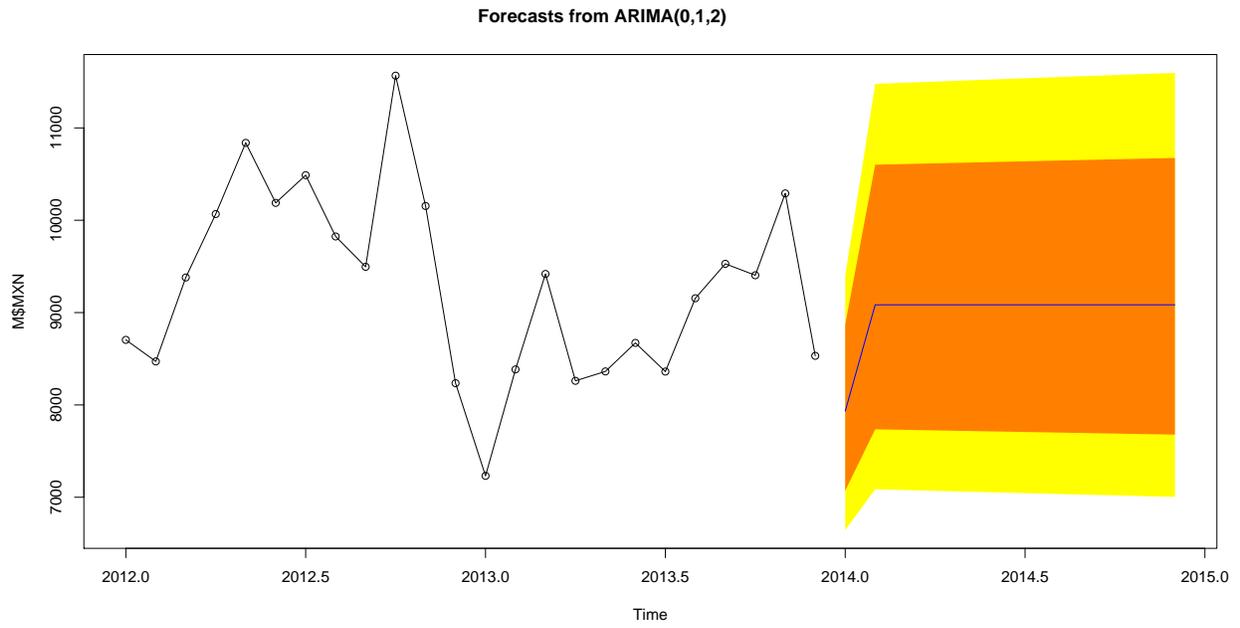
## Verificación del modelo

El siguiente paso es la verificación del modelo, para lo cual usaremos el código del apéndice A, de donde obtendremos el estadístico de Ljung-Box para los residuales, el ACF, la gráfica Normal cuantil cuantil y el resultado de la prueba de normalidad de Shapiro-Wilk. La respuesta de R a la prueba de Ljung-Box es la siguiente

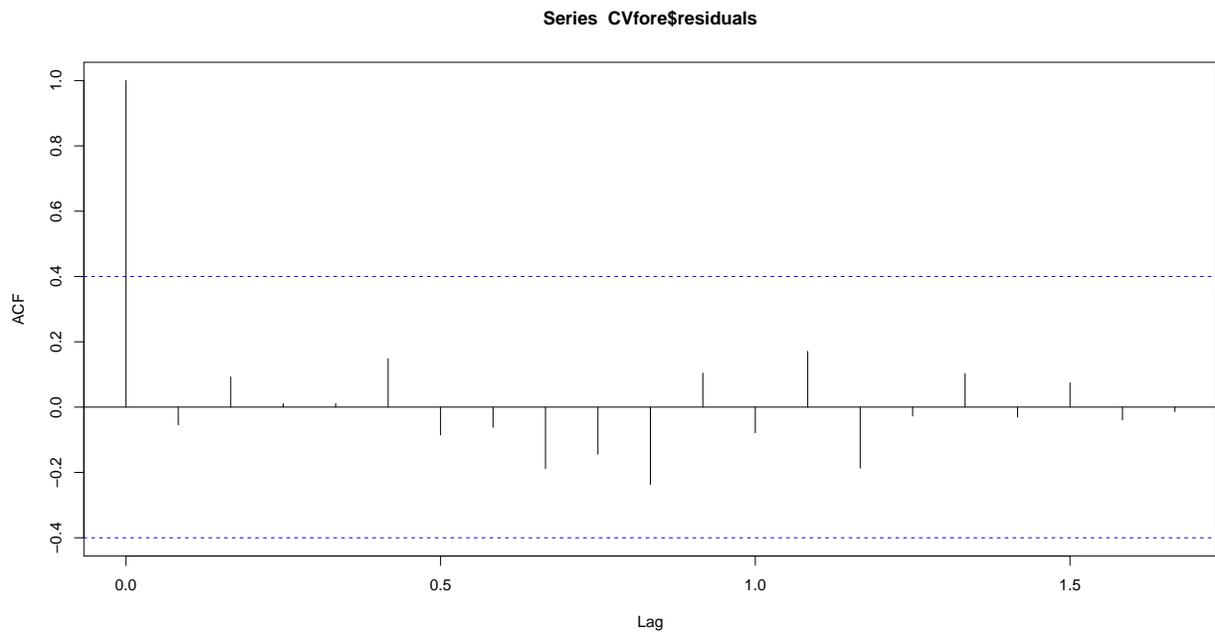
```
Box-Ljung test
```

```
data:  CVfore$residuals
X-squared = 12.6464, df = 20, p-value = 0.892
```

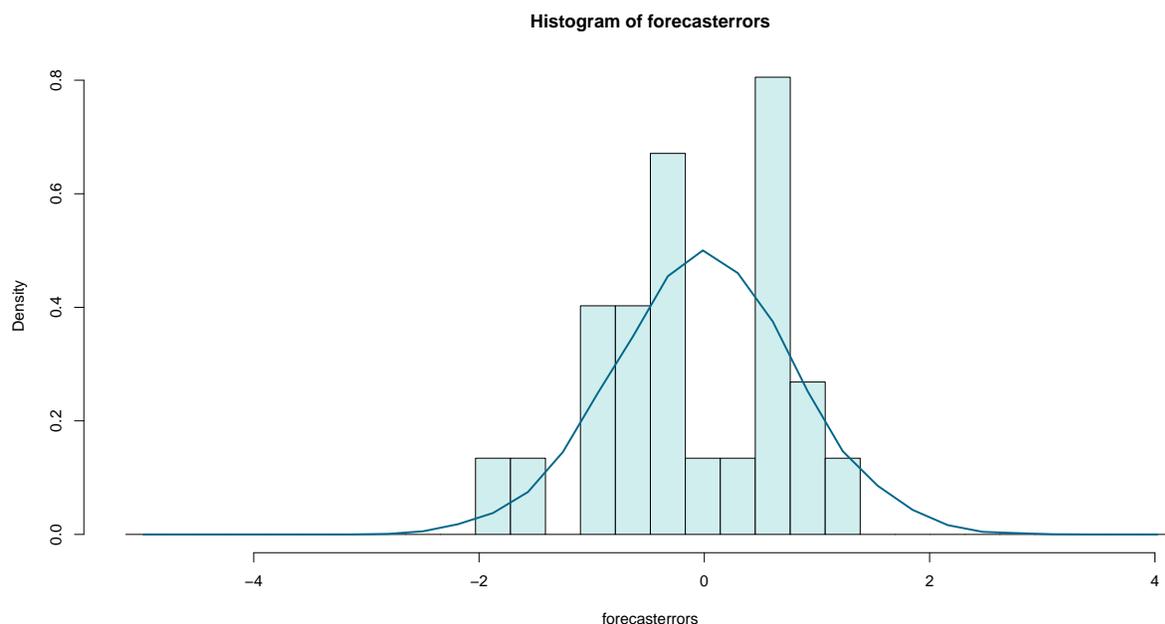
La prueba de Ljung-Box es usada para examinar la hipótesis nula de la independencia en una serie de tiempo.



**Figura 3.5:** Pronóstico de ventas 2014



**Figura 3.6:** ACF de los residuales del pronóstico



**Figura 3.7:** Histograma de los residuos con una curva normal superpuesta

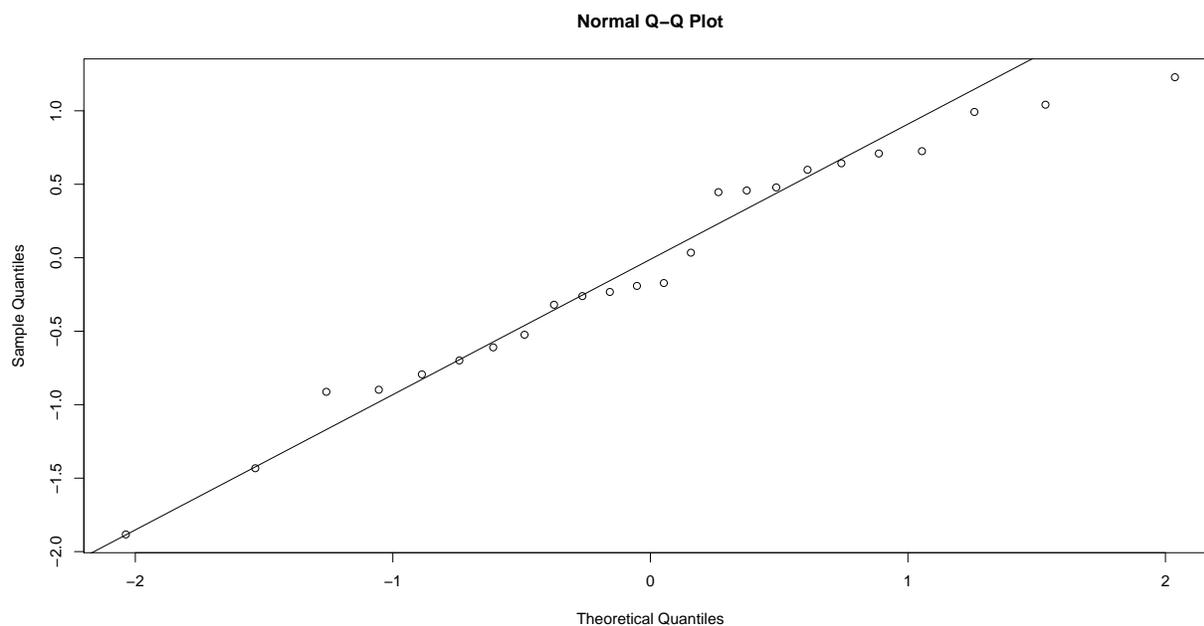
En el ACF (gráfica 3.6) se observa que ninguna de las autocorrelaciones exceden los intervalos confianza, por otra parte la respuesta a la prueba de Ljung-Box arroja un p-value de 0.892. Por lo tanto podemos concluir que hay muy poca evidencia de que las autocorrelaciones de los residuales (los errores del pronóstico) no son cero.

Para asegurarse de que el modelo predictivo no se puede mejorar, es una buena idea comprobar si los errores de pronóstico se distribuyen normalmente con media cero y varianza constante, para lo cual podemos hacer varias pruebas. La primera es trazar un histograma de los errores del pronóstico, con una curva normal superpuesta con media 0 y la misma desviación estándar de los residuales (gráfica 3.7). La segunda es aplicar la prueba de Shapiro-Wilk de Normalidad, misma que tiene como hipótesis nula el ajuste de normalidad de los datos. A continuación vemos la respuesta de R a la prueba de Shapiro-Wilk con los residuales de nuestro pronóstico.

#### Shapiro-Wilk normality test

```
data: CVfore$residuals
W = 0.9653, p-value = 0.554
```

La tercera prueba para detectar si nuestros residuales se distribuyen de manera normal, es mediante una gráfica Normal cuantil cuantil, la cual muestra los cuantiles teóricos vs cuantiles muestrales de los residuales de nuestra serie de tiempo (Gráfica 3.8).



**Figura 3.8:** Gráfica Normal cuantil cuantil de los residuales de la serie de tiempo

De las pruebas anteriores podemos concluir que nuestros datos se distribuyen normalmente con media 0 y varianza constante, por lo que podemos decir que el modelo elegido no se puede mejorar y el pronóstico obtenido es el mejor que se puede obtener por este modelo.

# Capítulo 4

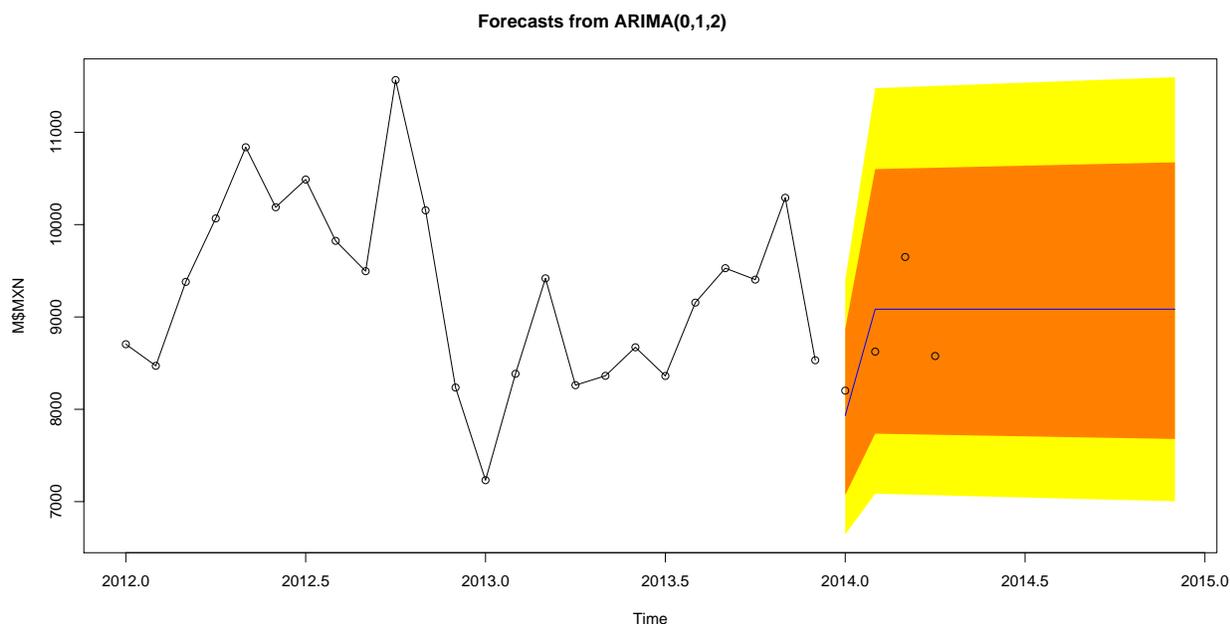
## Discusión, Conclusiones y Recomendaciones

### Discusión

En el caso de estudio se expuso a una compañía de manufactura se encuentra a unos meses de la junta de planeación de cierre de año, por lo que se necesita conocer cuál sería el pronóstico de las ventas de los próximos 12 meses. El área de Business Intelligence de la compañía (BI), es la encargada de entregar el pronóstico. Para llevar a cabo la proyección, los analistas de BI cuentan con la información de las ventas mensuales de los últimos 2 años de la empresa. El primer paso para realizar el pronóstico, es el tratamiento de los datos para garantizar la estacionariedad de la serie. Por lo tanto, a la información histórica se le aplicó una transformación de Box-Cox, la cual con el valor adecuado de  $\lambda$  ( $\lambda=0.25$ ) y una diferenciación ( $d=1$ ) alcanza la estacionariedad. Una vez que hemos determinado el orden de diferenciación "d", debemos identificar los órdenes p y q del proceso  $ARIMA(p,d,q)$ . Para ello se hizo la inspección del ACF y PACF, de lo que observamos en el PACF podemos decir que solo 2 autocorrelaciones son distintas de 0, por lo tanto  $p=2$ . De la observación del ACF podemos ver que ninguna autocorrelacion parcial es distinta de 0, por lo que  $q=0$ . Con el análisis del ACF y PACF podemos hacernos una idea del modelo que describe nuestra serie, o al menos de cuáles son los primeros candidatos que debemos probar. Los modelos que se propusieron para la inspección de los siguientes:  $ARIMA(0,1,1)$ ,  $ARIMA(1,1,1)$  y  $ARIMA(0,1,2)$ . Los valores del AIC y BIC se resumen en el cuadro 4.1, de donde se observa que el modelo  $ARIMA(0,1,2)$  es el que presenta el menor AIC y BIC. Se realizó la verificación

	ARIMA(0,1,1)	ARIMA(1,1,1)	ARIMA(0,1,2)
AIC	70.64	69.08	65.78
BIC	72.92	72.48	69.18

**Cuadro 4.1:** AIC y BIC para  $ARIMA(0,1,1)$ ,  $ARIMA(1,1,1)$  y  $ARIMA(0,1,2)$



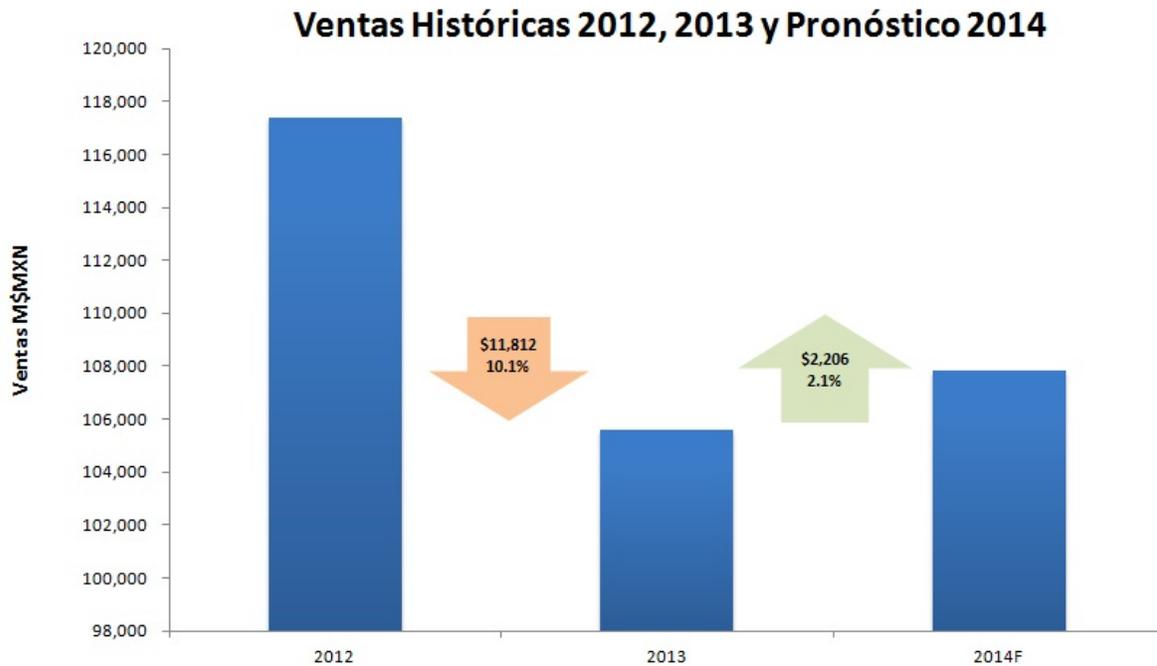
**Figura 4.1:** Pronóstico de ventas 2014 y Ventas reales 2014

del modelo ARIMA(0,1,2) por medio de las siguientes pruebas a los residuales: estadístico de Ljung-Box , el ACF, la gráfica Normal cuantil cuantil y el resultado de la prueba de normalidad de Shapiro-Wilk. Todas las pruebas indican que el modelo elegido no se puede mejorar y el pronóstico obtenido es el mejor que se puede obtener por medio de modelos ARIMA. El pronóstico de las ventas del año 2014 lo encontramos en la gráfica de la figura 3.5 donde podemos observar el comportamiento de los valores futuros de las ventas, se reviso que no se espera otra caída en las ventas para el 2014, si no un incremental de 2.1% vs el año 2013. También se mostro la gráfica 3.5 para dar a conocer la calendarización mensual. Como se reviso en el caso de estudio, este pronóstico fue solicitado por el equipo de liderazgo para tener una base solida de donde partir para fijar el objetivo de ventas de la empresa el próximo año. Con el transcurso del tiempo se puede ir actualizando y comprando las diferencias entre los valores pronosticados y los valores reales. En la gráfica de la imagen 4.1 podemos ver los valores reales de los primeros 4 meses del 2014. El código para generar la gráfica 4.1 lo encontramos en la última sección del apéndice A. En la tabla 4.2 encontramos la comparación entre los resultados del pronóstico y los valores reales. De donde podemos ver que el error entre los valores reales y los valores pronosticados oscila en un  $\pm 6\%$ . Conforme pasa el tiempo es necesario actualizar el pronóstico con los nuevos datos reales y hacer una recalibración periódica de las variables de la serie de tiempo.

El equipo de BI presentó varias gráficas al equipo de liderazgo para mostrar sus resultados. La gráfica 4.2, muestra la fuerte caída que hubo en las ventas entre el año 2012 y 2013 correspondiente a una baja del 10.1% en las ventas es decir, \$11,812 M de pesos. El primer punto a destacar del pronóstico de BI, es que no espera otra caída en las ventas para el

Mes	Valor real	Valor Pronosticado	Error
Jan 2014	8,202	7,932	3.3 %
Feb 2014	8,624	9,084	-5.3 %
Mar 2014	9,649	9,084	5.9 %
Apr 2014	8,577	9,084	-5.9 %

**Cuadro 4.2:** Comparación entre el pronóstico y las ventas reales 2014



**Figura 4.2:** Gráfica de Ventas 2012, 2013 y pronóstico 2014

próximo año, si no un incremental de \$2,206M de pesos en las ventas correspondiente al 2.1 % vs el año 2013. También se mostro la gráfica 3.5 para dar a conocer la calendarización mensual del pronóstico de las ventas en conjunto con los años anteriores. Estos datos serán utilizados por el equipo de liderazgo como base para fijar el objetivo de ventas de la empresa el próximo año.

## Conclusiones

En este documento, se mostró como realizar un pronóstico de ventas por medio de la metodología Box-Jenkins, de la siguiente manera: En el capítulo 1 se revisaron algunos conceptos básicos de probabilidad, como el espacio muestra y la  $\sigma$ -álgebra, conceptos con los

---

cuales logramos definir un conjunto que agrupa a todos los eventos de un mismo fenómeno aleatorio para los cuales se pueden definir o calcular sus probabilidades. También se definió a los espacios de probabilidad como el modelo matemático creado para estudiar a los experimentos aleatorios. Otro concepto revisado fue el de la variable aleatoria, que es una transformación  $X$  del espacio de resultados  $\Omega$  al conjunto de los números reales. Se explico a grandes rasgos a la función de probabilidad, a la función de distribución y se definió que un proceso estocástico es una colección de variables aleatorias  $\{X_t, t \in T\}$  indexadas por un conjunto  $T$  y definidas en algún espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ . El capítulo 1 nos da la base teórica necesaria para poder definir a las series de tiempo. En el capítulo 1 se definió como serie de tiempo  $\{X_t\}_{t=1}^T$  de datos de longitud  $T$  a la realización de un proceso estocástico. La estrategia de pronóstico revisada en este documento se encuentra enfocada a las series de tiempo estacionarias, por lo que se explicaron las condiciones necesarias para llamar estacionaria a una serie de tiempo así como los distintos tipos de estacionariedad. En el análisis de series de tiempo, existen dos representaciones muy útiles para expresar una serie de tiempo. Los modelos autoregresivos (AR) y los modelos de medias móviles (MA). Por lo que se explico que los modelos AR expresan la serie de tiempo como una función de valores previos y como los modelos MA representan un proceso estocástico como una suma finita ponderada de choques aleatorios independientes. Se explicó la generalización de los modelos AR y MA que consiste en combinar ambas clases de modelos para obtener lo que se conoce como modelos autoregresivos y de promedios móviles (ARMA). Los Modelos autoregresivos e integrados de promedios móviles (ARIMA) pueden ser vistos como una generalización de los modelos ARMA, ya que lo que en realidad hace es aplicar al operador diferencia  $\nabla^d$  es eliminar una posible tendencia polinomial. Con el capítulo 1, logramos entender teóricamente a las series de tiempo y sus característica. En el capítulo 2 se presentó la estrategia de construcción de modelos para series de tiempo estacionarias de Box-Jenkins, en el cual aplicamos todos los fundamentos revisados en los capítulos previos. Se estudiaron las 4 etapas de la metodología de Box y Jenkins que se resumen a continuación:

1. Identificación. En esta etapa se busca determinar los órdenes de los polinomios autoregresivo y de promedios móviles, es decir, aquellos que reproducen las características de la serie. En otras palabras la determinación de los valores  $p$ ,  $d$  y  $q$  de un modelo ARIMA.
2. Estimación. En esta etapa se realizan los cálculos de los parámetros a través de la función de verosimilitud la cual nos indica qué tan verosímil es que una muestra observada sea función de los posibles valores de parámetro. Llevar al máximo la verosimilitud dará los valores del parámetro para los cuales es más creíble que la muestra observada se haya generado, es decir, los valores de parámetro que concuerdan de manera más cercana con los datos observados.
3. Verificación. Aquí se verifica que el modelo proporcione un ajuste adecuado y se satisfagan los supuestos.
4. Uso del modelo. El uso del modelo se explica en el capítulo 3 con un caso de estudio,

---

donde se aplican de manera práctica cada una de las fases de la metodología de Box y Jenkins.

El presente documento es aplicable a cualquier industria donde las series de tiempo por la metodologías de Box-Jenkins puedan ser aplicadas. Es decir, donde se necesiten realizar pronósticos de corto y mediano plazos para datos económicos como el precio, inventario, producción y ventas, entre otros. En el campo de la ingeniería química un buen pronóstico sobre producción, es el que nos dará la pauta para determinar en qué punto del futuro será necesario nuevo equipo y de que capacidad sería el mismo (datos básicos para la planeación y diseño del equipo).

La metodología presentada no sustituye la guía y *expertise*<sup>1</sup> de los directores comerciales, si no que la complementa y orienta, para que puedan dar los ajustes finales de acuerdo a sus estrategias y objetivos comerciales.

## Recomendaciones

Las técnicas de elaboración de de pronósticos expuestas en este documento, implican la manipulación de datos históricos para generar pronósticos por medio del análisis de parametros estadísticos. Sin embargo el criterio del analista con respecto al valor final del pronóstico se encuentra muy al margen. El hecho de descartar el criterio en la interpretación del método cuantitativo para pronosticar, es definitivamente una visión errónea, ya que el buen juicio es un componente esencial en las técnicas de pronóstico. Los pronósticos de juicio o ajustes de pronóstico se basan generalmente en el dominio del conocimiento. El dominio del conocimiento es cualquier información relevante diferente de la serie de tiempo. De manera interesante, la investigación muestra que, cuando hay disponibles datos históricos, la modificación de los pronósticos producida por medios analíticos tiende a reducir la precisión de los pronósticos<sup>2</sup>. Esto pudiera atribuirse a alguna desviación por parte del pronosticador, debido quizá a una tendencia a ser demasiado optimista o a subestimar la incertidumbre futura. Vale la pena tomar en consideración estas técnicas, ya que con frecuencia los ejecutivos consideran su propio juicio superior a otros métodos de pronóstico, ya que consideran que su conocimiento del producto, del mercado y del cliente, así como su visión e información interior le proporciona el mejor pronóstico.

Se recomienda tener claras las siguientes preguntas antes de iniciar un proceso para realizar un pronóstico:

- ¿Por qué se requiere de los pronósticos?
- ¿Quién usará el pronóstico y cuáles son sus requerimientos específicos?

---

<sup>1</sup>especialista en algún conocimiento

<sup>2</sup>S. Makridakis, The Art and Science of Forecasting, International Journal 01 Forecasting, Vol. 2 (1986), p. 45.

- 
- ¿Qué nivel de detalle o agregación se requiere y cuál es el horizonte de tiempo apropiado?
  - ¿Qué datos hay disponibles y serán éstos suficientes para generar el pronóstico requerido?
  - ¿Qué tan preciso se puede esperar que sea el pronóstico?
  - ¿Se hará el pronóstico a tiempo para ayudar en el proceso de toma de decisiones?
  - ¿Hay disponible un proceso de retroalimentación para evaluar el pronóstico después de efectuado y para evaluar de acuerdo con ello el proceso de pronóstico?

El proceso real de pronóstico puede comenzar cuando se formulan y responden adecuadamente las preguntas anteriores. Se debe de tener una retroalimentación una vez que el proceso de pronóstico está en marcha, para determinar si se obtuvo la precisión suficiente y si la administración encuentra útil y de costo efectivo el pronóstico en el proceso de toma de decisiones.

Se recomienda hacer una revisión periodica en el proceso de pronóstico y reconsiderar los procedimientos que se están utilizando. Los pasos a seguir pudieran ser los siguientes:

- a) Se agregan al banco de datos los valores reales más recientes.
- b) Después de esta actualización de datos, se recalculan los parámetros utilizados en el modelo de pronóstico. Por ejemplo, los valores óptimos de las constantes de ponderación empleados, ya que tal vez estos pudieran cambiar de forma considerable, al agregar valores de datos más recientes. O los coeficientes en un análisis de regresión pudieran cambiar al ajustar diferentes valores de datos en una ecuación.
- c) Se examina el modelo de pronóstico con nuevos parámetros para una adecuada precisión. Si se juzga que esta precisión es suficiente, el modelo se usa como antes, hasta el siguiente periodo de actualización. Si la precisión del pronóstico se juzga inadecuada o marginal, se pueden revisar los patrones en los nuevos datos con la posibilidad de elegir un nuevo procedimiento de pronóstico. Este proceso continúa hasta que se juzga adecuada la precisión del modelo elegido, de acuerdo con la precisión de pronóstico de periodos históricos anteriores.

La idea consiste en establecer límites dentro de los cuales se espera que se ubiquen los errores generados por el pronóstico, si el proceso de pronóstico es adecuado. Mientras que los errores se ubiquen dentro de estos límites aceptables, el proceso de pronóstico continúa, tan pronto como un error cae fuera del margen aceptable, la atención debe enfocarse en el proceso de pronóstico y deben retomarse los pasos de actualización y revision.

# Bibliografía

- [1] PETER J. BROCKWELL, RICHARD A. DAVIS, *Introduction to Time Series and Forecasting*, primera edición, Springer, New York Estados Unidos, 1996, pp. 45-59, 179-203.
- [2] GEORGE C CANAVOS, *Probabilidad y estadística Aplicaciones y métodos*, Primera edición, Mc Graw Hill, Naucalpan Mexico, 1996, pp. 29,53,57.
- [3] CHRIS CHATFIELD, *The Analysis of Time Series: An Introduction*, sexta edición, Chapman Hall/CRC, Florida Estados Unidos, 2003, pp. 33-36, 68, 77-79.
- [4] VICTOR M. GUERRERO, *Análisis estadístico de series de tiempo económicas*, segunda edición, Thomson, DF México, 2004, pp. 5-13, 102-110, 128-134, 201-210.
- [5] JAMES D. HAMILTON, *Time Series Analysis*, primera edición, Princeto University Pres, New Jersey Estados Unidos, 1994, pp. 76-77.
- [6] JOHN E. HANKE, DEAN W WICHERN, *Pronósticos en los negocios*, novena edición, Pearson, DF México, 2010, pp. 62-69,79-81, 404-411, 481-483, 503-509.
- [7] S. MAKRIDAKIS, *The Art and Science of Forecasting*, International Journal 01 Forecasting, Vol. 2 (1986), Francia, p. 45
- [8] SHELDON M. ROSS, *A First Course in Probability*, novena edición, Pearson, New Jersey Estados Unidos, 2009, pp. 22-29, 58-65, 117-123, 232-240.
- [9] SHELDON M. ROSS, *Introduction to Probability and Statistics for Engineers and Scientists*, quinta edición, Academic Press, San Diego Estados Unidos, 2014, pp. 57-59, 89-95, 111-118, 235-244.
- [10] ROBERT H. SHUMWAY, *Time series analysis and its applications with R examples*, tercera edición, Springer, California Estados Unidos, 2011, pp. 22,57, 102-105, 108-111, 121-140, 144-154.
- [11] WILLIAM W. S. WEI, *Time series analysis, univariate and multivariate methods*, segunda edición, Pearson, Estados Unidos, 2006 pp. 6-15,17-22, 68-74, 88-94, 108-116, 136-130.

# Apéndice A

## Códigos en R

En este apéndice encontraremos los códigos en R que se mencionan a lo largo del documento

### Librerías, vectores, gráficas y estacionariedad

Debido a que R es un software muy usado en la comunidad estadística, constantemente se desarrollan de nuevas librerías que son agregadas a los repositorios de R. Los paquetes que se usaron para el desarrollo de este documento pueden ser instalados directamente de los repositorios de R de manera gratuita. En el menú de R, dentro de la opción de Paquetes. Se encuentra la opción de instalación, ahí seleccionando cualquier servidor es posible descargar los 3 paquetes necesarios, que son: MASS, fUnitRoots y forecast. Una vez instaladas estas librerías deben de ser llamadas al inicio del código.

Código para mandar a llamar a las librerías una vez instaladas

```
#Librerías necesarias
library(MASS)
library(fUnitRoots)
library(forecast)
#Fin librerías necesarias
```

Código para la construcción del vector que contiene las ventas del periodo 2012 y 2013

```
CV=c(8704.83 ,
8471.09 ,
9380.29 ,
10067.47 ,
10838.13 ,
10188.27 ,
```

---

```
10488.25 ,
9824.35 ,
9495.84 ,
11565.98 ,
10154.73 ,
8235.75 ,
7232.29 ,
8384.41 ,
9418.43 ,
8261.11 ,
8362.74 ,
8671.63 ,
8361.37 ,
9155.42 ,
9528.27 ,
9404.26 ,
10290.89 ,
8531.74 )
```

Código para convertir el vector en una serie de tiempo indexada por fecha y para la construcción de la gráfica de datos históricos.

```
CV = ts(CV, start=c(2012,1), frequency=12)
plot(CV,type="o",pch=20,ylab="Ventas (M$MXN)")
```

Código para la prueba de estacionariedad de Dickey-Fuller aumentada

```
adfTest(CV)
```

## Transformada de Box-Cox

Código para la determinación del parámetro  $\lambda$  de la transformada de Box-Cox

```
#Crea objeto necesario para ocupar en la funci'on Box-Cox
t=1:length(x)
plot(t,x)
trend = lm(x~t)
tCV=1:length(CV)
plot(tCV,CV,type="o",pch=20,ylab="")
CVobj = lm(CV~tCV)
#Fin Crear objeto

#Inicio transformada box cox

boxcox(CVobj, lambda = seq(-3, 3.5, 1/20), plotit = TRUE,
```

---

```

eps = 1/50, xlab = expression(lambda),ylab = "log-Likelihood")

temp2=boxcox(CVobj, lambda = seq(-3, 3.5, 1/20),
plotit = FALSE, eps = 1/50, xlab = expression(lambda),ylab = "log-Likelihood")
##Escoger lambda con el valor mas alto de log_likelihood
lambda2=temp2$x[match(c(max(temp2$y)),temp2$y)]
lambda2
# fin transformada box cox

```

## Transformación de datos y diferenciación

Código para aplicar la transformada de Box-Cox a los datos de la serie de tiempo original.

```

#Transformación de datos
CV2=(CV^(lambda2)-1)/lambda2
plot(CV2,type="o",pch=20,ylab="")
adfTest(CV2)
#Fin Transformación de Datos

```

Código para primera diferenciación y prueba de estacionariedad.

```

#Primera diferencia
CV2d1<-diff(CV2)
plot(CV2d1,type="o",pch=20,ylab="")
adfTest(CV2d1)
#Fin diferencia y prueba de estacionariedad

```

## ACF, PACF y Ajuste de modelo ARIMA

Código para crear el ACF y PACF de la serie de tiempo

```

#ACF & PACF
windows();par(mfrow=c(1,2)) #2 graficos en una ventana
ic=.95#intervalo de confianza
acf(CV2,lag=length(CV2),ci=ic); pacf(CV2,lag=length(CV2),ci=ic)
#Fin ACF & PACF

```

Código para ajustar los datos a un modelo Arima específico, en este caso (0,1,1), (1,1,1) y (0,1,2).

```

#Ajuste Orden Arima
arima(CV2,order=c(0,1,1))

```

---

```
arima(CV2,order=c(1,1,1))
arima(CV2,order=c(0,1,2))
#Fin ajuste ARIMA
```

## Pronóstico, verificación y actualización

Código para el pronóstico de la serie de tiempo.

```
#Forecast
windows();par(mfrow=c(1,1))
CVari <- arima(CV2, order=c(0,1,2))
CVfore <- forecast.Arima(CVari, h=24, lambda=lambda2,
                        seasonal=list(order=c(0,1,2),period=12))
CVfore$x=((lambda2*CVfore$x)+1)^(1/lambda2)
#Fin Forecast
```

Código para las pruebas de residuales (Verificación).

```
#Pruebas ajuste
acf(CVfore$residuals, lag.max=20) #ACF
Box.test(CVfore$residuals, lag=20, type="Ljung-Box") #Prueba de Ljung-Box
#Inicio codigo para crear Histograma con Normal superpuesta
plotForecastErrors <- function(forecasterrors)
{
  # make a histogram of the forecast errors:
  mybinsize <- IQR(forecasterrors)/4
  mysd <- sd(forecasterrors)
  mymin <- min(forecasterrors) - mysd*4
  mymax <- max(forecasterrors) + mysd*4
  # generate normally distributed data with mean
  #0 and standard deviation mysd
  mynorm <- rnorm(10000, mean=0, sd=mysd)
  mymin2 <- min(mynorm)
  mymax2 <- max(mynorm)
  if (mymin2 < mymin) { mymin <- mymin2 }
  if (mymax2 > mymax) { mymax <- mymax2 }
  # make a red histogram of the forecast errors,
  #with the normally distributed data overlaid:
  mybins <- seq(mymin, mymax, mybinsize)
  hist(forecasterrors, col="red", freq=FALSE, breaks=mybins)
  # freq=FALSE ensures the area under the histogram = 1
  # generate normally distributed data with mean 0 and standard
  #deviation mysd
  myhist <- hist(mynorm, plot=FALSE, breaks=mybins)
```

---

```

    # plot the normal curve as a blue line on top
    # of the histogram of forecast errors:
    points(myhist$mids, myhist$density, type="l", col="blue", lwd=2)
  }
#fin Histograma con Normal superpuesta
plotForecastErrors(CVfore$residuals) #crear Histograma con Normal superpuesta
shapiro.test(CVfore$residuals) #Prueba de Normalidad de Shapiro

qqnorm(CVfore$residuals) #Grafica Normal cuantil cuantil de muestra
qqline(CVfore$residuals) #Grafica Normal cuantil cuantil teorica

#####

Código para la actualización de valores reales.

# Actualizaci'on
plot(CVfore)
#puntos reales
CVupd=c(8202.09 ,
8624.56 ,
9649.79 ,
8577.12 )
CVupd = ts(CVupd, start=c(2014,1), frequency=12)
points(CVupd, cex = 1, col = "black")
#Fin Actualizaci'on

```