

## PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito parcial para  
obtener el Título de INGENIERO DE SISTEMAS

# USO DE LA MINERÍA DE DATOS Y SERIES TEMPORALES PARA PRONOSTICAR EL PRECIO DEL CAFÉ SEGÚN SU TIPO

[www.bdigital.ula.ve](http://www.bdigital.ula.ve) Por

Br. Jesús Miguel Herrera Díaz

Tutor: Prof. Francisco Barilla

Noviembre 2019

©2019 Universidad de Los Andes, Mérida, Venezuela



C.C. Reconocimiento

# **Uso de la minería de datos y series temporales para pronosticar el precio del café según su tipo**

Br. Jesús Miguel Herrera Díaz

Proyecto de Grado — Control y Automatizacion, 70 páginas  
Escuela de Ingeniería de Sistemas, Universidad de Los Andes, 2019

**Resumen:** Este trabajo se enfoca en la realización de un modelo, que mediante el uso de técnicas de minería de datos y de series de tiempo, sea capaz de predecir el precio del café según su tipo. Para esto se usarán los datos subministrados por el ICO, el cual es un órgano que se encarga de almacenar la variación del precio, producción, exportación entre otros datos relacionados al café desde 1990 hasta la actualidad. Haciendo uso de estos datos estos datos, luego de su validación y observando su grado de incidencia se construirán varios modelos con el fin de poder seleccionar el mejor de ellos por cada café.

**Palabras clave:** Minería de datos, Series de tiempo, Precio, Café, Predicción

Este trabajo fue procesado en L<sup>A</sup>T<sub>E</sub>X.

C.C. Reconocimiento

# Índice

<b>Índice de Tablas</b>	vii
<b>Índice de Figuras</b>	viii
<b>Agradecimientos</b>	xi
<b>1 Introducción</b>	1
1.1 Antecedentes . . . . .	2
1.2 Planteamiento del Problema . . . . .	3
1.3 Objetivos . . . . .	3
1.3.1 Objetivo General . . . . .	3
1.3.2 Objetivos Específicos . . . . .	3
1.4 Metodología . . . . .	4
1.5 Alcances . . . . .	7
<b>2 Marco teórico</b>	8
2.1 Organización . . . . .	8
2.1.1 International Coffee Organization . . . . .	8
2.2 Tipos de café . . . . .	11
2.2.1 café arabica . . . . .	11
2.2.2 Café robusta . . . . .	12
2.3 Minería de datos . . . . .	12
2.3.1 Definición . . . . .	12
2.3.2 Área de aplicación . . . . .	12
2.3.3 Objetivos . . . . .	13

2.3.4	Minería de datos en las organizaciones . . . . .	13
2.4	Reglas de asociación . . . . .	15
2.4.1	Algoritmo Apriori . . . . .	15
2.5	Series de tiempo . . . . .	16
2.5.1	Definición . . . . .	16
2.5.2	Componentes de una serie de tiempo . . . . .	16
2.6	Modelos Arima . . . . .	17
2.6.1	Modelos autorregresivos . . . . .	18
2.6.2	Modelo de medias móviles . . . . .	18
2.6.3	Modelos de Box-Jenkins . . . . .	19
2.6.4	Error Porcentual Absoluto Medio . . . . .	20
<b>3</b>	<b>Marco Metodológico</b>	<b>22</b>
3.1	Aplicación de las reglas de asociación . . . . .	22
3.2	Fases 1 y 2 de la Metodología CRISP-DM . . . . .	23
3.2.1	Compresión y análisis del problema . . . . .	23
3.2.2	Entendimiento de datos . . . . .	23
3.3	Fase 3: Preparación de los datos . . . . .	23
3.3.1	Estacionariedad . . . . .	25
3.4	Fase 4: Modelación . . . . .	35
3.4.1	Identificación . . . . .	35
3.4.2	Estimación . . . . .	41
3.5	Fase 5: Evaluación . . . . .	44
3.5.1	Evaluación del modelo . . . . .	44
3.5.2	Pronóstico . . . . .	49
<b>4</b>	<b>Análisis de resultados</b>	<b>59</b>
4.1	Análisis de las reglas de asociación . . . . .	59
4.1.1	Relaciones 1 a 1 más relevantes . . . . .	60
4.1.2	Relaciones 2 a 1 más relevantes . . . . .	60
4.1.3	Relaciones 3 a 1 más relevantes . . . . .	61
4.1.4	Relaciones 4 a 1 más relevantes . . . . .	62

4.2	Análisis de las series temporal . . . . .	62
4.2.1	Café suaves colombiano Arabico . . . . .	62
4.2.2	Café suaves brasileños Arabico . . . . .	63
4.2.3	Café robusta . . . . .	64
4.2.4	Otros café suaves Arabico . . . . .	64
4.2.5	Indicador compuesto del OIC . . . . .	66
<b>5</b>	<b>Conclusiones y Recomendaciones</b>	<b>67</b>
	<b>Bibliografía</b>	<b>69</b>

www.bdigital.ula.ve

C.C. Reconocimiento

# Índice de Tablas

3.1	Variables con su descripción . . . . .	24
3.2	Valores reales con valores pronosticados del cafe colombiano . . . . .	54
3.3	Valores reales con valores pronosticados del cafe Brasileño . . . . .	55
3.4	Valores reales con valores pronosticados del cafe Robusta . . . . .	56
3.5	Valores reales con valores pronosticados de otros cafe suaves arabicos .	57
3.6	Valores reales con valores pronosticados del coeficiente compuesto del OIC	58

www.bdigital.ula.ve

# Índice de Figuras

1.1	Visión general CRISP-DM . . . . .	4
1.2	Fases de la Metodología CRISP-DM . . . . .	5
2.1	Fases de la Metodología Box - Jenkins . . . . .	19
2.2	Error Porcentual Absoluto Medio . . . . .	20
3.1	Base de Dato del precio del café segun el OIC . . . . .	25
3.2	Serie de tiempo Café colombiano . . . . .	26
3.3	funcion de Autocorrelación . . . . .	27
3.4	Serie de tiempo diferenciada del Café colombiano . . . . .	27
3.5	Serie de tiempo Café brasileño . . . . .	28
3.6	funcion de Autocorrelación . . . . .	28
3.7	Serie de tiempo diferenciada del Café Brasileño . . . . .	29
3.8	Serie de tiempo Café Robusto . . . . .	30
3.9	funcion de Autocorrelación . . . . .	30
3.10	Serie de tiempo diferenciada del Café Robusta . . . . .	31
3.11	Serie de tiempo Café otras arabicas . . . . .	31
3.12	funcion de Autocorrelación . . . . .	32
3.13	Serie de tiempo diferenciada del otros café arabica . . . . .	33
3.14	Serie de tiempo del Coeficiente compuesto del OIC . . . . .	33
3.15	funcion de Autocorrelación . . . . .	34
3.16	Serie de tiempo diferenciada de otros café arabicas . . . . .	34
3.17	Función de autocorrelación del café colombiano . . . . .	36
3.18	Función de autocorrelación parcial del café colombiano . . . . .	36
3.19	Función de autocorrelación del café brasileño . . . . .	37

3.20	Función de autocorrelación parcial del café brasileño . . . . .	37
3.21	Función de autocorrelación del café Robusta . . . . .	38
3.22	Función de autocorrelación parcial del café Robusta . . . . .	38
3.23	Función de autocorrelación de otros café suaves arabicos . . . . .	39
3.24	Función de autocorrelación parcial de otros café suaves arabicos . . . . .	40
3.25	Función de autocorrelación del café otros suaves arabicos . . . . .	40
3.26	Función de autocorrelación parcial del café otros suaves arabicos . . . . .	41
3.27	Parametros del modelo arima(3,1,2) para el precio del café colombiano	41
3.28	Parametros del modelo arima(3,1,2) para el precio café otros brasileño .	42
3.29	Parametros del modelo arima(0,1,5) para el precio del café robusta . . .	43
3.30	Parametros del modelo arima(3,1,2) para el precio de otros café suaves arabicos . . . . .	43
3.31	Parametros del modelo arima(2,1,2) para el precio compuesto del OIC .	44
3.32	Análisis residual del modelo arima(3,1,2) para el precio del café colombiano	45
3.33	Análisis residual del modelo arima(3,1,2) para el precio del café brasileño	46
3.34	Análisis residual del modelo arima(0,1,5) para el precio de café robusta	47
3.35	Análisis residual del modelo arima(3,1,2) para el precio de otros café suaves arabicos . . . . .	48
3.36	Análisis residual del modelo arima(2,1,2) para el precio compuesto del OIC . . . . .	48
3.37	Grafica de pronóstico por 30 meses del café colombiano . . . . .	50
3.38	Grafica de pronóstico por 30 meses del café brasileño . . . . .	50
3.39	Grafica de pronóstico por 30 meses del café robusta . . . . .	51
3.40	Grafica de pronóstico por 30 meses de otros café suave arabicos . . . .	52
3.41	Grafica de pronóstico por 30 meses del indicador compuesto del OIC .	52
4.1	Relación soporte-confianza de las 291 reglas de asociación . . . . .	59
4.2	Grafica de la serie de tiempo del precio del café colombiano con su respectivo pronóstico . . . . .	63
4.3	Grafica de la serie de tiempo del precio del café brasileño con su respectivo pronóstico . . . . .	64

4.4	Grafica de la serie de tiempo del precio del café robusta con su respectivo pronóstico . . . . .	65
4.5	Grafica de la serie de tiempo del precio de otros café con su respectivo pronóstico . . . . .	65
4.6	Grafica de la serie de tiempo del precio del indicador compuesto del OIC con su respectivo pronóstico . . . . .	66

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

C.C. Reconocimiento

# Capítulo 1

## Introducción

Los granos del café son uno de los principales productos de origen agrícola que se comercializan en los mercados internacionales siendo el segundo producto más comercializado después del petróleo en el mundo, originando a menudo una gran contribución a los rubros de exportación en las regiones productoras. El cultivo del café está culturalmente ligado a la historia y al progreso de muchos países en donde se ha producido por más de un siglo. La estadística del café estima que 125 millones de personas que viven del cultivo del café, incluyendo 25 millones de pequeños productores. Cada año se beben alrededor 400 000 millones de tazas de café en el mundo. Por lo tanto, en juego hay muchos intereses económicos y sociales extremadamente importantes.

Los mayores exportadores del café son los sudamericanos. Colombia, Brasil y Honduras han exportado, desde hace décadas, millones de toneladas de este producto a todo el mundo. A inicios del siglo XXI los precios del café subieron (en promedios mensuales segun el índice OIC entre 78 a 101,44 céntimos de dólar estadounidense por libra). Esta subida fue causada por un aumento del consumo en Rusia y China, así como una cosecha entre un 10 % y un 20 % inferior a la registrada en los años anteriores. Actualmente muchos agricultores de café pueden vivir de sus productos, pero no en todas las etapas, pues el aumento del precio del petróleo encarece los costes de transporte, la torrefacción y el empaquetado de los granos de café, produciendo una

variación en los precios finales.

## 1.1 Antecedentes

(Figueroa et al. 2019), elaboraron dos modelos que fueron estimados mediante regresión de componentes principales. Tomando como variables explicativas la exportación y los distintos precios del café. Con el fin de analizar el efecto de los precios de los grupos de café clasificados por la Organización Internacional del Café (OIC), así como los precios de la bolsa de valores de Nueva York y de Londres, en la producción y las exportaciones del aromático. Concluyendo que los precios internacionales fueron los que influyeron en mayor medida en la variabilidad de la producción mundial y las exportaciones de café.

(Hernandez 2019), modeló la señal del precio del café colombiano por medio de dos sistemas dinámicos no lineales ajustando sus parámetros con un algoritmo de colonia artificial de abejas (ABC). Los resultados obtenidos permitieron reconstruir el pasado, capturar los atributos de la señal original y hacer una predicción en términos de un ensamble de modelos.

(Moreno 2018), planteó un modelo basado en autómatas celulares para la identificación del precio interno diario del café colombiano. Los parámetros del modelo fueron ajustados con la ayuda de un algoritmo genético y una función objetivo que captaba algunas características como tendencias, cambios de dirección y valores pico de la señal original. Concluyendo que la dinámica del precio del café era determinada por más de un elemento con comportamiento caótico, ya que eran necesarios diez autómatas para generar una predicción fiable

(Nieto & Velásquez 2018), midieron el impacto del precio del café sobre la producción de este, durante el periodo 2002 a 2017. A través de un modelo multivariado, obteniendo como resultado del modelo, que el precio internacional generó un impacto negativo sobre la producción.

Durante los últimos años el crecimiento en el consumo de café ha traído consigo un aumento en la producción de café, ocasionando que el precio de las diferentes cepas vaya variando mensualmente, observado que el café es el segundo producto con mayor

exportación en el mercado internacional solo por detrás del petróleo, y conociendo que los datos históricos representan una información útil para creación de modelos temporales, es de interés investigar cómo puede variar el precio de cada café a través del tiempo, con el fin de poder analizar la dinámica del precio en este rubro con el pasar de los años.

## 1.2 Planteamiento del Problema

La Organización Internacional del Café (OIC) es una organización intergubernamental que se ocupa de asuntos cafeteros. La OIC reúne a gobiernos de países exportadores e importadores para abordar, mediante la cooperación internacional, los desafíos con que ha de enfrentarse el sector cafetero mundial. Con el fin de fortalecer el sector cafetero mundial y promover su expansión en un entorno basado en el mercado para beneficio de todos los participantes en el sector.

Al ser el café un producto de alto impacto en el sector económico en los últimos 200 años, y a que muchos países se benefician económicamente de su producción, debido a su alto consumo a nivel mundial; Conociendo el precio del café desde 1990 hasta la actualidad gracias a los datos recogidos por el OIC, es de sumo interés realizar un estudio de minería de datos y series temporales, con el cual poder realizar una predicción del precio del café según su tipo.

## 1.3 Objetivos

### 1.3.1 Objetivo General

Pronosticar mediante el uso de minería de datos y series temporales el precio del café según su tipo.

### 1.3.2 Objetivos Específicos

- Indagar sobre los países más productores y más comercializadores de café.

- Estudiar la base de datos y sus variables.
- Limpiar y validar los datos.
- Indicar que variables o factores económicos inciden en los precios del café.
- Obtener varios modelos de pronóstico de precios del café.
- Seleccionar el mejor modelo
- Predecir los precios del café por tipos usando el modelo seleccionado.
- Determinar cuales son los tipos de café más comercializados y que mantienen mejores indices de precios a través del tiempo.

## 1.4 Metodología

www.bdigital.ula.ve

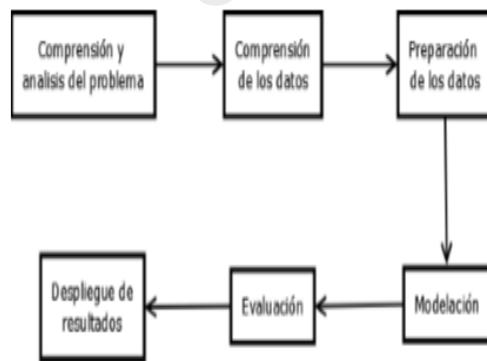


Figura 1.1: Visión general CRISP-DM

Para este trabajo se utilizara la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), la cual es una de las metodologías más usadas en minería de datos. En la figura 1.1 se muestra una visión general de las fases.

C.C. Reconocimiento

CRISP-DM, fue creada por el grupo de empresas SPSS, NCR y Daimer Chrysler en el año 2000, es actualmente la guía de referencia más utilizada en el desarrollo de proyectos de minería de datos. Estructura el proceso en seis fases. La sucesión de fases, no es necesariamente rígida. Cada fase es descompuesta en varias tareas generales de segundo nivel. Las tareas generales se proyectan a tareas específicas, pero en ningún momento se propone como realizarlas, es decir, CRISP-DM establece un conjunto de tareas y actividades para cada fase del proyecto pero no especifica cómo llevarlas a cabo. Estas actividades dependerán del propósito del proyecto que se desarrolle.(Moine et al. 2010)



Figura 1.2: Fases de la Metodología CRISP-DM

CRISP-DM, es el método que se adapta mejor para el desarrollo de este proyecto, y por ende, se tomará como referencia en este proyecto. En la figura 1.2 se detallan las fases que comprenden dicha metodología

Según (Chapman & Clinton 2000) las fases de la metodología CRISPDPM se explican a continuación:

1. **Análisis del problema:** Incluye la comprensión de los objetivos y requerimientos del proyecto, con el fin de convertirlos en objetivos técnicos y en una planificación, pensando en el cómo abordar el problema y las diferentes opciones con las que se podrían trabajar.
2. **Comprendión de datos:** Comprende la recolección inicial de datos, de modo que sea posible establecer un primer contacto con el problema, identificando la calidad de los datos y estableciendo las relaciones más evidentes que permitan establecer las primeras hipótesis. En esta fase se debe analizar y predefinir la manera en la que se abordará el problema en materia de software e incluso técnicas, ya que la siguiente fases dependerá de estas decisiones.
3. **Preparación de datos:** Incluye las tareas generales de selección de datos a los que se va a aplicar la técnica de modelado (variables y muestras), limpieza de los datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato. Esta fase se encuentra relacionada con la fase de modelado, puesto que en función de la técnica de modelado que vaya a ser utilizada, los datos necesitan ser procesados para que puedan ser aplicados a ese software y técnicas en particular. Por lo tanto, las fases de preparación y modelado interactúan de forma sistemática.
4. **Modelado:** Se seleccionan de manera definitiva las técnicas de modelado más apropiadas para el proyecto en función de los siguientes criterios: ser apropiada al problema, disponer de datos adecuados y conocimiento de la técnica. Los parámetros utilizados en la generación del modelo dependen de las características de los datos.
5. **Evaluación:** se evalúa el modelo, no desde el punto de vista de los datos, sino desde el cumplimiento de los criterios de éxito del problema. Se debe revisar el proceso seguido, teniendo en cuenta los resultados obtenidos para poder repetir algún paso en el que, a la vista del desarrollo posterior del proceso, se hayan podido cometer errores. Si el modelo generado es válido en función de los criterios de éxito establecidos en la primera fase, se procede a la aplicación del modelo.

## 1.5 Alcances

En esta tesis se cumplen importantes metas como lo es obtener modelos de pronósticos de precios del café internacional, de esta forma se puede saber cual es el tipo de café más comercializado a nivel mundial y cual mantiene un mejor precio a través del tiempo.

Otro importante alcance es la determinación de las variables o factores económicos que tienen una mayor incidencia en los precios del café.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

C.C. Reconocimiento

# Capítulo 2

## Marco teórico

### 2.1 Organización

Según Koontz & Weihrich (1999), la organización se define como la identificación, clasificación de actividades requeridas, conjunto de actividades necesarias para alcanzar objetivos, asignación de un grupo de actividades a un administrador con poder de autoridad, delegación, coordinación y estructura organizacional.

#### 2.1.1 International Coffee Organization

##### Reseña histórica

La Organización Internacional del Café fue establecida en 1963, a raíz de la entrada en vigor en 1962, del primer Convenio Internacional del Café por un período de cinco años, y viene funcionando desde entonces al amparo de sucesivos Convenios.ICO (2013)

Tras una serie de convenios a corto plazo entre países productores, se creó un Grupo de Estudio del Café, encargado de considerar la negociación de un Convenio en el que participasen tanto los países exportadores como los importadores. Las tareas del Grupo de Estudio tuvieron por feliz conclusión la negociación, en la sede de las Naciones Unidas en Nueva York, del Convenio Internacional del Café de 1962. Un segundo Convenio, de cinco años de vigencia, fue negociado en 1968. En uno y otro de los citados Convenios figuraban estipulaciones en las que se establecía un sistema

de cuotas en virtud del cual se retiraban del mercado las cantidades de café que representaban un exceso de la oferta con respecto a la demanda de los consumidores. Otras disposiciones sirvieron para iniciar políticas de producción y diversificación para limitar la oferta de café, emprendiéndose asimismo actividades de promoción para aumentar el consumo. ICO (2013)

El Convenio Internacional del Café de 1976 fue negociado en 1975, contra el telón de fondo de una situación radicalmente diferente de la que reinaba en el mercado durante las negociaciones de los Convenios de 1962 y 1968, fechas en las que el exceso de la oferta con respecto a la demanda de los consumidores ejercía presión en sentido descendente sobre los precios. En 1975, y debido primordialmente a las graves heladas que aquejaron al Brasil, principal país productor, se plantearon dudas en cuanto a la capacidad de la oferta para satisfacer la demanda en el futuro inmediato, con el consiguiente y acusado aumento de los precios. Consideraciones de ese orden hicieron que los Miembros, al negociar el Convenio de 1976 no se limitaran a conservar muchas de las estipulaciones cuya eficacia había quedado demostrada durante la vigencia de los anteriores sino que concertasen además otras, encaminadas a fortalecer y mejorar el funcionamiento de la Organización. ICO (2013)

Una de las principales novedades del Convenio de 1976 consistió en estipular la suspensión de cuotas en caso de alza excesiva de los precios y su restablecimiento en caso de baja excesiva. En virtud de ese sistema se implantó el régimen de cuotas en 1980. ICO (2013)

Las principales características del Convenio de 1983, al igual que en todos los anteriores Convenios, la Organización recopilaba y difundía datos acerca de todas las cuestiones relativas al café, con miras a facilitar la rápida puesta en práctica de los artículos de carácter económico del Convenio y la corrección de los desequilibrios que pudieran presentarse. Actuaba también la Organización como centro de estudios e investigaciones económicas de la producción, distribución y consumo del café. La información estadística obtenida de los Miembros y la resultante del funcionamiento

del sistema de control fue informatizada, con miras a lograr la mayor rapidez de acceso y análisis. Se creó una base de datos, llamada COFFEELINE, que facilita al público un servicio de información sobre muy variados aspectos del café. ICO (2013)

El proceso de negociación de un nuevo Convenio ganó renovado ímpetu con el descenso de los precios hasta niveles sin precedentes que caracterizó a los años cafeteros de 1990/91 y 1991/92, y el Consejo acordó prorrogar una vez más el Convenio hasta el 30 de septiembre de 1993. Decidió al mismo tiempo el Consejo crear un Grupo de Trabajo, al que encargó que llevase a cabo un amplio examen de todas las propuestas e ideas que se formulasesen por lo que se refiere a una futura cooperación en cuestiones cafeteras. ICO (2013)

Durante el Convenio de 2001 la OIC obtuvo US 45,2 millones en financiación para 20 proyectos y puso en práctica un Programa de Mejora de la Calidad del Café para mejorar la calidad de la oferta del café mundial. El Director Ejecutivo introdujo un informe mensual sobre el mercado de café para aumentar la transparencia del mercado y se publicó una Guía para promover el consumo de café que formó parte de un plan de acción para promover el consumo.ICO (2013)

Los 77 países Miembros del Consejo Internacional del Café, reunidos en Londres, aprobaron el 28 de septiembre de 2007, el texto del Acuerdo Internacional del Café de 2007, que es el séptimo de los convenios cafeteros. El Acuerdo fue oficialmente adoptado por el Consejo en virtud de la Resolución Número 431 y entró en vigor el 2 de febrero 2011. El Acuerdo fortalecerá las funciones de la OIC como foro para consultas intergubernamentales, facilitará el comercio internacional mediante al aumento de la transparencia y del acceso a la pertinente información, y fomentará una economía cafetera sostenible en beneficio de todos los interesados y, en especial, de los pequeños caficultores en los países productores de café. ICO (2013)

## Misión

La misión de la OIC es fortalecer el sector cafetero mundial y promover su expansión sostenible en un entorno basado en el mercado para beneficio de todos los participantes en el sector. La OIC contribuye de una manera práctica al desarrollo de un sector cafetero mundial sostenible y a reducir la pobreza en los países en desarrollo.

## 2.2 Tipos de café

### 2.2.1 café arabica

La especie más conocida y valorada en el mundo es el café arábica, se cultiva desde hace más de 1000 años, presenta un aroma muy perfumado, dulce y con un toque ligeramente ácido, en el café arábica distinguimos además una crema color avellana claro que tiende al rojizo, de gran persistencia gracias a los aceites que caracterizan esta variedad.

La variedad arábica es originaria de Etiopía, pero su expansión ha dado lugar a diferentes regiones productoras y multitud de matices organolépticos, en cafés Candelas queremos contarte como es el proceso de extracción de este tipo de café en dos de los principales países exportadores: Brasil y Colombia.

#### Arábico brasileño

El café arábica cultivado en Brasil se caracteriza por ser un café de cuerpo medio, sabor equilibrado y ligeramente dulzón, además presenta notas de frutos secos y cacahuetes con cierto tono achocolatado. Es un producto con gran capacidad de maridaje a la perfección el sabor suave y el dulce, siendo así ideal para el disfrute del paladar.

#### Arábico colombiano

En general el café colombiano se considera un café suave, de aroma delicado y gusto afrutado convirtiéndose en una opción ideal para quienes prefieren un café suave y gustoso.

### 2.2.2 Café robusta

es una especie de café (género Coffea) originaria del África occidental. Crece sobre todo en África y en Brasil, donde se conoce con el nombre de Conillon. También se halla en el sureste asiático, donde los colonialistas franceses introdujeron el cultivo en Vietnam a fines del s. XIX, y de allí pasó a Brasil. Es más fácil de cuidar que el café arabica y más barato de producir teniendo un precio de mercado internacional por debajo del valor del dólar. La especie robusta tiene entre el doble y el triple de cafeína que el cafe arabica así como un elevado porcentaje de materiales pesados.

## 2.3 Minería de datos

### 2.3.1 Definición

Es el proceso de descubrir conocimiento interesante de grandes cantidades de datos almacenadas en bases de datos, data warehouses u otro repositorio de información. Ham & Kamber (2001)

### 2.3.2 Área de aplicación

- Medicina: biomedicina, efectos colaterales de medicamentos, análisis de secuencias genéticas y predicción
- Finanzas: aprobación de créditos, predicción de quiebras, predicción en el mercado de valores, detección de fraudes, detección de acceso no autorizado a datos de crédito
- Agricultura: clasificación de enfermedades en tomates y soya
- Social: datos demográficos, tendencias electorales
- Mercadotecnia y ventas: identificación de sub-grupos socioeconómicos que muestran comportamiento inusual, patrones de compra, análisis de productos.

- Publicidad: ediciones personalizadas.
- Minería web, minería de blogs, minería de opinión, minería de sentimientos.

### 2.3.3 Objetivos

La minería de datos persigue dos objetivos de manera general, según Carrillo & Viserda (2007):

- Descripción: el principal producto del proceso de la minería de datos es el descubrimiento de reglas. Estas mostrarán nuevas relaciones entre las variables o excepciones de acuerdo a la empresa en que se utilice este proceso, lo cual enriquecerá el análisis y la descripción de la forma en que trabaja una compañía y ayudará en la planificación y en el diseño de futuros cambios.
- Predicción: una vez descubiertas reglas importantes, estas pueden ser utilizadas para estimar algunas variables de salida. Puede ser en el caso de secuencias en el tiempo, o bien; en la identificación e interrupción a tiempo de una futura mala experiencia de crédito. En esta tarea, se complementan las técnicas estadísticas tradicionales con aquellas provenientes de la inteligencia artificial.

### 2.3.4 Minería de datos en las organizaciones

La minería de datos es conocida por ser usada en diferentes organizaciones con diversos fines, para la extracción de datos y su análisis mediante técnicas estadísticas de grandes bases de datos brindando conocimiento que le sea de utilidad que lleve a mejorar algún aspecto de importancia para la organización. Por ejemplo, en las empresas es ampliamente usada para prospectar clientes, para saber quiénes son los mejores candidatos para un programa o un producto en especial, para analizar el producto o parte de un proceso de producción.

El tema de minería de datos está ligado al servicio, aporta respuestas a lo que está pasando en mercadeo y servicio, es una puerta para las peticiones, quejas y reclamos, si

se trabaja adecuadamente esta información, se pueden establecer cambios en servicios y mejorar las operaciones de la empresa.

### **Minería de datos en los procesos de toma de decisiones**

En la actualidad las organizaciones son impulsadas por el deseo de un cambio continuo; específicamente una mejora continua, por ello, las empresas privadas tanto como las públicas deben tener la capacidad de ser adaptativas, flexibles y abiertas a cambios constantes, contando con la facilidad para aprender cómo resolver problemas y generar conocimiento, y así poder establecer nuevos métodos en pro de la resolución de los mismos.

El interés del mundo profesional por la minería de datos se ve reflejado en el variado y amplio conjunto de empresas que han lanzado al mercado productos de data mining, entre las cuales se encuentran las tres empresas más importantes en herramientas de análisis: SPSS (Clementine), SAS (Enterprise Miner) e IBM (Intelligent Miner). Desde el punto de vista de los expertos se habla de la minería de datos como uno de los campos emergentes y con gran futuro en la investigación de marketing.

Es importante resaltar que poseer datos no es el elemento esencial en una toma de decisión acertada; es extraer de los datos información que al ser analizada resulte en conocimiento que pueda ser usado para la implementación, esto es lo que proporciona el apoyo necesario para la toma de una decisión argumentada, que oriente a la empresa hacia el cumplimiento de sus metas y objetivos. El nivel de satisfacción del cumplimiento del objetivo de este proyecto se mide a través de qué tan valioso es el conocimiento que se va a obtener a partir de la extracción y análisis de información, es decir, en qué medida va a ayudar este conocimiento a alcanzar las metas del OIC que es fortalecer el sector cafetero mundial y promover su expansión sostenible en un entorno basado en el mercado para beneficio de todos los participantes en el sector.

## 2.4 Reglas de asociación

Según (Rodrigo 2018), los algoritmos de reglas de asociación tienen como objetivo encontrar relaciones dentro un conjunto de transacciones, en concreto, items o atributos que tienden a ocurrir de forma conjunta. En este contexto, el término transacción hace referencia a cada grupo de eventos que están asociados de alguna forma. Una regla de asociación se define como una implicación del tipo “si X entonces Y” ( $X \Rightarrow Y$ ), donde X e Y son itemsets o items individuales. El lado izquierdo de la regla recibe el nombre de antecedente o left-hand-side (LHS) y el lado derecho el nombre de consecuente o right-hand-side (RHS).

### 2.4.1 Algoritmo Apriori

Según (Alcalde 2018), el algoritmo a priori es un algoritmo utilizado sobre bases de datos transaccionales, que permite encontrar de forma eficiente ”conjuntos de ítems frecuentes”, los cuales sirven de base para generar reglas de asociación. Procede identificando los ítems individuales frecuentes en la base y extendiéndolos a conjuntos de mayor tamaño siempre y cuando esos conjuntos de datos aparezcan suficientemente seguidos en dicha base de datos.

#### Algoritmo

1. Generar  $L_1$ (itemsets frecuentes de longitud 1)
2. Repetir hasta que no se encuentren más itemsets nuevos:
  - (a) Generar el conjunto  $C_{k+1}$  de itemsets candidatos a partir de  $L_k$ , combinando solo aquellos que se diferencien en el último ítem.
  - (b) Calcular el soporte de cada candidato.
  - (c) Eliminar los candidatos infrecuentes.
  - (d) Incrementar k en 1.

## 2.5 Series de tiempo

### 2.5.1 Definición

Según Carvajal & Murillo (2003), una serie de tiempo se le denomina a cualquier variable que conste de datos reunidos, registrados u observados sobre incrementos sucesivos de tiempo. Por lo tanto, se concluye que es una secuencia ordenada de observaciones sobre una variable en particular.

Una serie estacionaria es aquella cuyos momentos al origen y a la media no varían a través del tiempo. Estas situaciones se presentan cuando los patrones de demanda que influyen sobre la serie son relativamente estables.

### 2.5.2 Componentes de una serie de tiempo

La descomposición clásica es un método que se basa en el supuesto que la serie de datos se pueden desagregar en componentes como: tendencia, ciclo, estacionalidad e irregularidad; que se describen a continuación Carrillo & Viserda (2007):

#### Tendencia

Una serie de tiempo con tendencia es aquella que contiene un componente de largo plazo que representa el crecimiento o declinación de la serie a través de un período amplio.

#### Estacional

Se define como estacional una serie de tiempo con un patrón de cambio a si mismo año tras año. Por lo regular, el desarrollo de una técnica de pronóstico estacional comprende la selección de un método multiplicativo o uno de adición y estimar después índices estacionales a partir de la historia de la serie.

### Ciclo

El efecto cíclico se define como la fluctuación en forma de onda alrededor de la tendencia. Los patrones cíclicos tienden a repetirse en los datos cada dos, tres o más años. Es difícil establecer un modelo para estos patrones cíclicos, ya que no son estables.

### Irregular

El componente irregular de la serie de tiempo es el factor residual, es decir, “todo lo que sobra” y toma en consideración las desviaciones de los valores reales de la serie de tiempo en comparación con los esperados; es el elemento aleatorio.

## 2.6 Modelos Arima

Según Herrera & Hernández (2002), Los modelos ARIMA constituyen una herramienta de amplio espectro de aplicación en la modelación económica. Mediante su utilización, es posible entender el comportamiento de ciertas variables a través del tiempo y de esta forma poder realizar pronósticos de corto y mediano plazo, a partir únicamente de la historia de dicho proceso.

En su forma más general el modelo ARIMA(p,d,q),podría escribirse como:

$$Y'_t = \theta_1 Y'_{t-1} + \theta_2 Y'_{t-2} + \dots + \theta_p Y'_{t-p} + a_t - \phi_1 a_{t-1} + \phi_2 a_{t-2} + \dots + \phi_q a_{t-q} \quad (2.1)$$

En donde,  $Y'_t=Y_t-Y_{t-1}$ .

Es la combinación de los modelos autorregresivo y el de medias móviles, con la particularidad de incluir un proceso de restablecimiento (el cual se denomina integración) de inestabilidad original presente en una serie de tiempo. El modelo ARIMA(p,d,q) es capaz de operar sobre cualquier serie de tiempo ya que se incluye el proceso de estabilización de la serie, por medio del parámetro d, que simboliza el grado de diferenciación aplicado a la serie de tiempo.

### 2.6.1 Modelos autorregresivos

La palabra ARIMA significa Modelos Autorregresivos Integrados de Medias Móviles. Definimos un modelo como autorregresivo si la variable endógena de un período  $t$  es explicada por las observaciones de ella misma correspondientes a períodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda  $Y_t$  puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error (innovación).

Los modelos autorregresivos se abrevian con la palabra AR tras la que se indica el orden del modelo: AR(1), AR(2),...etc. El orden del modelo expresa el número de observaciones retasadas de la serie temporal analizada que intervienen en la ecuación.

El término de error de los modelos de este tipo se denomina generalmente ruido blanco cuando cumple las tres hipótesis básicas tradicionales:

- media nula
- varianza constante
- covarianza nula entre errores correspondientes a observaciones diferentes

La expresión genérica de un modelo autorregresivo, no ya de un AR(1) sino de un AR( $p$ ) sería la siguiente:

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + a_t \quad (2.2)$$

### 2.6.2 Modelo de medias móviles

Un modelo de los denominados de medias móviles es aquel que explica el valor de una determinada variable en un período  $t$  en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente. Estos modelos se denotan normalmente con las siglas MA, seguidos, como en

el caso de los modelos autorregresivos, del orden entre paréntesis. Así, un modelo con q términos de error MA(q) respondería a la siguiente expresión:

$$Y_t = a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} + \dots - \phi_q a_{t-q} \quad (2.3)$$

### 2.6.3 Modelos de Box-Jenkins

En 1970, Box y Jenkins desarrollaron un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un papel fundamental. Una parte importante de esta metodología está pensada para liberar al investigador económetra de la tarea de especificación de los modelos dejando que los propios datos temporales de la variable a estudiar nos indiquen las características de la estructura probabilística subyacente. Esta metodología se basa en tratar de determinar cuál es el modelo probabilístico que rige el proceso a lo largo del tiempo.

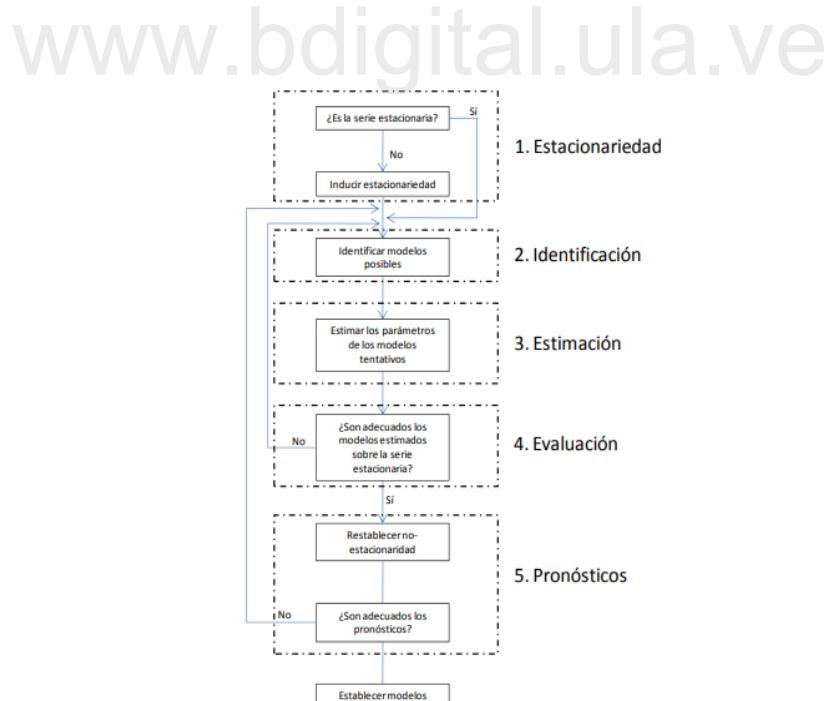


Figura 2.1: Fases de la Metodología Box - Jenkins

La imagen 2.1 muestra los pasos para realizar esta metodología. Cada una de estas

etapas se desarrolla a continuación:

- **Estacionariedad:** El primer paso de la metodología Box-Jenkins es determinar si la serie de tiempo es estacionaria. En caso de que no sea estable a lo largo del tiempo, es necesario aplicar una transformación a la serie para inducirla a ello.
- **Identificación:** Si la función de autocorrelación simple y parcial calculadas de la serie que se desea ajustar a un modelo es semejante a algunas o varias autocorrelación simple y parcial teóricas, entonces se puede decir que el modelo teórico es un modelo tentativo para la serie.
- **Estimación:** En esta etapa se estiman los coeficientes del modelo escogido tentativamente en el paso anterior. Para esto se utiliza el criterio de máxima verosimilitud o el criterio de los mínimos cuadrados
- **Evaluación del modelo:** En este paso se comprueba la eficiencia del modelo y se decide si es estadísticamente adecuado. Para que sea estadísticamente adecuado los residuos son independientes entre sí, indicando que son completamente aleatorios.
- **Pronóstico:** Pronosticar valores futuros de la serie de tiempo, acotando que las series de tiempo solo sirven para predecir evento en corto plazo de tiempo.

#### 2.6.4 Error Porcentual Absoluto Medio

$$MAPE = \frac{\sum_{t=1}^n \frac{|A_t - F_t|}{|A_t|}}{n}$$

Figura 2.2: Error Porcentual Absoluto Medio

El Error Porcentual Absoluto Medio (MAPE o Mean Absolute Percentage Error) es un indicador del desempeño del Pronóstico de Demanda que mide el tamaño del error (absoluto) en términos porcentuales. El hecho que se estime una magnitud del error porcentual lo hace un indicador frecuentemente utilizado por los encargados de

elaborar pronósticos debido a su fácil interpretación. Incluso es útil cuando no se conoce el volumen de demanda del producto dado que es una medida relativa. La fórmula para el cálculo del MAPE o Error Porcentual Absoluto Medio se puede ver en 2.2

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

C.C. Reconocimiento

# Capítulo 3

## Marco Metodológico

### 3.1 Aplicación de las reglas de asociación

Es de interés observar la acción que puede tener una variable sobre otra, para el caso de estudio, éste método implicaría ver las relaciones existentes entre cada precio de café como consecuencia de los cambios que presenten las mismas. Mediante la implementación de las ya descritas reglas de asociación, definidas en el capítulo anterior, serán de utilidad para encontrar y determinar patrones existentes entre los diferentes precios de café por tipo.

Para generar las reglas de asociación se usara el algoritmo Apriori, en el marco teórico se definió teóricamente su funcionamiento y los pasos para ejecutarlo. Para hacer posible la utilización de este algoritmo aplicado a nuestro caso (datos numéricos) es necesario hacer una normalización estandar del conjunto de datos, para luego poder realizar una transformación. Por eso, se crea una función determinista a partir de la cual se puede asociar correctamente el comportamiento de los diferentes precio de café y el indicador compuesto OIC. Dicha función, se encarga de generar una matriz de ceros y uno) dependiendo del valor del precio normalizado que presente el café. Si el valor es positivo, la función sustituye ese valor asignando un 1 a esa posición, y en caso contrario le asignará un 0.

Luego de obtener la matriz de ceros se procede a aplicar el algoritmo Apriori, el cual arrojó en sus resultados 291 conjuntos de reglas de asociación con su respectivo soporte y confianza, los cuales los mas significativo van a ser analizados en el próximo capítulo.

## 3.2 Fases 1 y 2 de la Metodología CRISP-DM

### 3.2.1 Compresión y análisis del problema

Para dar un pronóstico del precio del café por tipo se deben analizar los datos registrados en la base de datos del OIC, mediante técnicas de minería de datos con la finalidad de encontrar un modelo por tipo de café.

### 3.2.2 Entendimiento de datos

Los datos proporcionados por la organización internacional del café (OIC), corresponden al período de enero de 1990 hasta junio de 2018, que corresponde a 4 tipos de café y un indicador compuesto manejado por la misma organización. Los datos corresponden a una base de datos que registra las actividades realizadas por los países exportadores e importadores de café perteneciente a la organización y están digitalizadas en Excel de Microsoft Office; la planilla es manejada en su web pública y gratuitamente, datos anteriores a 1990 son privados ya que se debe cancelar 250 libras esterlinas para conseguirlos. La tabla 3.1 se presenta una breve descripción de los datos

## 3.3 Fase 3: Preparación de los datos

En esta fase se procede a trabajar en conjunto con la metodología de Box-Jenkins, ya que necesitamos que la serie de tiempo sea estacionaria para poder identificar un modelo. La figura 3.1 muestra como esta distribuida la base de datos

Al analizar la tabla 3.1 se observa que las variables al lado de los años no aportan información útil para la prueba, debido a que son promedios anuales y la idea es tener la mayor cantidad de datos posibles para crear una serie de tiempo, por eso se

Tabla 3.1: Variables con su descripción

Nombre del campo	Tipo de campo	Descripcion
Fecha	Fecha	Precio exacto del evento
ICO indicador compuesto	Numero Real	Promedio ponderado de los precios en centavo de dolar indicativos de los cuatro grupos distintos de cafe.
Cafe arabico colombiano suave	Numero Real	Precio en centavo de dolar estadounidense del cafe arabico colombiano
Cafe arabico Brasileiro suave	Numero Real	Precio en centavo de dolar estadounidense del cafe arabico brasileiro
Cafe robustas	Numero Real	Precio en centavo de dolar estadounidense del cafe robustas
Otros cafe arabicas suave	Numero Real	Precio en centavo de dolar estadounidense de otros cafe arabicas

tomaran en cuanta solo los datos mensuales, por tanto, se eliminan estos promedios. A parte el campo de fecha es modificado para que posea la siguiente estructura mes/año respectivamente.

Para la creación del modelo solo se usarán los datos subministrados hasta diciembre del 2015, el resto de los datos que corresponde a 30 meses posteriores se utilizara para validar el modelo.

Fecha	ICO composite indicator	Colombian Milds	Other Milds	Brazilian Naturals	Robustas
<b>1990</b>	<b>71,53</b>	<b>96,53</b>	<b>89,46</b>	<b>82,97</b>	<b>53,60</b>
January	62,75	82,07	76,02	70,36	49,47
February	67,01	91,55	83,95	77,59	50,06
March	75,25	103,24	94,73	86,17	55,76
April	75,34	101,79	94,71	87,45	55,95
May	73,30	99,14	92,97	86,31	53,62
June	69,91	96,01	89,15	82,94	50,67
July	68,36	92,45	86,65	78,94	50,07
August	74,10	103,30	94,43	90,25	53,78
September	75,55	102,21	95,39	92,20	55,70
October	73,89	97,20	91,58	85,78	56,21
November	70,10	92,38	84,72	77,46	55,48
December	72,83	97,06	89,18	80,17	56,46
<b>1991</b>	<b>66,80</b>	<b>89,76</b>	<b>84,98</b>	<b>72,91</b>	<b>48,62</b>
January	69,38	91,55	85,93	75,59	52,83
February	70,55	94,21	89,21	79,39	51,87
March	72,47	99,36	93,56	83,83	51,37
April	71,45	97,27	91,96	81,58	50,93

Figura 3.1: Base de Dato del precio del café segun el OIC

### 3.3.1 Estacionariedad

Como primer paso del estudio de series temporales, se realiza una representación gráfica de cada serie para ver el comportamiento de los datos y a partir de este punto decidir sobre la estacionariedad, después se le aplicara algún tipo de suavizamiento a la serie de tipo logarítmico; con el fin de lograr transformar los valores para que representen una media y varianza constante a través del tiempo.

Luego de aplicar el suavizado a la serie, para comprobar la estacionariedad de la misma, se procede a realizar el test de Dickey-Fuller, también conocido como la prueba de la raíz unitaria. Un artículo publicado por Moffatt, M. (2019), menciona que esta prueba se utiliza para comprobar si la serie es estacionaria o no, partiendo de una hipótesis nula  $H_0$  que alega la existencia de que una raíz unitaria se encuentra presente en la serie temporal. Dicha prueba consta de 2 parámetros fundamentales que determinan el rechazo o aceptación de  $H_0$ , los cuales son: el valor del estadístico de prueba Dickey-Fuller, el cual debería ser negativo para cumplir con la hipótesis alternativa  $H_a$ , junto con el valor de p-value, que debería ser menor a 0.05. ( $p\text{-value} < 0.05$ ) . Cuanto más negativo es el valor del estadístico de prueba ADF, más fuerte es el rechazo de la hipótesis de que hay una raíz unitaria. Para realizar estos paso se hara por separado para poder detallar cada serie de una mejor forma.

En caso de no ser estacionaria, se procede a observar la gráfica de la serie temporal a ver si presenta algún tipo comportamiento donde se logre comprobar la existencia de algún componente estacional, de tendencia, o en su defecto ambos. En caso de ser así, es necesario eliminar dicha tendencia lineal o periodicidad mediante el uso de una diferenciación.

### Café suaves Colombiano Arabico



Figura 3.2: Serie de tiempo Café colombiano

En la figura 3.2 se observa los datos no presenta ninguna estacionariedad ya que se observa varios picos de subida y bajada a lo largo de la serie. Ademas se percibe en el grafico 3.3 de la función de autocorrelación se percibe que los lag disminuyen lentamente pero siguen estando fuertemente correlacionados.

Para dar una mayor confianza de que la serie no es estacionaria, se aplica el test de Dickey-Fuller, el cual se realiza mediante el software libre R, con el siguiente comando `"adf.test(as.xts(PCP), alternative = "stationary")"`. Esta acción genera que el p-value dice 0.4854 y el coeficiente Dickey-Fuller -2.2159, estos valores nos indica primero que la serie no es estacionaria ya que el p-value es mayor a 0.05, por lo tanto existe un raíz unitaria que impide que la serie sea estacionaria.

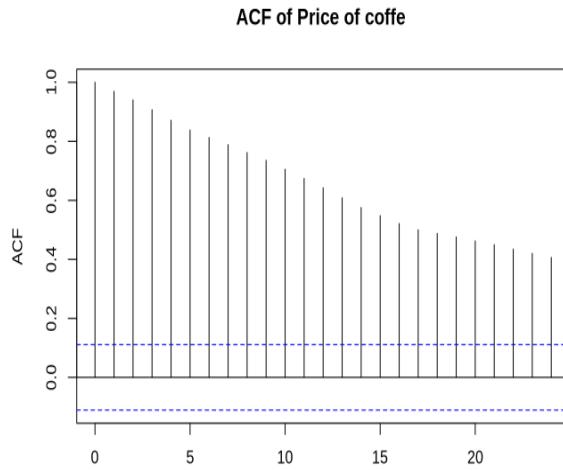


Figura 3.3: función de Autocorrelación

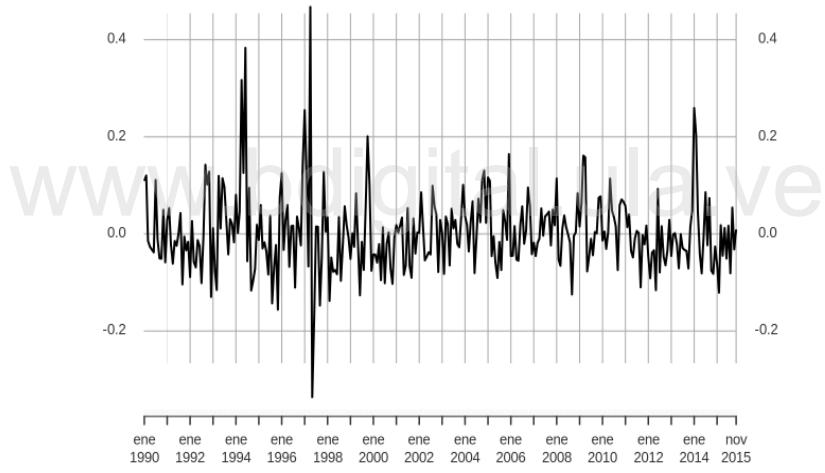


Figura 3.4: Serie de tiempo diferenciada del Café colombiano

Ahora se de realizar una diferenciación regular, debido a que la serie no fue estacionaria, no presenta ninguna estacionalidad. Después se aplica el test de Dickey-Fuller el cual genero un p-value de 0.01 y un estadístico Dickey-Fuller de -6.477, al ser menor que 0.05 el pvalue y al ser el estadístico negativo se puede asegurar que la serie posee una media y una varianza constante en el tiempo. En la figura 3.4 se detalla como queda la representación de los datos al haberse aplicado la diferenciación regular.

### Café suaves Brasileño Arabico



Figura 3.5: Serie de tiempo Café brasileño

En la figura 3.5 se observa que los datos no presenta ninguna estacionariedad ya que se observa varios picos de subida y bajada a lo largo de la serie. Ademas se percibe en el grafico 3.6 de la función de autocorrelación se percibe que los lag disminuyen lentamente pero siguen estando fuertemente correlacionados.

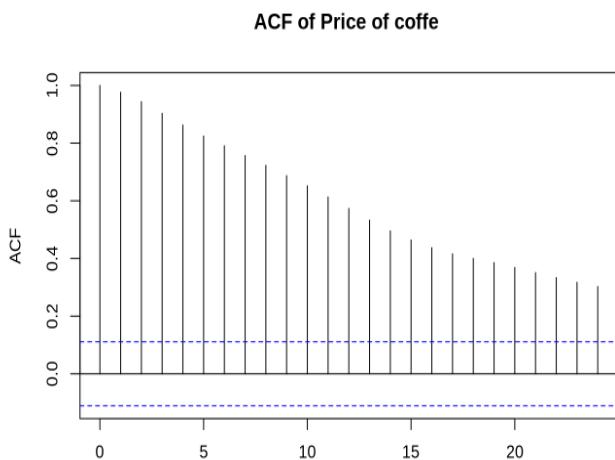


Figura 3.6: funcion de Autocorrelación

Para dar una mayor confianza de que la serie no es estacionaria, se aplica el test de

Dickey-Fuller, el cual se realiza mediante el software libre R, con el siguiente comando ”`adf.test(as.xts(PCP), alternative = "stationary")`”. Esta acción genera que el p-value dice 0.2984 y el coeficiente Dickey-Fuller -2.6594, estos valores nos indica primero que la serie no es estacionaria ya que el p-value es mayor a 0.05, por lo tanto existe un raíz unitaria que impide que la serie sea estacionaria.

Ahora se debe realizar una diferenciación regular, debido a que la serie no fue estacionaria, y no presenta ninguna estacionalidad. Después se aplica el test de Dickey-Fuller el cual genero un p-value de 0.01 y un estadístico Dickey-Fuller de -6.449, al ser menor que 0.05 el pvalue y al ser el estadístico negativo se puede asegurar que la serie posee una media y una varianza constante en el tiempo. En la figura se observa 3.7 como queda la representación de los datos al haberse aplicado la diferenciación regular.

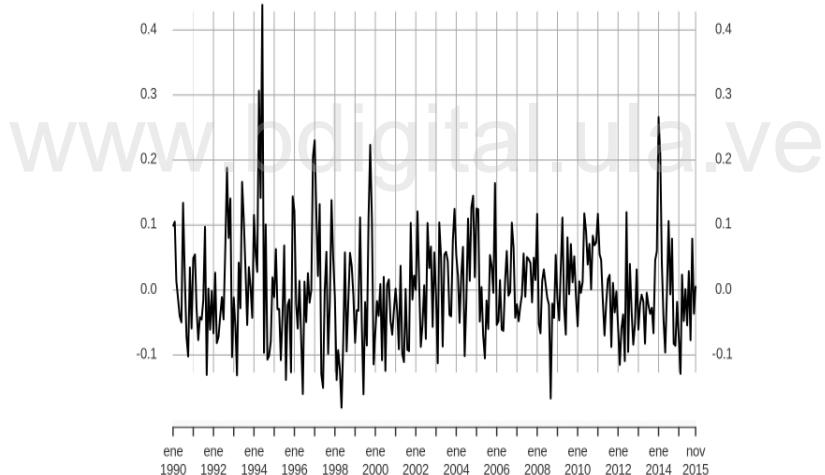


Figura 3.7: Serie de tiempo diferenciada del Café Brasileño

### Café Robusta

En la figura 3.8 observa que los datos no presenta ninguna estacionariedad ya que se observa varios picos de subida y bajada a lo largo de la serie. Ademas se percibe en el grafico 3.9 de la función de autocorrelación se percibe que los lag disminuyen lentamente pero siguen estando fuertemente correlacionados.



Figura 3.8: Serie de tiempo Café Robusto

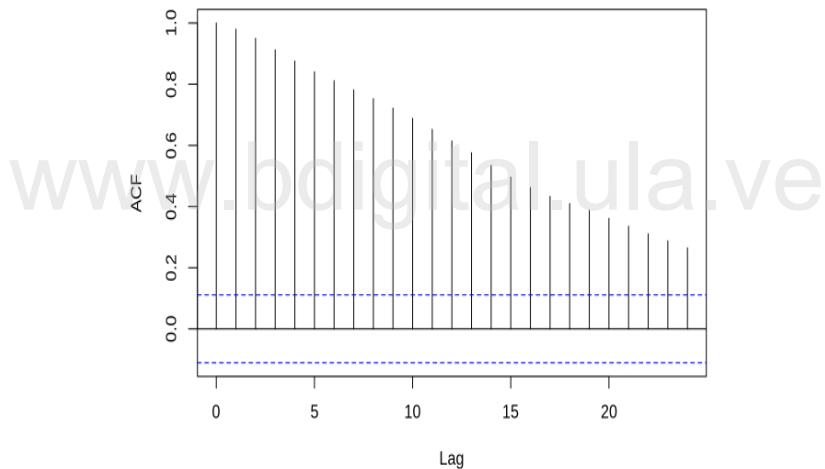


Figura 3.9: función de Autocorrelación

Para dar una mayor confianza de que la serie no es estacionaria, se aplica el test de Dickey-Fuller, el cual se realiza mediante el software libre R, con el siguiente comando ”`adf.test(as.xts(PCP), alternative = "stationary")`”. Esta acción genera que el p-value dice 0.4369 y el coeficiente Dickey-Fuller -2.3331, estos valores nos indica primero que la serie no es estacionaria ya que el p-value es mayor a 0.05, por lo tanto existe un raíz unitaria que impide que la serie sea estacionaria.

Ahora se debe realizar una diferenciación regular, debido a que la serie no fue

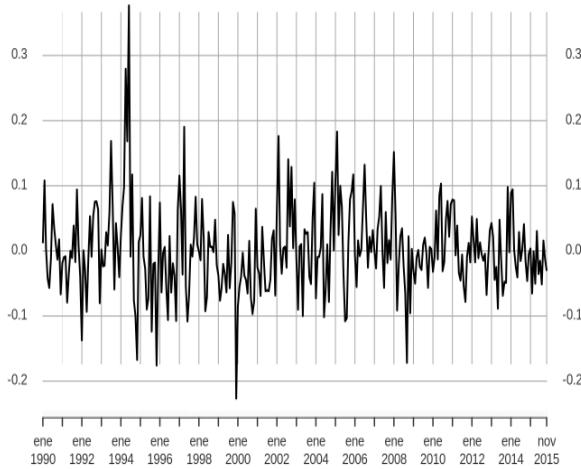


Figura 3.10: Serie de tiempo diferenciada del Café Robusta

estacionaria, y no presenta ninguna estacionalidad. Después se aplica el test de Dickey-Fuller el cual genera un p-value de 0.01 y un estadístico Dickey-Fuller de -6.3603, al ser menor que 0.05 el pvalue y al ser el estadístico negativo se puede asegurar que la serie posee una media y una varianza constante en el tiempo. En la figura 3.10 se observa como queda la representación de los datos al haberse aplicado la diferenciación regular.

### **otros café suave arabicas**

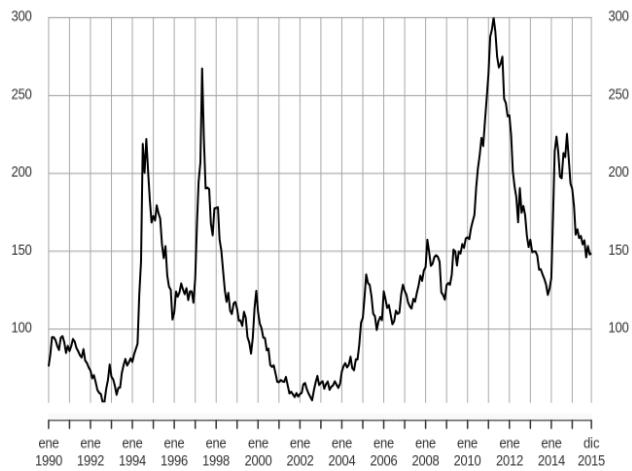


Figura 3.11: Serie de tiempo Café otras arabicas

En la figura 3.11 se observar que los datos no presenta ninguna estacionariedad ya que se observa varios picos de subida y bajada a lo largo de la serie. Ademas se percibe en el grafico 3.12 de la función de autocorrelación se percibe que los lag disminuyen lentamente pero siguen estando fuertemente correlacionados.

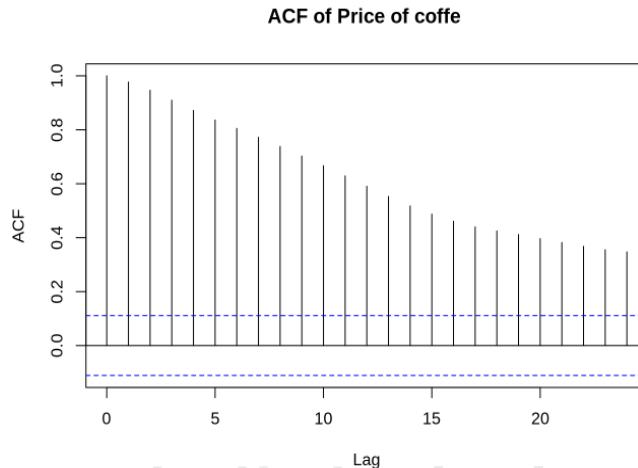


Figura 3.12: funcion de Autocorrelación

Para dar una mayor confianza de que la serie no es estacionaria, se aplica el test de Dickey-Fuller, el cual se realiza mediante el software libre R, con el siguiente comando ”`adf.test(as.xts(PCP), alternative = "stationary")`”. Esta acción genera que el p-value diese 0.2699 y el coeficiente Dickey-Fuller -2.727, estos valores nos indica primero que la serie no es estacionaria ya que el p-value es mayor a 0.05, por lo tanto existe un raíz unitaria que impide que la serie sea estacionaria.

Ahora se debe realizar una diferenciación regular, debido a que la serie no fue estacionaria, y no presenta ninguna estacionalidad. Después se aplica el test de Dickey-Fuller el cual genero un p-value de 0.01 y un estadístico Dickey-Fuller de -6.0767, al ser menor que 0.05 el pvalue y al ser el estadístico negativo se puede asegurar que la serie posee una media y una varianza constante en el tiempo. En la figura 3.13 se observa como queda la representación de los datos al haberse aplicado la diferenciación regular.

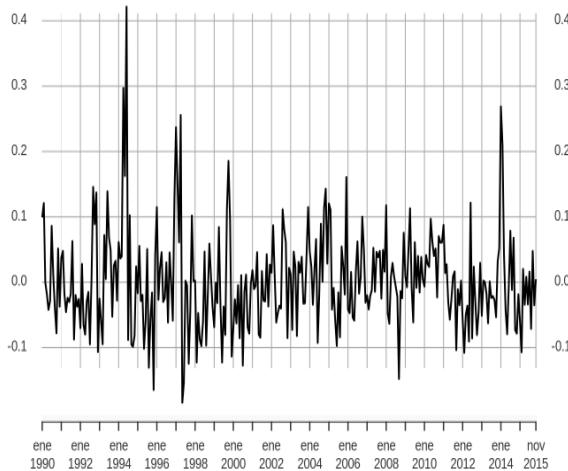


Figura 3.13: Serie de tiempo diferenciada del otros café arabica

#### Coeficiente compuesto del OIC



Figura 3.14: Serie de tiempo del Coeficiente compuesto del OIC

En la figura 3.14 se observa que los datos no presentan ninguna estacionariedad ya que se observan varios picos de subida y bajada a lo largo de la serie. Además se percibe en el gráfico 3.15 de la función de autocorrelación se percibe que los lag disminuyen lentamente pero siguen estando fuertemente correlacionados.

Para dar una mayor confianza de que la serie no es estacionaria, se aplica el test de

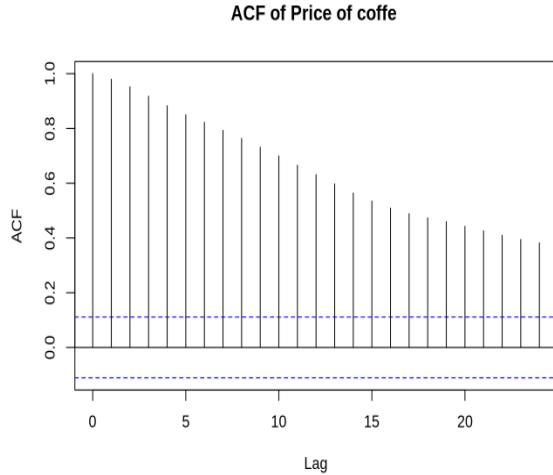


Figura 3.15: función de Autocorrelación

Dickey-Fuller, el cual se realiza mediante el software libre R, con el siguiente comando  
`"adf.test(as.xts(PCP), alternative = "stationary")"`. Esta acción genera que el p-value  
diese 0.2699 y el coeficiente Dickey-Fuller -2.727, estos valores nos indica primero que  
la serie no es estacionaria ya que el p-value es mayor a 0.05, por lo tanto existe un  
raíz unitaria que impide que la serie sea estacionaria.

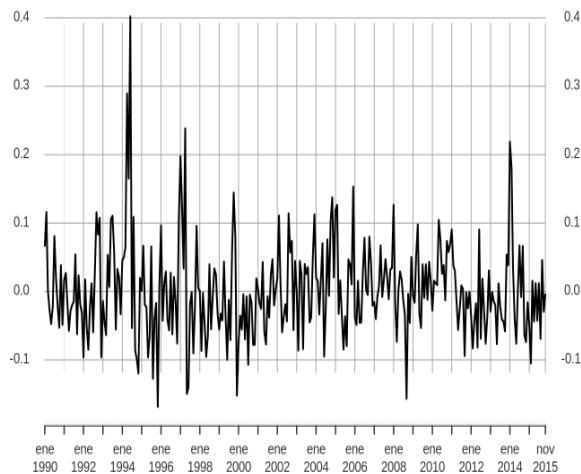


Figura 3.16: Serie de tiempo diferenciada de otros café arabicas

Ahora se debe realizar una diferenciación regular, debido a que la serie no fue

estacionaria, y no presenta ninguna estacionalidad. Después se aplica el test de Dickey-Fuller el cual genero un p-value de 0.01 y un estadístico Dickey-Fuller de -2.5108, al ser menor que 0.3611 el pvalue y al ser el estadístico negativo se puede asegurar que la serie posee una media y una varianza constante en el tiempo. En la figura 3.16 se observa como queda la representación de los datos al haberse aplicado la diferenciación regular.

## 3.4 Fase 4: Modelación

### 3.4.1 Identificación

Depende la cantidad de veces que se diferenció en el proceso anterior se identifica el orden de integración  $I(d)$  del modelo ARIMA, se procede a la identificación de los parámetros  $p$  y  $q$  mediante los gráficos ACF y PACF respectivamente. Cabe anotar, que en el procedimiento de identificación de  $p$  y  $q$ , se consideran aquellos rezagos estadísticamente significativos, es decir, se consideran los  $p$  retardos de las relaciones directas que existen entre las observaciones, donde se determina el orden  $p$  del modelo ARIMA a partir del primer retardo hasta que la función se anula o no logra superar la franja o límites de confianza inferior o superior.

Para esto se tomara en cuenta un principio de parsimonia de un máximo de 6 parámetros por series, con el fin de que las serie temporales no posean tantos parámetros redundantes.

#### Café suaves Colombiano Arabico

Al haber diferenciados solo una vez se puede decir que  $d=1$ , por lo tanto, solo queda encontrar el parámetro  $p$  y  $q$ ; para  $p$  observamos la función de autocorrelación en la figura 3.18, en el lag 5 y 17 corta con los límites inferiores.

Para encontrar el parámetro  $q$  se utiliza la función de autocorrelación, la cual se puede observar en la figura 3.17, en el lag 1, 5 y 17. Como todas las combinaciones posibles nos da mayor a 6 parámetros nos basamos en el criterio AIC, el cual es una

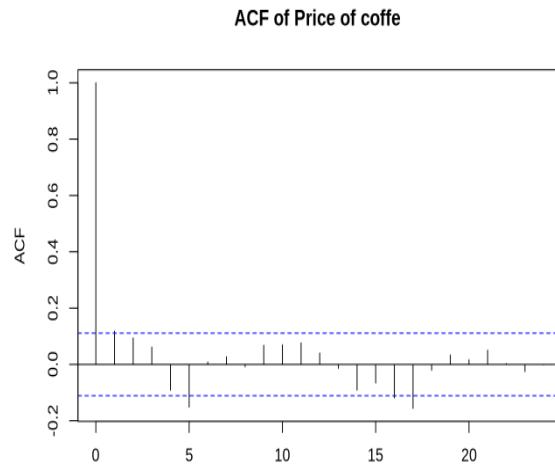


Figura 3.17: Función de autocorrelación del café colombiano

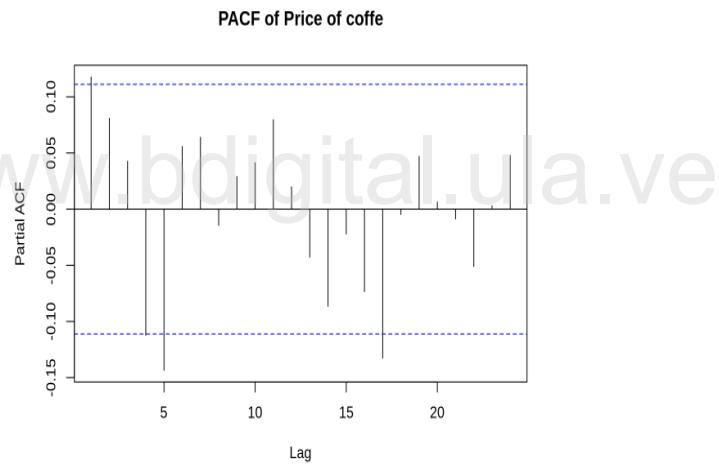


Figura 3.18: Función de autocorrelación parcial del café colombiano

medida de la calidad relativa de un modelo estadístico, para un conjunto dado de datos. Mientras el AIC sea más bajo la calidad del modelo es mejor, por eso en este caso se tomo  $p=3$  y  $q=2$ ; por lo tanto el modelo arima para la serie de tiempo del precio del café colombiano es ARIMA(3,1,2).

### Café suaves Brasileño Arabico

Al haber diferenciados solo una vez se puede decir que  $d=1$ , por lo tanto, solo queda encontrar el parámetro  $p$  y  $q$ ; para  $p$  observamos la función de autocorrelación en la figura 3.20, no se aprecia que exista un corte en los límites.

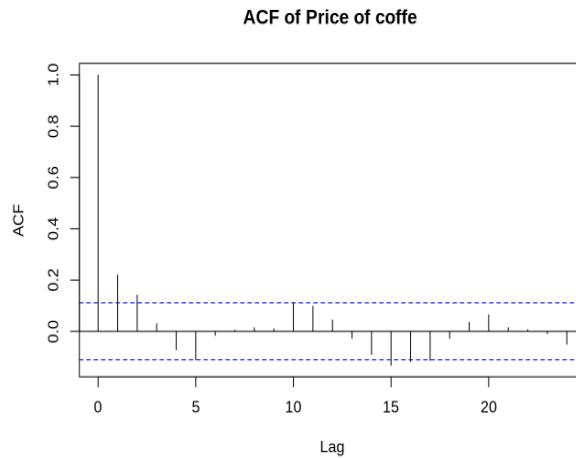


Figura 3.19: Función de autocorrelación del café brasileño

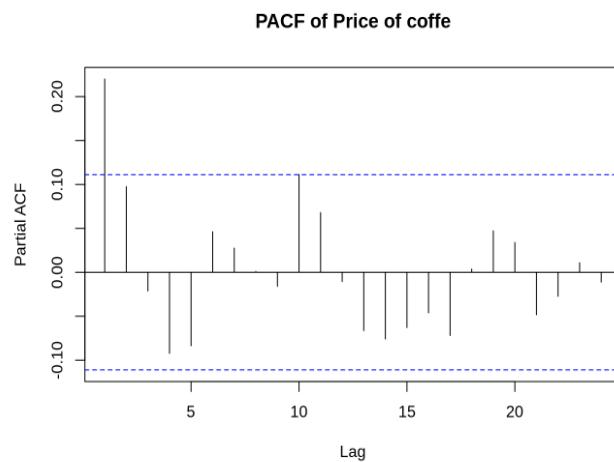


Figura 3.20: Función de autocorrelación parcial del café brasileño

Para encontrar el parámetro  $q$  se utiliza la función de autocorrelación, la cual se puede observar en la figura 3.19, tampoco se percibe un corte claro en los límites. Nos

basamos en el criterio AIC, por eso en este caso se tomo  $p=3$  y  $q=2$ ; por lo tanto el modelo arima para la serie de tiempo del precio del café brasileño es ARIMA(3,1,2).

### Café robusta

Al haber diferenciados solo una vez se puede decir que  $d=1$ , por lo tanto, solo queda encontrar el parámetro  $p$  y  $q$ ; para  $p$  observamos la función de autocorrelación en la figura 3.22, en el lag 5,15 y 19 corta con los límites inferiores.

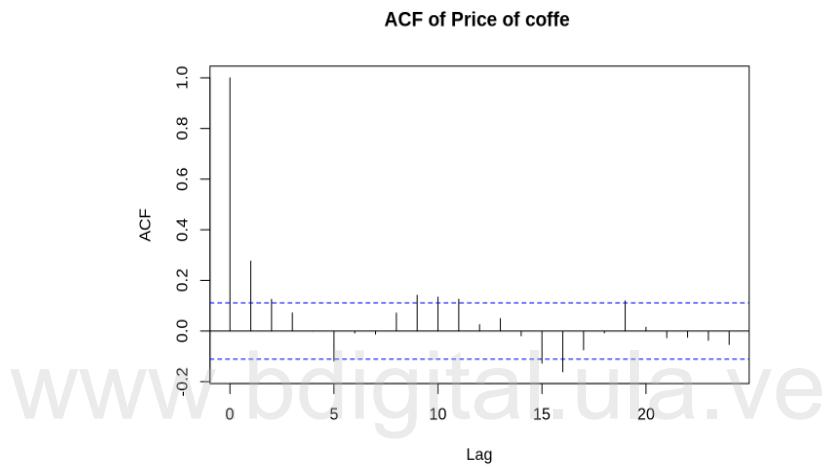


Figura 3.21: Función de autocorrelación del café Robusta

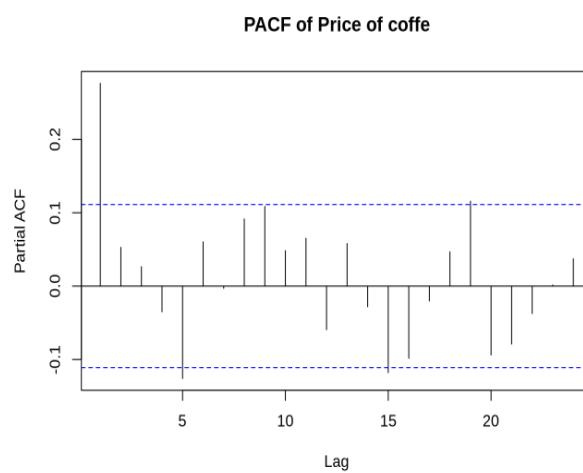


Figura 3.22: Función de autocorrelación parcial del café Robusta

Para encontrar el parámetro  $q$  se utiliza la función de autocorrelación, la cual se puede observar en la figura 3.21, en el lag 1, 9, 10 y 19. Como todas las combinaciones posibles nos da mayor a 6 parámetros nos basamos en el criterio AIC, por eso en este caso se tomo  $p=0$  y  $q=5$ ; por lo tanto el modelo arima para la serie de tiempo del precio del café robusto es ARIMA(0,1,5).

### Otros café suaves Arabico

Al haber diferenciados solo una vez se puede decir que  $d=1$ , por lo tanto, solo queda encontrar el parámetro  $p$  y  $q$ ; para  $p$  observamos la función de autocorrelación en la figura 3.24, se aprecia que el lag 17 existe un corte el límite inferior.

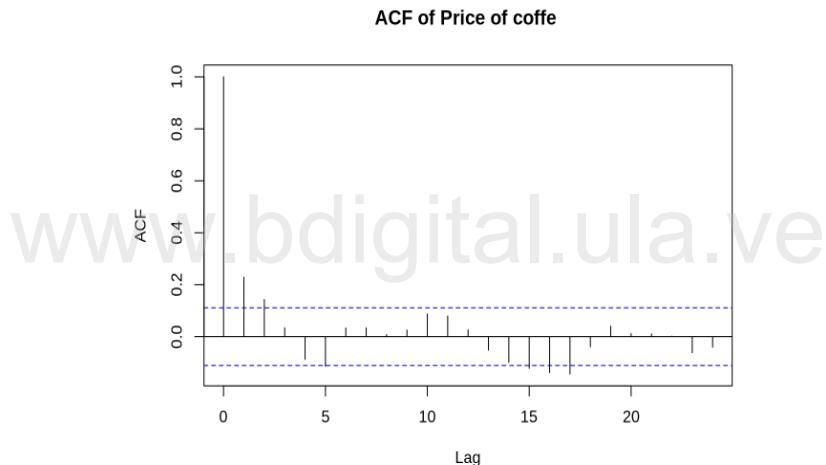


Figura 3.23: Función de autocorrelación de otros café suaves arabicos

Para encontrar el parámetro  $q$  se utiliza la función de autocorrelación, la cual se puede observar en la figura 3.23, se percibe cortes en los lags 1,2,15,16,17. Por eso es mejor basarse en el criterio AIC para cumplir el principio de parsimonia, por eso en este caso se tomo  $p=3$  y  $q=2$ ; por lo tanto el modelo arima para la serie de tiempo del precio del café otros suaves arabicos es ARIMA(3,1,2).

### Coeficiente compuesto del OIC

Al haber diferenciados solo una vez se puede decir que  $d=1$ , por lo tanto, solo queda encontrar el parámetro  $p$  y  $q$ ; para  $p$  observamos la función de autocorrelación en la

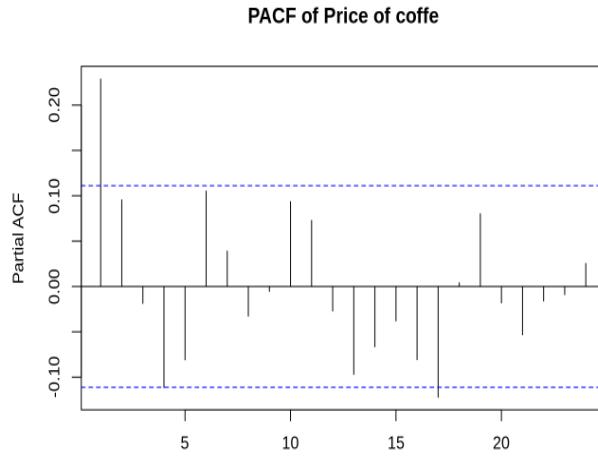


Figura 3.24: Función de autocorrelación parcial de otros café suaves arabicos

figura 3.26, no se aprecia que exista un corte los límites .

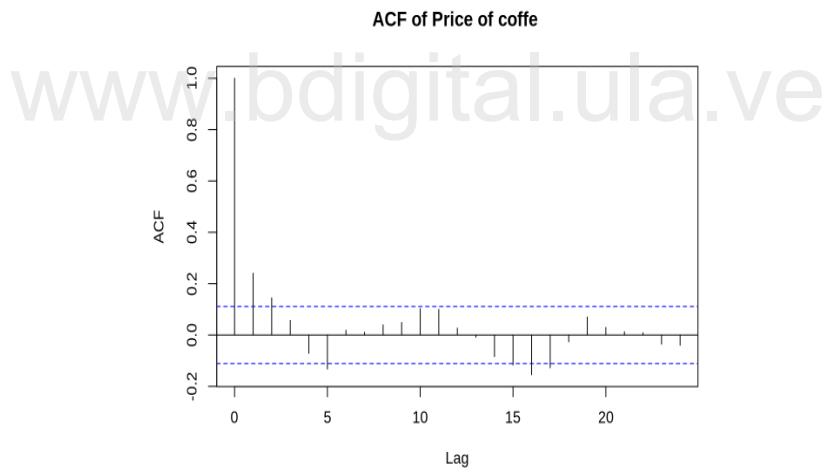


Figura 3.25: Función de autocorrelación del café otros suaves arabicos

Para encontrar el parámetro  $q$  se utiliza la función de autocorrelación, la cual se puede observar en la figura 3.25, se percibe cortes en los lags 1,2,5,16,17. Por eso es mejor basarse en el criterio AIC para cumplir el principio de parsimonia, por eso en este caso se tomo  $p=2$  y  $q=2$ ; por lo tanto el modelo arima para la serie de tiempo del precio compuesto del OIC es ARIMA(2,1,2).

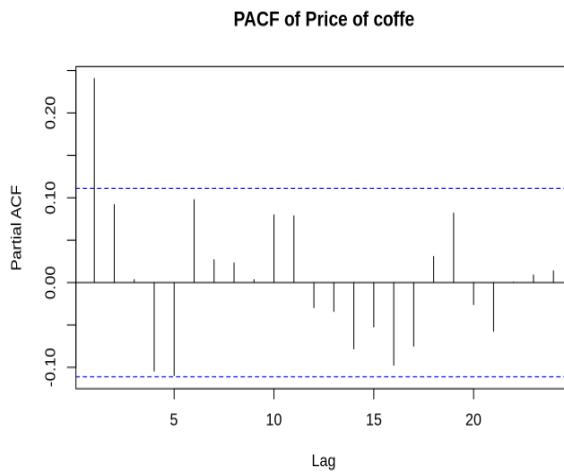


Figura 3.26: Función de autocorrelación parcial del café otros suaves arabicos

### 3.4.2 Estimación

En esta etapa se estiman los coeficientes de los términos autorregresivos y de media móvil incluidos en el modelo, cuyo número de rezagos p y q ya han sido identificados en la etapa anterior. En el caso de estudio, la herramienta para identificarlo será R, mediante el uso del comando ARIMA(p,d,q).

#### Café suaves colombiano Arabico

Al aplicar el comando ”Arima(as.xts(PCP), order = c(3,1,2))” en donde as.xts(PCP), son los datos mensuales del precio del café colombiano de 1990 hasta diciembre del 2015, y el parámetro order indica la cantidad de paramáetros (ma,i,ar) en el modelo, en la figura 3.27, se aprecia el valor de cada coeficiente.

```

Coefficients:
      ar1      ar2      ar3      ma1      ma2  constant
      0.8382  -0.8725 -0.0380 -0.8816  1.0000   0.1863
      s.e.    0.0572   0.0558   0.0573   0.0384   0.0853   0.8027
  
```

Figura 3.27: Parametros del modelo arima(3,1,2) para el precio del café colombiano

Utilizando la ecuación 2.1, se obtiene que el modelo que representa esta serie es el siguiente:

$$Y'_t = 0.8382Y'_{t-1} - 0.8725Y'_{t-2} - 0.0380Y'_{t-3} + 0.1863 + 0.8816a_{t-1} - a_{t-2} \quad (3.1)$$

En donde,  $Y'_t = Y_t - Y_{t-1}$ .

### Café suaves Brasileño Arabico

Al aplicar el comando ”Arima(as.xts(PCP), order = c(3,1,2))” en donde as.xts(PCP), son los datos mensuales del precio del café brasileño de 1990 hasta diciembre del 2015, y el parámetro order indica la cantidad de parámetros (ma,i,ar) en el modelo, en la figura 3.28, se aprecia el valor de cada coeficiente.

```
Coefficients:
ar1      ar2      ar3      ma1      ma2  constant
0.3711  -0.7334  0.1658 -0.1823  0.8504   0.1762
s.e.    0.1559  0.1360  0.0984  0.1507  0.1125   0.7937
```

Figura 3.28: Parametros del modelo arima(3,1,2) para el precio café otros brasileño

Usando la ecuación 2.1, se obtiene que el modelo que representa esta serie es el siguiente:

$$Y'_t = 0.3711Y'_{t-1} - 0.7334Y'_{t-2} + 0.1658Y'_{t-3} + 0.1762 + 0.1823a_{t-1} - 0.8504a_{t-2} \quad (3.2)$$

En donde,  $Y'_t = Y_t - Y_{t-1}$ .

### Café Robusta

Al aplicar el comando ”Arima(as.xts(PCP), order = c(0,1,5))” en donde as.xts(PCP), son los datos mensuales del precio del café robusta de 1990 hasta diciembre del 2015, y el parámetro order indica la cantidad de paramáetros (ma,i,ar) en el modelo, en la figura 3.29, se aprecia el valor de cada coeficiente.

```
Coefficients:
      ma1     ma2     ma3     ma4     ma5  constant
      0.2308  0.2309  0.0249 -0.0115 -0.1485   0.0973
  s.e.  0.0567  0.0576  0.0586  0.0587  0.0517   0.4169
```

Figura 3.29: Parametros del modelo arima(0,1,5) para el precio del café robusta

Utilizando la ecuación 2.1, se obtiene que el modelo que representa esta serie es el siguiente:

$$Y'_t = 0.0973 - 0.2308a_{t-1} - 0.2309a_{t-2} - 0.0249a_{t-3} + 0.0115a_{t-4} + 0.1485a_{t-5} \quad (3.3)$$

En donde,  $Y'_t = Y_t - Y_{t-1}$ .

### Otros café arabico suaves

Al aplicar el comando ”Arima(as.xts(PCP), order = c(3,1,2))” en donde as.xts(PCP), son los datos mensuales del precio de otros café suaves arabico de 1990 hasta diciembre del 2015, y el parámetro order indica la cantidad de parámetros (ma,i,ar) en el modelo, en la figura 3.30, se aprecia el valor de cada coeficiente.

```
Coefficients:
      ar1     ar2     ar3     ma1     ma2  constant
      0.2713 -0.7600  0.1889 -0.1085  0.8829   0.2419
  s.e.  0.0912  0.1154  0.0671  0.0784  0.0841   0.8345
```

Figura 3.30: Parametros del modelo arima(3,1,2) para el precio de otros café suaves arabicos

Utilizando la ecuación 2.1, se obtiene que el modelo que representa esta serie es el siguiente:

$$Y'_t = 0.2713Y'_{t-1} - 0.76Y'_{t-2} + 0.1889Y'_{t-3} + 0.2419 + 0.1085a_{t-1} - 0.8829a_{t-2} \quad (3.4)$$

En donde,  $Y'_t = Y_t - Y_{t-1}$ .

### Indicador compuesto del OIC

Al aplicar el comando ”Arima(as.xts(PCP), order = c(2,1,2))” en donde as.xts(PCP), son los datos mensuales del indicador compuesto del OIC de 1990 hasta diciembre del 2015, y el parámetro order indica la cantidad de parámetros (ma,i,ar) en el modelo, en la figura 3.31, se aprecia el valor de cada coeficiente.

Coefficients:					
	ar1	ar2	ma1	ma2	constant
	0.8305	-0.6258	-0.6610	0.7035	0.1749
s.e.	0.2335	0.2204	0.2371	0.1557	0.5896

Figura 3.31: Parametros del modelo arima(2,1,2) para el precio compuesto del OIC

Usando la ecuación 2.1, se obtiene que el modelo que representa esta serie es el siguiente:

$$Y'_t = 0.8305Y'_{t-1} - 0.6258Y'_{t-2} + 0.1749 + 0.6610a_{t-1} - 0.7035a_{t-2} \quad (3.5)$$

En donde,  $Y'_t = Y_t - Y_{t-1}$ .

## 3.5 Fase 5: Evaluación

### 3.5.1 Evaluación del modelo

En esta etapa, se comprueba que los residuos no tengan una estructura de dependencia y sigan un proceso de ruido blanco, es decir, que todos los posibles períodos de oscilación se presentan con igual intensidad (con media igual a 0 y varianza  $\sigma^2$ ) y no tienen correlación serial. Si los residuos muestran estructura de dependencia se modifica el modelo para incorporarla y se repite el procedimiento desde la etapas de identificación, basandonos en el AIC mas pequeño y cumpliendo el principio de parsimonia de un maximo de 6 parametros hasta obtener un modelo adecuado. Para esto, se utiliza como soporte el test de Ljung – Box, mediante una prueba efectuada para determinar la significancia de la hipótesis nula  $H_0$  que hace referencia a la NO autocorrelación de

los residuos, donde el valor valor de significancia  $\alpha$  ó p-value < 0.05 para aceptar  $H_0$ , es decir, que los residuos son independientes. De lo contrario se aceptaría la hipótesis alternativa  $H_1$  en la que se alega que si existe una dependencia entre ellos. La grafica que representa los p-value del estadístico Ljung-box, nos indicara si el modelo es o no aceptado, si ningún punto da inferio al limite  $\alpha < 0.05$

### Café suaves colombiano Arabico

Como se puede observar en la gráfica 3.32, residuos estandarizados se presentan de manera estacionaria a lo largo del tiempo (varianza y media constante), en la función de autocorrelación se observa que ningún lag sobrepasa los límites, pudiéndonos confirmar que los residuos son independientes entre sí. A parte el grafico Normal Q-Q plot de los residuos estandarizado se acopla muy bien a la relación lineal.

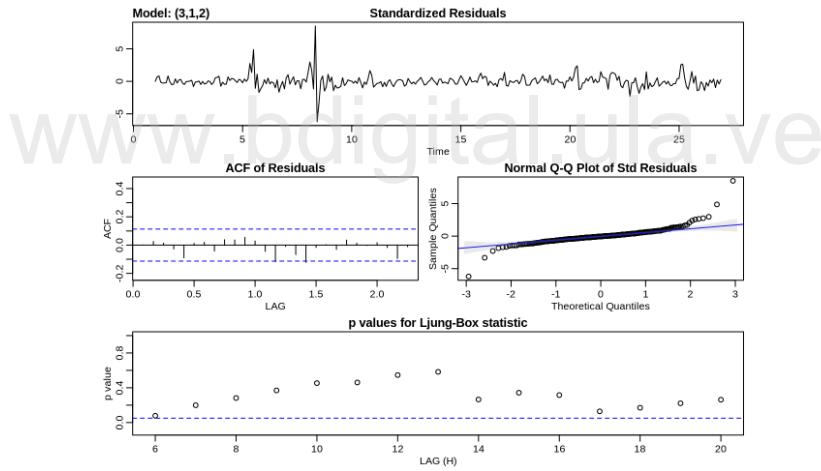


Figura 3.32: Análisis residual del modelo arima(3,1,2) para el precio del café colombiano

La grafica que representa los p-value del estadístico Ljung-box de esta serie nos muestra que todo los puntos están por encima de 0.05, entonces, se puede afirmar con completa seguridad que los residuos son independiente, por lo tanto el modelo es apto para continuar a la siguiente fase.

### Café suaves brasileño Arabico

Como se puede observar en la gráfica 3.33, residuos estandarizados se presentan de manera estacionaria a lo largo del tiempo (varianza y media constante), en la función de autocorrelación se observa que ningún lag sobrepasa los límites, pudiéndonos confirmar que los residuos son independientes entre sí. A parte el grafico Normal Q-Q plot de los residuos estandarizado se acopla muy bien a la relación lineal.

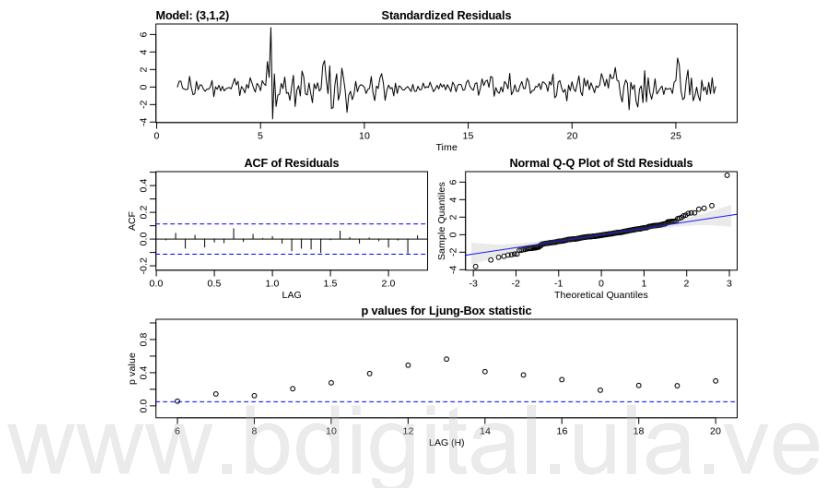


Figura 3.33: Análisis residual del modelo arima(3,1,2) para el precio del café brasileño

La grafica que representa los p-value del estadístico Ljung-box de esta serie nos muestra que todo los puntos están por encima de 0.05, entonces, se puede detallar que el lag 6 esta muy cerca del 0.05, sin embargo se puede afirmar con completa seguridad que los residuos son independiente, por lo tanto el modelo es apto para continuar a la siguiente fase.

### Café Robusta

Como se observa que en la gráfica 3.34, residuos estandarizados se presentan de manera estacionaria a lo largo del tiempo (varianza y media constante), en la función de autocorrelación se aprecia que el lag 1.6 sobrepasa el límites, mostrando que existe cierta dependencia de residuos. Además el grafico Normal Q-Q plot de los residuos estandarizado se acopla muy bien a la relación lineal.

La grafica que representa los p-value del estadístico Ljung-box de esta serie muestra

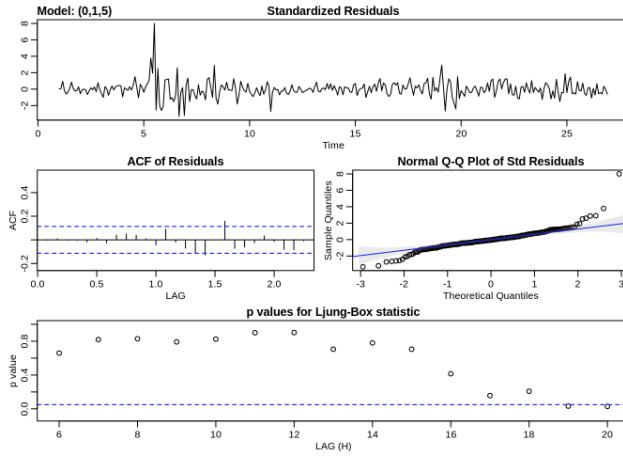


Figura 3.34: Análisis residual del modelo arima(0,1,5) para el precio de café robusta

que los lag 19 y 20 dan por debajo de 0.05, a pesar de que se probaron los otros modelos, estos demostraban una mayor dependencia a lo largo de la serie, por este motivo para efectos de este estudio se decide adoptar este modelo, el cual se dara una mejor explicación del por que de la elección en el capítulo siguiente.

### Otros café suaves arabicos

Como se puede observar en la gráfica 3.35, residuos estandarizados se presentan de manera estacionaria a lo largo del tiempo (varianza y media constante), en la función de autocorrelación se observa que ningún lag sobrepasa los límites, confirmando que los residuos son independientes entre sí. A parte el grafico Normal Q-Q plot de los residuos estandarizado se acopla muy bien a la relación lineal.

La grafica que representa los p-value del estadístico Ljung-box de esta serie nos muestra que todo los puntos están por encima de 0.05, entonces, se puede detallar que el lag 6 esta muy cerca del 0.05, sin embargo se puede afirmar con completa seguridad que los residuos son independiente, por lo tanto el modelo es apto para continuar a la siguiente fase.

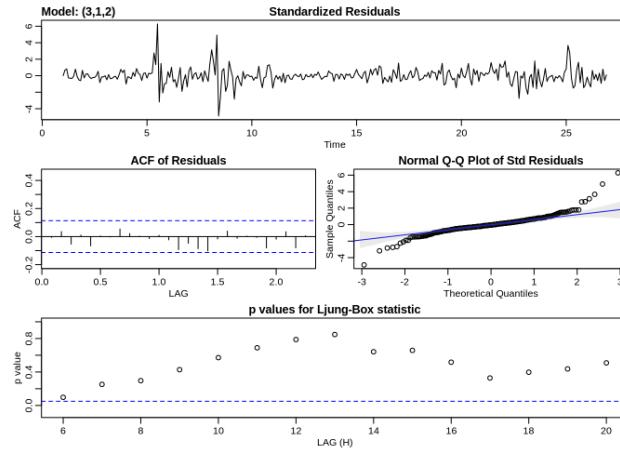


Figura 3.35: Análisis residual del modelo arima(3,1,2) para el precio de otros café suaves arabicos

### Indicador compuesto del OIC

Como se puede observar en la gráfica 3.36, residuos estandarizados se presentan de manera estacionaria a lo largo del tiempo (varianza y media constante), en la función de autocorrelación se observa que ningún lag sobrepasa los límites, confirmando que los residuos son independientes entre sí. A parte el grafico Normal Q-Q plot de los residuos estandarizado se acopla muy bien a la relación lineal.

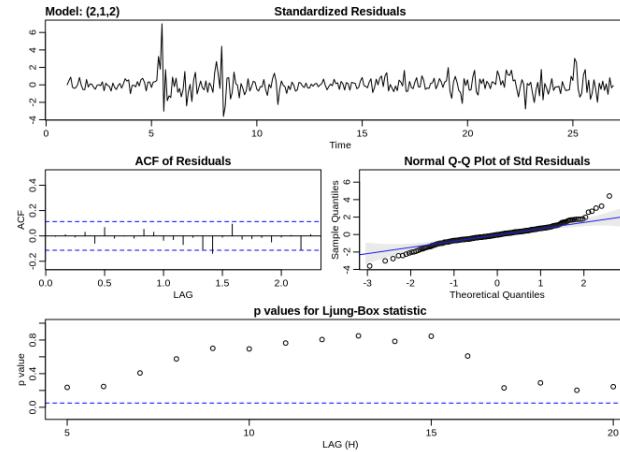


Figura 3.36: Análisis residual del modelo arima(2,1,2) para el precio compuesto del OIC

La grafica que representa los p-value del estadístico Ljung-box de esta serie nos muestra que todo los puntos están por encima de 0.05, igual se puede detallar que el lag 6 esta muy cerca del 0.05, sin embargo se puede afirmar con completa seguridad que los residuos son independiente, por lo tanto el modelo es apto para continuar a la siguiente fase.

### 3.5.2 Pronóstico

En esta última etapa, luego de obtener los valores pronosticados por el modelo, se construye una tabla que contiene las observaciones del conjunto de datos originales resultantes, los cuales se dejaron como fase o período de prueba con el propósito de compararlos con dichos valores resultantes arrojados por el pronóstico de ARIMA.

Para evaluar la precisión global de los pronósticos se hace uso del error porcentual absoluto medio, explicado en el marco teórico. Tomando en cuenta que solo se pronosticaran 30 valores por modelo.

#### Café suaves colombiano arabico

La tabla 3.2, muestra los valores reales y los valores pronosticados del precio del café colombiano, para ver la eficiencia del modelo se aplicará la medida MAPE, el cual es un estadístico para calcular el error porcentual absoluto medio, el cual fue explicado en el capítulo anterior. Esto generó un error de 6.97%.

En la figura 3.37, se graficó los valores de la tabla 3.2, siendo la curva de color rojo la que representa los valores reales, y la curva negra representa los valores pronosticados por el modelo, se aprecia que el intervalo observado es el periodo de enero de 2016 hasta junio de 2018.

#### Café suaves brasileño arabico

La tabla 3.3, muestra los valores reales y los valores pronosticados del precio del café brasileño, para ver la eficiencia del modelo se aplicará la medida MAPE, el cual es un estadístico para calcular el error porcentual absoluto medio, el cual fue explicado en el

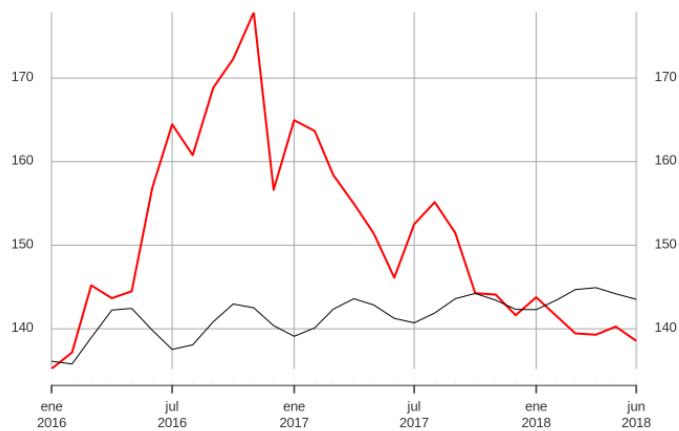


Figura 3.37: Grafica de pronóstico por 30 meses del café colombiano

capítulo anterior. Esto genero un error de 7.65%.

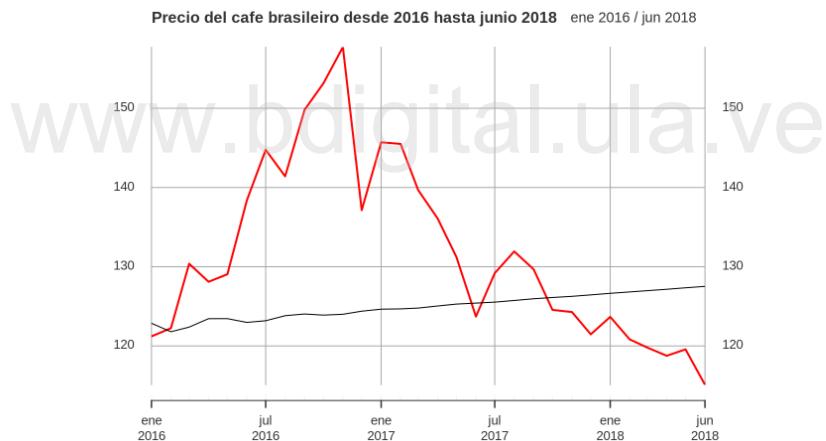


Figura 3.38: Grafica de pronóstico por 30 meses del café brasileño

En la figura 3.38, se graficó los valores de la tabla 3.3, siendo la curva de color rojo la que representa los valores reales, y la curva negra representa los valores pronosticados por el modelo, se aprecia que el intervalo observado es el periodo de enero de 2016 hasta junio de 2018.

### Café robusta

La tabla 3.4, muestra los valores reales y los valores pronosticados del precio del café robusta, para ver la eficiencia del modelo se aplicará la medida MAPE, el cual es un estadístico para calcular el error porcentual absoluto medio, el cual fue explicado en el capítulo anterior. Esto genero un error de 14.23%.

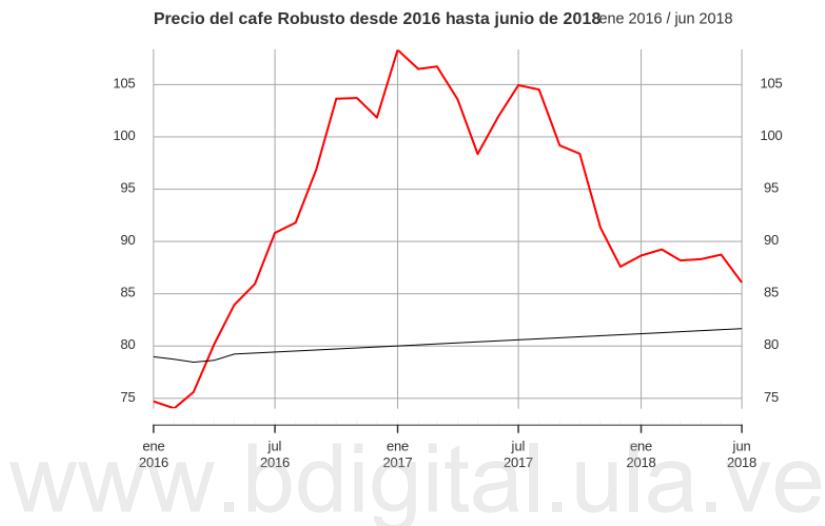


Figura 3.39: Grafica de pronóstico por 30 meses del café robusta

En la figura 3.39, se graficó los valores de la tabla 3.4, siendo la curva de color rojo la que representa los valores reales, y la curva negra representa los valores pronosticados por el modelo, se aprecia que el intervalo observado es el periodo de enero de 2016 hasta junio de 2018.

### Otros café suaves Arabicos

La tabla 3.5, muestra los valores reales y los valores pronosticados del precio de otros café arabico, para ver la eficiencia del modelo se aplicará la medida MAPE, el cual es un estadístico para calcular el error porcentual absoluto medio, el cual fue explicado en el capítulo anterior. Esto genero un error de 8.72%.

En la figura 3.40, se graficó los valores de la tabla 3.5, siendo la curva de color rojo la que representa los valores reales, y la curva negra representa los valores pronosticados por el modelo, se aprecia que el intervalo observado es el periodo de enero de 2016

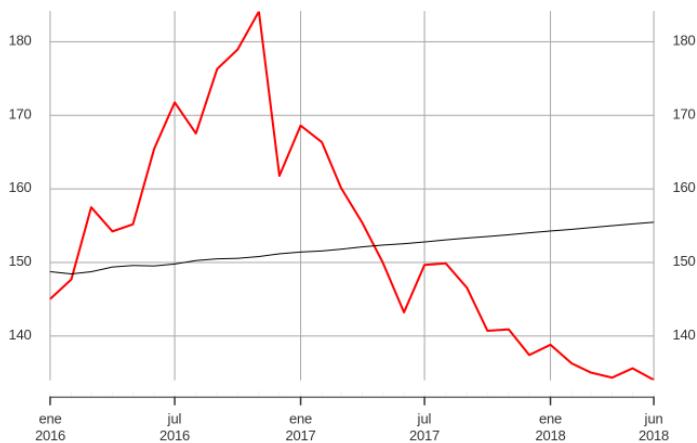


Figura 3.40: Grafica de pronóstico por 30 meses de otros café suave arabicos

hasta junio de 2018.

### Indicador compuesto del OIC

La tabla 3.6, muestra los valores reales y los valores pronosticados del precio del indicador compuesto del oic, para ver la eficiencia del modelo se aplicará la medida MAPE, el cual es estadístico para calcular el error porcentual absoluto medio, el cual fue explicado en el capítulo anterior. Esto genero un error de 7.85%.

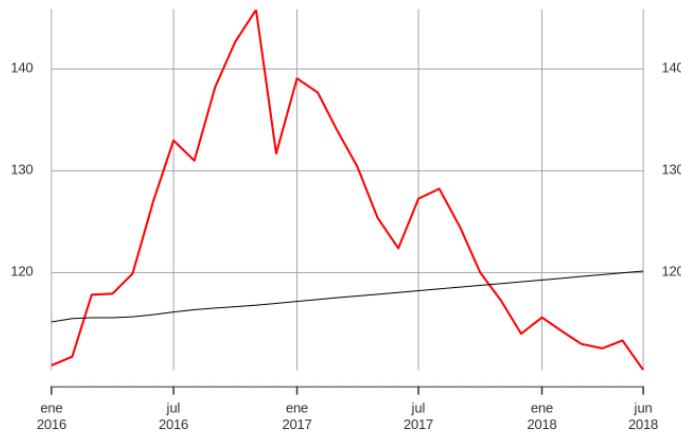


Figura 3.41: Grafica de pronóstico por 30 meses del indicador compuesto del OIC

En la figura 3.41, se graficó los valores de la tabla 3.6, siendo la curva de color rojo

la que representa los valores reales, y la curva negra representa los valores pronosticado por el modelo, se pueden apreciar que el intervalo observado es el periodo de enero de 2016 hasta junio de 2018.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

C.C. Reconocimiento

Tabla 3.2: Valores reales con valores pronosticados del cafe colombiano

$Y_r$	$Y_i$
135.21	136.12
137.17	135.79
145.20	138.95
143.66	142.23
144.49	142.44
156.86	139.83
164.46	137.53
160.78	138.08
168.85	140.84
172.28	142.96
177.85	142.51
156.64	140.38
164.94	139.10
163.67	140.11
158.40	142.35
154.97	143.60
151.41	142.85
146.12	141.25
152.21	140.71
155.15	141.89
151.47	143.60
144.26	144.23
144.09	143.42
141.62	142.32
143.77	142.29
141.50	143.45
139.45	144.69
139.29	144.92
140.26	144.19
138.55	143.52

Tabla 3.3: Valores reales con valores pronosticados del cafe Brasileño

$Y_r$	$Y_i$
121.21	122.85
122.24	121.79
130.38	122.37
128.10	123.44
129.05	123.44
138.38	122.97
144.76	123.18
141.41	123.82
149.80	124.03
153.15	123.89
157.72	124.00
137.14	124.39
145.70	124.64
145.50	124.68
139.67	124.78
136.09	125.04
131.21	125.28
123.71	125.41
129.19	125.53
131.93	125.74
129.67	125.96
124.55	126.12
124.28	126.26
121.47	126.45
123.67	126.64
120.83	126.82
119.80	126.98
118.76	127.15
119.57	127.34
115.10	127.52

Tabla 3.4: Valores reales con valores pronosticados del cafe Robusta

$Y_r$	$Y_i$
74.71	78.97
74.04	78.74
75.60	78.45
80.18	78.64
83.93	79.23
85.94	79.32
90.82	79.43
91.79	79.52
96.88	79.62
103.65	79.72
103.72	79.82
101.85	79.91
108.32	80.01
106.49	80.11
106.73	80.20
103.58	80.30
98.36	80.40
101.95	80.50
104.94	80.59
104.52	80.69
99.18	80.79
98.39	80.89
91.33	80.98
87.59	81.08
88.65	81.18
89.24	81.27
88.18	81.37
88.31	81.47
88.74	81.57
86.07	81.66

Tabla 3.5: Valores reales con valores pronosticados de otros cafe suaves arabicos

$Y_r$	$Y_i$
145.03	148.75
147.70	148.44
157.70	148.74
154.22	149.38
155.19	149.59
165.45	149.53
171.76	149.79
167.54	150.26
176.30	150.49
178.96	150.56
184.12	150.80
161.78	151.18
168.61	151.42
166.35	151.56
160.15	151.80
155.40	152.12
150.00	152.36
143.22	152.55
149.66	152.78
149.88	153.07
146.54	153.32
140.71	153.53
140.90	153.77
137.42	154.03
138.81	154.28
136.28	154.50
135.03	154.74
134.34	154.99
135.61	155.24
134.03	155.47

Tabla 3.6: Valores reales con valores pronosticados del coeficiente compuesto del OIC

$Y_r$	$Y_i$
110.89	115.16
111.75	115.49
117.83	115.58
117.93	115.58
119.91	115.66
127.05	115.88
132.98	116.14
131.00	116.36
138.22	116.52
142.68	116.65
145.82	116.80
131.70	116.98
139.07	117.18
137.68	117.36
134.07	117.54
130.39	117.70
125.40	117.87
122.39	118.04
127.26	118.23
128.24	118.40
124.46	118.58
120.01	118.75
117.26	118.93
114.00	119.10
115.60	119.28
114.19	119.45
114.19	119.63
112.99	119.80
113.34	119.98
110.40	120.15

# Capítulo 4

## Análisis de resultados

### 4.1 Análisis de las reglas de asociación

En esta sección se tiene planeado explicar las reglas con mayor relevancia, sin embargo, en la figura 4.1, se puede ver el conjunto de las 297 reglas graficada con sus respectiva confianza y soporte, notándose que existen muchas que tienen una confianza muy alta, pero tienen un soporte muy bajo, por lo cual no debe ser tomado muy en cuenta.

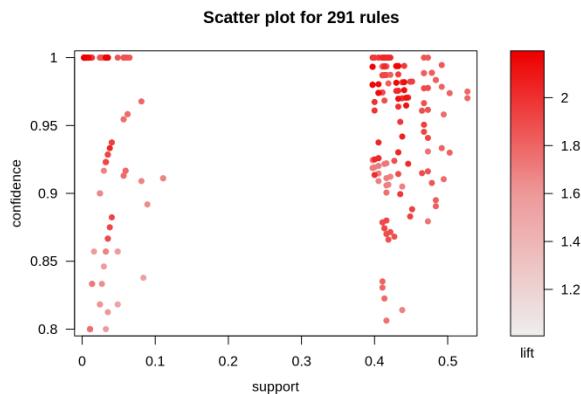


Figura 4.1: Relación soporte-confianza de las 291 reglas de asociación

#### 4.1.1 Relaciones 1 a 1 más relevantes

**café colombiano = 1 → Indicador compuesto del oic = 1**, con un 98.22% de confianza y un soporte de 44.86%, indica que, si sube de precio el café colombiano, entonces el indicador compuesto también aumenta su precio, teniendo esto sentido ya que este indicador es un promedio ponderado formado por los cuatro grupos en este caso de estudio.

**Indicador compuesto del oic = 0 → café colombiano = 0**, con un 98.35% de confianza y un soporte de 48.37%, es el contrarecíproco de la regla anterior.

**otros café = 1 → Indicador compuesto del oic = 1**, con un 98.24% de confianza y un soporte de 45.13%, indica que, si sube de precio otros café arabicos, entonces el indicador compuesto también aumenta su rendimiento, teniendo esto sentido ya que este indicador es un promedio ponderado formado por los cuatro grupos en este caso de estudio.

**Indicador compuesto del oic = 0 → otros café = 0**, con un 98.35% de confianza y un soporte de 48.37%, indica que, es el contrarecíproco de la regla anterior.

**café brasileño = 1 → otros café = 1**, con un 97.38% de confianza y un soporte de 50.27%, indica que, si el precio del cafe brasileño sube, entonces los otros cafe arabico tambien lo hara.

**otros café = 0 → café brasileño = 0**, con un 93% de confianza y un soporte de 50.27%, indica que, si el contrarecíproco de la regla anterior.

#### 4.1.2 Relaciones 2 a 1 más relevantes

A parti aca, debido a la cantidad de reglas encontradas, no se tomaran en cuenta los contrarecíprocos, pero estos tambien tiene una confianza y soporte elevados.

**Indicador compuesto del oic = 0 y el café brasileño = 0 → Otros café arabicos = 0**, con un 100% de confianza y un soporte de 47.86%, indica que, si el

café brasileño baja de precio y el indicador compuesto tambien, entonces lo otros tipos de café arabicos tambien bajara su precio.

**el café colombiano = 0 y el café robusta = 0 → el indicador compuesto = 0**, con un 100% de confianza y un soporte de 42.16%, indica que, si el café colombiano baja de precio y el café robusta baja tambien, entoces indicador compuesto bajara.

**otros café = 0 y el café robusta = 0 → el indicador compuesto = 0**, con un 100% de confianza y un soporte de 41.89%, indica que, si los otros café arabico bajan de precio y el café robusta baja tambien, entoces indicador compuesto bajara.

**café brasileño = 0 y el café robusta = 0 → otros cafe arabicos = 0**, con un 99.35% de confianza y un soporte de 41.35%, indica que, si los café brasileño baja de precio y el café robusta baja tambien, entoces los otro café arabico tambien lo hara.

**otros café = 0 y el café robusta = 0 → el café colombiano = 0**, con un 99.35% de confianza y un soporte de 41.62%, indica que, si los otros café arabico bajan de precio y el café robusta baja tambien, entoces el café colombiano bajara su precio.

#### 4.1.3 Relaciones 3 a 1 más relevantes

**el café colombiano = 1 , otros café = 1 y el café robusta = 1 → indicador compuesto OIC = 1** , con un 100% de confianza y un soporte de 40.54%, indica que, si los otros café arabico, el café robusta y el café colombiano aumenta su precio, entonces el indicador compuesto del OIC tambien lo hara.

**el café colombiano = 1 , café brasileño = 1 y el café robusta = 1 → indicador compuesto OIC = 1** , con un 100% de confianza y un soporte de 40.54%, indica que, si el café brasileño, el café robusta el café colombiano aumentan su precio, entonces el indicador compuesto del OIC tambien lo hara.

**el café brasileño = 1 , otros café = 1 y el café robusta = 1 → indicador compuesto OIC = 1 ,** con un 100% de confianza y un soporte de 40.54%, indica que, si los otros café arabico, el café robusta y el café brasileño aumenta su precio, entonces el indicador compuesto del OIC tambien lo hara.

#### 4.1.4 Relaciones 4 a 1 más relevantes

**el café colombiano= 1 , el café brasileño = 1 , otros café = 1 y el café robusta = 1 → indicador compuesto OIC = 1 ,** con un 100% de confianza y un soporte de 39.72%, indica que, si los otros café arabico, el café robusta, el café colombiano y el café brasileño aumenta su precio, entonces el indicador compuesto del OIC tambien lo hara.

**el café colombiano= 1, el café brasileño = 1 , otros café = 1, indicador compuesto OIC = 1 y el café robusta = 1 → otros café = 1,** con un 100% de confianza y un soporte de 39.72%, indica que, si el indicador compuesto del OIC, el café robusta, el café colombiano y el café brasileño aumenta su precio, entonces otros café arabico tambien lo hara.

**indicador compuesto OIC = 0, el café brasileño = 0 , otros café = 0 y el café robusta = 0 → el café colombiano= 0 ,** con un 100% de confianza y un soporte de 41.08%, indica que, si los otros café arabico, el café robusta, otros café arabico y el café brasileño baja su precio, entonces el café colombiano lo hara.

## 4.2 Análisis de las series temporal

### 4.2.1 Café suaves colombiano Arabico

El modelo hallado resulto ser un modelo muy bueno, ya que arrojo un error porcentual absoluto medio de 6.97% en los pronósticos de 30 muestras posteriores, además, se obtuvo un modelo con pocos parámetros, facilitando asi su uso para futuras visiones.

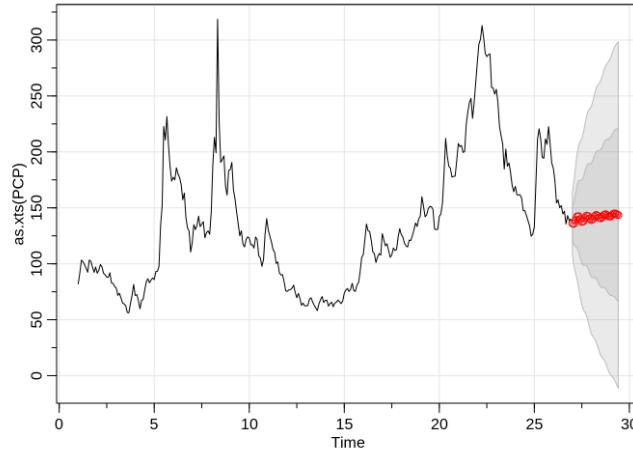


Figura 4.2: Grafica de la serie de tiempo del precio del café colombiano con su respectivo pronóstico

En la figura 4.2, se puede observar el modelo encontrado, con la predicción realizada, es de interés observar que mientras más datos sean necesario predecir el error, tiende a crecer, lo que esta sombreado indica el intervalo de 90 y 95 % de confianza en el cual estará ubicado el punto real según el modelo. También se observa una tendencia de subida de forma oscilatoria.

#### 4.2.2 Café suaves brasileños Arabico

El modelo hallado resultó ser un modelo muy bueno, ya que arrojo un error porcentual absoluto medio de 7.65% en los pronósticos de 30 muestras posteriores, además, se obtuvo un modelo con pocos parámetros, facilitando así su uso para futuras visiones.

En la figura 4.3, se puede observar el modelo encontrado, con la predicción realizada, es de interés observar que mientras más datos sean necesario predecir el error, tiende a crecer, lo que esta sombreado indica el intervalo de 90 y 95 % de confianza en el cual estará ubicado el punto real según el modelo. También se observa una tendencia de subida de forma lineal.

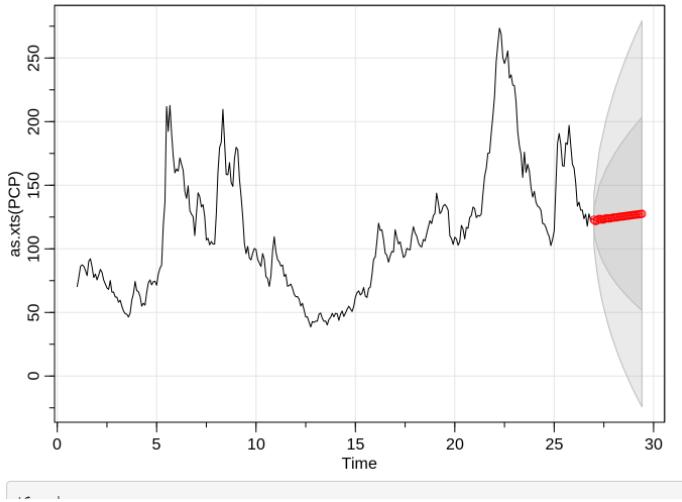


Figura 4.3: Grafica de la serie de tiempo del precio del café brasileño con su respectivo pronóstico

#### 4.2.3 Café robusta

El modelo encontrado para el café robusta, se considera que fue el mejor a pesar de que posee un error porcentual absoluto de 14.23 %, y de que los residuos se correlacionaban entre si, cumple la norma de parsimonia de máximo 6 parámetros. Realizando un estudio se encontró un modelo en que los residuos no estaban correlacionado y poseía un error porcentual absoluto medio de 10.34%, pero transformaba el modelo a uno con 20 parámetros, ocasionando un modelo difícil de comprender, siendo la mejora poco significativa.

En la figura 4.4, se puede observar el modelo encontrado, con la predicción realizada, es de interés observar que mientras más datos sean necesarios predecir el error, tiende a crecer, lo que esta sombreado indica el intervalo de 90 y 95 % de confianza en el cual estará ubicado el punto real según el modelo. También se observa una tendencia de subida de forma lineal.

#### 4.2.4 Otros café suaves Arabico

El modelo hallado resultó ser un modelo muy bueno, ya que arrojó un error porcentual absoluto medio de 8.72% en los pronósticos de 30 muestras posteriores, además, se

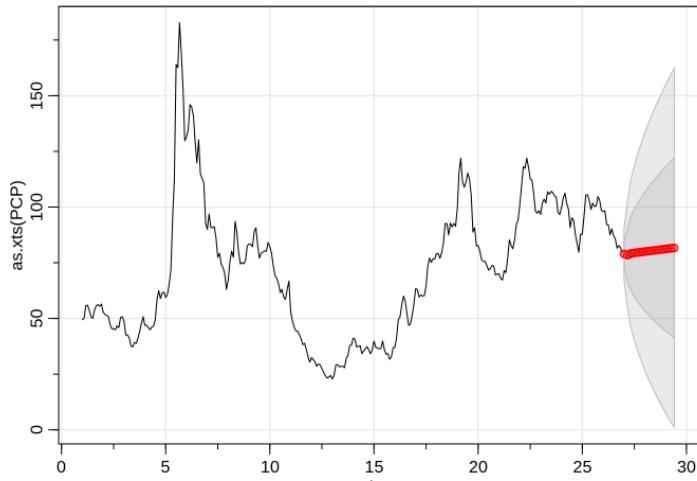


Figura 4.4: Grafica de la serie de tiempo del precio del café robusta con su respectivo pronóstico

obtuvo un modelo con pocos parámetros, facilitando así su uso para futuras visiones.

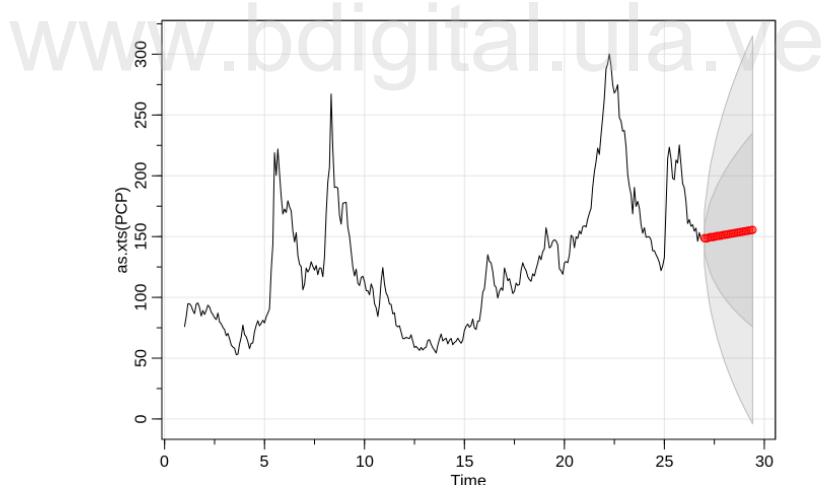


Figura 4.5: Grafica de la serie de tiempo del precio de otros café con su respectivo pronóstico

En la figura 4.5, se puede observar el modelo encontrado, con la predicción realizada, es de interés observar que mientras más datos sean necesarios predecir el error, tiende a crecer, lo que esta sombreado indica el intervalo de 90 y 95 % de confianza en el cual estará ubicado el punto real según el modelo. También se observa una tendencia de

subida de forma lineal.

#### 4.2.5 Indicador compuesto del OIC

El modelo hallado resulto ser un modelo muy bueno, ya que arrojo un error porcentual absoluto medio de 7.85% en los pronósticos de 30 muestras posteriores, además, se obtuvo un modelo con pocos parámetros, facilitando así su uso para futuras visiones.

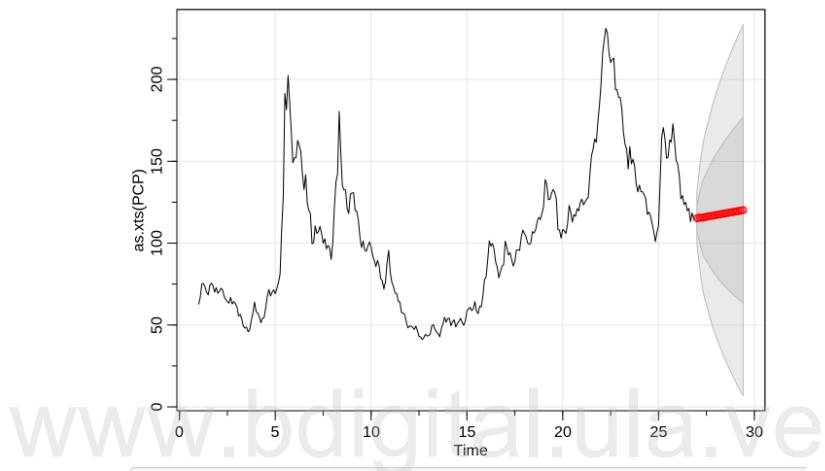


Figura 4.6: Grafica de la serie de tiempo del precio del indicador compuesto del OIC con su respectivo pronóstico

En la figura 4.6, se puede observar el modelo encontrado, con la predicción realizada, es de interés observar que mientras más datos sean necesarios predecir el error, tiende a crecer, lo que esta sombreado indica el intervalo de 90 y 95 % de confianza en el cual estará ubicado el punto real según el modelo. También se observa una tendencia de subida de forma lineal.

# Capítulo 5

## Conclusiones y Recomendaciones

Los modelos identificados en este trabajo, tienen una eficacia de predicción mayor al 85 %, debido a que los errores porcentuales absolutos medio, fueron inferiores a un 15%, siendo la serie temporal del precio del café robusta menos precisa de las 5 identificadas; sin embargo, considerando que se utilizó un principio de parsimonia de máximos de 6 parámetros por series, los modelos encontrados pueden decir que son buenos a corto plazo, con un máximo de 30 observaciones futuras.

Por otra parte, en las reglas de asociación se observan 291 reglas, de las cuales se hablaron de las más significantes, en cuanto a confianza y soporte. En general, el método de aprendizaje automático nos mostró la correlación que existe entre los precios de los diversos tipos de café estudiadas a lo largo de este trabajo. Mediante un antecedente y un consecuente (causa y efecto) resultante a partir de las observaciones que recibió el algoritmo y su posterior aprendizaje.

Finalmente, el estudio realizado en el presente trabajo de grado muestra que es posible predecir el precio de los distintos tipos de cafés, mediante el uso de estos modelos de análisis, así como también, medir cuan correlacionadas se encuentran éstas, facilitando así la toma de decisiones de inversión de capital en las distintas cepas de café.

Los análisis realizados en este proyecto son recomendables y útiles para todo

aquel inversionista, comerciante o agricultor de café, dado que con la ayuda de estas herramientas podrían determinar cuál es el mejor momento para invertir su capital en el negocio de café, cuándo predecir un decremento o alza en el precio. Se recomienda ver si existe la forma de obtener los datos diarios en vez de mensualmente, para así poder tener mayores observaciones que reforzaría los modelos para trabajos futuros. También al haber usado modelos arima, se debe considerar que estas predicciones solo tienen significacias para evento a corto plazo; por lo que si se desean modelos para largo plazo, se puede reajustar estos con los datos actualizados o simplemente implementar otro algoritmo de predicción.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

C.C. Reconocimiento

# Bibliografía

- Alcalde, A. (2018), Aprendizaje no supervisado y detección de anomalías: Reglas de asociación. Recuperado el 3 de noviembre del 2019.
- Carrillo, J. & Viserda, J. (2007), Minería de datos y aplicaciones, Master's thesis, Departamento de ingeniería telemática. Universidad Carlo III, Madrid.
- Carvajal, P. & Murillo, J. (2003), 'Estudio del pronóstico de la demanda de energía eléctrica, utilizando modelos de series de tiempo', *Scientia et Technica* (23), 37.
- Chapman, P. & Clinton, J. (2000), *The CRISP-DM Process*.
- Figueroa, E., Perez, F., Godinez, L. & Perez, R. A. (2019), 'Los precios de café en la producción y las exportaciones a nivel mundial', *Revista Mexicana de Economía y Finanzas Nueva Época* 14(1), 41–56.
- Ham, J. & Kamber, M. (2001), Scalable frequent-pattern mining methods: an overview, Master's thesis, Simon Fraser University, Canada.
- Hernandez, A. P. R. (2019), Diseño de un sistema dinámico no lineal ajustado por medio de un algoritmo de colonia artificial de abejas (abc) para la reconstrucción de la señal del precio diario interno del café colombiano, Master's thesis, Universidad Distrital Francisco José de Caldas, Bogota.
- Herrera, J. P. & Hernández, G. A. (2002), Metodología de un modelo arima condicionado para el pronóstico del pib, Master's thesis, Departamento Nacional de Planeación, Colombia.
- ICO, I. C. O. (2013). Recuperado el 25 de octubre del 2019.

Koontz, H. & Weihrich, H. (1999), *Administración, una perspectiva global*, Mexico.

Moine, J., silvia Gordillo & Haedo, A. (2010), Estudio comparativo de metodologías para minería de datos. Recuperado el 3 de octubre del 2019.

Moreno, K. G. M. (2018), Sintonización de un mecanismo basado en autómatas celulares para la reconstrucción de la dinámica producida por el precio del café colombiano, Master's thesis, Universidad Distrital Francisco José de Caldas, Bogota.

Nieto, M. F. R. & Velásquez, M. M. C. (2018), Incidencia del precio internacional del café sobre la producción del mismo estudio de caso café de colombia en el periodo 2002 y 2017, Master's thesis, Universidad Católica de Colombia, Bogota.

Rodrigo, J. A. (2018), Reglas de asociaciñ y algoritmo apriori con r. Recuperado el 3 de octubre del 2019.