



**Diseño metodológico para la evaluación de riesgo financiero y de mercado en la Plataforma  
AGRO de SFA CEBAR**

Sergio Andrés Sierra Luján

Tesis de maestría presentada para optar al título de Magíster en Métodos Cuantitativos para  
Economía y Finanzas

Director

Edwin Andrés Jiménez Echeverri, Magíster (MSc) en Administración Financiera

Universidad de Antioquia  
Facultad de Ciencias Económicas  
Maestría en Métodos Cuantitativos para Economía y Finanzas  
Medellín, Antioquia, Colombia  
2022



Maestría en Métodos Cuantitativos para Economía y Finanzas, Cohorte I.



**Repositorio Institucional:** <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - [www.udea.edu.co](http://www.udea.edu.co)

**Rector:** John Jairo Arboleda Céspedes

**Decano/Director:** Sergio Iván Restrepo Ochoa.

**Jefe departamento:** Lina María Grajales Vanegas

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

## **Dedicatoria**

Este trabajo se lo dedico a mi familia. Principalmente a mis padres quienes estuvieron apoyándome incondicional durante estos tres años de maestría.

También le quiero dedicar este trabajo a mi esposa Daniela Parra, quien nunca me dejó desfallecer en este proceso, quien se devolaba conmigo en las noches conmigo mientras estudiaba, y que con su amor y paciencia, me apoyó incondicionalmente en la culminación de la maestría.

## **Agradecimientos**

Quiero agradecerle a mi familia y a mi esposa con su apoyo incondicional en la finalización de esta etapa profesional de mi vida.

Así mismo agradecerle a mi tutor, el profesor Edwin Andrés Jiménez, quien se tomó el tiempo de guiarme en el proceso de desarrollo de la tesis y culminarla con honores.

Por su parte, agradecerla a SFA Cebal, por permitirme aplicar los conocimientos adquiridos en la maestría para mejorar los procesos de medición y evaluación de riesgos agropecuarios.

Finalmente, agradecerle a la Universidad de Antioquia, por aceptarme en la primera corte de la maestría y enseñarme herramientas cuantitativas aplicadas en el campo de la economía y finanzas.

## Tabla de contenido

Resumen ejecutivo .....	13
Abstract.....	14
1 Identificación del Problema .....	15
2 Objetivos .....	17
2.1 Objetivo general .....	17
2.2 Objetivos específicos.....	17
3 Marco Conceptual.....	18
3.1 Metodologías de medición del riesgo de mercado .....	18
3.1.1 Contexto histórico .....	18
3.1.1.1 Metodología de cálculo de volatilidad histórica .....	18
3.1.1.2 Metodología de cálculo de VaR sobre fluctuación de los precios .....	20
3.1.1.3 Metodología de cálculo de índice Beta .....	23
3.1.2 Contexto de monitoreo .....	23
3.1.3 Contexto predicción .....	24
3.1.3.1 Modelos Box y Jenkins .....	25
3.1.3.2 Redes neuronales .....	30
3.1.3.3 Modelos ARCH y GARCH .....	35
3.2 Metodologías de medición del riesgo financiero .....	37
3.2.1 Contexto histórico .....	37
3.2.1.1 Simulación Montecarlo .....	37
3.2.1.2 Análisis de sensibilidad a través de modelos de regresión .....	38
4 Metodología.....	43
4.1 Metodología para el diagnóstico .....	43
4.2 Metodología para la Intervención y evaluación .....	45

5	Diagnóstico .....	47
5.1	Estado actual de la Plataforma Agro .....	47
5.1.1	Componente temático.....	47
5.1.2	Componente tecnológico.....	48
5.2	Desarrollos futuros para la Plataforma Agro.....	49
5.3	Recomendaciones del equipo temático de riesgos ante la propuesta metodológica de la consultoría .....	50
5.4	Viabilidad del equipo de sistemas para la implementación de la propuesta de la consultoría .....	50
6	Plan de acción .....	52
6.1	Desarrollo de los algoritmos de las metodologías seleccionadas.....	52
6.1.1	Riesgo de Mercado.....	52
6.1.1.1	Cálculos de amenaza.....	53
6.1.1.1.1	Histórico .....	53
6.1.1.1.2	Monitoreo .....	61
6.1.1.1.3	Pronóstico .....	67
6.1.1.2	Cálculo de vulnerabilidad .....	92
6.1.1.2.1	Modelos de regresión para medir el grado de sensibilidad de los precios de los sistemas productivos agropecuarios .....	92
6.1.2	Riesgo Financiero.....	96
6.1.2.1	Cálculos de amenaza .....	97
6.1.2.2	Cálculo de vulnerabilidad.....	98
6.1.2.2.1	Simulación Montecarlo para medir el grado de sensibilidad de la rentabilidad de los sistemas productivos agropecuarios.....	98
6.2	Aplicación de las metodologías de evaluación de riesgo de mercado y financiero .....	101
6.2.1	Aplicación de metodología para missing data en la serie de precios de aguacate hass	

6.2.2	Riesgo de Mercado.....	102
6.2.2.1	Histórico.....	102
6.2.2.1.1	Indicador de Volatilidad .....	102
6.2.2.1.2	VaR de las fluctuaciones de los precios .....	103
6.2.2.1.3	Índice Beta.....	105
6.2.2.2	Monitoreo.....	106
6.2.2.2.1	Medias Móviles .....	106
6.2.2.2.2	Bandas de Bollinger.....	107
6.2.2.2.3	Indicador de Media Móvil de Convergencia/Divergencia (MACD).....	108
6.2.2.2.4	Índice de Fuerza Relativa (RSI) .....	108
6.2.2.3	Predicción .....	109
6.2.2.3.1	Modelo Box y Jenkins .....	109
6.2.2.3.2	Redes Neuronales LSTM.....	117
6.2.2.3.3	Modelo ARCH-GARCH .....	123
6.2.2.4	Análisis de vulnerabilidad.....	127
6.2.3	Riesgo financiero.....	131
6.2.3.1	Simulación Montecarlo .....	131
	Conclusiones y recomendaciones.....	135
	Referencias bibliográficas .....	138

## Lista de tablas

Tabla 1 Hiperparámetros tuneados para modelo LSTM para predicción de arroz paddy basmati	32
Tabla 2 Hiperparámetros tuneados para cada modelo LSTM.....	34
Tabla 3 Evaluación de las métricas de predicción para todos los modelos.....	34
Tabla 4 Estudios relacionados con la modelación LSTM.....	35
Tabla 5 Metodologías seleccionadas para la medición de riesgo de Mercado y Financiero .....	52
Tabla 6 Ejemplo de transformación de las series de precios.....	81
Tabla 7 Estructura del flujo de caja libre y del productor .....	98
Tabla 8 Resultados del modelo de predicción modelo Box y Jenkins .....	116
Tabla 9 Aplicación del método de ventanas para el entrenamiento de la red neuronal LSTM ...	117
Tabla 10 Resultados del modelo de predicción modelo Redes Neuronales LSTM .....	119
Tabla 11 Resultados del modelo de predicción modelo Redes Neuronales LSTM con tuneo en los hiperparámetros .....	121
Tabla 12 Comparación de los modelos de redes neuronales LSTM con y sin tuneo en los hiperparámetros .....	122
Tabla 13 Comparación de los modelos de Box-Jenkins y redes neuronales LSTM.....	122
Tabla 14 Flujo de caja libre para el sistema productivo de aguacate .....	132



## Lista de figuras

Figura 1 Algoritmo para estimación de modelo de Redes Neuronales .....	27
Figura 2 Resumen metodologías de pronósticos de precios de productos agropecuarios en el mercado spot .....	30
Figura 3 Hiperparametrización modelo LSTM para los precios del cerdo. ....	31
Figura 4 Plataforma Agro SFA CEBAR .....	49
Figura 5 Pruebas de Bondad de Ajuste .....	58
Figura 6 Intervalos para la medición de volatilidad de los rendimientos de los precios del aguacate hass .....	103
Figura 7 Aplicación de pruebas de bondad de ajuste para los rendimientos de los precios del aguacate hass en la central mayorista de Antioquia .....	104
Figura 8 VaR en riesgo de las fluctuaciones de los precios del aguacate hass en la central mayorista de Antioquia .....	105
Figura 9 Pruebas de raíces unitarias estacionarias .....	112
Figura 10 Pruebas de raíces unitarias estacionales.....	113
Figura 11 Modelo resumido para los precios del aguacate hass de la central mayorista Antioquia .....	114
Figura 12 ACF residuales del modelo.....	115
Figura 13 Arquitectura modelo de redes neuronales LSTM para la serie de precios de aguacate hass de la central mayorista de Antioquia.....	118
Figura 14 Arquitectura final modelo de redes neuronales LSTM para la serie de precios de aguacate hass de la central mayorista de Antioquia con calibración en los hiperparámetros .....	120
Figura 15 Rendimientos de los precios de aguacate hass para la central mayorista de Antioquia .....	124
Figura 16 Prueba de efecto ARCH sobre los rendimientos de los precios .....	125
Figura 17 Modelo de varianza para los rendimientos de los precios .....	125
Figura 18 Residuales modelo de varianza para los rendimientos de los precios .....	126

Figura 19 Pruebas sobre los residuales del modelo de varianza para los rendimientos de los precios .....	126
Figura 20 Pronostico de los rendimientos de los precios del aguacate hass para la central mayorista de Antioquia .....	127
Figura 21 Modelo de regresión lineal para los precios de la central mayorista de Antioquia con respecto a la cantidad de importaciones, exportaciones y tasa representativa del mercado .....	128
Figura 22 Prueba de Breusch-Pagan para la heterocedasticidad en el término de error .....	129
Figura 23 Modelo de regresión lineal con la corrección de los errores con la matriz de wald ....	129
Figura 24 Valor inflado de la varianza del modelo .....	130
Figura 25 Modelo de regresión con datos atípicos Yohai .....	130
Figura 26 Modelo de regresión con datos atípicos Huber.....	131
Figura 27 Intervalo de confianza para el Valor Presente Neto .....	133
Figura 28 Modelo de regresión lineal para medir el impacto de las variables sobre el VPN .....	134

## Lista de gráficos

Gráfico 1 Resultado de todos los modelos de predicción. ....	31
Gráfico 2 Comparación predicción y precio real del arroz paddy basmati .....	33
Gráfico 3 Serie de tiempo con datos faltantes .....	55
Gráfico 4 Gráfico de serie de precios con datos imputados .....	57
Gráfico 5 Resultado Bandas de Bollinger ambiente $R$ .....	62
Gráfico 6 Resultado Medias Móviles ambiente $R$ .....	64
Gráfico 7 Resultado MACD ambiente $R$ .....	65
Gráfico 8 Resultado gráfico RSI ambiente $R$ .....	66
Gráfico 9 Series de precios Aguacate hass central mayorista de Antioquia 2012-2022.....	101
Gráfico 10 Series de precios imputados para el Aguacate hass central mayorista de Antioquia 2012-2022.....	102
Gráfico 11 Histograma de los rendimientos de los precios del aguacate hass en la central mayorista de Antioquia .....	104
Gráfico 12 Medias móviles para los precios del aguacate hass en la central mayorista de Antioquia.....	106
Gráfico 13 Bandas de Bollinger para los precios del aguacate hass en la central mayorista de Antioquia.....	107
Gráfico 14 MACD para los precios del aguacate hass en la central mayorista de Antioquia.....	108
Gráfico 15 RSI para los precios del aguacate hass en la central mayorista de Antioquia .....	109
Gráfico 16 Monthplot para los precios del aguacate hass en la central mayorista de Antioquia. ....	110
Gráfico 17 ACF para los precios del aguacate hass en la central mayorista de Antioquia.....	111
Gráfico 18 PACF para los precios del aguacate hass en la central mayorista de Antioquia.....	111
Gráfico 19 Comparativo de la predicción de los precios del aguacate hass de la central mayorista de Antioquia con respecto a los precios reales con modelo Box-Jenkins .....	116
Gráfico 20 Comparativo de la predicción de los precios del aguacate hass de la central mayorista de Antioquia con respecto a los precios reales con modelo Redes neuronales LSTM .....	119

Gráfico 21 Comparativo de la predicción de los precios del aguacate hass de la central mayorista de Antioquia con respecto a los precios reales con modelo Redes neuronales LSTM calibrado 121

Gráfico 22 Predicción del precio del aguacate hass en la central mayorista de Antioquia para los periodos octubre – diciembre de 2022 ..... 123

Gráfico 23 Histograma de los valores presente netos simulados ..... 132

## **Resumen ejecutivo**

La compañía SFA CEBAR es una empresa líder en el sector financiero agropecuario especializada en la estructuración, evaluación de riesgo técnico, financiero y ambiental de proyectos agropecuarios y soluciones financieras para el sector agropecuario. Esta compañía desarrolló una plataforma llamada “Plataforma Agro”, cuyos objetivos son agilizar el acceso a la financiación bancaria por parte de los productores agropecuarios, y realizar una correcta medición financiera de cada negocio evaluado, con el fin de aportar elementos para la toma de decisiones por parte de los productores. Dentro de los módulos de esta Plataforma se encuentra el módulo de riesgos, donde se destaca la medición del riesgo de mercado y el riesgo financiero. En este contexto, la consultoría busca diseñar la metodología de evaluación del riesgo de mercado y financiero para la plataforma AGRO de SFA Cebar SAS, la cual será un soporte para la toma de decisiones de los productores agropecuarios y las entidades financieras. Para lo anterior, se realizó la validación del estado actual de la Plataforma Agro, luego se realizó la revisión de experiencias nacionales e internacionales sobre la medición de estos riesgos, posteriormente se seleccionaron y se diseñaron los algoritmos de las metodologías seleccionadas con base en la revisión de experiencias. Como resultado principal, se obtuvo una selección de 11 indicadores y metodologías, divididas en los contextos histórico, monitoreo y pronóstico, los cuales se entregaron al área de sistemas de información para su posterior implementación. Como conclusión principal de la consultoría se logra la inclusión de estos indicadores en la Plataforma Agro de SFA Cebar lo que generará un valor agregado a las compañías de financiamiento, debido a que tendrán una visión completa del producto sujeto a financiación para determinar el monto del préstamo y reconocer cuáles riesgos corren al momento de otorgar el crédito. Por otro lado, los indicadores y metodologías propuestos permiten prevenir las situaciones de riesgo a las cuales pueden estar expuestos los productores, de manera que puedan tomar decisiones para mitigar o transferir los riesgos antes de su materialización.

Palabras Clave: Finanzas agropecuarias, riesgo de mercado agropecuario, riesgo financiero agropecuario.

## **Abstract**

The company SFA CEBAR is a leading company in the agricultural financial sector specialized in structuring, technical, financial and environmental risk assessment of agricultural projects and financial solutions for the agricultural sector. This company developed a platform called "Plataforma Agro", whose objectives are to expedite access to bank financing by agricultural producers, and to carry out a correct financial measurement of each business evaluated, in order to provide elements for decision making. by the producers. Within the modules of this Platform is the risk module, where the measurement of market risk and financial risk stands out. In this context, the consultancy seeks to design the market and financial risk assessment methodology for the AGRO platform of SFA Cebbar SAS, which will be a support for decision-making by agricultural producers and financial entities. For the above, the validation of the current state of the Agro Platform was carried out, then the review of national and international experiences on the measurement of these risks was carried out, later the algorithms of the selected methodologies were selected and designed based on the review. of experiences. As a main result, a selection of 11 indicators and methodologies was obtained, divided into historical, monitoring and forecast contexts, which were delivered to the information systems area for subsequent implementation. The main conclusion of the consultancy is the inclusion of these indicators in the SFA Cebbar Agro Platform, which will generate added value for financing companies, since they will have a complete vision of the product subject to financing to determine the amount of the loan. and recognize the risks they run when granting credit. On the other hand, the proposed indicators and methodologies make it possible to prevent risk situations to which producers may be exposed, so that they can make decisions to reduce or transfer risks before they materialize.

**Keywords:** Agricultural finance, agricultural market risk, agricultural financial risk.

## **1 Identificación del Problema**

La compañía SFA CEBAR es una empresa líder en el sector financiero agropecuario especializada en la estructuración, evaluación de riesgo técnico, financiero y ambiental de proyectos agropecuarios y soluciones financieras para el sector agropecuario. Al ser esta empresa un referente en el sector debe estar actualizada en todos los aspectos relacionados con el desarrollo de su razón social tales como los cambios en las políticas de estructuración y evaluación de proyectos agropecuarios, las políticas de incentivos otorgados por FINAGRO (Fondo para el financiamiento del sector Agropecuario) y las normas de los bancos asociados para el otorgamiento de crédito, entre otros.

En el marco de las ideas anteriores, SFA CEBAR ha iniciado el desarrollo de una plataforma llamada “Plataforma Agro”, cuyos objetivos son agilizar el acceso a la financiación bancaria por parte de los productores agropecuarios, y realizar una correcta medición financiera de cada actividad evaluada, con el fin de aportar a la toma de decisiones de los productores. Esta plataforma se piensa para todos los eslabones de la cadena agropecuaria, en especial, a que los pequeños y medianos productores que presentan un bajo nivel de financiación bancaria, además, existe una necesidad inmediata de capacitar a los agricultores sobre las mejores prácticas de gestión de riesgos para decidir la elección de un cultivo en particular (Senapati, 2020).

La plataforma cuenta con varios módulos entre los que se destacan el módulo de evaluación técnica, evaluación financiera, evaluación de riesgos agroclimáticos, riesgos sanitarios, riesgos de mercado y riesgo financiero, Sin embargo, solo los módulos de evaluación técnica y financiera se encuentran desarrollados al 100%, mientras que los módulos de riesgos aún no cuentan con un desarrollo metodológico.

En la planeación inicial, los módulos de riesgo financiero y de mercado de la plataforma tienen como objetivo brindar a los productores información clara y precisa para mejorar su toma de decisiones, debido a que por falta de conocimiento “los agricultores están aumentando su producción y las áreas que siembran en respuesta a la volatilidad de los precios”, (Assouto, Houensou, & Semedo, 2020), generando de esta forma caídas en el precio por sobreoferta.

Así mismo, este módulo busca que las entidades financieras puedan desarrollar políticas de crédito con base en los resultados de los análisis de mercado realizados por la plataforma. De acuerdo con (Senapati, 2020), es evidente la necesidad de desarrollar la medición del riesgo para el sector agropecuario, ya que:

“proporciona a las agencias gubernamentales un esquema para saber cómo el entorno agrícola riesgoso afecta las decisiones de producción de los agricultores y el diseño de políticas como el seguro de cosechas, el seguro de cosechas basado en el clima y otras redes de seguridad que abordan efectivamente el problema de los agricultores.” (p. 1)

Por todo lo expuesto en los párrafos anteriores, la pregunta de investigación planteada es ¿Cuáles deben ser las metodologías que deben contener los módulos de riesgo financiero y de mercado de la plataforma agro CEBAR, de tal forma que brinde información pertinente, oportuna y eficaz para mejorar el proceso de toma de decisión de las entidades financieras y de los productores agropecuarios que interactúen con la plataforma?



## **2 Objetivos**

### **2.1 Objetivo general**

Diseñar la metodología de evaluación del riesgo de mercado y financiero para la plataforma AGRO de SFA Cebal SAS, la cual será un soporte para la toma de decisiones de los productores agropecuarios y las entidades financieras.

### **2.2 Objetivos específicos**

- Identificar experiencias nacionales e internacionales sobre la medición del riesgo de mercado y financiero en el sector agropecuario.
- Seleccionar aquellas metodologías que ofrezcan mejores resultados en la toma de decisiones de los productores agropecuarios y entidades financieras.
- Desarrollar los algoritmos de la metodología creada para la evaluación del riesgo de mercado y financiero para la plataforma AGRO de SFA Cebal.

### **3 Marco Conceptual**

La evaluación del riesgo de mercado y el riesgo financiero presenta tres momentos temporales importantes que se deben tener en cuenta en el momento de su evaluación: histórico, monitoreo y predicción. Esto con el propósito de que las decisiones que tomen las entidades financieras, en materia de financiación; y los productores agropecuarios, en materia de producción; estén mejor sustentadas.

En el marco de la idea mencionada anteriormente, el capítulo de marco teórico presentará la revisión de literatura de las metodologías según el contexto temporal de análisis, y también expondrá las propiedades, beneficios y contextos de la aplicación de éstas.

#### **3.1 Metodologías de medición del riesgo de mercado**

##### ***3.1.1 Contexto histórico***

En este primer momento de análisis, las metodologías que se desarrollaran son el cálculo de la volatilidad histórica de los precios, el valor en riesgo (VaR) sobre las fluctuaciones de los precios y el índice beta para comparar la volatilidad de un mercado local con respecto al mercado nacional. La bondad de éstas tres metodologías es que permite conocer cómo han variado los precios a lo largo de su historia y comparar las volatilidades de un mismo producto agropecuario en los diferentes mercados en los que se comercializa. Adicionalmente, permite determinar un intervalo de confianza en el que se podrá determinar el rango en el que pueden llegar a fluctuar dichos precios.

##### **3.1.1.1 Metodología de cálculo de volatilidad histórica**

Como punto inicial, se tiene la definición de Chicago Mercantile Exchange (CME GROUP) de la forma en cómo se tiene que calcular la volatilidad histórica, la cual, se describe (Piot-Lepetit & M'Barek, 2011) así:

“El cálculo que CME propone para la volatilidad histórica es la desviación estándar anualizada de la primera diferencia en los valores logarítmicos de los precios de liquidación de futuros cercanos. Matemáticamente, se puede escribir como:

$$Volatilidad = Desviación\ estandar \left( \ln \left( \frac{P_t}{p_{T-1}} \right) \right) * \sqrt{252}$$

” (p. 29)

De otro lado, se tiene la definición de (Benavides, 2004) que describe la volatilidad histórica simplemente como un cálculo de la varianza o desviación estándar de los rendimientos de la manera estadística habitual durante algún período histórico (marco de tiempo) y esto puede convertirse en un pronóstico para todos los períodos futuros.

A partir de las fórmulas que se acaban de describir, respecto a sus aplicaciones en el sector agropecuario se tiene el trabajo de Assouto et al (2020) quienes:

“Usaron la volatilidad específica del cultivo de maíz considerado para capturar el riesgo de precio del producto. Esta volatilidad se determina a partir de la desviación estándar de la primera diferencia del logaritmo de precios. La medida de volatilidad se anualiza multiplicándola por la raíz cuadrada de 12, debido al sigma que se calcula a intervalos mensuales” (p. 6)

Otra forma de calcular la volatilidad, es calculando el coeficiente de variabilidad, tal y como lo muestra (León-Castro et al., 2021) quienes:

“Como resultado del desarrollo acelerado de los productos agrícolas que se ofrecen en el mercado de futuros, existe una alta volatilidad en los precios. Desde la segunda mitad de 2008, los precios de diferentes commodities, especialmente agrícolas, han disminuido y la volatilidad de sus precios está aumentando. Este fenómeno se debe a los bajos niveles de inventarios, la depreciación del dólar estadounidense y la producción de biocombustibles, entre otros factores. Por lo tanto, es importante generar nuevos escenarios que consideren

la posible volatilidad del precio de varios productos, incluidos los productos agrícolas. Una de las formas clásicas de calcular la volatilidad es según el coeficiente de variación. Este se puede definir de la siguiente manera:

$$v(a_1, \dots, a_n) = \frac{\sigma(a_1, \dots, a_n)}{\mu(a_1, \dots, a_n)}$$

” (p. 164)

### 3.1.1.2 Metodología de cálculo de VaR sobre fluctuación de los precios

La metodología que permite calcular la volatilidad de precios es la que se describe en el artículo de (Wang, Zhao, & Huang, 2010) donde aplican la metodología de Value at Risk (VaR) sobre las variaciones de siete precios de productos agropecuarios en el mercado chino según la distribución que sigue cada precio. Para la aplicación de esta metodología, (Wang et al., 2010) la dividen en cuatro componentes que se describen a continuación:

Componente 1: Análisis descriptivo sobre las fluctuaciones de los precios.

En el componente de análisis descriptivo se toman los mismos indicadores que se mencionaron en la metodología anterior, exceptuando el uso del test de Jarque Bera, donde (Jarque & Bera, 1987) demuestran que este test sirve para definir tanto la simetría como la curtosis de una serie de datos, que se puede comparar en una distribución  $\chi^2$  para identificar la hipótesis de existencia de normalidad, empleando la siguiente fórmula:

$$JB = \frac{n - k + 1}{6} * \left( A_p^2 + \frac{1}{4} (K_p - 3)^2 \right) \sim \chi^2; \begin{cases} JB \leq \chi^2 \rightarrow \text{Serie normal} \\ JB \geq \chi^2 \rightarrow \text{Serie no normal} \end{cases}$$

La prueba de Jarque – Bera se interpreta sobre la condición del p-valor en contraste del valor asignado al JB en distribución  $\chi^2$ , si el valor de JB es menor que el P-valor entonces se acepta la hipótesis de que la serie de datos presume normalidad. En caso contrario es una serie anormal de datos.

Componente 2: Pruebas de bondad de ajuste para determinar la distribución de cada serie de precios.

Posterior al análisis descriptivo de las series, los autores observaron que no todas las series de tiempo estudiadas se distribuían normalmente. Es por ello que para cada producto se realizó una prueba de bondad de ajuste con los estadísticos  $chi^2$ , Kolgomorov-Smirnov y Anderson Darling para determinar mejor distribución paramétrica y evitar así sesgos en los cálculos del VaR.

La prueba  $chi^2$  “se usa para comparar la posible diferencia entre las frecuencias observadas en la distribución de una variable con respecto a las esperadas” (Gómez, Danglot, & Vega, 2003). El estadístico que se usa para la comparación es:

$$x^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

Y la prueba de hipótesis de este test corresponde a:

$$H_0: F_0 = F_e$$

$$H_a: F_0 \neq F_e$$

Por su parte, el estadístico de Kolmogorov-Smirnov “pertenece a la clase superior de estadísticas EDF y esta clase de estadísticas se basa en la mayor diferencia vertical entre la distribución hipotética y empírica” (Mohd & Bee, 2011) en el cual el estadístico de prueba se define como:

$$T = \sup_x [F^*(x) - F_n(x)]$$

Donde “sup” significa supremum, es decir, el más grande, y  $F^*(x)$  es la función de distribución hipotética, mientras que  $F_n(x)$  es el EDF estimado con base en la muestra aleatoria. En la prueba de normalidad de KS,  $F^*(x)$  se toma como una distribución normal con media conocida,  $u$ , y desviación estándar,  $\sigma$ .

La estadística de prueba KS está diseñada para probar:

$$H_0: F(x) = F^*(x)$$

$$H_a: F(x) \neq F^*(x)$$

Como última prueba de bondad de ajuste, se tiene el test de Anderson-Darling el cual “pertenece a la clase cuadrática del estadístico EDF, el cual se basa en la diferencia al cuadrado de  $[F_n(x) - F^*(x)]^2$ ” (Mohd & Bee, 2011). Esta prueba se define como:

$$w_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \psi(F^*(x)) dF^*(x)$$

Y la hipótesis para esta prueba se expresa así:

$$H_0: F(x) = F^*(x)$$

$$H_a: F(x) \neq F^*(x)$$

Componente 3: Estimación de parámetros de las distribuciones definidas.

Luego de realizar las pruebas de bondad de ajuste y definir la función de distribución que mejor se adapta al comportamiento de los precios, los autores estimaron los parámetros de cada distribución mediante el método de estimación de máxima verosimilitud el cual se define matemáticamente por (Montgomery & Runger, 2008) así:

“Suponga que  $X$  es una variable aleatoria con distribución de probabilidad  $f(x; \theta)$ , donde  $\theta$  es un solo parámetro desconocido. Sean  $x_1, x_2, \dots, x_n$  los valores observados en una muestra aleatoria de tamaño  $n$ . Entonces la función de verosimilitud de la muestra es:

$$L(\theta) = f(x_1; \theta) * f(x_2; \theta) * \dots * f(x_n; \theta)$$

Obsérvese que la función de máxima verosimilitud es ahora una función exclusiva del parámetro desconocido  $\theta$ . El estimador de máxima verosimilitud de  $\theta$  es el valor de  $\theta$  que maximiza la función de verosimilitud  $L(\theta)$ ”

Componente 4: Cálculo del VaR.

Como componente final de la metodología propuesta, los autores calcularon el rango de fluctuación de los precios con un 95% de confianza con el Valor en Riesgo (VaR) teniendo en cuenta las distribuciones definidas en el componente anterior.

El VaR se define como “en una condición normal del mercado financiero y un cierto nivel de confianza, cuál sería el tamaño de la pérdida si se mide en el peor de los casos esperados de un período de tiempo dado” (Wang et al., 2010) y matemáticamente se describe por (Wang et al., 2010) de la siguiente forma:

“El conjunto  $X$  se define como una variable aleatoria para describir la pérdida de la cartera,  $F(x)$  es la función de distribución de probabilidad y el nivel de confianza es, por lo tanto:

$$VaR(\alpha) = -\min(x|F(x) \geq \alpha), \quad \text{luego}$$

$$P(\Delta X > VaR) = 1 - \alpha$$

Aquí,  $\Delta X$  es la pérdida de cartera en el período de tenencia y el VaR es el valor en riesgo bajo el nivel de confianza  $\alpha$ . Como se puede ver, el VaR en realidad calcula los cuartiles superior e inferior de  $F(x)$  bajo el nivel de confianza  $\alpha$ ”

### 3.1.1.3 Metodología de cálculo de índice Beta

“El Beta representa la sensibilidad de los rendimientos de las acciones a los cambios en el mercado” (Insana, 2022). En este sentido, Insana (2022) define que:

“Beta representa la pendiente de la relación lineal entre el activo y rendimientos del mercado. Teniendo una serie de tiempo bivariado  $(R_i, R_{Mi})_{i=1}^2$  compuesta por el exceso de rendimiento de un valor y por el exceso de rendimiento del mercado, podemos evaluar beta utilizando un método clásico de mínimos cuadrados ordinarios (OLS):

$$\hat{\beta} = \frac{cov(R_i, R_{Mi})}{var(R_{Mi})}$$

” (p.2)

### 3.1.2 Contexto de monitoreo

El contexto de monitoreo tiene como propósito hacer un seguimiento en tiempo real del comportamiento de los precios. Para ello se realizó una adaptación de los indicadores técnicos de bolsa de valores a los precios del sector agropecuario. Estos indicadores permitirán a las entidades financieras y a los productores ver cómo se están comportando los precios en el presente y el impacto que pueden llegar a tener éstos sobre los flujos de caja de los sistemas productivos.

En este sentido, las metodologías que se adaptaron fueron: medias móviles, bandas de bollinger, MACD y RSI, para los cuales el marco teórico que los sustenta se describe a continuación:

Como punto inicial, se tiene el trabajo de (Macijauskas, 2012) quien utilizó las medias móviles para analizar las tendencias de varios activos financieros en épocas de incertidumbre, encontrando

como resultado que “el método SMA, cuando se utiliza como indicador de tendencia para los principales índices de acciones y REIT, puede reducir significativamente la desviación estándar y las medidas de reducción máxima” (p.114)

Por otro lado, se tiene el trabajo de (Thi Nhung et al., 2021) quienes:

“Utilizando datos primarios recopilados de entrevistas en profundidad con 11 expertos prácticos que han estado trabajando en la industria de los productos básicos durante un promedio de 7 años y utilizando la herramienta de las Bandas de Bollinger durante al menos 3 años, el artículo muestra que los signos de las Bandas de Bollinger en las previsiones de precios o la toma de decisiones para comprar o vender en el futuro comercio de productos agrícolas en Vietnam es muy apreciada debido a su precisión.” (p.56)

Finalmente, se tiene el trabajo doctoral de (Heymans, 2008), quien demostró la eficiencia de los indicadores técnicos, RSI, bandas de bollinger, MACD y medias móviles, para detectar puntos clave de entrada y salida en el mercado de commodities de Sur África.

En dicho trabajo se puede observar que la aplicación de los índices mencionados en el monitoreo del precio, son una propuesta novedosa en el sentido de que se pueden tomar las señales de estos indicadores como puntos de tendencias del precio de los sistemas productivos agropecuarios.

### ***3.1.3 Contexto predicción***

Finalmente, las metodologías que se van a desarrollar en este contexto tienen como propósito tratar de describir el comportamiento futuro de los precios de los productos del sector agropecuario, tanto de la media como de su volatilidad. Las metodologías presentadas se pueden aplicar a cualquier serie de tiempo, por esta razón se tuvieron en cuenta para la predicción de precios.



### 3.1.3.1 Modelos Box y Jenkins

Como metodología inicial de pronóstico, se encuentra la metodología Box-Jenkins, la cual es una familia de modelos que se aplican según el comportamiento de la serie de tiempo. Dentro de la familia mencionada, están los modelos Autorregresivos (AR), Medias Móviles (MA), Autorregresivos de Medias Móviles (ARMA) y Autorregresivos Integrados de Medias Móviles (ARIMA).

Los procesos autorregresivos según (Cowpertwait & Metcalfe, 2011) “son, como sugiere su nombre, regresiones sobre sí mismos. Específicamente, un proceso autorregresivo de p-ésimo orden  $Y_t$  satisface la ecuación”:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

“El valor actual de la serie  $Y_t$  es una combinación lineal de los p valores pasados más recientes de sí misma más un término error que incorpora todo lo nuevo en la serie en el tiempo t que no se explica por los valores pasados” (Cowpertwait & Metcalfe, 2011)

Por su parte, “en el caso de que sólo un número finito de pesos  $\psi$  sea distinto de cero, tenemos lo que se llama un proceso de promedio móvil” (Cowpertwait & Metcalfe, 2011) cuya ecuación es:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_p e_{t-p}$$

“Llamamos a esta serie una media móvil de orden q y abreviamos el nombre a MA(q). La terminología media móvil surge del hecho de que  $Y_t$  se obtiene aplicando los pesos 1,  $-\theta_1$ ,  $-\theta_2$ , ...,  $-\theta_q$  a las variables  $e_t$ ,  $e_{t-1}$ ,  $e_{t-2}$ , ...,  $e_{t-q}$  y luego moviendo los pesos y aplicándolos a  $e_{t+1}$ ,  $e_t$ ,  $e_{t-1}$ , ...,  $e_{t-q+1}$  para obtener  $Y_{t+1}$ ” (Cowpertwait & Metcalfe, 2011)

Luego, “Si asumimos que la serie es en parte autorregresiva y en parte media móvil, obtenemos un modelo de series de tiempo bastante general” (Cowpertwait & Metcalfe, 2011). A este proceso se le denomina ARMA y se representa matemáticamente así:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_p e_{t-p}$$

Como último modelo de la metodología Box-Jenkins, se tiene el modelo ARIMA donde Cowpertwait y Metcalfe (2011) explican que:

“Se dice que una serie de tiempo  $Y_t$  sigue un modelo de promedio móvil autorregresivo integrado si la diferencia d-ésima  $W_t = \nabla^d Y_t$  es un proceso ARMA estacionario. Si  $W_t$  sigue un modelo ARMA (p, q), decimos que  $Y_t$  es un proceso ARIMA (p, d, q)”

Los términos matemáticos de un proceso ARIMA (p,d,q) con  $W_t = Y_t - Y_{t-1}$  siguen la siguiente ecuación:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

O en términos de la serie observada sería:

$$Y_t - Y_{t-1} = \phi_1 (Y_{t-1} - Y_{t-2}) + \phi_2 (Y_{t-2} - Y_{t-3}) + \dots + \phi_p (Y_{t-p} - Y_{t-p-1}) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Ahora bien, como primera metodología de pronóstico se revisó el artículo de (Zou, Xia, Yang, & Wang, 2007) quienes pronosticaron los precios del tomate en el mercado chino con el uso de los modelos ARIMA y Redes Neuronales Artificiales con ventanas de pronóstico diario, semanal y mensual. Para llevar a cabo este proceso metodológico de pronóstico, los autores proponen el paso a paso que se describe a continuación:

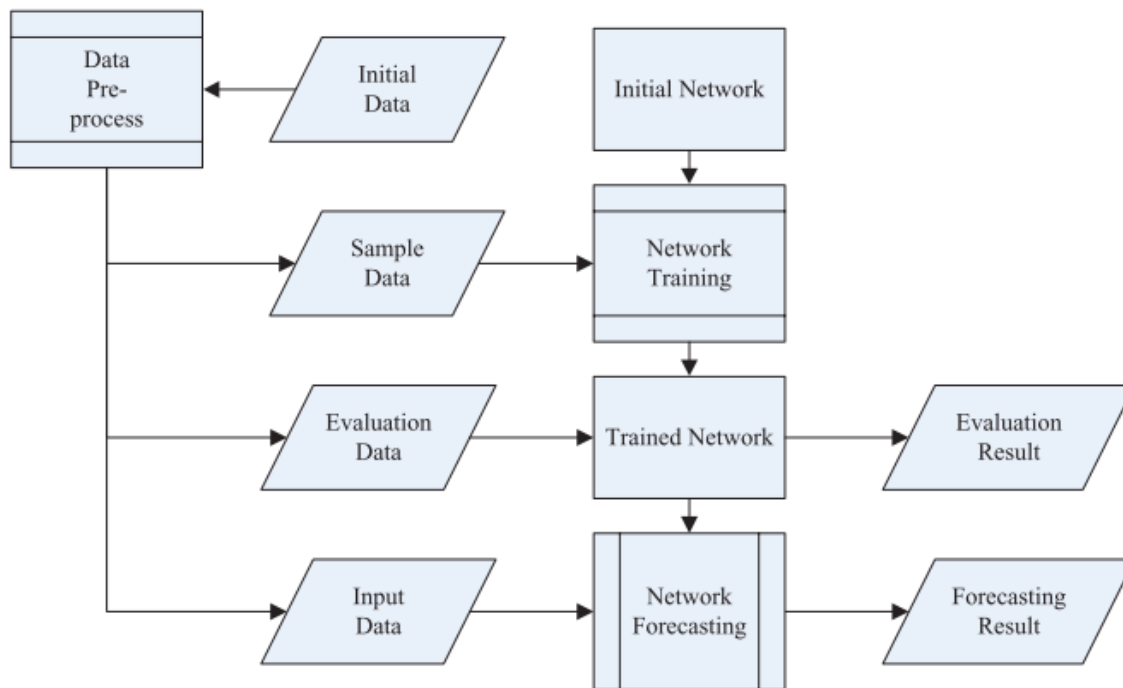
Paso 1: Estimación del modelo ARIMA para cada periodicidad establecida.

Paso 2: Estimación del Modelo de Redes Neuronales Artificiales para cada periodicidad establecida.

Antes de la descripción del algoritmo de las Redes Neuronales, (Zou et al., 2007) definen este modelo matemático o computacional como aquel que “intenta simular la estructura y / o aspectos funcionales de redes neuronales biológicas. Consiste en un grupo interconectado de neuronas artificiales y procesa la información utilizando un enfoque conexionista de la computación”.

Dentro de los tipos de Redes Neuronales, el usado por los autores es el feed-forward el cual es definido por (Zou et al., 2007) como aquel modelo donde “las conexiones entre las unidades no forman un ciclo. En esta red, la información se mueve en una sola dirección, hacia adelante, desde los nodos de entrada, a través de los nodos ocultos (si los hay) y hacia los nodos de salida”.

**Figura 1** Algoritmo para estimación de modelo de Redes Neuronales



Nota: Tomado de (Zou et al., 2007)

Paso 3: Comparación de modelos.

Como paso final, los autores comparan los modelos estimados en cada una de las periodicidades definidas. Para ello, se hace uso de la fórmula de error relativo que se define como:

$$P_{error} = \frac{|X_f - X_{obs}|}{X_{obs}} * 100$$

Una metodología similar a la mencionada anteriormente se encuentra en el trabajo de (Li, Xu, & Li, 2010), donde también se usan los modelos ARIMA y Redes Neuronales Artificiales para pronosticar precios, aunque en este caso no se hace con el precio del tomate, sino con los precios de los granos en el mercado chino.

Cabe resaltar que en esta metodología, a diferencia de la anterior, utilizan un tercer modelo cuya característica es la combinación de los modelos propuestos inicialmente, obteniendo así un modelo híbrido.

En este sentido, los componentes propuestos por los autores en la metodología descrita en su artículo son los que se describen a continuación:

Componente 1: Selección y estimación de modelo.

- Modelos ARIMA

(Li et al., 2010) explican que en el modelo ARIMA “el valor futuro de una variable es una combinación lineal de valores y errores pasados. Generalmente, una serie de tiempo no estacional se puede modelar como una combinación de valores pasados y errores”.

- Redes Neuronales Artificiales

Para los autores, la ventaja principal de las Redes Neuronales es la capacidad que tienen para modelar relaciones complejas no lineales, sin supuestos a priori de la naturaleza de la relación. “Este modelo realiza un mapeo funcional no lineal de las observaciones pasadas  $(X_{t-1}, X_{t-2}, \dots, X_{t-n})$  al valor futuro  $X_t$ , *i. e* y se expresa de la siguiente forma:

$$X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-p}, w) + et$$

*Donde:*

*w es un vector de todos los parámetros*

*f es una función determinada por la estructura de la red y los pesos de conexión.*

- Combinación de modelos ARIMA y Redes Neuronales Artificiales (ANN) con el método Ponderaciones Iguales (EW).

“Se han realizado una variedad de estudios que muestran que una combinación de pronósticos, a menudo supera a los pronósticos obtenidos con un solo método” (Li et al., 2010). Es por ello, que los autores decidieron combinar los resultados del modelo ARIMA y ANN para validar la hipótesis descrita haciendo uso de la fórmula que se describe a continuación.

$$\hat{y}(t) = \sum_{i=1}^n w_i y_i(t)$$

Lo difícil de la combinación de modelos según los autores, es determinar los pesos  $w_i$  adecuados. Sin embargo, en la literatura existente se encuentran varios métodos para determinar los mismos. En ese mismo sentido, los autores decidieron hacer uso del método de ponderaciones iguales (EW), que utiliza un promedio aritmético de los pronósticos individuales, es decir, los pesos asignados son exactamente iguales.

Componente 2: Pronóstico y evaluación de modelos.

En el pronóstico y evaluación de modelos, los autores estimaron los parámetros para cada modelo y realizaron el respectivo pronóstico sobre dos etapas de la serie: muestra de entrenamiento y muestra de evaluación. Con los resultados obtenidos, se calcularon los criterios de evaluación Error Absoluto Medio (MAE), Error Cuadrático Medio (MSE) y Error Porcentual Absoluto Medio (MAPE).

Finalmente, se cita el trabajo de (Xiong, Li, & Bao, 2018) quienes proponen un modelo híbrido con los procedimientos de Descomposición de Tendencias Estacionales basados en Loess (STL) y Máquinas de Aprendizaje Extremo (ELM) para la previsión a corto, medio y largo plazo de los precios de las hortalizas estacionales.

**Figura 2** Resumen metodologías de pronósticos de precios de productos agropecuarios en el mercado spot

Author	Models	Type	Prediction horizon (Prediction strategy)	Whether deal with seasonal	Time scale
Dieng (2008)	Naive, exponential smoothing, and ARIMA	Tomato, potato, and onion	1	No	Monthly
Adanacioglu and Yercan (2012)	SARIMA	Tomato	4 (Iterative Strategy)	SARIMA	Monthly
Martin-Rodriguez (2013)	Restricted evolving spline model	Tomato	1	Spline functions	Daily
Nasira and Hemegeetha (2012)	BP neural network	Tomato	1	No	Weekly
Amiri et al. (2011)	Geostatistical models, neuro fuzzy approach, and Winter method	Potato and onion	1	No	Quarterly
Melin et al. (2007)	Modular neural networks	Tomato	1	No	Monthly
Hemegeetha and Nasira (2013)	BP neural network, RBF neural network	Tomato	1	No	Weekly
Colino et al. (2011)	VARs, Bayesian VARs	Hog	1, 2, and 3 (-)	Seasonal dummy variables	Quarterly
Davenport and Funk (2015)	Characteristic based clustering-based prediction model	Grain	1, 2, and 3 (Iterated strategy)	SARIMA	Monthly
Luo et al. (2011)	BP neural network, RBF neural network	Lentinus edodes	1	No	Monthly
Li et al. (2013)	Chaotic neural network	Egg	1,2,3,4, and 5 (-)	No	Weekly
Jumah and Kunst (2008)	Restricted rank model and unrestricted rank model	Barley and wheat.	1,2,..., 19, and 10 (-)	Seasonal dummy variables	Quarterly
Jha and Sinha (2014)	Time-delay neural networks (TDNN)	Oilseeds	1,3,6, and 12 (Iterative Strategy)	Seasonality-adjusted method	Monthly
Kömm and Küsters (2015)	Zero-inflated models, ARIMA(1,1,0)-GARCH(1,1) model	Milk-based Commodities	1,2,3,4, and 5 (-)	No	Weekly
Onour and Sergi (2011)	Thin tailed the normal distribution, fat-tailed Student t-distribution models	Wheat, rice, beef, groundnut, sugar, and coffee	1	No	Monthly
Ribeiro and Oliveira (2011)	Hybrid model based on neural networks and Kalman filter	Sugar.	1	No	Monthly
Ramirez and Fadiga (2003)	GARCH	Soybean, sorghum, and wheat	1	Centered multiplicative moving-average procedure	Quarterly
Shih et al. (2009)	Case-based reasoning approach	Broiler	1	No	Monthly
Ticlavilca et al. (2010)	Multivariate relevance vector machine	Cattle, hog, and corn	1, 2, and 3 (Multi-input Multi-output strategy)	No	Monthly
West and Wong (2014)	The factor model based on principal components extracted from commodity prices themselves	Corn, cotton, soybean meal, soybean oil, and wheat	1,3, and 12 (Direct strategy)	No	Monthly
Zhang et al. (2005)	Moving average, linear regression, and neural networks	Aquatic product	1	No	Annual, quarterly, monthly, weekly, and 3-daily
Zou et al. (2007)	Hybrid model based on ARIMA and neural networks	Wheat	1	No	Monthly

Nota: Tomado de (Xiong et al., 2018).

### 3.1.3.2 Redes neuronales

En la aplicación de las redes neuronales para la predicción de precios de productos del sector agropecuario, se tiene el trabajo de (Chuluunsaikhan et al., 2020), quienes desarrollaron la predicción de los precios del cerdo en Corea del Sur desde el año 2010 a 2019, utilizando modelos de la estadística clásica, machine learning y Deep learning. En éste se hizo énfasis en la estructura de las redes neuronales LSTM obteniendo como resultado la Figura 3.

Figura 3 Hiperparametrización modelo LSTM para los precios del cerdo.

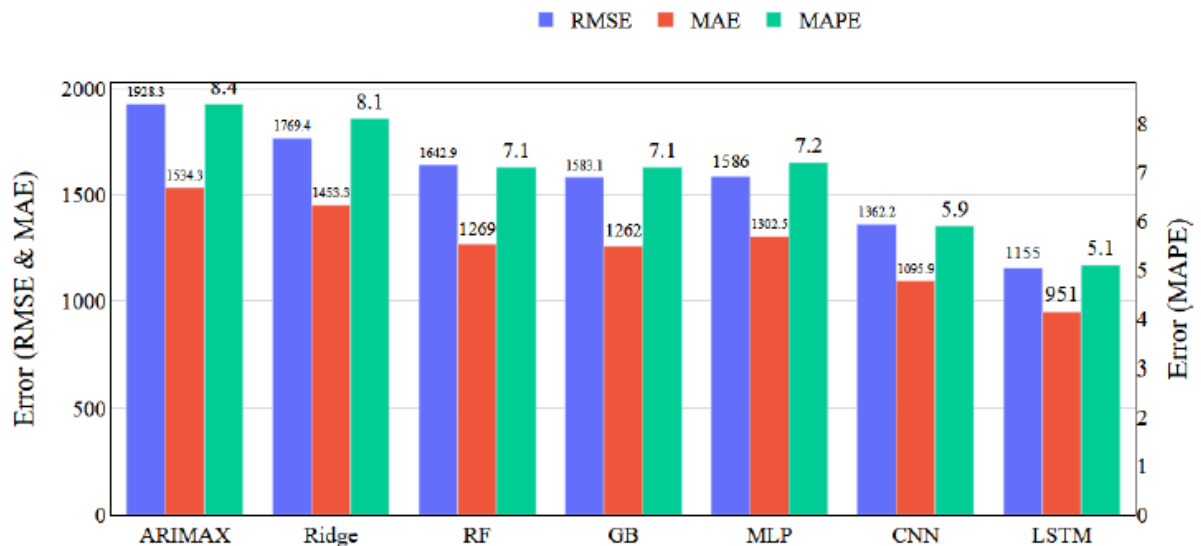
Case No.	Hyperparameters				Errors		
	Window	Batches	Epochs	Learning Rate	RMSE	MAE	MAPE
1	10	32	30	0.001	1152.886	917.175	5.105
2	20	32	30	0.001	1195.985	977.666	5.378
3	15	32	30	0.001	1057.874	822.069	4.564
4	15	32	15	0.001	1254.486	1031.557	5.716
5	15	32	45	0.001	1109.823	904.309	5.012
6	15	64	30	0.001	1033.131	811.057	4.498
7	15	128	30	0.001	1042.166	820.804	4.551
8	15	256	45	0.001	1200.07	977.025	5.37
9	15	32	30	0.002	1007.432	793.786	4.411

RMSE: Root Mean Squared Error; MAE: Mean Absolute Error; MAPE: Mean Absolute Percentage Error.

Nota: Tomado de (Chuluunsaikhan et al., 2020)

En el trabajo citado, se utilizaron los modelos estadísticos machine learning y Deep learning para realizar la predicción del precio del cerdo, y se obtuvieron los resultados que se observan en el Gráfico 1.

Gráfico 1 Resultado de todos los modelos de predicción.



Nota: Tomado de (Chuluunsaikhan et al., 2020)

En el gráfico anterior se observa que para los precios del cerdo en Corea del Sur, el mejor modelo de predicción fue el LSTM con un MAPE del 5,1%, un MAE de 951 y un RMSE de 1.115.

Por otro lado, se tiene el trabajo de (Grewal & Daneshyari, 2022), quienes desarrollaron un aplicativo web para la predicción de precios del sector agropecuario para los productores de la India. En el artículo utilizaron como ejemplo el arroz paddy basmati para aplicar el modelo LSTM, obteniendo como resultado la siguiente arquitectura de la red neuronal y el comparativo de la predicción vs precio real, representados en la Tabla 1 y Gráfico 2.

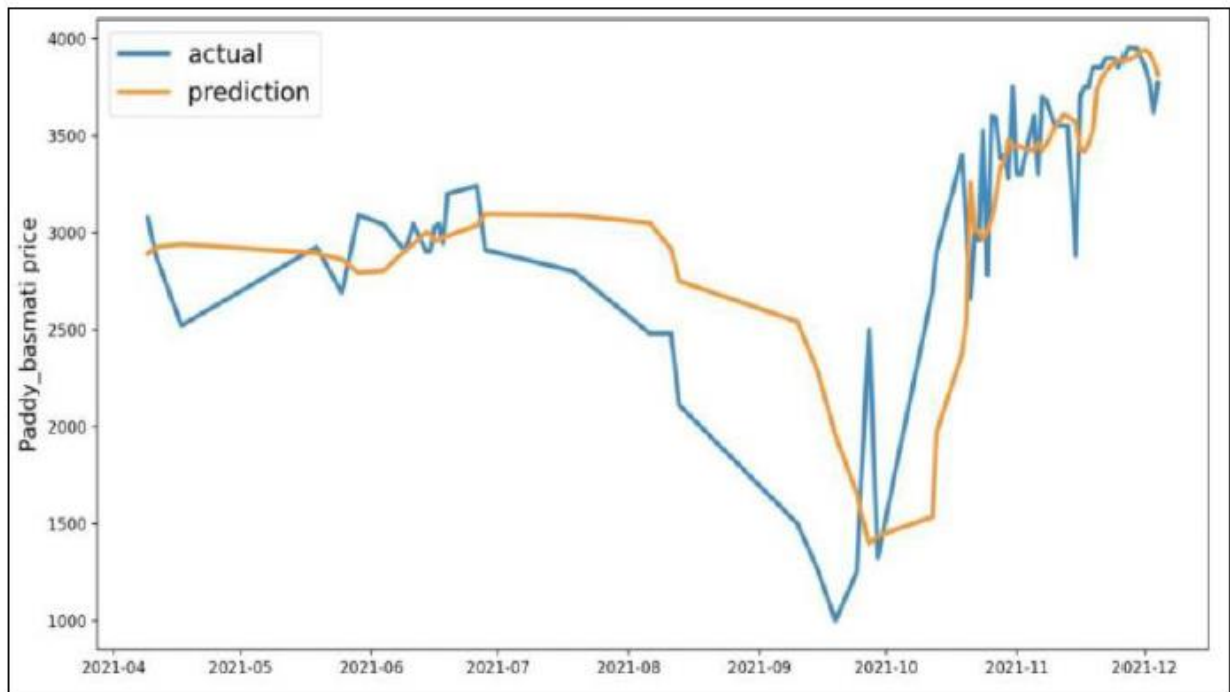
Tabla 1 Hiperparámetros tuneados para modelo LSTM para predicción de arroz paddy basmati

<b>Parameters</b>	<b>Inputs</b>
neurons	100
activation function	linear
dropout	20%
loss function	MSE
optimizer	adam
window length	5
training epoch	20
batch size	32

Nota: Tomado de (Grewal & Daneshyari, 2022)



Gráfico 2 Comparación predicción y precio real del arroz paddy basmati



Nota: Tomado de (Grewal & Daneshyari, 2022)

Por su parte, se tiene el modelo estudio de (Murugesan et al., 2021) donde:

“se emplean cinco versiones de técnicas de aprendizaje profundo LSTM para la tarea de predicción de precios de cinco productos básicos agrícolas en un conjunto de datos de series de tiempo univariadas de arroz, trigo, plátano y maní que abarca desde enero de 2000 hasta julio de 2020” (p. 1)

Al aplicar las variantes del modelo LSTM sobre estos sistemas productivos se obtuvieron los resultados representados en la Tabla 2, donde se observa los tuneos de los hiperparámetros y los resultados de la Tabla 3, donde se observan las medidas de desempeño de cada modelo utilizado.

Tabla 2 Hiperparámetros tuneados para cada modelo LSTM.

MODEL	PARAMETERS
<b>Basic LSTM</b>	50 output units, single dense hidden layer, 'relu' activation, 'adam' optimization, 'mse' loss
<b>Stacked LSTM</b>	50 output units, two dense hidden layer, 'relu' activation, 'adam' optimization, 'mse' loss
<b>Bi directional LSTM</b>	50 output units, single two-way dense hidden layer, 'relu' activation, 'adam' optimization, 'mse' loss
<b>CNN LSTM</b>	CNN: 64 filters, 2 kernels, pool_size = 0.5 LSTM: 50 output units, single dense hidden layer, 'relu' activation, 'adam' optimization, 'mse' loss
<b>Convolutional LSTM</b>	CNN: 64 filters, (1,2) kernels LSTM: single dense hidden layer, 'adam' optimization, 'mse' loss

Nota: Tomado de (Murugesan et al., 2021)

Tabla 3 Evaluación de las métricas de predicción para todos los modelos

Commodity	Model	MAE	MSE	MAPE	RMSE	R <sup>2</sup>
<b>Rice</b>	Basic LSTM	0.18999	0.36999	0.00121	0.192353	0.96873
	Bi LSTM	0.259999	0.103999	0.001668	0.322490	0.91213
	Stacked LSTM	0.269999	0.109000	0.001726	0.330151	0.90790
	CNN LSTM	0.289999	0.092999	0.001857	0.304959	0.92142
	Conv LSTM	0.20999	0.469999	0.001345	0.216794	0.96029
<b>Wheat</b>	Basic LSTM	0.130000	0.230000	0.000794	0.151657	0.96811
	Bi LSTM	0.150000	0.270000	0.000919	0.163431	0.96778
	Stacked LSTM	0.190000	0.450000	0.001163	0.212132	0.96631
	CNN LSTM	0.230000	0.065000	0.001405	0.254950	0.95467
	Conv LSTM	0.360000	0.238000	0.002187	0.487852	0.94050
<b>Gram</b>	Basic LSTM	0.300000	0.152000	0.002071	0.389871	0.96848
	Bi LSTM	0.150000	0.270000	0.001038	0.164316	0.97617
	Stacked LSTM	0.230000	0.650000	0.001599	0.254950	0.97799
	CNN LSTM	0.290000	0.103000	0.001996	0.320936	0.96542
	Conv LSTM	0.340000	0.136000	0.002350	0.368781	0.95074
<b>Banana</b>	Basic LSTM	0.200000	0.480000	0.001538	0.219089	0.96683
	Bi LSTM	0.179999	0.037999	0.001385	0.194935	0.96492
	Stacked LSTM	0.210000	0.073000	0.001626	0.270185	0.96518
	CNN LSTM	0.260000	0.074000	0.002002	0.272029	0.96511
	Conv LSTM	0.199999	0.047999	0.001531	0.219080	0.96683
<b>Groundnut</b>	Basic LSTM	0.220000	0.068000	0.001470	0.260768	0.99890
	Bi LSTM	0.319999	0.143999	0.002193	0.379473	0.99219
	Stacked LSTM	0.269999	0.170990	0.001773	0.413521	0.99219
	CNN LSTM	0.239999	0.0679999	0.001640	0.260768	0.998905
	Conv LSTM	0.240000	0.070000	0.0016338	0.264575	0.99887

Nota: Tomado de (Murugesan et al., 2021)

Finalmente, al igual que para los modelos ARIMA, en los modelos LSTM se encontró una tabla en el trabajo (Murugesan et al., 2021) donde se relacionan estudios de aplicación de los modelos LSTM en diferentes precios de productos agropecuarios, el cual se puede observar en la Tabla 4.

Tabla 4 Estudios relacionados con la modelación LSTM

Name of the authors	Name of the commodities	Deep Learning Models used for prediction	Results
R L et al. (2021)	Cottonseed, Castor seed, Rape mustard seed, Guar seed, soybean seed	LSTM Base line models: ARIMA, TDNN	The LSTM model provided a better forecast.
Ouyang et al. (2019)	Cotton, Sugar, bean, bean II, soya bean oil, cardamom, strong Wheat, Corn, Coffee, cocoa, Frozen orange juice	LSTNet Base Line Models: CNN, RNN, ARIMA, VAR	The LSTNet performed better results over the r baseline methods on average.
Kurumatani K. (2020)	Cabbage, Tomato, Lettuce	LSTM (Recurrent neural network)	The LSTM performed the best result.
Jin et al. (2019)	Chinese cabbage, Radishes	LSTM	The optimum performance was obtained by the LSTM.
Prakash & Farzana, (2019)	Tomato	LSTM	The LSTM is one of the most effective models for dealing with nonlinear patterns in prediction.
Chen et al. (2021)	Chicken, Chili, Tomatoes	LSTM Baseline models: ARIMA, SVR, Prophet, XGBoost	Among the five baseline models, the LSTM was forecasted to produce the best results.

Nota: Tomado de (Murugesan et al., 2021)

### 3.1.3.3 Modelos ARCH y GARCH

Como modelo inicial del estudio de la predicción de la volatilidad se tienen los modelos ARCH-GARCH, conocidos generalmente como procesos o modelos Auto regresivos de heterocedasticidad condicionada, los cuales según (Pérez, 2006), y (Casas & Cepeda, 2008) son usualmente usados en el análisis de volatilidad de la información, así como también su variante, los modelos GARCH

(Auto regresivos de heterocedasticidad condicionada generalizados); éstos fundamentan su estudio en la condición de la varianza de las innovaciones de la volatilidad, a partir de las siguientes consideraciones:

$$\sigma_t^2 = f(y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-k})$$

Dado que, para determinar una función de la información sobre las volatilidades es posible reescribir el proceso con una función lineal, se toma como ejemplo el primer valor de rezago:

$$\sigma_t^2 = f(y_{t-1}) = \alpha_0 + \alpha_1(y_{t-1} - \mu)^2$$

Siendo su forma general:

$$\sigma_t^2 = f(y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-k}) = \alpha_0 + \sum_{k=1}^q \alpha_1(y_{t-k} - \mu)^2$$

Donde  $\alpha_0$  y  $\alpha_1$  son valores constantes positivos mayores que cero, y con valores  $U_t \sim N(\mu, \sigma_t^2)$ .

Si,  $Y_t = \mu + \sigma_t U_t$  asumiendo que es un valor ruido blando, entonces:

$$y_t | y_{t-1}, y_{t-2}, y_{t-3}, \dots \sim N(\mu, \sigma_t^2)$$

$$Var(y_t | y_{t-1}) = \alpha_0 + \alpha_1(y_{t-1} - \mu)^2$$

(Silveira, Maciel, Mattos, & Ballini, 2017) hacen uso del modelo TARARCH para explicar la variabilidad de las series de precios de soja y maíz dado que “los rendimientos de los precios de futuros de las materias primas agrícolas generalmente no siguen una distribución normal, dada la presencia de una asimetría distinta de cero y una curtosis mayor que tres” (p. 410). Razón por la cual los modelos GARCH se adaptan mejor a estas condiciones.

Además, “teniendo en cuenta que los mercados reaccionan asimétricamente a las buenas y malas noticias, se elige un modelo de heterocedasticidad condicional autorregresiva de umbral (TARARCH) para capturar el proceso de volatilidad” (Silveira et al., 2017).

Finalmente, se tienen metodologías de estimación de volatilidad de precios similares a los mencionados anteriormente. Dentro de estas metodologías, se encuentra la de (Onour & Sergi, 2011) donde se estima la volatilidad de los precios del trigo, arroz, azúcar, carne vacuna, café y maní en el mercado global con los modelos GARCH. Asimismo, (Sekhar, 2004) calcula para el mercado indio la volatilidad de los precios del arroz, trigo, aceite de cacahuete, aceite de coco, algodón y azúcar, con los modelos ARCH-GARCH. Por último, se encuentra el trabajo

desarrollado por (A. Reztis & Stavropoulos, 2010) quienes hicieron uso de la familia de modelos ARCH para calcular la volatilidad del precio del ganado en el mercado griego.

## **3.2 Metodologías de medición del riesgo financiero**

Al igual que el riesgo de mercado, el riesgo financiero también se evalúa desde los contextos temporales histórico, monitoreo y predicción. No obstante, el riesgo financiero difiere de las metodologías presentadas en el riesgo de mercado en el contexto histórico. Para los contextos monitoreo y predicción las metodologías son iguales.

En el marco de la idea anteriormente mencionada, se presentan a continuación las metodologías para la evaluación del riesgo financiero desde el contexto de monitoreo.

### **3.2.1 Contexto histórico**

En el riesgo financiero, las metodologías desarrolladas para el contexto histórico tienen como propósito analizar el impacto que tienen sobre la rentabilidad, las variables críticas que afectan las finanzas de los sistemas productivos agropecuarios. El beneficio de estas metodologías está en que las entidades financieras y los productores agropecuarios, pueden cuantificar el impacto de estas variables críticas y tomar medidas para la mitigación del impacto de estos riesgos sobre la situación financiera de los sistemas productivos. En el caso de la banca, para poder determinar tasas y anillos de seguridad para el otorgamiento de créditos a los productores agropecuarios; mientras que, para los productores, analizar la posibilidad de adquirir algún tipo de seguro agropecuario.

#### **3.2.1.1 Simulación Montecarlo**

Con respecto a las metodologías para la medición del riesgo financiero, se cuenta con la metodología de Simulación Montecarlo, y como referente de la aplicación de Simulación Montecarlo en la medición de riesgo financiero, se tiene en cuenta el trabajo de (López, 2019), quien aplicó esta metodología a través de @Risk para analizar el efecto que tiene el precio de la papa, su rendimiento, los costos de fertilizantes, semillas, pesticidas, mano de obra y maquinaria

sobre el margen bruto de la actividad agropecuaria para zonas de Perú y Lima, obteniendo como resultado una matriz de percentiles de riesgo que expresan la probabilidad de que el margen bruto se encuentre en un valor específico.

(Miguez, 2014), presenta la herramienta llamada Sistema de Evaluación de Riesgo Agropecuario (SERA) que usa la simulación Montecarlo para medir el efecto de las variables de ingresos y costos, sobre el margen neto de cultivos de soja, girasol, maíz y sorgo, con lo cual consigue un gráfico de histograma del comportamiento del margen bruto, luego de la simulación, y, un semáforo que alerta sobre las probabilidades de comportamiento de dicho margen.

En esta misma dirección del uso de la simulación Montecarlo para crear modelos de gestión de riesgo, se encuentra el trabajo de (Berger, Pena, & Melón, 2015) quienes modelaron a través de simulación Montecarlo, el efecto que tendría la oferta forrajera, el precio de los animales, la producción y los costos de producción, sobre el Valor Presente Neto de un proyecto de ganadería de cría y levante. A diferencia de los trabajos presentados anteriormente, éste muestra cómo las variables de clima afectan la oferta forrajera, la cual será una variable de entrada en la simulación.

### **3.2.1.2 Análisis de sensibilidad a través de modelos de regresión**

De manera adicional, respecto a los análisis multivariados para los riesgos de mercado y financiero, se encuentra el trabajo de (Herrera et al., 2018) quienes usaron el Modelo de Regresión Lineal y el Tobit para explicar los ingresos, la productividad y diversificación en función de las variables: propiedad de la tierra, género, edad, área, cooperativa, asistencia rural, regiones, renta social, beneficios, ingresos fuera de la finca, fuerza laboral contratada y grado de escolarización; encontrando que:

...los pequeños agricultores que forman parte de una cooperativa agrícola o miembros de una asociación afectan positivamente los ingresos, la productividad y la diversificación. Se muestra que la edad de los jefes de hogar tiene una relación no lineal en los tres casos, mientras que el jefe de hogar siendo mujer presentó un efecto negativo en todas las regresiones. Aunque la asistencia técnica reciente mostró impactos negativos en los ingresos y productividad, la asistencia del agricultor afectó positivamente la probabilidad

de que un pequeño agricultor diversificara su producción y, por lo tanto, estar menos sujeto a los desequilibrios de precios (p. 455)

Por otro lado, se tiene el trabajo de (Rehman, Jingdong, Chandio, & Hussain, 2017) quienes explicaron la relación del PIB agropecuario y la producción pecuaria en Pakistán, a través del método de mínimos cuadrados (MCO) y la prueba de Dickey-Fuller aumentada (ADF), cuyos resultados se interpretaron utilizando la prueba de cointegración de Johansen. Los resultados obtenidos fueron los siguientes:

...la leche, grasa, huevos, huesos y cordero tienen una relación positiva y significativa con el PIB agropecuario, mientras que la producción de carne de vacuno, carne de ave, lana, pelo, pieles y cueros tiene una relación negativa e insignificante con el PIB agrícola de Pakistán. Por lo tanto, el estudio sugiere que el gobierno de Pakistán inicia nuevos esquemas de financiamiento para el desarrollo del sector ganadero. (p. 168)

(Tadesse, Algeri, Kalkuhl, & von Braun, 2014) explican el comportamiento de los picos de los precios de maíz, trigo y soya según los choques de producción, precios del petróleo, stock de inventarios, tasa de crecimiento, especulación de los contratos de futuros y el índice de crisis financiera, esto, por medio del modelo de mínimos cuadrados ordinarios, mostrando con los resultados que:

...los picos de precios se correlacionan negativamente con choques de oferta (anticipado) y se correlacionan positivamente con el crecimiento económico (demanda). Los resultados muestran un efecto positivo y significativo de actividades especulativas excesivas sobre los picos de los precios de los alimentos, aunque se controla la anticipación de los shocks de oferta y demanda (...) Los resultados sugieren además el papel de las fluctuaciones de la producción anticipada como una causa importante del pico de los precios de los alimentos a corto plazo. (p.123)

Posteriormente, (Kadigi et al., 2020) realizaron el pronóstico de los rendimientos y precios para los productos maíz, sorgo y arroz, mediante la simulación Montecarlo multivariada, donde tuvieron

en cuenta las variables de superficie cultivada, costo medio, alto y bajo; tasa de inflación de los costos de producción y la tasa de descuento para VPN. Los resultados determinan que:

... el rendimiento de los cultivos muestra una tendencia creciente para los tres cultivos en la región de Dodoma. Asimismo, se espera que el rendimiento del arroz aumente marginalmente en Morogoro con una tendencia a la baja para el maíz y el sorgo; mientras tanto, para los precios de los tres cultivos se proyecta que aumente para las dos regiones. Generalmente, los resultados sobre la viabilidad económica en términos de VPN revelaron una alta probabilidad de éxito para todos los cultivos en Dodoma a pesar de un mayor riesgo relativo para el arroz. Los resultados en Morogoro presentaron una alta probabilidad de éxito para el arroz y el sorgo, siendo el maíz el que indica un mayor riesgo y una probabilidad del 2,41% de VPN negativo. (p.2)

Ahora bien, (Barten & Vanlout, 1996) aplican el teorema de Cobweb para vincular la oferta que reacciona al precio rezagado con la demanda que reacciona al precio actual. Para el modelado de la oferta se aplicó el modelo Acreage Allotment y para la demanda, el modelo de Demanda Inversa. Los resultados de cada uno de los modelos utilizados para la oferta y la demanda, se combinan con un modelo de simulación de dinámica endógena, y como consecuencia, se obtiene una dinámica bastante amortiguada y un autovalor dominante negativo.

Por su parte, (Eissa & Al Refai, 2019) estudian la dinámica de los precios del petróleo con el maíz, cebada y aceite de colza a través de los modelos lineales y modelos de retardo distribuido autorregresivo y no lineal (ARDL). Los resultados del trabajo muestran que

Al aplicar el modelo lineal, la cebada, el maíz y el aceite de colza no se mueven a largo plazo con los precios del petróleo. Cuando usamos el modelo ARDL no lineal para superar el problema de simetría y volver a estimar la relación dinámica, nuestros resultados muestran que la cebada, el maíz y el aceite de colza se mueven junto con los precios del petróleo a largo plazo (...) Los efectos de los multiplicadores dinámicos muestran que los precios de la cebada, el maíz y el aceite de colza responden rápida y fuertemente a las caídas cíclicas de los precios del petróleo en el corto plazo, pero se ajustan completamente al nuevo equilibrio cuando el proceso es relativamente prolongado (p. 1)



(A. N. Rezitis & Tsionas, 2019) hacen uso del modelo de corrección de error de panel multivariante (PVECM) “para investigar la transmisión de los precios asimétricos entre los segmentos de granjas, procesadores y minoristas de la cadena europea de suministro de alimentos para el período 2005-2016 en los mercados europeos” (p.216) teniendo como resultado que, en el corto y largo plazo, los precios minoristas responden con mayor fuerza a los aumentos de los precios de procesadores y distribuidores.

(Siddig & Grethe, 2014) combinan los modelos de análisis métrico de transmisión de precios (PTA) y el equilibrio general computable (CGE) con el objetivo de determinar el grado de transmisión de los precios internacionales sobre los mercados nacionales y encuentran los siguientes aspectos:

...(1) La transmisión del precio del mercado mundial al mercado interno será mayor cuanto más sea la participación comercial del sector y cuanto mayor incidencia tenga sobre la economía en general; (2) el aumento de los precios de exportación junto con elasticidades CET bajas y elasticidades de producción altas pueden incluso resultar en efectos negativos sobre los precios al consumidor interno; (3) cuanto mayor el factor movilidad, menor es la transmisión de precios; (4) cuanto mayor las elasticidades comerciales, mayor es la transmisión de precios; (5) cuanto mayor son las elasticidades de producción (nivel superior), mayor es la sustitución de opciones que tienen los productores y menor es la transmisión de precios; y (6) el traspaso de un precio mundial en aumento sería mayor bajo un régimen de tipo de cambio fijo en comparación con uno flexible, si la moneda nacional se aprecia debido al impacto de los precios del mercado mundial.

Finalmente, (Bayramoglu & Chakir, 2016) utilizaron “un modelo econométrico estructural para medir los efectos de los precios de los cultivos en la demanda de agroquímicos, insumos y asignación de tierras” y estimaron una función de ganancias de múltiples resultados utilizando datos de panel a nivel de granja. Los resultados demuestran que:

... un aumento en el precio de la colza, que es la principal materia prima para la producción de biodiésel en Francia, tiene un efecto positivo y significativo sobre la demanda de insumos agroquímicos (...) estos resultados sugieren que los cambios en el precio de la

colza, impulsados en parte por las políticas de biocombustibles, inducen cambios en la demanda de productos químicos y la asignación de tierras que pueden tener efectos potencialmente adversos en el medio ambiente.

## **4 Metodología**

El desarrollo de la metodología de evaluación de riesgo de mercado y financiero se fundamenta en la descripción y comprobación de las metodologías aplicadas en el marco nacional e internacional, por lo cual el presente trabajo se enmarca en un estudio descriptivo, y de acuerdo con (Méndez, 1988), el método de investigación usado es de tipo analítico cuantitativo, el cual inicia con la identificación de cada una de las metodologías obtenidas en la revisión de experiencias nacionales e internacionales, con el objetivo de establecer los casos de éxito y aplicarlos en la Plataforma Agro. Las fuentes de información son secundarias: artículos, documentos y textos de apoyo. Con respecto al tratamiento de la información, se aplicaron modelos de series de tiempo para completar series con datos faltantes y corrección de datos atípicos.

### **4.1 Metodología para el diagnóstico**

Para el desarrollo del diagnóstico, se realizaron sesiones con los equipos temáticos de la Plataforma Agro y con cada uno de ellos se enfocaron las sesiones de la siguiente forma:

- Equipo de expertos en sistemas agrícolas y pecuarios:

Las sesiones con este primer panel de expertos, se orientaron a identificar los parámetros que utiliza el módulo técnico de la Plataforma Agro para la proyección de producción y rendimientos en los sistemas agrícolas, y los parámetros de proyección de animales y producción de leche para los sistemas pecuarios. Adicionalmente, se validaron las fuentes de información utilizadas para la estimación de los parámetros y las metodologías utilizadas para los cálculos de estas proyecciones. Por su parte, también se analizaron en estas sesiones los porcentajes de desarrollo temático en el módulo técnico y los pasos a seguir.

- Equipo de expertos financieros:

Con este segundo panel de expertos, se validó todo el proceso para la evaluación financiera de los sistemas productivos agrícolas y pecuarios. Esta validación inició desde la forma en cómo se construyen los ingresos y costos de producción para estos sistemas, la proyección de inversiones en capital de trabajo y activos fijos operativos, la proyección de la deuda financiera, la construcción

de los estados financieros y el flujo de caja libre, los indicadores para la evaluación financiera de los sistemas productivos y los indicadores para el seguimiento del desempeño de éstos.

Adicional a la validación de todo el proceso de evaluación financiera, se validó la viabilidad de la aplicación de las metodologías propuestas para la evaluación de riesgo de mercado y financiero, y qué aspectos se debían tener en cuenta con base en la información y la estructura del módulo financiero de la Plataforma.

- Equipo de riesgos agropecuarios:

En el equipo de riesgos agropecuarios, las sesiones se centraron en revisar los avances asociados a la medición del riesgo de mercado y el riesgo financiero, el porcentaje de avance y cómo las metodologías propuestas pueden mejorar la evaluación de los riesgos para los sistemas agrícolas y pecuarios.

Con base en estas sesiones, se realizó un primer informe donde se describe el estado actual de desarrollo de la plataforma desde el punto de vista temático en los módulos: técnico, financiero y evaluación de riesgos, identificando el porcentaje de desarrollo, planes de acción, pasos a seguir y desarrollos futuros. Esto permitió a la presente consultoría, direccionar la búsqueda de fuentes de información y definir los desarrollos metodológicos que se adapten al contexto de la agricultura colombiana.

Posterior a estas sesiones, se realizaron otras con el equipo de sistema de información de la empresa con el fin de identificar el estado de desarrollo de la plataforma en términos de programación, cronograma de actividades por parte del equipo, desarrollos futuros, y un aspecto importante que es validar la viabilidad técnica de implementar en el módulo de riesgos las metodologías propuestas en la presente consultoría. De estas sesiones, se realizó un segundo informe donde se documenta la opinión del equipo respecto a las metodologías propuestas, su viabilidad respecto a la programación e implementación en el corto o mediano plazo en la Plataforma Agro.

Con base en las sesiones realizadas con los equipos temáticos y de sistemas de información de la empresa, se realizó un informe de diagnóstico donde se relacionan los siguientes aspectos:

- Estado actual de desarrollo de la Plataforma Agro desde el punto de vista temático y de programación.
- Desarrollos futuros de la Plataforma Agro.
- Opiniones y recomendaciones por parte del equipo temático sobre las metodologías propuestas para la evaluación del riesgo de mercado y del riesgo financiero en los sistemas agrícolas y pecuarios.
- Consideraciones del equipo de desarrollo sobre la implementación en el mediano plazo de las metodologías propuestas para la evaluación de los riesgos de mercado y financiero, en el módulo de riesgos.

#### **4.2 Metodología para la Intervención y evaluación**

A partir de los resultados del diagnóstico, se procedió con el desarrollo del plan de intervención, el cual consta de las siguientes actividades:

- a. Se inició con la revisión de la literatura sobre experiencias nacionales e internacionales en la modelación del riesgo de mercado y del riesgo financiero en el sector agropecuario, con el fin de identificar las metodologías que se han aplicado en el sector agropecuario y su impacto en la medición de estos riesgos, así como aquellas experiencias exitosas por parte de entidades gubernamentales y privadas que hayan aplicado estas metodologías de medición. Para ambos casos se identificaron los aspectos que se deben tener en cuenta a la hora de la aplicación.
- b. Se desglosó cada metodología extraída de la revisión, identificando el proceso de aplicación, la información y supuestos que se deben tener en cuenta en el momento de su implementación; esto permite conocer el grado de complejidad en la administración de las fuentes de información para desarrollar el proceso metodológico.
- c. Se evaluaron y seleccionaron las mejores metodologías, para lo cual se tuvo en cuenta el nivel de adaptación al contexto de la agricultura colombiana y las fuentes de información con las que cuenta la Plataforma Agro.
- d. Se diseñaron los algoritmos que reflejan el proceso de las metodologías seleccionadas y su forma de aplicación; dichos algoritmos se desarrollaron en los ambientes de R y Python.

- e. Se presentaron las metodologías a los equipos: financiero, riesgos agropecuarios y sistemas de información, para la validación y retroalimentación dado el caso. para luego entregar estos algoritmos al área de sistemas de la empresa SFA CEBAR SAS, indicándoles cómo funcionan para su correcta implementación en la plataforma Agro.

Sobre la base del desarrollo del marco teórico presentado, la selección de varias metodologías se debe a la forma en cómo se va a evaluar el riesgo según los contextos temporales. Y para cada uno de estos contextos, la idea es tener más de una metodología para poder contrastar y comparar los resultados, con el fin de argumentar mejor las decisiones de las entidades financieras y los productores agropecuarios.

## 5 Diagnóstico

Partiendo de la metodología propuesta para la realización del diagnóstico, se presenta el informe del mismo donde se abordan los siguientes aspectos: estado actual de desarrollo de la Plataforma Agro desde el punto de vista temático y de programación, desarrollos futuros de la Plataforma Agro, opiniones y recomendaciones por parte del equipo temático sobre las metodologías propuestas para la evaluación del riesgo de mercado y del riesgo financiero, en los sistemas agrícolas y pecuarios y consideraciones del equipo de desarrollo sobre la implementación en el mediano plazo de las metodologías propuestas para la evaluación de los riesgos de mercado y financiero, en el módulo de riesgos.

### 5.1 Estado actual de la Plataforma Agro

Como punto inicial del diagnóstico, se tiene la descripción del estado actual de la Plataforma Agro desde el frente temático y tecnológico. A continuación, se presentan los resultados obtenidos en cada uno de estos frentes.

#### 5.1.1 *Componente temático*

Como primer resultado del diagnóstico realizado, se tiene el estado de avance que se explica a continuación:

- Equipo de expertos en sistemas agrícolas y pecuarios: con este primer equipo se evidenció que ya se encuentra en la Plataforma la información técnica de 40 sistemas agrícolas y un sistema pecuario. Esta información consta de las matrices de coeficientes técnicos y los parámetros productivos, reproductivos, sanitarios y nutricionales, los cuales, se obtuvieron a partir de la información histórica y la propia experiencia del equipo en campo.
- Equipo de evaluación financiera: con este segundo equipo, se evidenció que el módulo financiero se encuentra en un 90% de desarrollo. Este 90% lo conforma el desarrollo del CAPEX y del capital de trabajo neto operativo, la proyección de financiación, estados financieros (estado de situación financiera, estado de resultados y estado de flujo de

efectivo), flujo de caja libre e indicadores de evaluación financiera, actividad, rentabilidad, liquidez e inductores de valor. Con respecto al 10% faltante, está en desarrollo la determinación del cupo del crédito con base en el flujo de caja libre.

- Equipo de evaluación de riesgos: Finalmente, en las sesiones realizadas con el equipo de evaluación de riesgos agropecuarios, se evidenció que existen avances asociados a la medición de riesgo agroclimático, sanitario, de mercado y financiero. No obstante, no se observó una metodología definida para la medición y evaluación de éstos, en especial los asociados a riesgo de mercado y financiero, lo cual, deja en evidencia la importancia del desarrollo de esta consultoría.

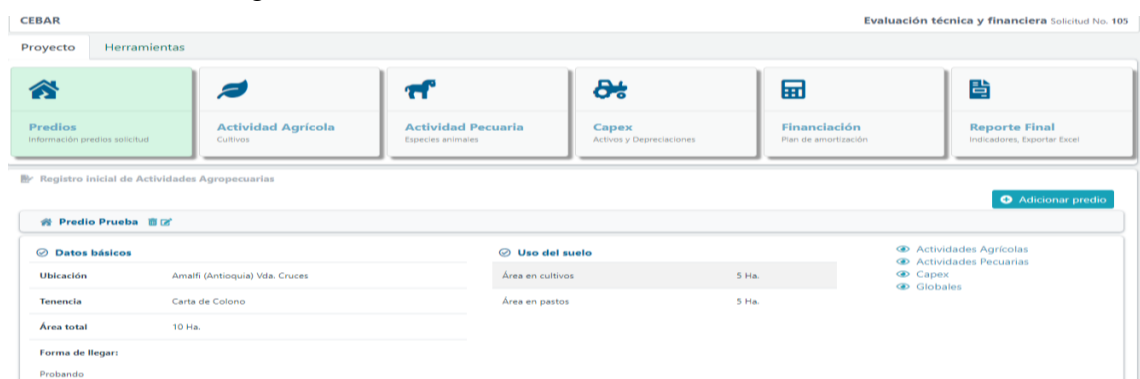
### **5.1.2 Componente tecnológico**

Con respecto al componente tecnológico de la Plataforma Agro, tal y como se observa en la Figura 4, cuenta con 6 módulos los cuales tienen como propósito:

- Predios: Módulo donde se ingresa la información asociada con la información de distribución de las áreas e inventarios.
- Actividad Agrícola: Módulo donde se ingresa la información asociada con los rendimientos y matrices de coeficientes técnicos de las actividades productivas.
- Actividad Pecuaria; Módulo donde se ingresa la información asociada a los parámetros técnicos productivos y matrices de coeficientes técnicos de las actividades pecuarias
- Capex: Módulo donde se ingresa la información de los activos fijos actuales del negocio agropecuario y las nuevas inversiones que se realizarán sobre éstos.
- Financiación: Módulo donde se ingresan los datos asociados al capital de trabajo y la proyección de la deuda tanto para las obligaciones actuales como para la solicitud que se está realizando.
- Reportes: Módulo donde se muestran los resultados del negocio agropecuario, donde se proyectan los estados financieros, el flujo de caja, indicadores de viabilidad financiera, indicadores financieros e informe de análisis de resultados.



**Figura 4** Plataforma Agro SFA CEBAR



Nota: Tomado de SFA CEBAR

Las sesiones realizadas con el equipo de sistemas, dejaron en evidencia que, en términos de desarrollo, la Plataforma Agro ya es funcional en un 100% para las 45 actividades productivas agropecuarias.

No obstante, faltan otros sistemas productivos que se describirán en los desarrollos futuros. En este sentido, el estado actual de la Plataforma Agro incluyendo lo mencionado, se encuentra en un 80% de desarrollo.

## 5.2 Desarrollos futuros para la Plataforma Agro

Con respecto a los desarrollos futuros para la Plataforma Agro, se concluyó que hace falta:

Desde lo temático:

- Recopilar y organizar la información asociada a matrices de coeficientes técnicos y parámetros para los sistemas pecuarios: cerdos, aves y piscicultura.
- Determinación del cupo de crédito con base en el flujo de caja libre.
- Terminar la metodología de medición de riesgo agroclimático.
- Implementación de los indicadores y metodologías propuestos para la medición del riesgo de mercado y del riesgo financiero obtenidos en la presente consultoría.

Desde lo tecnológico:

- Implementar los sistemas pecuarios que va a desarrollar el equipo de temáticos en la Plataforma Agro.
- Implementar la metodología para determinar el cupo de crédito.
- Preparar el ambiente de desarrollo para la puesta en marcha de las metodologías propuestas para la medición de riesgo del mercado y del riesgo financiero.

### **5.3 Recomendaciones del equipo temático de riesgos ante la propuesta metodológica de la consultoría**

En las sesiones que se llevaron a cabo con los equipos temáticos de evaluación financiera y de riesgos que participan en el desarrollo de la Plataforma Agro se enunciaron varias recomendaciones en el momento de iniciar con la búsqueda de metodologías y al momento de presentar las metodologías seleccionadas. A continuación listo las recomendaciones recibidas:

- En la medida de lo posible, clasificar las metodologías halladas en los contextos de histórico, monitoreo y predicción, con el fin de tener la trazabilidad completa para la medición de los riesgos de mercado y financiero.
- Tener presente qué tipo de restricciones pueden llegar a presentar las metodologías halladas al momento de la aplicación. Así mismo, validar la viabilidad de éstas cuando se lleven a algoritmos para la programación.
- Finalmente, validar si las metodologías seleccionadas al final del ejercicio son aplicables para cualquier sistema productivo.

### **5.4 Viabilidad del equipo de sistemas para la implementación de la propuesta de la consultoría**

Finalmente, en las sesiones realizadas con el equipo de sistemas de información de la empresa, al cual se le está realizando la consultoría, se obtuvo el visto bueno para la futura implementación siempre y cuando los algoritmos estuviesen desarrollados en los ambientes de R o Python. Es de

resaltar que en este punto y con base en los desarrollos ya programados por el equipo, las metodologías seleccionadas no entrarían en la fase de implementación. Es decir que el alcance de la consultoría y cumpliendo con el objetivo general, solo llega hasta la entrega de algoritmos en los lenguajes sugeridos.

## 6 Plan de acción

En primer lugar, se tiene la selección de las metodologías para la medición del riesgo de mercado y del riesgo financiero, las cuales se detallan en la Tabla 5.

**Tabla 5** Metodologías seleccionadas para la medición de riesgo de Mercado y Financiero

Componente del riesgo	Definición	Metodologías		
		Histórico	Monitoreo	Pronóstico
Amenaza	Caída de los precios de los productos asociados a su volatilidad.	Indicador de volatilidad. VaR de las fluctuaciones de los precios. Índices Beta.	- Bandas de Bollinger.	Metodología de Box-Jenkins. - Redes Neuronales.
Vulnerabilidad	La sensibilidad de la rentabilidad y liquidez de los sistemas agropecuarios a la volatilidad de los precios.	- Modelo de regresión. - Simulación Montecarlo.	- Medias Móviles. - MACD. - RSI.	

Nota: Elaboración propia a partir de la revisión de literatura.

### 6.1 Desarrollo de los algoritmos de las metodologías seleccionadas

En segundo lugar, se tiene el desarrollo de los algoritmos para el cálculo de las metodologías seleccionadas en la Tabla 5, las cuales se crearon en el ambiente R en combinación con hojas de cálculo de Excel para el modelo financiero, y, Python para el proceso de los algoritmos de Machine learning.

#### 6.1.1 Riesgo de Mercado

Los algoritmos asociados a riesgo de mercado, presentan la forma como se deben calcular los indicadores seleccionados y las consideraciones que se deben de tener en cuenta en el momento de su desarrollo.

### **6.1.1.1 Cálculos de amenaza**

A continuación, se mostrarán los algoritmos desarrollados para el cálculo de las metodologías seleccionadas para la medición del riesgo de mercado en el contexto histórico.

#### ***6.1.1.1.1 Histórico***

##### **6.1.1.1.1.1 Indicador de Volatilidad**

Inicialmente, se debe crear una variable que contenga la serie de precios a los cuales se les calculará el indicador de volatilidad. Para ello, el código a aplicar será el siguiente:

```
Precio <- df_precios[,n]
```

*Donde:*

*df\_precios* = *Es la base de datos de los precios de todos los productos agropecuarios.*

*n* = *Columna donde se encuentra el precio del producto seleccionado.*

Luego, se establece una segunda variable donde se calcula el rendimiento o variación del precio en términos discretos, cuya fórmula está determinada por:

$$\text{Rendimiento Discreto} = \left( \frac{P_t}{P_{t-1}} \right) - 1$$

En términos del código, la variable creada debe seguir la siguiente sintaxis:

```
Rendimiento_Precio <- diff(Precio)/Precio[-nrow(Precio)]
```

Finalmente, en la variable creada se encuentran los rendimientos discretos del precio seleccionado, a los cuales se les deberá aplicar la fórmula que se muestra a continuación:

$$\sigma = \sqrt{\frac{\sum_{t=1}^n (r_i - m)^2}{n - 1}}$$

En términos del código, la sintaxis a utilizar es la siguiente:

```
Indicador_Volatilidad <- sd(Rendimiento_Precio)  
print(Indicador_Volatilidad)
```

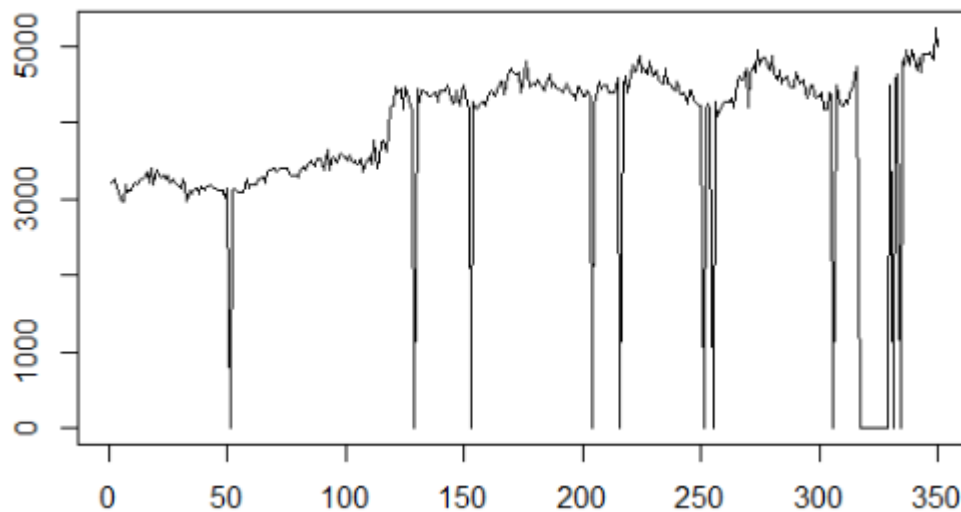
En ese orden de ideas, con la sintaxis mencionada se puede obtener el resultado del indicador de volatilidad sobre las variaciones de los precios, los cuales se podrán comparar con los precios de los demás productos agropecuarios aplicando el mismo algoritmo.

### **Consideraciones:**

- Consideración 1: Previo a la aplicación de esta metodología, es necesario validar la existencia de missing data dentro de las series de precios. En caso de existir, se recomienda utilizar el siguiente algoritmo para imputar los datos faltantes:

En el Gráfico 3 se observa una serie de tiempo con datos faltantes en diferentes momentos del tiempo.

**Gráfico 3** Serie de tiempo con datos faltantes



Nota: Elaboración propia a partir de los cálculos realizados en el software R.

En primer lugar, se desarrolla una matriz que contenga en cada columna los datos faltantes, y, sobre éstas se crean las variables dummy para ubicar en qué momento del tiempo no se reportaron datos. En términos de la sintaxis en el lenguaje R, el código a aplicar es el siguiente:

```
Dummy_1 <- matrix(0,nrow = nrow(Precio), ncol = sum(Precio==0))
```

```
k <- 1
```

```
for (j in 1:nrow(Precio)){
```

```
  if(Precio [j] == 0){
```

```
    Dummy_1[j,k] <- 1
```

```
    k <- k+1
```

```
  }
```

```
}
```

Luego de tener la matriz con las variables dummy, se procede con la estimación de un modelo ARMA(p,q), donde se tendrán en cuenta los datos faltantes como datos atípicos

de tipo aditivo (AO). En vista de que se desconoce el modelo ARMA(p,q) de la serie de precios, se aproxima usando un modelo ARI(p,1). Sobre esta base, el código quedaría de la siguiente forma:

```
p <- round(nrow(Precio)^(1/3))
mod_falt_CP <- Arima(Precio, order=c(p,1,0), include.drift=F, method=c("CSS-
ML"), xreg=Dummy_1)
summary(mod_falt_CP)
tsdiag(mod_falt_CP)
```

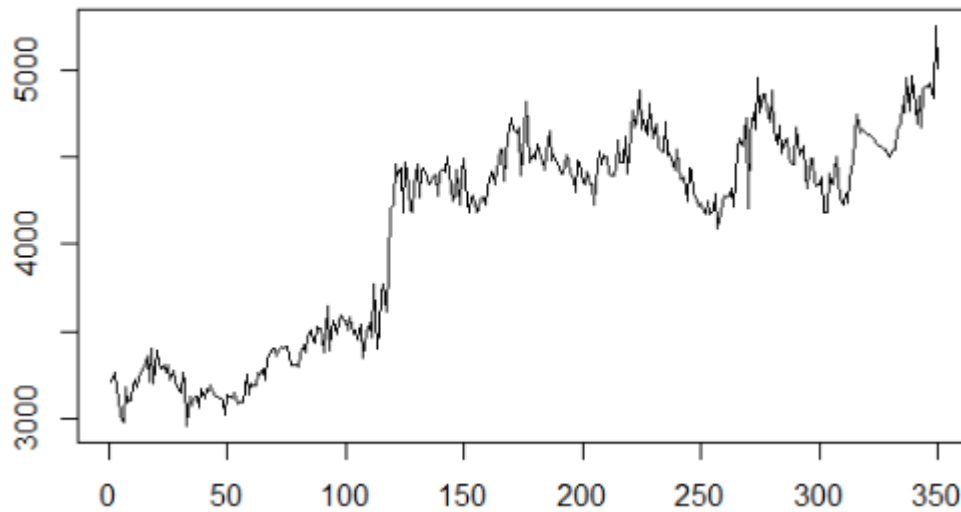
Posteriormente, se inicia con la imputación de los datos. Para este proceso, el código que se deberá aplicar es el siguiente:

```
k <- 1
for (i in 1:nrow(Precio)){
  if(Precio == 0){
    Precio[i] <- Precio[i] - coef(mod_falt_CP)[p+k]
    k <- k+1
  }
}
```

Con la aplicación de este último código, el resultado de imputación para los datos faltantes que fueron expuestos en el Gráfico 3, se presenta en el Gráfico 4.



**Gráfico 4 Gráfico de serie de precios con datos imputados**



Nota: Elaboración propia a partir de los cálculos realizados en el software R.

- Consideración 2: En caso tal de que en el ambiente R se tengan abiertas las librerías TSA, FitAR o alguna que se relacione con series de tiempo, es necesario cambiar el código de la estimación del modelo para imputación de datos

```
mod_falt_CP <- Arima(Precio, order=c(p,l,0), include.drift=F, method=c("CSS-  
ML"), xreg=Dummy_1)
```

por

```
mod_falt_CP <- forecast::Arima(Precio, order=c(p,l,0), include.drift=F,  
method=c("CSS-ML"), xreg=Dummy_1)
```

#### **6.1.1.1.2 VaR de las fluctuaciones de los precios**

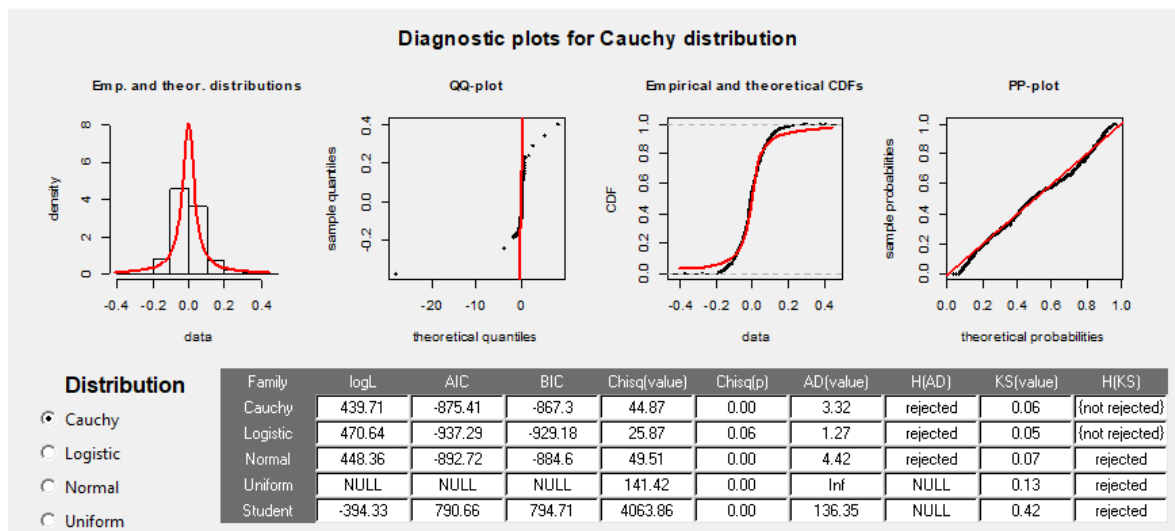
Este indicador se interpreta como la probabilidad de que las variaciones de los precios oscilen en un rango y en una periodicidad determinada. En términos del algoritmo, primero se deben calcular los rendimientos de la serie de precios tal y como se describe en el indicador anterior.

Seguidamente, es necesario definir la mejor función de distribución de los rendimientos de los precios y calcular los parámetros de la distribución seleccionada. Para lo anterior, se crea una variable y almacena en la función `fit.cont()`, la cual realizará las pruebas de bondad de ajuste de Kolmogorov-Smirnov, Andersen y Chi2, sobre diferentes distribuciones continuas, y con ello determinará cuál es la mejor según el criterio AIC. En términos del código, la sintaxis a aplicar es la que se describe a continuación:

```
Ajuste <- fit.cont(Rendimiento_Precio)
```

Al aplicar el código, el resultado obtenido es el que se observa en la Figura 5.

**Figura 5** Pruebas de Bondad de Ajuste



Nota: Elaboración propia

Por último, se calcula el Valor en Riesgo (VaR) con base en la función de distribución recomendada por la función `fit.cont`. Para esto, el código a emplear es el siguiente:

```
Intervalos<-
```

```
matrix(c(qcauchy(0.025,Ajuste$fittedParams[1],Ajuste$fittedParams[2]),qcauchy(0.975,Ajuste$fittedParams[1],Ajuste$fittedParams[2])),ncol = 2)
```

```
colnames(Intervalos) <- c("Intervalo 2,5%", "Intervalo 97,5%")
```

```
rownames(Intervalos) <- "95% de confianza"
```

```
print(Intervalos)
```

De esta manera, se obtiene el indicador VaR sobre las fluctuaciones del precio seleccionado, cuyo resultado muestra la probabilidad de que el precio esté variando sobre los intervalos calculados.

### **Consideraciones:**

- Consideración 1: Para aplicar la función `fit.cont()`, deberá importar y cargar la librería que se menciona a continuación:

```
if(!require("rriskDistributions")) install.packages("rriskDistributions")  
library(rriskDistributions)
```

- Consideración 2: Según la distribución seleccionada por la función bajo el criterio AIC, deberá ajustar la línea de código donde se calcula la distribución. Es decir, bajo el código que se observa, si la distribución ajustada es un Cauchy su función será *qcauchy*; en caso de ser una normal, la función deberá ser *qnorm* y de ser una beta, la función deberá ser *qbeta* y así sucesivamente, de acuerdo con la distribución que corresponda.

```
Intervalos <-  
matrix(c(qcauchy(0.025,Ajuste$fittedParams[1],Ajuste$fittedParams[2]),qcauchy(0.975,  
Ajuste$fittedParams[1],Ajuste$fittedParams[2])),ncol = 2)
```

#### **6.1.1.1.1.3 Índices Beta**

En este indicador, deberá calcular los rendimientos de la serie de precios tal y como se describe en el indicador de volatilidad. Luego, deberá calcular el promedio del precio en el mercado local y mercado nacional del producto seleccionado. Se debe tener en cuenta que con las bases de datos

que se tienen, el cálculo de este índice funcionará para los precios que reporta SIPSA en centrales mayoristas.

Para lo anteriormente descrito, los códigos que se deben aplicar son los siguientes:

```
Precios_Mercado <- sqldf("SELECT Año, Mes, Semana, producto, avg(precio_medio) as  
Precio_Medio FROM Query WHERE id_producto = n GROUP BY Año, Mes, Semana")
```

```
Precio_central <- sqldf("SELECT Año, Mes, Semana, producto, avg(precio_medio) as  
Precio_Medio FROM Query WHERE id_producto = n AND id_lugar_venta = m GROUP BY Año,  
Mes, Semana")
```

*Donde:*

*n = id del producto que reporta SIPSA.*

*m = id del lugar de venta/central que reporta SIPSA.*

Posteriormente, se extrae el precio medio de las matrices y se aplica la fórmula que se describe a continuación:

$$\beta = \frac{\text{cov}(\text{Rendimientos}_{\text{Mercado}}, \text{Rendimientos}_{\text{Central}})}{\text{var}(\text{Rendimientos del Mercado})}$$

En términos del código bajo el ambiente R, la sintaxis es la siguiente:

```
Precios <- cbind(Precios_Mercado[,5],Precio_central[,5])
```

```
colnames(Precios) <- c("Precios_Mercado", "Precio_Central")
```

```
Precios <- as.data.frame(Precios)
```

```
Beta <- cov(Precio_central[,5],Precios_Mercado[,5])/var(Precios_Mercado[,5])
```

```
Beta
```

El resultado obtenido en la variable *Beta* se interpreta de la siguiente forma: si el Beta es mayor a uno, significa que los rendimientos del precio del mercado local son más volátiles que los rendimientos del precio nacional; si el Beta es menor a uno, significa que los rendimientos del precio del mercado local son menos volátiles que los rendimientos del precio del mercado nacional. Finalmente, si el beta es igual a uno, significa que los rendimientos del precio del mercado local son neutrales con respecto a los rendimiento del precio del mercado nacional.

### **Consideraciones:**

- Consideración 1: Si se desea calcular el promedio del precio local y nacional bajo el código presentado a través de la función *sqldf*, deberá importar y cargar la siguiente librería:

```
if(!require("sqldf")) install.packages("sqldf")  
library(sqldf)
```

En caso contrario, debe tener en cuenta que este indicador solo requiere dos series, rendimiento del precio mercado local y rendimiento del precio mercado nacional. Por lo tanto, el procesamiento de datos para llegar a estos resultados se puede desarrollar de forma diferente a la propuesta.

- Consideración 2: Para el cálculo del precio promedio nacional, se deben seleccionar las centrales que tengan los datos completos con el fin de evitar sesgos en el cálculo del indicador.

#### ***6.1.1.1.2 Monitoreo***

##### **6.1.1.1.2.1 Bandas de Bollinger**

Este indicador tiene interpretación para cada una de las bandas. La banda superior representa el límite máximo al que podría llegar el precio con un 95% de confianza; por su parte, la banda inferior representa el límite mínimo al que podría llegar el precio con un 95% de probabilidad.

Para el cálculo de las bandas de Bollinger y los demás indicadores de monitoreo en el ambiente de programación R, existe una librería llamada *TTR*. Para hacer uso de ésta, se debe instalar y cargar el paquete mencionado bajo el siguiente código:

```
if(!require("TTR")) install.packages("TTR")  
library(TTR)
```

Después de la instalación del paquete, se usa la función *BBands* para calcular las bandas de Bollinger, la cual debe ser guardada en una variable. Los argumentos que recibe esta función son la serie de precios que se desea consultar y el número de desviaciones estándar para la amplitud de las bandas. En este orden de ideas, el código aplicado es el siguiente:

```
bb <- BBands(Precio, sd=2.0)
```

*donde:*

*Precio: Es la serie de precios consultada.*

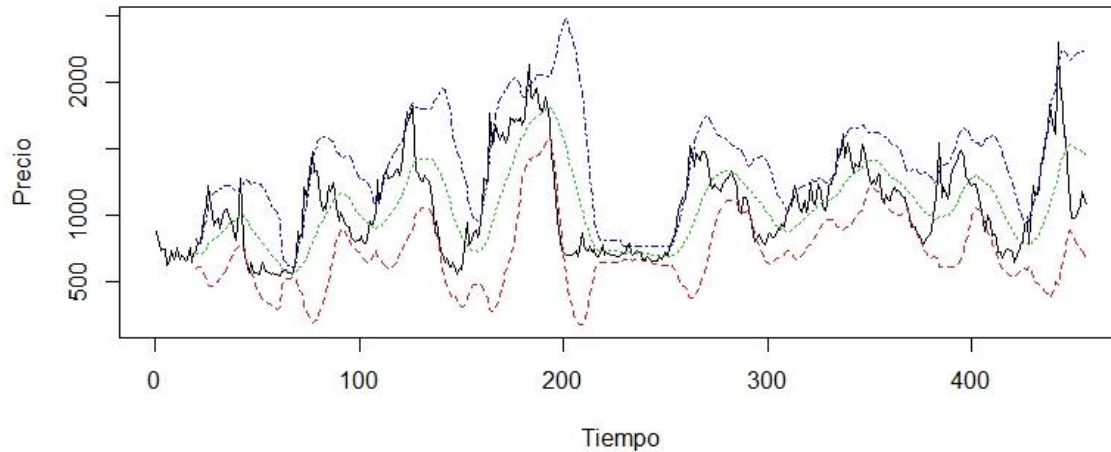
*sd = desviación estándar que define la amplitud de las bandas (Se asumen dos desviaciones estándar, dado que cuentan con el 95% de información según la regla empírica)*

Finalmente, se procede a graficar los resultados alojados en la variable *bb*. Para esto, se crea una variable donde se concatenen los precios con los resultados de la variable *bb*, y se representa haciendo uso de la función *matplot*. En términos del código, la sintaxis que se debe aplicar corresponde a la siguiente:

```
Plot_bb <- data.frame(Precio, bb20[,1:3])  
matplot(Plot_bb, type = "l", xlab = "Precio", ylab = "Tiempo")
```

El resultado obtenido luego de aplicar este código es el que se observa en el Gráfico 5.

Gráfico 5 Resultado Bandas de Bollinger ambiente R



Nota: Elaboración propia a partir de los cálculos realizados en el software R.

### Consideración:

- En el argumento `sd` de la función `Bbands`, se recomienda trabajar con dos desviaciones estándar.

#### 6.1.1.1.2.2 Medias Móviles

El indicador de medias móviles permite determinar la tendencia del comportamiento de los precios. El cálculo de éste requiere hacer uso de la función `SMA`, la cual tiene dos argumentos que son la serie de precios y el periodo desde donde se va a calcular el promedio. Ahora bien, para calcular las medias móviles, es necesario crear dos variables haciendo uso de la misma función, y, asignar dos periodos diferentes. En términos de código, se hace uso de la siguiente sintaxis:

```
sma20 <- SMA(Precios[,2], n=20)
```

```
sma40 <- SMA(Precios[,2], n=40)
```

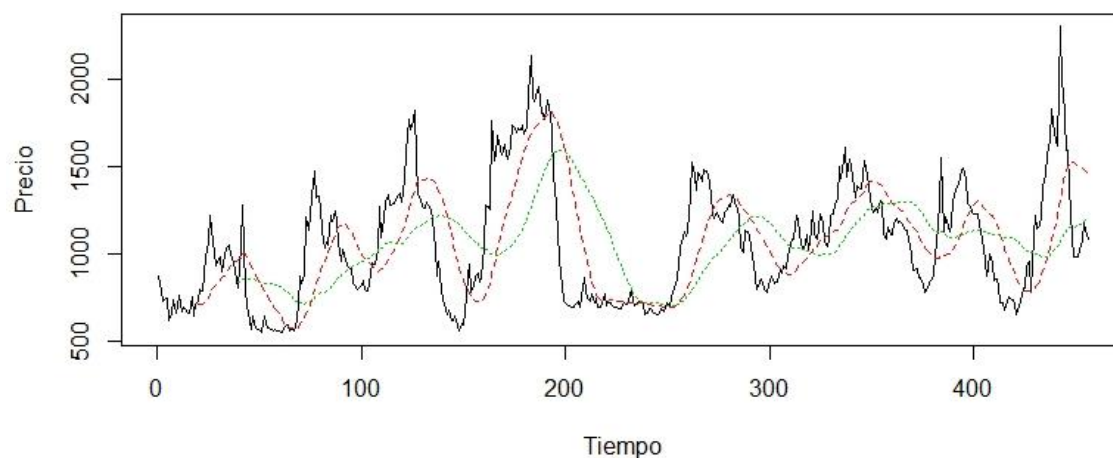
Finalmente, se grafican los resultados almacenados en las variables creadas. Para ello, se crea una variable donde se almacene el precio y las medias móviles, luego, se usa la función `matplot` para mostrar los resultados de forma gráfica. En este sentido, el código a utilizar corresponde al siguiente:

```
mm <- data.frame(Precios, sma20, sma40)
```

```
matplot(mm, type = "l", ylab = "Precio", xlab = "Tiempo")
```

Como resultado de aplicar este código, se obtiene lo que puede observarse en el Gráfico 6

Gráfico 6 Resultado Medias Móviles ambiente R



Nota: Elaboración propia

### Consideración:

- El número de periodos para el cálculo de las medias móviles es una variable de entrada que se puede elegir según el análisis. No obstante, se recomiendan los periodos 20 y 40 como estándar en caso de no tener claros los periodos a utilizar.

#### 6.1.1.1.2.3 Indicador de Media Móvil de Convergencia/Divergencia (MACD)

Este es un indicador de momentum que captura la tendencia y muestra la relación entre dos medias móviles de los precios, de modo que permite hacer seguimiento a la tendencia. Dicho indicador según (Kahn, 2010) se interpreta de la siguiente forma:

“La teoría subyacente es que a medida que un mercado se mueve hacia arriba, el más corto de dos promedios móviles está por encima del promedio más largo. Esto se debe a que el promedio más corto reacciona más rápido a los movimientos de precios. El promedio más largo siempre se quedará atrás. Cuando el promedio más corto cruza por debajo del más largo, es una señal de que la tendencia podría estar revirtiéndose” (p.288)



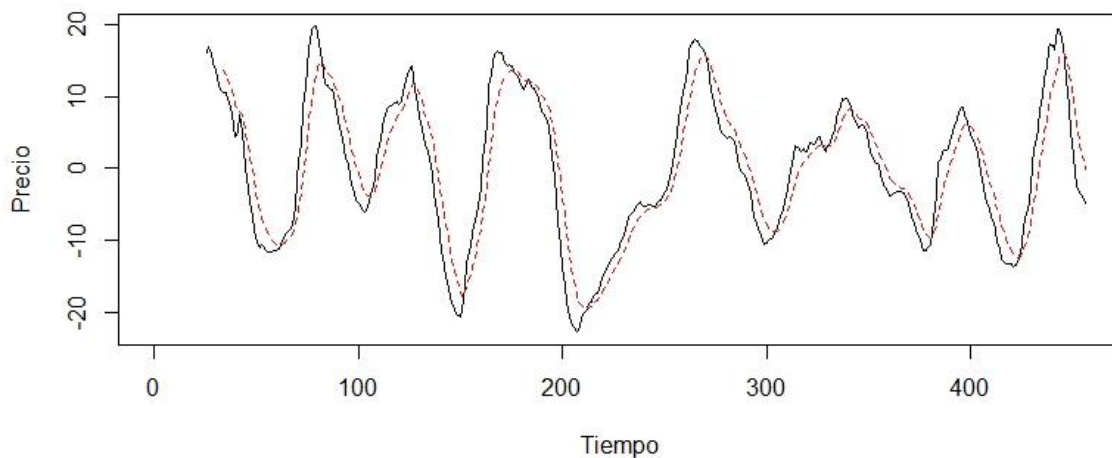
Para el cálculo del indicador MACD, debe hacerse uso de la función MACD del paquete TTR. Los argumentos de esta función son la serie de precios, número de periodos media móvil rápida, número de periodos media móvil lenta, número de periodos media móvil de señal y la función de cálculo. El cálculo del MACD implica crear una variable donde se almacenen los resultados de la función MACD. De acuerdo con lo anterior, el código a utilizar es el siguiente:

```
MACD_Precios <- MACD(Precio, nFast=12, nSlow=26, nSig=9, maType=EMA)
```

Por último, se representa gráficamente los resultados almacenados en la variable creada, utilizando la función `matplot`. En ese orden de ideas, el código a utilizar corresponde al siguiente y el resultado es el que se observa en el Gráfico 7:

```
matplot(MACD_Precios)
```

Gráfico 7 Resultado MACD ambiente R



Nota: Elaboración propia.

### Consideración:

- En los argumentos `nFast` y `nSlow`, se recomienda tener los valores de 12 y 26 para la función MACD.

-

#### 6.1.1.1.2.4 Índice de Fuerza Relativa (RSI)

Este indicador tiene como objetivo medir la fuerza del precio mediante la comparación de los movimientos individuales al alza o a la baja de los sucesivos precios de cierre. La forma de interpretar este indicador es de la siguiente forma: si la línea se encuentra por encima del 70%, hay una tendencia alta a que el precio retorne a su media, es decir, el precio tenderá a la baja; por su parte, si se encuentra por debajo del 30%, hay una tendencia alta a que el precio retorne a su media, es decir que, el precio tenderá al alza.

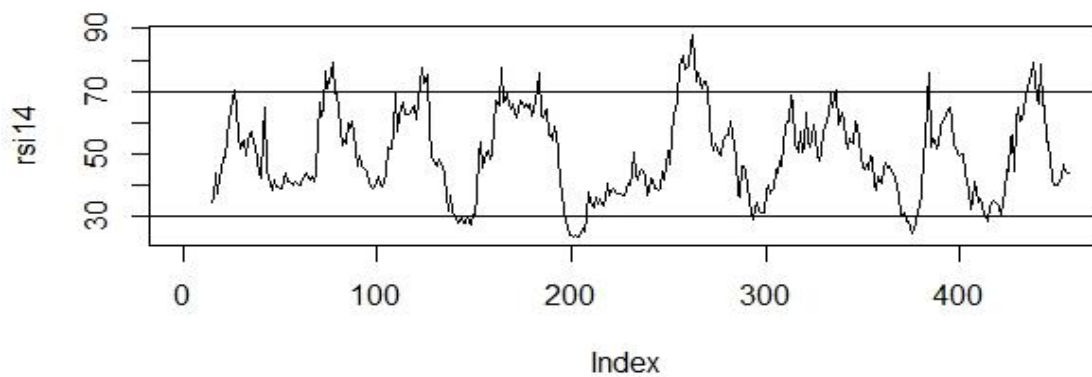
Para el cálculo del RSI, se debe utilizar la función RSI del paquete TTR. Los argumentos que tiene en cuenta esta función son la serie de precios y el periodo de la media móvil. Sobre esta base, se deberá crear una variable y almacenar el código que se muestra a continuación:

```
rsi14 <- RSI(Precios[,2], n = 14)
```

Posteriormente, se procede a graficar los resultados tal y como se evidencia en el Gráfico 8. Para esto, el código a implementar consta de dos líneas, la primera donde se realiza el gráfico, y la segunda, donde se agregan las líneas en los límites del 30 y 70.

```
plot(rsi14, type = "l")  
abline(h=c(30,70))
```

Gráfico 8 Resultado gráfico RSI ambiente R



Nota: Elaboración propia

### **6.1.1.1.3 Pronóstico**

Es importante mencionar que el módulo de pronósticos no tiene como resultado un indicador como la comparación de los contextos históricos y de monitoreo. Por el contrario, los resultados finales en este módulo son la aplicación de metodologías junto con un vector que consta de un periodo determinado y una variable definida.

#### **6.1.1.1.3.1 Metodología de Box-Jenkins**

Para comenzar se debe validar si la serie de tiempo a pronosticar tiene raíz unitaria tanto en el componente estacionario como en el componente estacional. Para lo anterior, se deben aplicar las pruebas de raíces unitarias de Dickey-Fuller aumentado y el Test de Hegy. A continuación, se describe el código que permite determinar la existencia de raíces unitarias:

- Dickey-Fuller aumentado:

En primer lugar, se debe convertir la serie de precios en una serie de tiempo; para ello el código a implementar es el siguiente:

```
df <- ts(Precio, start = c(2012, 1), frequency = 51)
```

Luego se debe estimar el máximo valor de los rezagos para la prueba, utilizando la siguiente fórmula:

```
maxlag=floor(12*(length(df)/100)^(1/4))
```

Posteriormente, se debe crear una variable que contenga la prueba de raíz unitaria. Para ello, se debe utilizar la función `ur.df` del paquete `URCA`. Dicha función tiene 4 argumentos. El primero de ellos es la serie de tiempo que se desea evaluar; el segundo, `TYPE`, describe el comportamiento de la serie y los valores que recibe este argumento son: "none", "drift" or "trend"; el tercer argumento es el número de rezagos, cuyo valor será la variable `maxlag` creada en el paso anterior; y el último argumento es el criterio de decisión, para el cual se recomienda utilizar el criterio AIC.

```
ru_z <- ur.df(df, type = c("none"), lags=maxlag, selectlags = c("AIC"))
```

Después de realizar la prueba se debe analizar el estadístico. Para esto, el código que se presenta permite reconocer si la serie tiene al menos una raíz unitaria o no. Si la variable Order\_d es igual a uno, significa que la serie tiene al menos una raíz unitaria, de lo contrario, significa que la serie no tiene raíces unitarias.

```
Estadistico <- ru_z@teststat[,1]
Valores_criticos <- ru_z@cval[1,]
if (abs(Estadistico)<=abs(Valores_criticos[2])){
  Order_d = 1
} else{
  Order_d = 0
}
summary(ru_z)
```

- Hegy Test:

Para el test de hegy, se debe utilizar la función hegy.test del paquete uroot. Esta función contiene 2 argumentos que son: (1) La serie de tiempo a analizar; determinístico, el cual es un vector de longitud tres que contiene ceros o unos para indicar, respectivamente, si se incluyen variables ficticias constantes, de tendencia o estacionales en la ecuación de regresión de la prueba; (2) y el lag.method, cuya recomendación es utilizar el AIC.

```
Prueba_RE <- hegy.test(df, deterministic = c(1,1,0), lag.method = "AIC")
```

Luego de realizar la prueba, se crea una condicional para determinar si existen raíces unitarias estacionales. Si el valor de la variable Order\_D es igual a uno, significa que la serie tiene al menos una raíz unitaria estacional; de lo contrario la serie no presenta raíces unitarias estacionales.

```
if (sum(Prueba_RE$pvalues < 0.05)==0){
  Order_D <- 0
}
```

```
} else {  
  Order_D <- 1  
}  
Prueba_RE
```

Posterior a las pruebas de raíces unitarias estacionarias y estacionales, se procede con la realización de la prueba de Box y Cox, con el fin de regular la varianza de la serie, lo cual implica una transformación de los datos según el resultado de la prueba. Para este proceso se debe crear una variable y utilizar la función `BoxCox.lambda` del paquete `forecast`, la cual usa 3 argumentos; el primero es la serie de tiempo a analizar; el segundo es el método, cuya recomendación es utilizar la de “guerrero”, y el tercer argumento son los intervalos de confianza, para los cuales se sugiere un rango entre -2 y 2.

```
Lambda <- round(BoxCox.lambda(df, method = c("guerrero"), lower = -2, upper = 2),1)
```

Después de calcular el lamda para la transformación de la serie que permita estabilizar la varianza de ésta, se modifica la serie haciendo uso de la función `BoxCox` del paquete `forecast`. Los argumentos que recibe esta función son dos: el primero es la serie que se está analizando y el segundo es el lamda calculado.

```
df_cox <-forecast::BoxCox(df,Lambda)
```

Con todas las pruebas realizadas, se debe llevar a cabo la estimación del modelo. Para lo anterior, se propone un algoritmo que permite evaluar y seleccionar los 5 mejores modelos que cumplan con los criterios de la normalidad y el ruido blanco en el término de error. El número de modelos a evaluar son el resultado de los órdenes máximos que se definan en los parámetros  $p$ ,  $q$ ,  $P$  y  $Q$ . A modo de ejemplo, si para todos estos parámetros se define que el orden máximo es 3, se estarían evaluando en el algoritmo  $(p+1) * (q+1) * (P+1) * (Q+1)$  modelos, es decir, un total de 256 modelos.

Inicialmente se deben definir los órdenes máximos de los parámetros. Como sugerencia y preservando el concepto de parsimonia, se recomienda un máximo de 3. Luego, se debe crear una matriz donde se almacenen los resultados del modelo con una dimensión de 9 columnas por 256 filas.

$p=3$

$q=3$

$P=3$

$Q=3$

$d=Order\_d$

$D=Order\_D$

$maxfilas=(p+1)*(q+1)*(P+1)*(Q+1)$

$ic\_mod=matrix(rep(-99, times=maxfilas*9), nrow=maxfilas, ncol=9)$

Después de definir el número de modelos a evaluar y la matriz donde se almacenarán los resultados, se da paso a la aplicación de un ciclo for anidado para que se evalúe cada modelo. En la primera parte del ambiente de los for anidados, se encuentra la función tso del paquete tsoutliers, el cual permitirá detectar los datos atípicos de la serie de tiempo según su característica: aditivos, cambio de nivel o temporales.

Posteriormente, se encuentra una serie de líneas de código que permitirán almacenar en una matriz, las variables dummy donde se presentaron los datos atípicos y en qué momento del tiempo se presentaron. Con esta matriz, se procede a estimar el modelo con la función ARIMA del paquete forecast; los argumentos de esta función son: La serie de tiempo no transformada, el orden del modelo, la matriz de datos atípicos, el método máxima verosimilitud a utilizar y el lamda de Box y Cox para la transformación.

Al momento de estimar el modelo, el algoritmo pasa a calcular los errores y a evaluar si éstos son ruido blanco y tienen un comportamiento normal. La prueba utilizada para evaluar si los errores son ruido es el test de Ljung y Box, mientras que para la normalidad, se utilizaron las pruebas de Jarque-Bera y Shapiro.

Finalmente, el algoritmo almacena los resultados en la matriz creada inicialmente, y, se elabora una segunda matriz donde se almacenen los 5 mejores modelos con mayores valores en el criterio AIC y cumplan con un *Ljung\_Box*  $\geq 0.05$  y un *Jarque\_Bera*  $\geq 0.05$ .

```

k=1
for(i in 0:p) {
  for(j in 0:q) {
    for(s in 0:P) {
      for(m in 0:Q) {
        out <- tso(df_cox,types=c("AO", "LS", "TC"), tsmethod="arima", discard.method = "bottom-
up",
          args.tsmethod=list(order=c(i,d,j),seasonal=list(order=c(s,D,m), period=51)))

        drift <- seq(df_subastar_comparacion)
        Outliers_det <- data.frame()
        Outliers_det_name <- data.frame()
        Outliers_LS <- data.frame()
        Outliers_LS_name <- data.frame()

        Contador_1 <- 0
        Contador_2 <- 0
        for (y in 1:nrow(out$outliers)){
          if (out$outliers[y,1] == "LS"){
            Contador_1 <- Contador_1 + 1
            Outliers_LS[1:nrow(df_subastar_comparacion),Contador_1] <-
1*(seq(df_subastar_comparacion[,1])>=out$outliers[y,2])
            Outliers_LS_name[1,Contador_1] <- paste(out$outliers[y,1],out$outliers[y,2])
          } else {
            Contador_2 <- Contador_2 + 1

```

```

        Outliers_det[1:nrow(df_subastar_comparacion),Contador_2] <-
1*(seq(df_subastar_comparacion[,1])==out$outliers[y,2])
        Outliers_det_name[1,Contador_2] <- paste(out$outliers[y,1],out$outliers[y,2])
    }
}
if(nrow(Outliers_LS)==0){
    Total_Outliers <- cbind(Outliers_det)
    colnames(Total_Outliers) <- c(Outliers_det_name)
}else if(nrow(Outliers_det)==0){
    Total_Outliers <- cbind(Outliers_LS)
    colnames(Total_Outliers) <- c(Outliers_LS_name)
} else {
    Total_Outliers <- cbind(Outliers_LS,Outliers_det)
    colnames(Total_Outliers) <- c(Outliers_LS_name,Outliers_det_name)
}

mod_interv1 <- Arima(df, order=c(i,d,j), seasonal = list(order=c(s,D,m), period = 51),xreg =
as.matrix(Total_Outliers), method=c("CSS-ML"), lambda = Lambda)

res1_CSS_ML <- residuals(mod_interv1)
res1_est <- res1_CSS_ML/(mod_interv1$sigma2^.5)
SHAPIRO <- shapiro.test(res1_est)
lb <- Box.test(res1_CSS_ML, lag = nrow(df_subastar_comparacion)/4, type = "Ljung-Box")
jb <- jarqueberaTest(res1_CSS_ML)
JB_Test <- jb@test$p.value
Shapiro <- SHAPIRO$p.value
LB_Test <- lb$p.value

numpar=length(coef(mod_interv1))
T=length(residuals(mod_interv1))
ic_mod[k,1]=i

```



```

ic_mod[k,2]=j
ic_mod[k,3]=s
ic_mod[k,4]=m
ic_mod[k,5]=-2*(mod_interv1$loglik/T)+(2*numpar)/T # AIC
ic_mod[k,6]=-2*(mod_interv1$loglik/T)+(numpar*log(T))/T # SIC
ic_mod[k,7]= Shapiro
ic_mod[k,8]= LB_Test
ic_mod[k,9]= JB_Test

print(c(i,j,s,m))

k=k+1
}
}
}
}
p_=ic_mod[,1]
q_=ic_mod[,2]
P_=ic_mod[,3]
Q_=ic_mod[,4]
AIC=ic_mod[,5]
SIC=ic_mod[,6]
Normalidad=ic_mod[,7]
Ljung_Box=ic_mod[,8]
Jarque_Bera=ic_mod[,9]
Crit_Inf=cbind(p_, q_, P_, Q_,AIC, SIC,Normalidad,Jarque_Bera,Ljung_Box)
Crit_Inf=data.frame(Crit_Inf)
aic=Crit_Inf[order(Crit_Inf$AIC),]
Modelos_Seleccionados <- subset(aic, Ljung_Box >= 0.05 & Jarque_Bera >= 0.05)
Modelos_Seleccionados <- Modelos_Seleccionados[1:5,]
print(Modelos_Seleccionados)

```

Con los modelos seleccionados se procede a medir el poder predictivo de los mismos, de manera que se debe aplicar el algoritmo que se describe a continuación:

En primer lugar, se debe separar la serie en una muestra de entrenamiento y una muestra de evaluación. Se recomienda un 80%-20%, siendo el primero la muestra de entrenamiento y el segundo la muestra de evaluación. Luego, se debe crear una matriz con una dimensión de 5\*9, la cual tendrá los órdenes del modelo y los criterios de evaluación MAPE, RMSE, MAE, sesgo en media y sesgo en varianza.

```
Resultados_Pronostico <- matrix(nrow = 5,ncol=9)
colnames(Resultados_Pronostico) <- c("p","q","P","Q","MAPE","RMSE",
"MAE","Sesgo_Media", "Sesgo_Varianza")
train <- ts(df[1:313],start = c(2012, 2),frequency = 51)
test <- df[314:326]
```

Después de esto, se debe crear un for anidado el cual realizará un pronóstico de cada modelo y lo confrontará con la muestra de evaluación. Asimismo, calculará los indicadores de evaluación. En este sentido, la primera parte del algoritmo se ocupa de estimar los parámetros de cada modelo con los datos de muestra, y también de reconocer los datos atípicos tal y como se describió anteriormente.

El siguiente paso es estimar el modelo y realizar el pronóstico. Para ello se debe utilizar la función forecast del paquete forecast, cuyos argumentos son los siguientes: la variable donde se creó el modelo (la ventana de pronóstico será igual al tamaño de la muestra de evaluación); los intervalos de confianza que se desean calcular (se recomiendan intervalos del 80% y 95%); el valor de lamda para la transformación de la serie que permita estabilizar la varianza; y la matriz con los datos atípicos.

La última parte del algoritmo se encarga de calcular los criterios de evaluación y seleccionar el mejor modelo. El criterio seleccionado para escoger el modelo es aquel que tenga menor sesgo en la media.

```

for (i in 1:5) {
  order_p <- Modelos_Seleccionados[i,1]
  order_q <- Modelos_Seleccionados[i,2]
  order_P <- Modelos_Seleccionados[i,3]
  order_Q <- Modelos_Seleccionados[i,4]
  Lambda <- round(BoxCox.lambda(df, method = c("guerrero"), lower = -2, upper = 2),1)
  df_cox <-forecast::BoxCox(df,Lambda)
  out <- tso(df_cox,types=c("AO", "LS", "TC"), tsmethod="arima", discard.method = "bottom-
up",
args.tsmethod=list(order=c(order_p,d,order_q),seasonal=list(order=c(order_P,D,order_Q),
period=51)))
  drift <- seq(df)
  Outliers_det <- data.frame()
  Outliers_det_name <- data.frame()
  Outliers_LS <- data.frame()
  Outliers_LS_name <- data.frame()
  Contador_1 <- 0
  Contador_2 <- 0
  for (y in 1:nrow(out$outliers)){
    if (out$outliers[y,1] == "LS"){
      Contador_1 <- Contador_1 + 1
      Outliers_LS[1:nrow(df_Precio),Contador_1]<-
1*(seq(df_Precio[,3])>=out$outliers[y,2])
      Outliers_LS_name[1,Contador_1] <- paste(out$outliers[y,1],out$outliers[y,2])
    } else {
      Contador_2 <- Contador_2 + 1
      Outliers_det[1:nrow(df_Precio),Contador_2] <-
1*(seq(df_Precio[,3])=out$outliers[y,2])
      Outliers_det_name[1,Contador_2] <- paste(out$outliers[y,1],out$outliers[y,2])
    }
  }
}

```

```

}
if(nrow(Outliers_LS)==0){
  Total_Outliers <- cbind(Outliers_det)
  colnames(Total_Outliers) <- c(Outliers_det_name)
}else if(nrow(Outliers_det)==0){
  Total_Outliers <- cbind(Outliers_LS)
  colnames(Total_Outliers) <- c(Outliers_LS_name)
} else {
  Total_Outliers <- cbind(Outliers_LS,Outliers_det)
  colnames(Total_Outliers) <- c(Outliers_LS_name,Outliers_det_name)
}
Lista <- list()
for (k in 1:ncol(Total_Outliers)){
  if(sum(Total_Outliers[1:313,k])==0){
    Lista <- c(Lista,k)
  }
}
outliers_train <- Total_Outliers[1:313,-unlist(Lista)]
outliers_test <- Total_Outliers[314:326,-unlist(Lista)]

Mod_Eva_pron <- Arima(train, order=c(order_p,d,order_q), seasonal =
list(order=c(order_P,D,order_Q), period = 51),xreg = as.matrix(outliers_train),
method=c("ML"), lambda = Lambda)

z_pred <- forecast(Mod_Eva_pron, h=length(test), level=c(80, 95),lambda = Lambda,
xreg=as.matrix(outliers_test),biasadj=TRUE)

#Sesgo en Media
ecm <- mean((test-ts(z_pred$mean))^2)
prop_sesgom <- (mean(z_pred$mean)-mean(test))^2/ecm
#Sesgo en varianza
num_pron <- length(z_pred$mean)
sigmap=(((num_pron-1)/num_pron)*var(z_pred$mean))^0.5

```

```

sigmar=(((num_pron-1)/num_pron)*var(test))^0.5
prop_var=(sigmap-sigmar)^2/ecm
#Evaluación
Resultados_Pronostico[i,1] <- order_p
Resultados_Pronostico[i,2] <- order_q
Resultados_Pronostico[i,3] <- order_P
Resultados_Pronostico[i,4] <- order_Q
Resultados_Pronostico[i,5] <- mape(z_pred$mean, test)
Resultados_Pronostico[i,6] <- rmse(z_pred$mean, test)
Resultados_Pronostico[i,7] <- mae(z_pred$mean, test)
Resultados_Pronostico[i,8] <- prop_sesgom
Resultados_Pronostico[i,9] <- prop_var
}
Condicion <- min(Resultados_Pronostico[,8])
Modelo_Final <- Resultados_Pronostico[Resultados_Pronostico[,8]==Condicion,]
print(Resultados_Pronostico)

```

Finalmente, al tener el mejor modelo seleccionado, el último paso de este algoritmo consiste en realizar el pronóstico en los periodos posteriores al último dato reportado. Para ello el algoritmo a implementar consta de lo siguiente:

Primero, se estima el modelo con los outliers a través de la función ARIMA. Seguidamente, se utiliza la función forecast para realizar el pronóstico en la venta de tiempo que se desee, y, por último, se grafican e imprimen los resultados de los pronósticos con sus intervalos de confianza. Es de resaltar que el detalle no se explica en este punto debido a que los algoritmos son iguales a los anteriores, con la diferencia que aquí se está estimando con el mejor de todos.

```

Lambda <- round(BoxCox.lambda(df, method = c("guerrero"), lower = -2, upper = 2),1)
df_cox <-forecast::BoxCox(df,Lambda)

```

```

out <- tso(df_cox,types=c("AO", "LS", "TC"), tsmethod="arima", discard.method = "bottom-up",
args.tsmethod=list(order=c(Modelo_Final[1],                                d,
Modelo_Final[2]),seasonal=list(order=c(Modelo_Final[3], D, Modelo_Final[4]), period=51)))
drift <- seq(df_Precio)
Outliers_det <- data.frame()
Outliers_det_name <- data.frame()
Outliers_LS <- data.frame()
Outliers_LS_name <- data.frame()
Contador_1 <- 0
Contador_2 <- 0
for (y in 1:nrow(out$outliers)){
  if (out$outliers[y,1] == "LS"){
    Contador_1 <- Contador_1 + 1
    Outliers_LS[1:nrow(df_Precio),Contador_1] <- 1*(seq(df_Precio[,3])>=out$outliers[y,2])
    Outliers_LS_name[1,Contador_1] <- paste(out$outliers[y,1],out$outliers[y,2])
  } else {
    Contador_2 <- Contador_2 + 1
    Outliers_det[1:nrow(df_Precio),Contador_2] <- 1*(seq(df_Precio[,3])==out$outliers[y,2])
    Outliers_det_name[1,Contador_2] <- paste(out$outliers[y,1],out$outliers[y,2])
  }
}
if(nrow(Outliers_LS)==0){
  Total_Outliers <- cbind(Outliers_det)
  colnames(Total_Outliers) <- c(Outliers_det_name)
}else if(nrow(Outliers_det)==0){
  Total_Outliers <- cbind(Outliers_LS)
  colnames(Total_Outliers) <- c(Outliers_LS_name)
} else {
  Total_Outliers <- cbind(Outliers_LS,Outliers_det)
  colnames(Total_Outliers) <- c(Outliers_LS_name,Outliers_det_name)
}

```

```

Modelo_Final <- Arima(df, order=c(Modelo_Final[1], d, Modelo_Final[2]), seasonal =
list(order=c(Modelo_Final[3], D, Modelo_Final[4]), period = 51), xreg =
as.matrix(Total_Outliers), method=c("CSS-ML"), lambda = Lambda, biasadj=TRUE)
out_final_for <- matrix(0,nrow = length(test),ncol = ncol(Total_Outliers))
for (i in 1:ncol(Total_Outliers)){
  if(sum(Total_Outliers[c(nrow(Total_Outliers)-1,nrow(Total_Outliers)),i])>1){
    out_final_for[,i] <- 1
  }
}
colnames(out_final_for) <- colnames(Total_Outliers)
z_pred <- forecast(Modelo_Final, lambda = Lambda, xreg =
as.matrix(out_final_for),h=length(test), level=c(80, 95))
summary(Modelo_Final)
autoplot(z_pred)
print(z_pred)

```

#### 6.1.1.1.3.2 Redes Neuronales Long short-term memory (LSTM)

Para la aplicación de las redes neuronales recurrentes se recomienda el uso del lenguaje de programación Python. En este sentido, lo primero que se debe hacer es cargar los paquetes requeridos para trabajar con redes neuronales recurrentes que son:

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import tensorflow as tf
from pandas.tseries.offsets import MonthEnd
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
import tensorflow.keras.backend as K

```

```
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras.layers import LSTM
from keras.wrappers.scikit_learn import KerasRegressor
import sklearn.metrics as metrics

import numpy
from sklearn.model_selection import GridSearchCV
from keras.models import Sequential
from keras.layers import Dense
from keras.wrappers.scikit_learn import KerasClassifier
from keras.backend import clear_session
from keras.constraints import maxnorm
from keras.layers import Dropout
```

Luego de importar los paquetes, se procede a cargar la base de datos y a separar los mismos en muestra de entrenamiento y de prueba. Al igual que en la metodología Box y Jenkins, se recomienda un 80%-20%. Dicho proceso de separación se puede hacer a través de la fecha, tal y como se observa en el código:

```
b = 24
train = dataset.loc[0:len(dataset)-b, ['Precio']]
test = dataset.loc[len(dataset)-b:len(dataset)+1, ['Precio']]
```

Ahora bien, para las redes neuronales recurrentes es importante la normalización de los datos. Para ello se utiliza la función `MinMaxScaler()` la cual se debe guardar en una variable; luego se transforman los datos haciendo uso del siguiente código:

```
sc = MinMaxScaler()
train_sc = sc.fit_transform(train)
test_sc = sc.transform(test)
```



Dado que las redes neuronales necesitan para su cálculo de una variable independiente y varias variables explicativas, en esta estructura se deben transformar las series de precios. Para lograrlo, se deben definir los  $n$  periodos de pronóstico. Con este valor de  $n$  se particiona la serie de los precios, de tal manera que esos  $n$  precios serán las variables independientes y el  $n+1$  será la variable dependiente. A modo de ejemplo, supongamos que se tiene una serie con 20 datos y se desea pronosticar 8 periodos, entonces las variables independientes y dependientes quedarían expresadas como se observa en la Tabla 6.

**Tabla 6** Ejemplo de transformación de las series de precios

Variable Dependiente (Y)	Variables Independientes (X)
Precio 9	Precio 1, Precio 2, Precio 3, Precio 4, Precio 5, Precio 6, Precio 7, Precio 8.
Precio 10	Precio 2, Precio 3, Precio 4, Precio 5, Precio 6, Precio 7, Precio 7, Precio 8.
Precio 11	Precio 3, Precio 4, Precio 5, Precio 6, Precio 7, Precio 8, Precio 9, Precio 10.
Precio 12	Precio 4, Precio 5, Precio 6, Precio 7, Precio 8, Precio 9, Precio 10, Precio 11.
Precio 13	Precio 5, Precio 6, Precio 7, Precio 8, Precio 9, Precio 10, Precio 11, Precio 12.
Precio 14	Precio 6, Precio 7, Precio 8, Precio 9, Precio 10, Precio 11, Precio 12, Precio 13.
Precio 15	Precio 7, Precio 8, Precio 9, Precio 10, Precio 11, Precio 12, Precio 13, Precio 14.
Precio 16	Precio 8, Precio 9, Precio 10, Precio 11, Precio 12, Precio 13, Precio 14, Precio 15.
Precio 17	Precio 9, Precio 10, Precio 11, Precio 12, Precio 13, Precio 14, Precio 15, Precio 16.

Precio 18	Precio 10, Precio 11, Precio 12, Precio 13, Precio 14, Precio 15, Precio 16, Precio 17.
Precio 19	Precio 11, Precio 12, Precio 13, Precio 14, Precio 15, Precio 16, Precio 17, Precio 18.
Precio 20	Precio 12, Precio 13, Precio 14, Precio 15, Precio 16, Precio 17, Precio 18, Precio 19.

Nota: Elaboración propia.

En el marco de la idea anterior, el código de Python que nos permite realizar esta función es el siguiente:

```
train_sc_df = pd.DataFrame(train_sc, columns=['Scaled'], index=train.index)
test_sc_df = pd.DataFrame(test_sc, columns=['Scaled'], index=test.index)
```

*a = 12*

```
for s in range(1, a+1):
    train_sc_df['shift_{}'.format(s)] = train_sc_df['Scaled'].shift(s)
    test_sc_df['shift_{}'.format(s)] = test_sc_df['Scaled'].shift(s)
```

Con la estructura de datos preparada, se procede con la creación de los arreglos en formato numpy para ingresar los datos al modelo. Lo anterior, se debe realizar con el siguiente código:

```
X_train = train_sc_df.dropna().drop('Scaled', axis=1)
y_train = train_sc_df.dropna()[['Scaled']]
X_test = test_sc_df.dropna().drop('Scaled', axis=1)
y_test = test_sc_df.dropna()[['Scaled']]
X_train = X_train.values
X_test = X_test.values
y_train = y_train.values
y_test = y_test.values
```

```
X_train_t = X_train.reshape(X_train.shape[0], 1, a)
X_test_t = X_test.reshape(X_test.shape[0], 1, a)
```

Teniendo los datos preparados, se continúa con la definición de la arquitectura de la red neuronal. En vista de la complejidad de cada serie de precios, se sugiere revisar cada serie y adecuar la estructura que mejor logre explicar el comportamiento de precios. De acuerdo con lo anterior, una arquitectura estándar que se puede utilizar es la que se describe a continuación:

```
K.clear_session()
model = Sequential()

model.add(LSTM(20, input_shape=(1, a)))
model.add(Dropout(0.2))
model.add(Dense(1))
```

Esta red neuronal tiene 3 capas: una de entrada con 20 neuronas y las otras dos con una neurona. Al momento de definir la arquitectura, deberá compilar el modelo bajo los siguientes parámetros: función de pérdida y el optimizador. Para el primer parámetro, deberá utilizar el 'mean\_squared\_error', dado que es el más recomendado para problemas de regresión, y para el segundo parámetro, se sugiere emplear el adam debido a que su estructura permite llegar al valor mínimo de la función de pérdida de forma ligera. No obstante, como se presentará después, estos datos serán calibrados con validación cruzada.

```
model.compile(loss='mean_squared_error', optimizer='adam')
```

Teniendo el modelo compilado, se procede a ajustar el mismo con los datos de entrenamiento. Para esto, el código a utilizar es el que se presenta a continuación:

```
early_stop = EarlyStopping(monitor='loss', patience=1, verbose=1)
model.fit(X_train_t, y_train, epochs=100,
```

```
batch_size=1, verbose=0, callbacks=[early_stop])
```

Donde el parámetro epoch, representa el número de épocas para refinar el modelo; el batch\_size corresponde a la muestra a tomar del total de conjunto de datos; el verbose es la forma en cómo se presenta el progreso de las épocas; y el callback se encarga del entrenamiento cuando la función de pérdida no parece mejorar.

Antes de predecir el modelo, se deben tunear los hiperparámetros con el fin de obtener la mejor estructura de la red neuronal que permitirá predecir el comportamiento de los precios, para lo cual, se utilizó el siguiente algoritmo de validación cruzada con unas grillas determinadas para cada hiperparámetro.

- a. Tuneo del Bach y el número de épocas.

```
def grid_search_info(grid_result):  
    print(f"Mejor: {grid_result.best_score_} donde {grid_result.best_params_}")  
    means = grid_result.cv_results_['mean_test_score']  
    stds = grid_result.cv_results_['std_test_score']  
    params = grid_result.cv_results_['params']  
    for mean, stdev, param in zip(means, stds, params):  
        print(f"{mean:.5f} +/- {stdev:.5f} donde: {param}")  
    return None  
  
def create_model():  
    model = Sequential()  
    model.add(LSTM(20, input_shape=(1, a)))  
    model.add(Dropout(0.2))  
    model.add(Dense(1))  
    model.compile(loss='mean_squared_error', optimizer='adam', run_eagerly=True)  
    return model  
  
seed = 7  
numpy.random.seed(seed)
```

```

K.clear_session()
model = KerasRegressor(build_fn=create_model, verbose=0)

batch_size = [10, 20, 40, 60, 80, 100]
epochs = [10, 50, 100]
param_grid = dict(batch_size=batch_size, epochs=epochs)
grid = GridSearchCV(estimator=model, param_grid=param_grid, n_jobs=-1, cv=3)
grid_result = grid.fit(X_train_t, y_train)

grid_search_info(grid_result)

```

b. Tuneo del algoritmo de optimización.

```

def create_model(optimizer='adam'):
    model = Sequential()
    model.add(LSTM(20, input_shape=(1, a)))
    model.add(Dropout(0.2))
    model.add(Dense(1))
    model.compile(loss='mean_squared_error', optimizer=optimizer, run_eagerly=True)
    return model

```

```

K.clear_session()
model = KerasRegressor(build_fn=create_model, epochs=100, batch_size=10,
verbose=0)

optimizer = ['SGD', 'RMSprop', 'Adagrad', 'Adadelta', 'Adam', 'Adamax', 'Nadam']
param_grid = dict(optimizer=optimizer)
grid = GridSearchCV(estimator=model, param_grid=param_grid, n_jobs=-1, cv=3)
grid_result = grid.fit(X_train_t, y_train)

```

```
grid_search_info(grid_result)
```

- c. Tuneo de la regularización dropout.

```
def create_model(dropout_rate=0.0):
```

```
    model = Sequential()
```

```
    model.add(LSTM(20, input_shape=(1, a)))
```

```
    model.add(Dropout(dropout_rate))
```

```
    model.add(Dense(1))
```

```
    model.compile(loss='mean_squared_error', optimizer='Adam', run_eagerly=True)
```

```
    return model
```

```
K.clear_session()
```

```
model = KerasRegressor(build_fn=create_model, epochs=100, batch_size=10,  
verbose=0)
```

```
dropout_rate = [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]
```

```
param_grid = dict(dropout_rate=dropout_rate)
```

```
grid = GridSearchCV(estimator=model, param_grid=param_grid, n_jobs=-1, cv=3)
```

```
grid_result = grid.fit(X_train_t, y_train)
```

```
grid_search_info(grid_result)
```

- d. Tuneo del número de neuronas en capas ocultas.

```
def create_model(neurons=1):
```

```
    model = Sequential()
```

```
    model.add(LSTM(neurons, input_shape=(1, a)))
```

```
    model.add(Dropout(0.0))
```

```
    model.add(Dense(1))
```

```
model.compile(loss='mean_squared_error', optimizer='Adam', run_eagerly=True)
return model
```

```
K.clear_session()
```

```
model = KerasRegressor(build_fn=create_model, epochs=100, batch_size=10,
verbose=0)
```

```
neurons = [1, 5, 10, 15, 20, 25, 30]
```

```
param_grid = dict(neurons=neurons)
```

```
grid = GridSearchCV(estimator=model, param_grid=param_grid, n_jobs=-1, cv=3)
```

```
grid_result = grid.fit(X_train_t, y_train)
```

```
grid_search_info(grid_result)
```

e. Tuneo de capas ocultas.

```
def create_model(input_neurons = 12, hidden_layers=1, init_mode = 'uniform',
activation='relu'):
```

```
model = Sequential()
```

```
model.add(LSTM(25, input_shape=(1, a)))
```

```
for i in range(hidden_layers):
```

```
    # Add one hidden layer
```

```
    model.add(Dense(input_neurons, activation = activation, kernel_initializer =
init_mode))
```

```
model.add(Dropout(0.0))
```

```
model.add(Dense(1))
```

```
model.compile(loss='mean_squared_error', optimizer='Adam', run_eagerly=True)
```

```
return model
```

```

K.clear_session()
model = KerasRegressor(build_fn=create_model, epochs=100, batch_size=10,
verbose=0)

input_neurons = [2, 4, 8, 16, 32]
hidden_layers = [2, 3, 4]
init_mode = ['uniform', 'lecun_uniform', 'normal', 'zero', 'glorot_normal', 'glorot_uniform',
'he_normal', 'he_uniform']
activation = ['softmax', 'softplus', 'softsign', 'relu', 'tanh', 'sigmoid', 'hard_sigmoid', 'linear']
param_grid = dict(input_neurons = input_neurons ,hidden_layers=hidden_layers,
init_mode = init_mode, activation = activation)
grid = GridSearchCV(estimator=model, param_grid=param_grid, n_jobs=-1, cv=3)
grid_result = grid.fit(X_train_t, y_train)

grid_search_info(grid_result)

```

Finalmente, se tiene el modelo calibrado y se realizan las pruebas con la muestra de entrenamiento y a su vez se determina el poder de pronóstico a través de los indicadores de evaluación MAPE y RMSE. El código a aplicar es el siguiente:

```

K.clear_session()
model = Sequential()
model.add(LSTM(25, input_shape=(1, a)))
model.add(Dense(16, activation = 'linear', kernel_initializer = 'he_normal'))
model.add(Dense(16, activation = 'linear', kernel_initializer = 'he_normal'))
model.add(Dense(16, activation = 'linear', kernel_initializer = 'he_normal'))
model.add(Dense(16, activation = 'linear', kernel_initializer = 'he_normal'))
model.add(Dropout(0.0))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')

```



```

MAE = metrics.mean_absolute_error(y_test, y_pred)
MSE = metrics.mean_squared_error(y_test, y_pred)
RMSE = np.sqrt(MSE)
MAPE = metrics.mean_absolute_percentage_error(y_test,y_pred)

```

### 6.1.1.1.3.3 Modelo ARCH y GARCH

El algoritmo para el modelo ARCH y GARCH inicia con determinar si los rendimientos de las series de precios tienen efecto ARCH. Para lograrlo, se debe modelar el rendimiento del precio como un ruido blanco y luego elevar los errores al cuadrado y aplicar la prueba de Ljung-Box. Para esto, el algoritmo que se utiliza es el siguiente:

```

par(mfrow=c(2,1))
acf(Rendimiento_Precio^2)
pacf(Rendimiento_Precio^2)

mod_med1 <- arima(Rendimiento_Precio, order=c(0,0,0))
tsdiag(mod_med1, gof.lag=12)
K=c(1:12)

resMM1 <- residuals(mod_med1)

Box.test(resMM1^2, lag = 12, type = "Ljung-Box")

```

Luego de comprobar que los rendimientos tienen efecto ARCH, se procede con la determinación del modelo de varianza. El algoritmo a utilizar es el siguiente:

```

mod2norm = garchFit(formula = ~garch(1,1), data=Rendimiento_Precio, trace = FALSE)
summary(mod2norm)

# Resultados graficos del modelo
plot(mod2norm, which=2) # para elegir la opci?n 2.

```

```
plot(mod2norm@sigma.t, type='l')
```

```
par(mfcol=c(2,1)) # Grafico de los retornos y de volatilidad
```

```
plot(Rendimiento_Precio,type='l')
```

```
plot(mod2norm@sigma.t,type='l')
```

```
# Obtencion de la volatilidad
```

```
volatilnorm=mod2norm@sigma.t
```

Al tener el modelo de varianza se realiza el diagnóstico del modelo desde la normalidad y el ruido blanco en los errores, con el siguiente algoritmo:

```
# Diagnósticos
```

```
# cálculo de los residuales estandarizados
```

```
res_estnorm=mod2norm@residuals/mod2norm@sigma.t
```

```
plot(res_estnorm)
```

```
abline(-3, 0)
```

```
abline(3, 0)
```

```
# prueba de no autocorrelacion en los residuales estandarizados
```

```
par(mfrow=c(2,1))
```

```
Acf(res_estnorm, lag.max=15)
```

```
Pacf(res_estnorm, lag.max=15)
```

```
Box.test(res_estnorm,lag=15,type="Ljung")
```

```
# prueba de no autocorrelacion en los residuales estandarizados al cuadrado
```

```
par(mfrow=c(2,1))
```

```
Acf(res_estnorm^2, lag.max=15)
```

```
Pacf(res_estnorm^2, lag.max=15)
```

```
Box.test(res_estnorm^2,lag=15,type="Ljung")
```

```
# cálculo de estadísticos descriptivos de los residuales
```

```
basicStats(res_estnorm)
```

```
# Prueba de normalidad
```

```
par(mfcol=c(2,1))
```

```
plot(res_estnorm,type='l')
```

```
hist(res_estnorm)
```

```
par(mfcol=c(2,1))
```

```
densityPlot(as.timeSeries(res_estnorm))
```

```
qqnorm(res_estnorm) # Obtain normal probability plot (ideal case: a straight line)qqline(sresi) #
```

```
Impose a straight line on the QQ-plot.
```

```
qqline(res_estnorm)
```

Finalmente, luego de validar los supuestos, se procede con el ejercicio de predicción de la volatilidad. Para lo cual, le algoritmo a utilizar es el siguiente:

```
pron=predict(mod2norm, n.ahead = 12, trace = F, plot=T, nx=100)
```

```
# tabla con los pronsticos de media y volatilidad
```

```
pron
```

```
(std_pron=ts(pron[,3])) # volatilidad pronosticada
```

```
 finales=100 # N?mero de datos al final de la serie que se van a graficar con los pronósticos
```

```
 long=length(Rendimiento_Precio)
```

```
 inic=long-finales
```

```
 liminf=ts(c(Rendimiento_Precio[inic:long], pron[,4]), frequency=1)
```

```
 limsup=ts(c(Rendimiento_Precio[inic:long], pron[,5]), frequency=1)
```

```

pron_med=ts(c(Rendimiento_Precio[inic:long], pron[,1]), frequency=1)
real=ts(Rendimiento_Precio[inic:long], frequency=1)
volat_pred=ts(c(volatilnorm[inic:long],std_pron), frequency=1) # unión de la volatilidad histórica
estimada y la pronosticada
volatil=ts(volatilnorm[inic:long], frequency=1)

# grafico con los pronsticos de media y volatilidad
par(mfrow=c(2,1))
ts.plot(real, pron_med, liminf, limsup, main="Pronósticos de Log-Retornos del precio del
aguacate")
lines(liminf, col="blue")
lines(limsup, col="blue")
lines(pron_med, col="red")
lines(real, col="black")
ts.plot(volat_pred, volatil, main="Pronósticos Volatilidad de Log-Retornos del precio del
aguacate")
lines(volat_pred, col="red")
lines(volatil, col="black")

```

### **6.1.1.2 Cálculo de vulnerabilidad**

#### ***6.1.1.2.1 Modelos de regresión para medir el grado de sensibilidad de los precios de los sistemas productivos agropecuarios***

##### **6.1.1.2.1.1 Modelos de regresión**

Como punto inicial, se deben determinar las variables que inciden sobre el comportamiento de los precios, siendo las siguientes:

- Tasa de cambio: Si aumenta la tasa de cambio, se espera aumento en las exportaciones, lo cual hace que disminuya la oferta y, por tanto, se esperaría un aumento en los precios de los productos de los sistemas productivos.
- Precios internacionales: Si aumentan los precios internacionales, se estimularían las exportaciones y, por ende, se esperaría un aumento en los precios de los productos de los sistemas productivos.
- Producción e inventario: Si aumenta la oferta del producto asumiendo una demanda constante, se espera una reducción en los precios de los productos de los sistemas productivos.
- Exportaciones: Si aumentan las exportaciones, se espera una disminución en la oferta y, como consecuencia un aumento en los precios de los productos de los sistemas productivos.
- Importaciones: Si aumentan las importaciones, se espera un aumento en la oferta y, por tanto, una reducción en los precios de los productos de los sistemas productivos.

Cabe resaltar que las explicaciones económicas de las variables tomadas para el análisis de regresión cumplen con el supuesto de *ceteris paribus*, es decir, que el efecto de una variable dependiente con respecto a la variable independiente solo tendrá interpretación si las demás variables permanecen constantes.

Luego se procede con la estimación del modelo de regresión según el sistema productivo que se esté analizando. Para esto, se utiliza la función `lm` que viene por defecto en el lenguaje de programación R.

```
Modelo_Econometrico_1 <-  
lm(log(Precio_Interno)~log(Precio_Internacional)+log(TRM)+log(Importaciones)+log(exporta  
ciones)+log(producción)+log(inventario), data = Base_General)  
summary(Modelo_Econometrico_1)
```

Por último, se procede con las pruebas de heterocedasticidad, multicolinealidad y datos atípicos. Los códigos para implementar en la realización de estas 3 validaciones son los siguientes:

- Heterocedasticidad:

Dado que uno de los supuestos de los modelos de regresión es que los errores son homocedásticos, es necesario hacer la validación de éstos, pues de no hacerlo, las interpretaciones de la significancia de los parámetros estarán sesgadas. Por esta razón, se emplea el test de Breusch-Pagan para llevar a cabo esta validación, En este caso, la hipótesis a comprobar es:

$$H_0: \text{Existe homocedasticidad } \sigma_t^2 = \sigma^2 \forall t = 1, \dots, t$$

$$H_1: \text{Existe heterocedasticidad en función de } Z$$

Cuyo estadístico de prueba está definido por

$$BP = \text{Suma de cuadrados de la } \frac{\text{regresion}}{2}$$

$$BP > \chi_{p,\alpha}^2$$

En caso tal que se rechace la hipótesis, se deberá corregir la heterocedasticidad a través de la matriz de White, la cual se define como:

$$\text{Var}_w(\widehat{\beta}_{MCO}) = (X'X)^{-1} \left( \sum_{t=1}^T e_t^2 X_t X_t' \right) (X'X)^{-1}$$

En términos del algoritmo en R, se debe utilizar la función `bptest` del paquete `lmtest`, la cual cuenta con tan solo un argumento que es la variable donde se guardó el modelo de regresión. En este sentido, el código a emplear es el siguiente:

```
bptest(Modelo_Econometrico_1)
```

Si rechaza la prueba, significa que el modelo tiene problemas de heterocedasticidad, los cuales se deben corregir por medio de la matriz de wald. En este caso, la función a emplear es `coeftest`, el cual tiene 3 argumentos a saber: el modelo econométrico, los grados de libertad del modelo y la matriz de wald a emplear (se recomienda el HC4).

```
coeftest(Modelo_Econometrico_1,df=Modelo_Econometrico_1$df,vcov=vcovHC,type='HC4')
```

- Multicolinealidad:

Para la multicolinealidad, se calculó el valor inflado de la varianza (FIV), para lo cual se deben estimar  $n$  modelos de regresión según el número de variables explicativas del modelo, donde la variable dependiente para estas regresiones será cada variable independiente del modelo inicial. Luego se calcula para cada regresión su respectivo  $R^2$ , para posteriormente aplicar la fórmula del valor inflado de la varianza que está dada por:

$$FIV_i = \frac{1}{1 - R_i^2}$$

Finalmente, se analiza el valor de la multicolinealidad por cada variable. Si el resultado es mayor a 10, implica que hay multicolinealidad en el modelo, por tanto, se deben aplicar metodologías de regresión Ridge y Lasso.

Para la regresión Ridge la estimación de las betas está dada por:

$$\hat{\beta}_{ridge} = (X'X + kI_p)^{-1}X'y$$

Donde

$$I_p = \text{matriz de identidad } p \times p$$

Mientras que para la regresión Lasso la estimación de las betas está dada por:

$$\min_{\beta \in R^p} \left( \frac{1}{N} \|y - X\beta\|_2^2 \right) \text{ Sujeto a } \|\beta\|_1 \leq t$$

Con respecto al algoritmo en R, se debe utilizar la función `vif` del paquete `regclass`, el cual solo cuenta con un argumento y es el modelo econométrico. En ese orden de ideas, el código a aplicar es como sigue:

```
vif(Modelo_Econometrico_1)
```

- Datos atípicos:

Para los datos atípicos, se recomienda el uso de la función `influence.measures` del paquete `stats`. Esta función solo recibe un argumento que es el modelo econométrico y su código en R está representado de la siguiente forma:

```
influence.measures(Modelo_Econometrico_1)  
influence.measures(Modelo_Econometrico_1)$is.inf  
influencePlot(Modelo_Econometrico_1)
```

En caso tal de detectar datos atípicos, se deben estimar regresiones robustas de Yohai y Huber, con el fin de incluir en los modelos de regresión los efectos de los datos atípicos. En este sentido, las funciones a utilizar son `robustRegBS` y `robustRegH` del paquete `robustreg`, cuyos argumentos son: la especificación del modelo; la matriz donde se encuentran los datos del modelo; el `tune` o calibrador, cuya recomendación es mantener en 1.345; las iteraciones máximas del modelo y si se desea visualizar la tabla Anova.

```
## Regresión Yohai
```

```
Modelo_1 <- robustRegBS(log(Precio_Interno)~  
log(Precio_Internacional)+log(TRM)+log(Importaciones)+log(exportaciones)+log(producción)  
+log(inventario), data = Base_General, tune = 4.685, max.it = 1000, tol = 1e-5, anova.table = T)
```

```
## Regresión Huber
```

```
Modelo_2 <- robustRegH(log(Precio_Interno)~  
log(Precio_Internacional)+log(TRM)+log(Importaciones)+log(exportaciones)+log(producción)  
+log(inventario), data = Base_General, tune = 1.345, max.it = 1000, anova.table = T)
```

Con todas estas validaciones se tendrá como resultado los coeficientes que explican el impacto que tendrán las variables explicativas con respecto al comportamiento de los precios, teniendo en cuenta el concepto de *ceteris paribus*.

### **6.1.2 Riesgo Financiero**



### ***6.1.2.1 Cálculos de amenaza***

Para el riesgo financiero, las variables de análisis asociadas a su volatilidad son: Incremento en el precio del producto que se está comercializando, incremento en los precios de los insumos, incremento en los precios de la mano de obra, incremento de las tasas de interés y devaluación o revaluación de las tasas de cambio.

Sobre todas las variables mencionadas, se aplicarán las metodologías de amenaza descritas en la página 52 del capítulo de riesgo de mercado. Respecto a los análisis en el contexto de monitoreo y pronóstico, solo se aplicarán las metodologías descritas en el literal mencionado sobre las variables de precios de insumos y mano de obra; esto se debe a que las variables macroeconómicas ya cuentan con análisis de este nivel por parte de entidades especializadas como la Superintendencia Financiera y el Banco de la República.

Ahora bien, sobre las consideraciones mencionadas anteriormente, se procede a describir la metodología a utilizar para elegir los precios de los insumos y mano de obra que se van a analizar por cada sistema productivo. Para esto, el proceso metodológico a seguir es el siguiente:

#### **Paso 1: Construcción de los costos de producción**

En primer lugar, se deben construir los costos de producción de cada uno de los sistemas de producción. Para ello, se deberá multiplicar la matriz de coeficientes técnicos por los precios de los factores de producción. Con respecto al ejercicio de Pareto, no es necesario multiplicarlo por el total de hectáreas sembradas debido a que, en el momento de calcular los pesos, el número de hectáreas no tendrá algún efecto sobre la participación de los costos de insumos y mano de obra sobre el total.

#### **Paso 2: Cálculo de los pesos relativos**

En segundo lugar, se calcula la participación que tiene cada ítem de insumos y mano de obra respecto al total de costos. Para lo aquí mencionado la fórmula a utilizar es la que se describe a continuación:

$$Participación = \frac{Costo\ del\ insumo\ o\ mano\ de\ obra}{\sum Costos\ de\ insumos\ y\ mano\ de\ obra}$$

### Paso 3: Construcción del Pareto

A continuación se construye el Pareto de los ítems de insumo y mano de obra que serán seleccionados para el análisis de monitoreo y pronóstico. Luego de tener las participaciones de cada uno de los costos para cada uno de los sistemas productivos, se ordenan los mismos de manera descendente (de mayor a menor) y se calcula la participación acumulada.

Finalmente, todos los insumos que se encuentren dentro de la participación del 80% sobre el total de costos, serán aquellos productos a los cuales se les aplicará la metodología de monitoreo y pronóstico mencionados anteriormente.

#### *6.1.2.2 Cálculo de vulnerabilidad*

##### *6.1.2.2.1 Simulación Montecarlo para medir el grado de sensibilidad de la rentabilidad de los sistemas productivos agropecuarios*

###### **6.1.2.2.1.1 Simulación Montecarlo**

Inicialmente la Plataforma Agro ya cuenta con la proyección del flujo de caja libre y flujo de caja del productor, tal y como se observa en la Tabla 7, la cual es la base para el cálculo de la rentabilidad del productor y la variable de análisis para la simulación Montecarlo.

**Tabla 7** Estructura del flujo de caja libre y del productor

<b>Componente</b>	<b>Formula</b>
Ingresos operacionales	Precio x cantidad
Costos de producción	[Precios x Insumos] + [Precios x Mano de Obra] + [Precios x Indirectos]
<b>UTILIDAD BRUTA</b>	<b>Ingresos operacionales - Costos de producción</b>
Gastos operacionales	Sumatoria de gastos por concepto de servicios públicos, administración, viáticos, comunicaciones, entre otros.
<b>EBITDA</b>	<b>Utilidad bruta - Gastos operacionales</b>
Gasto de depreciaciones	Descrita en el componente de entrada del modelo financiero
<b>UTILIDAD OPERATIVA (EBIT)</b>	<b>EBITDA - Gasto de depreciaciones</b>
Gasto financiero	Sumatoria de los intereses de la deuda financiera.

<b>UTILIDAD ANTES DE IMPUESTOS</b>	<b>Utilidad operativa - Gasto financiero</b>
Impuestos	Utilidad antes de impuesto * 30%
<b>UTILIDAD NETA</b>	<b>Utilidad antes de impuestos - impuestos</b>
Gasto de depreciaciones	Descrita en el componente de entrada del modelo financiero
<b>GENERACIÓN INTERNA DE FONDOS (GIF)</b>	<b>Utilidad neta + Gasto de depreciaciones</b>
Gasto financiero	Sumatoria de los intereses de la deuda financiera.
<b>FLUJO DE CAJA BRUTO</b>	<b>Generación interna de fondos + Gasto financiero</b>
Variación del capital de trabajo neto operativo	KTNO período actual - KTNO período anterior
<b>EFFECTIVO GENERADO POR LA OPERACIÓN (EGO)</b>	<b>Flujo de caja bruto – Variación del capital de trabajo neto operativo</b>
Adquisición de propiedad planta y equipo	Descrita en el componente de entrada del modelo financiero
<b>FLUJO DE CAJA LIBRE</b>	<b>Efectivo generado por la operación – Adquisición de propiedad planta y equipo</b>
Crédito	Valor financiado
Servicio a la deuda	Sumatoria de los intereses más el abono a capital
<b>FLUJO DE CAJA DEL PRODUCTOR</b>	Flujo de caja libre + Crédito – Servicio a la deuda

Nota: Elaboración propia.

El indicador que se utiliza para la medición de la rentabilidad de los productores es la tasa interna de retorno (TIR). La cual se estima como:

$$TIR \rightarrow 0 = I_0 - \sum_{t=1}^n \frac{FC_t}{(1 + TIR)^t}$$

Donde

$I_0$  = Inversión inicial.

$t$  = Período en el cual se presenta cada flujo de caja para el productor

$FC_t$  = Flujo de caja del cultivo en el periodo  $t$ .

Luego de estas definiciones, se pasa a la determinación de las funciones de distribución de las variables críticas que influyen en la rentabilidad. Para ello se aplican las pruebas de Kolgomorov y Smirnov, Anderson y Darling, y Chi 2. En términos del algoritmo en R, se utiliza el mencionado en la página 57 del capítulo de riesgo de mercado.

Posterior a esto, se inicia con la simulación Montecarlo, para lo cual, se generan n valores aleatorios de las variables: precio del producto que se está comercializando, precios de los insumos, precios de la mano de obra, tasas de interés y devaluación o revaluación de las tasas de cambio con base en su función de distribución de probabilidad, las cuales se guardan en una matriz. Cada valor aleatorio se ingresa al modelo financiero y se capturan los resultados de ingresos netos y rentabilidad, los cuales se almacenarán en un vector. Luego, se concatena la matriz de valores aleatorios de las variables con el vector de resultados.

Finalmente, se procede al análisis de resultados. Para esto, se estimarán los modelos de regresión lineal con base en la matriz obtenida en el proceso descrito anteriormente. La regresión propuesta se debe realizar para la rentabilidad:

*log(Rentabilidad)*

$$\begin{aligned} &= \alpha + \beta_1 \log(\text{precio factor insumo } n) \\ &+ \beta_2 \log(\text{precio factor mano de obra } n) + \beta_3 \log(\text{precio del producto}) \\ &+ \beta_4 \log(\text{Tasa de interés}) + \beta_5 \log(\text{Tasa representativa del mercado}) + \varepsilon \end{aligned}$$

Donde los precios de los insumos y mano de obra pueden ser varios, y todos éstos se incluirán en el modelo propuesto.

Cada uno de los  $\beta$  estimados son el efecto o impacto que tienen cada una de las variables de todos los riesgos sobre la rentabilidad. Es importante resaltar que se utiliza el logaritmo sobre cada una de las variables para analizar la sensibilidad en términos porcentuales.

## 6.2 Aplicación de las metodologías de evaluación de riesgo de mercado y financiero

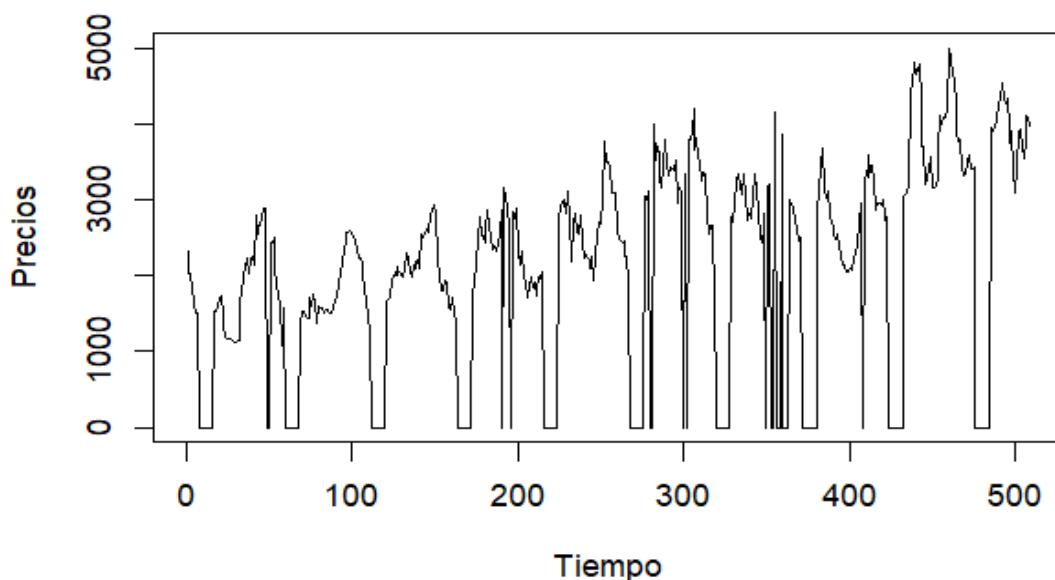
Presentados los algoritmos para la aplicación de las metodologías de medición del riesgo de mercado y del riesgo financiero en el numeral 6.1, se procede a presentar la aplicación de la misma para un productor de aguacate hass en Antioquia. Adicionalmente, se presentarán los análisis de los resultados obtenidos con el fin de ver la utilidad de las metodologías seleccionadas.

### 6.2.1 Aplicación de metodología para missing data en la serie de precios de aguacate hass

Las bases de datos utilizadas para la serie de precios de aguacate hass es la información que provee el DANE a través del Sistema de información de precios para el sector agropecuario (SIPSA). En vista que SIPSA presenta información para varias centrales de abastos, se seleccionó la central mayorista de Antioquia como referencia para el análisis.

El primer ejercicio que se debe realizar es completar la base de datos en caso de que falten datos. Como se observa en el Gráfico 9, la serie de precios presenta varios missing data que se deben de imputar antes de aplicar las metodologías.

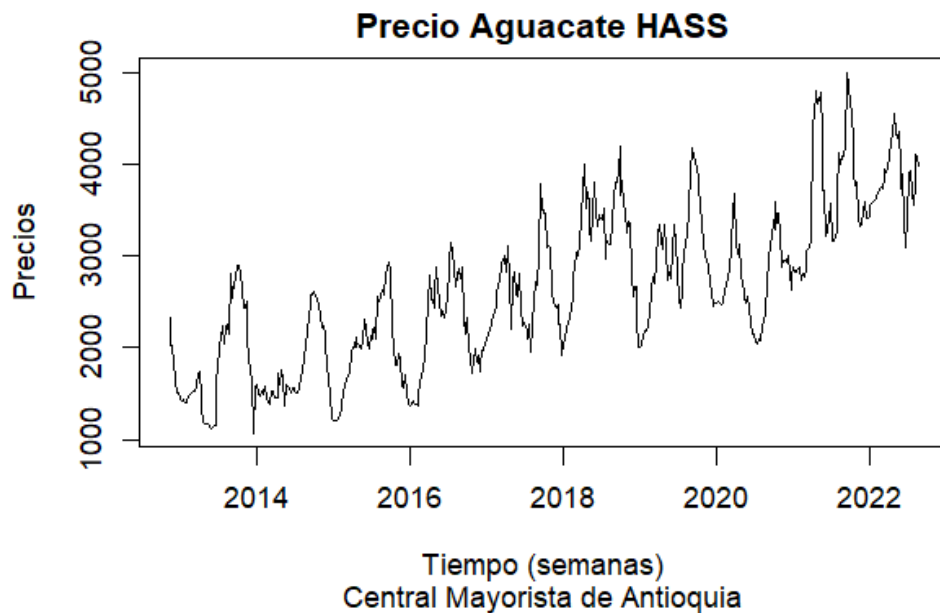
Gráfico 9 Series de precios Aguacate hass central mayorista de Antioquia 2012-2022



Nota: Elaboración propia a partir de los datos de SIPSA-DANE

Para la imputación de datos se aplicó la metodología descrita en el numeral 6.1.1.1.1 en el apartado de consideraciones, obteniendo como resultado el Gráfico 10. En el mismo se observa que el precio del aguacate hass presenta una estacionalidad y una deriva positiva en lo corrido del tiempo.

Gráfico 10 Series de precios imputados para el Aguacate hass central mayorista de Antioquia 2012-2022



Nota: Elaboración propia a partir de los datos de SIPSA-DANE

Con la información completada, se procede a aplicar las metodologías de medición de riesgo de mercado y financiero en los diferentes contextos temporales: histórico, monitoreo y predicción.

## 6.2.2 *Riesgo de Mercado*

### 6.2.2.1 *Histórico*

#### 6.2.2.1.1 *Indicador de Volatilidad*

En primer lugar, se calculó el indicador de volatilidad histórica para la serie de los rendimientos de los precios del aguacate hass ya imputados. Este ejercicio se hace sobre la desviación estándar de los rendimientos debido a que permite comparar los resultados entre diferentes centrales.

El resultado obtenido fue de 0.26, es decir, que la volatilidad de los rendimientos de la serie de precios del aguacate hass oscila cerca del 26%. La pregunta que surge es ¿Esta volatilidad es baja, media o alta? Para dar respuesta a esta pregunta, se realizó el mismo ejercicio para el precio del aguacate hass en todas las centrales que reporta el SIPSA y se calcularon 3 intervalos de volatilidad según los resultados obtenidos para todas las centrales. Estos intervalos se presentan en la Figura 6.

Figura 6 Intervalos para la medición de volatilidad de los rendimientos de los precios del aguacate hass

	Limite inferior	limite superior
Volatilidad baja	0.0446205	0.1408297
Volatilidad media	0.1408297	0.2370388
Volatilidad alta	0.2370388	0.3332480

Nota: Elaboración propia a partir de los datos de SIPSA-DANE

Así las cosas, se concluye que la volatilidad de los rendimientos del precio del aguacate hass de la central mayorista de Antioquia con respecto a los rendimientos de los precios de las demás centrales mayoristas es alta.

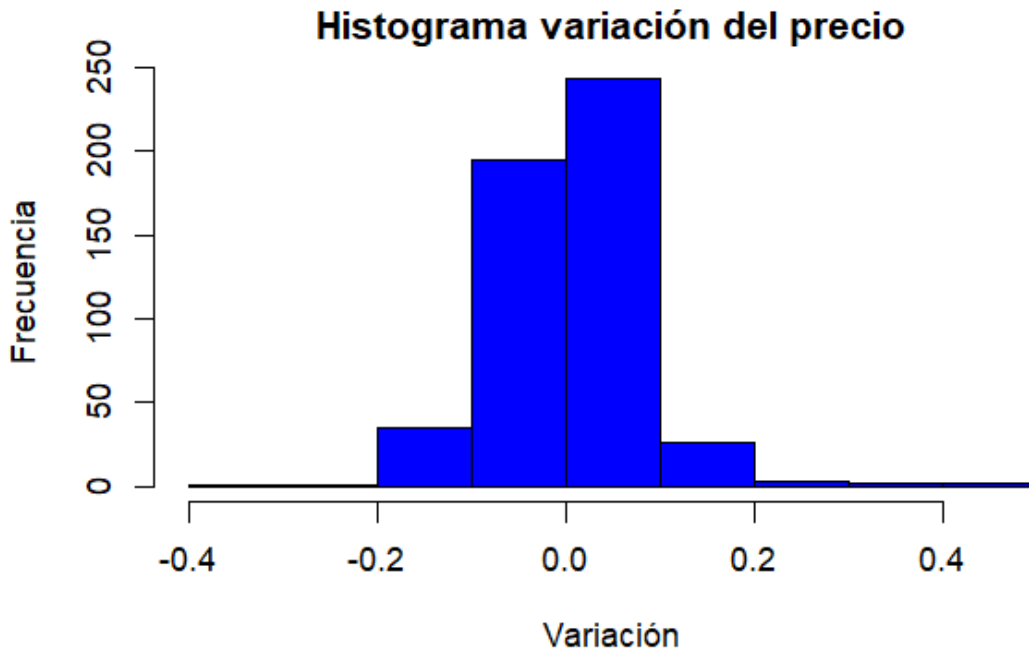
#### ***6.2.2.1.2 VaR de las fluctuaciones de los precios***

A continuación se procede a complementar este análisis con las dos metodologías restantes: VaR sobre las fluctuaciones de los precios y el índice Beta. Con respecto al primero, se realizó el histograma de los rendimientos de los precios analizados, obteniendo como resultado el Gráfico 11.

Para poder obtener el VaR, es importante conocer qué distribución siguen estos rendimientos, y para ello se aplicaron las pruebas de bondad de ajuste Chi2, Kolgomorv-smirnov y Andersen-

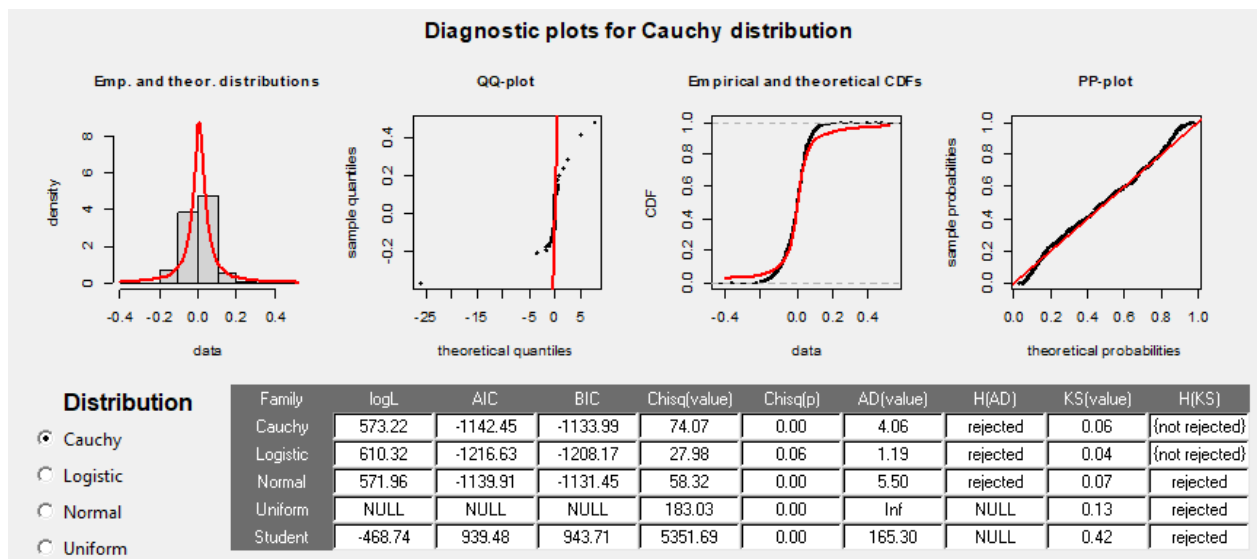
Darling. Como resultado se obtuvo que los rendimientos siguen una distribución Cauchy, cuyo resultado se puede observar en la Figura 7.

Gráfico 11 Histograma de los rendimientos de los precios del aguacate hass en la central mayorista de Antioquia



Nota: Elaboración propia a partir de los datos de SIPSA-DANE

Figura 7 Aplicación de pruebas de bondad de ajuste para los rendimientos de los precios del aguacate hass en la central mayorista de Antioquia





Nota: Elaboración propia.

Los estimadores obtenidos para esta distribución fueron: una locación de 0.0067 y una escala de 0.0362. Con ellos ya se puede calcular el VaR donde el resultado es el intervalo de confianza en que puede llegar a oscilar la volatilidad de los rendimientos. En este sentido, los resultados obtenidos están en la Figura 8.

Figura 8 VaR en riesgo de las fluctuaciones de los precios del aguacate hass en la central mayorista de Antioquia

	Intervalo 2,5%	Intervalo 97,5%
95% de confianza	-0.453705	0.4671293

Nota: Elaboración propia.

De la Figura 8 se puede concluir que las fluctuaciones de los precios del aguacate hass son altas, lo cual, va en la misma dirección de la conclusión del indicador de volatilidad histórica.

### **6.2.2.1.3 Índice Beta**

Finalmente, se calculó el índice Beta para esta serie de precios, obteniendo como resultado un valor de 1.10. Lo cual quiere decir que los precios del aguacate hass de la central mayorista de Antioquia son un 10% más volátiles que los precios promedios de todos los mercados.

Como conclusión general de este análisis histórico, podemos decir que los precios de la central mayorista de Antioquia son altamente volátiles en comparación con las demás centrales mayoristas. Para una entidad financiera es importante este dato para determinar la sensibilidad del flujo de caja para soportar la deuda financiera de un productor de aguacate hass. Mientras que, para el productor, ayuda a considerar el momento en que debe vender, dado que la volatilidad de los precios tiene impactos fuerte sobre los ingresos.

Las consideraciones que deben hacer las entidades financieras y los productores, sobre la base de los resultados obtenidos en el contexto histórico, se calculan y se miden en el riesgo financiero, el cual, se desarrollará en el numeral 6.2.3.

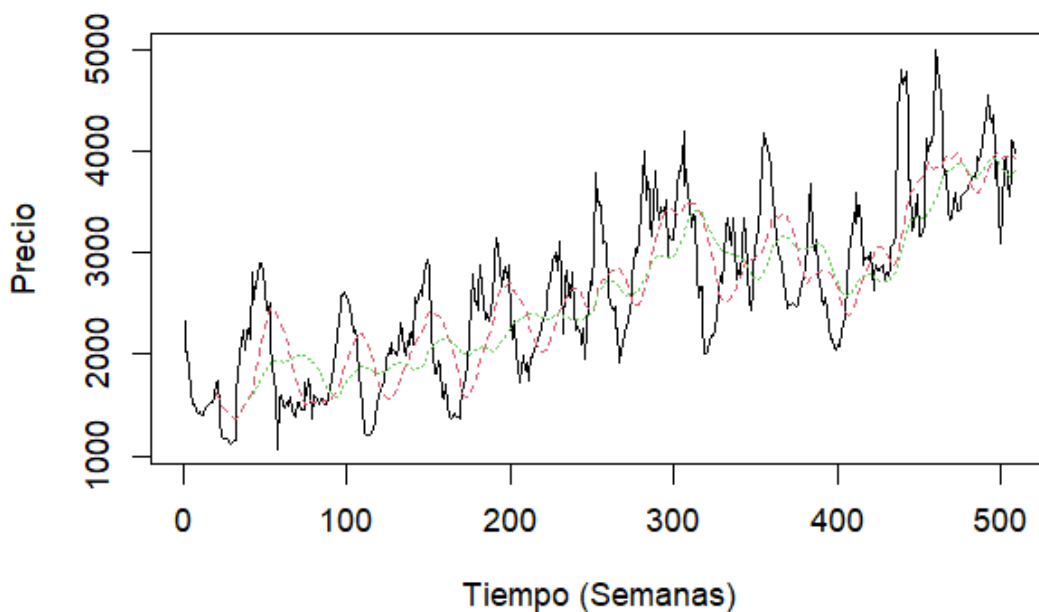
### 6.2.2.2 Monitoreo

A continuación se desarrollan las metodologías para la medición del riesgo de mercado en el contexto del monitoreo. Es importante recordar que el monitoreo permite realizar un seguimiento en tiempo real del comportamiento del precio y ver si éste podría llegar a tocar precios techos o base según las dinámicas del mercado.

#### 6.2.2.2.1 Medias Móviles

Al aplicar la metodología del numeral 6.1.1.1.2.2, se obtiene como resultado el Gráfico 12 en el cual se puede observar que el precio está por encima de la media móvil de largo plazo (la verde), y esta media móvil tiene una tendencia a la baja. Lo anterior significa que el precio del aguacate tendrá una tendencia a la baja en las semanas que siguen.

Gráfico 12 Medias móviles para los precios del aguacate hass en la central mayorista de Antioquia

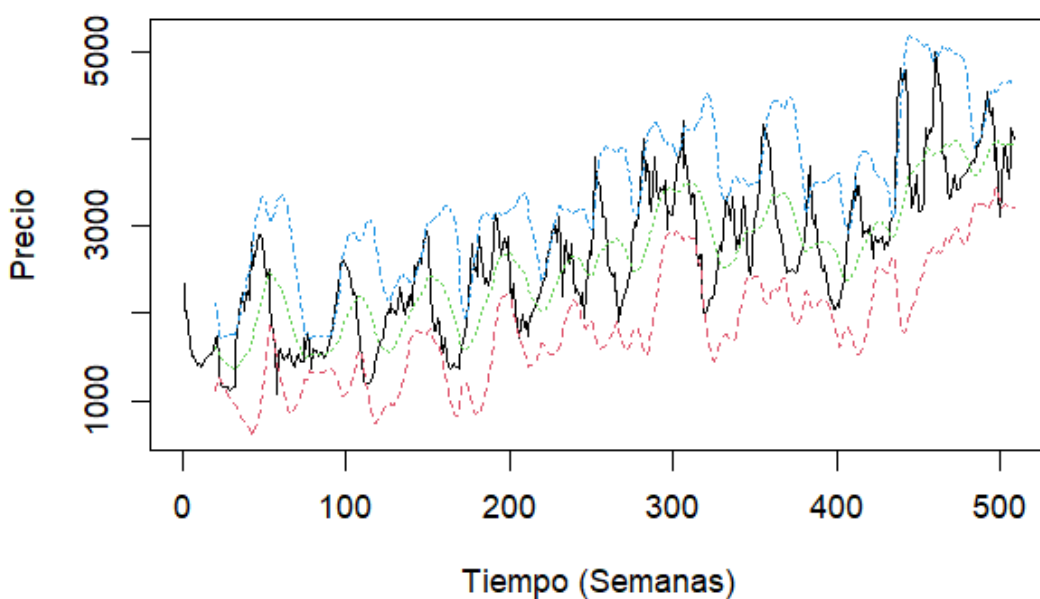


Nota: Elaboración propia a partir de los datos de SIPSA-DANE

#### 6.2.2.2.2 *Bandas de Bollinger*

Con respecto a las bandas de bollinger, se observa que en la última semana de análisis el precio se encuentra dentro del intervalo, lo cual pone en evidencia que no se han presentado situaciones fuertes en el mercado para que el precio presente una tendencia clara al alza o a la baja, es decir, que el precio se ha comportado como se esperaría en ese periodo de tiempo.

Gráfico 13 Bandas de Bollinger para los precios del aguacate hass en la central mayorista de Antioquia

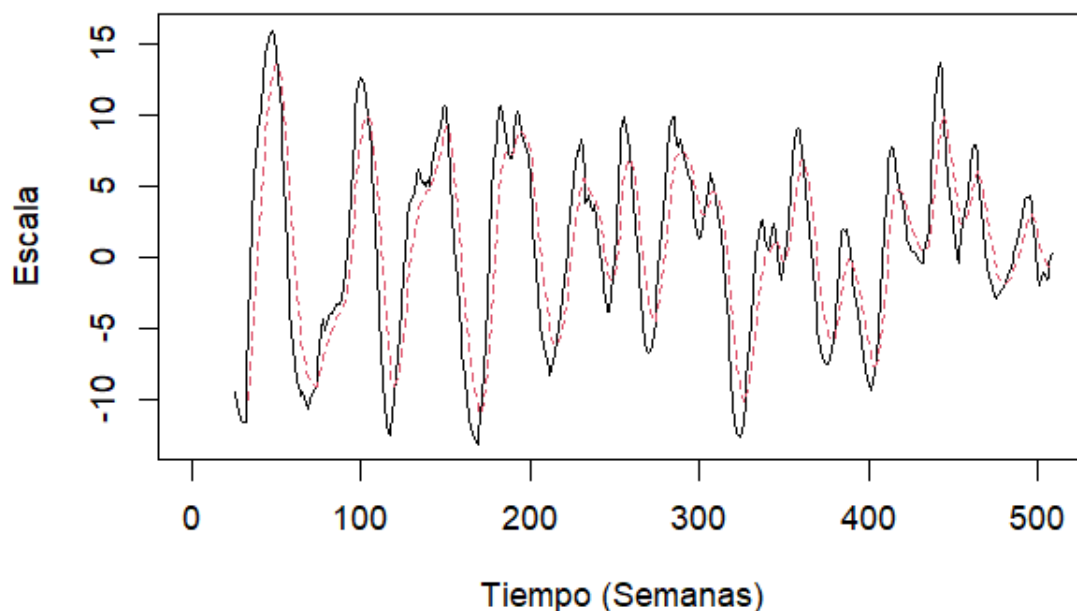


Nota: Elaboración propia a partir de los datos de SIPSA-DANE

### 6.2.2.2.3 *Indicador de Media Móvil de Convergencia/Divergencia (MACD)*

Al aplicar la metodología descrita en el numeral 6.1.1.1.2.3, se observa que la línea MACD (punteada de rojo) va a superar la línea señal (línea negra) de arriba hacia abajo. Esto muestra entonces que el precio va a tomar una tendencia a la baja en las próximas semanas.

Gráfico 14 MACD para los precios del aguacate hass en la central mayorista de Antioquia

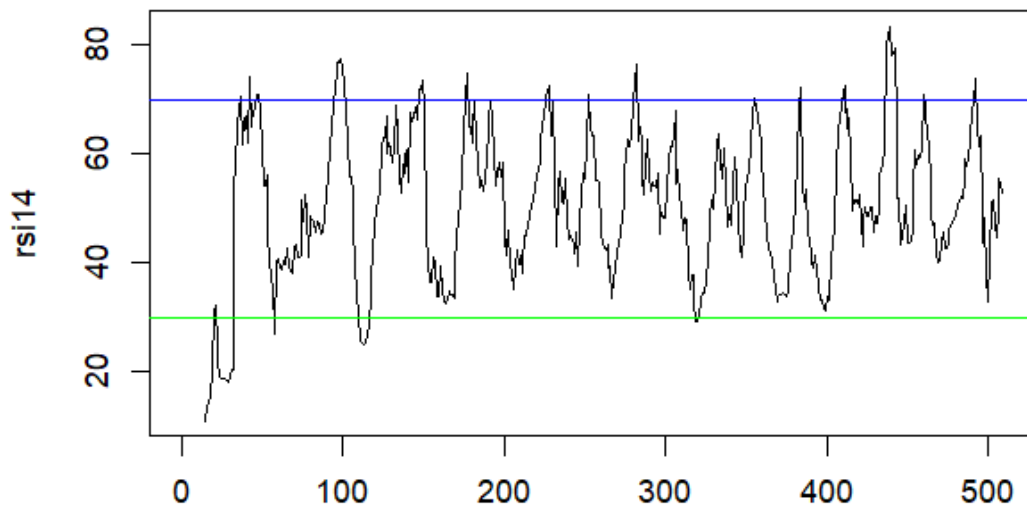


Nota: Elaboración propia a partir de los datos de SIPSA-DANE

### 6.2.2.2.4 *Índice de Fuerza Relativa (RSI)*

Finalmente, con respecto al RSI, se observa en el Gráfico 15 que el precio en las últimas semanas ha presentado una tendencia a la baja. No obstante, analizando la última semana el RSI no da una señal clara de la tendencia del comportamiento del precio.

Gráfico 15 RSI para los precios del aguacate hass en la central mayorista de Antioquia



Nota: Elaboración propia a partir de los datos de SIPSA-DANE.

Concluyendo sobre los cuatro indicadores de monitoreo, se observa que el precio en las semanas posteriores al período analizado tendrá una tendencia hacia la baja. Lo cual, para las entidades financieras es una señal para revisar si el flujo de caja del cliente agropecuario de aguacate soportará esta caída. Mientras que, para los productores, sirve para analizar si existe la posibilidad de vender el producto antes de que comience la etapa de tendencia a la baja del precio.

### 6.2.2.3 Predicción

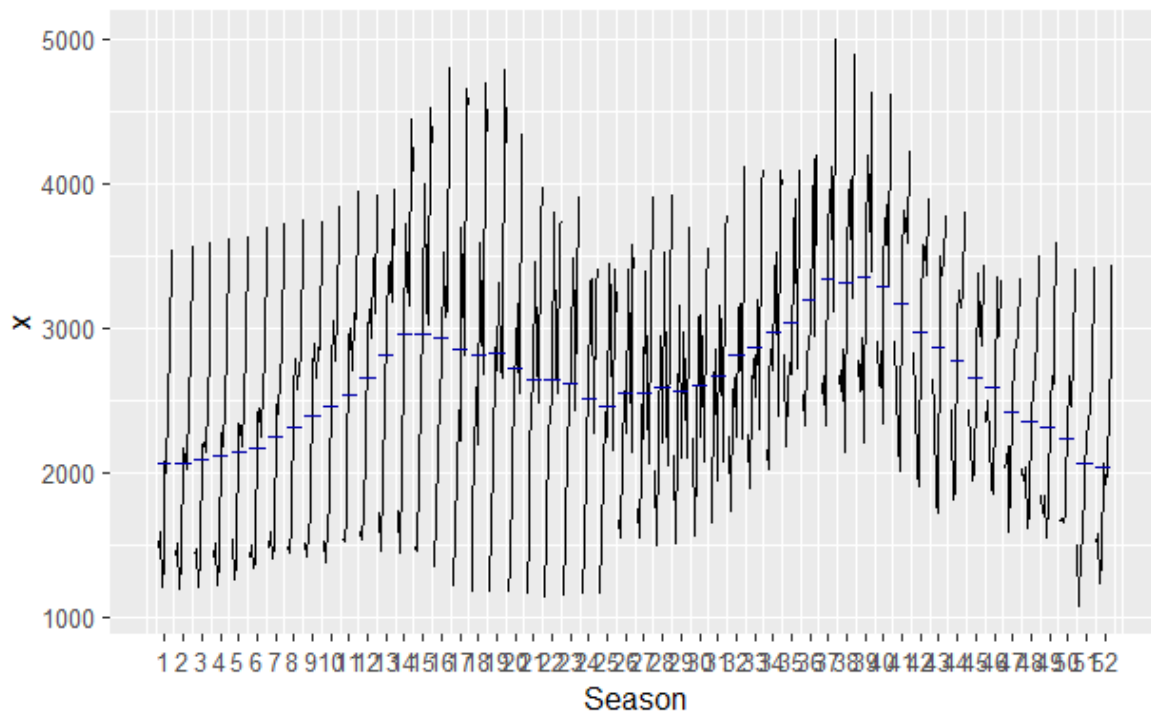
A continuación, se desarrollan las metodologías seleccionadas para el contexto de la predicción. Los resultados que se obtendrán son el posible comportamiento futuro del precio del aguacate hass, desde su media y volatilidad. Es por ello que se tienen dos metodologías de predicción sobre la media (Box-Jenkins y Redes neuronales LSTM) y sobre la volatilidad (ARCH-GARCH)

#### 6.2.2.3.1 Modelo Box y Jenkins

Partiendo del comportamiento expresado en el Gráfico 10, se procedió a graficar la serie de tiempo para entender cómo se comporta cada semana desde el año 2012 hasta septiembre de 2022 y si la serie presenta algún componente de estacionalidad que se deba tener en cuenta para la modelación.

Así, se obtuvo como resultado el Gráfico 16 en el cual se puede observar que en efecto existe un patrón estacional que se debe de tener en cuenta para la modelación. Por otro lado, se observa que no es una serie estacionaria en media. En este sentido, el grafico nos muestra que debemos aplicar pruebas de raíces unitarias estacionarias y estacionales para tener en cuenta en la modelación box y Jenkins.

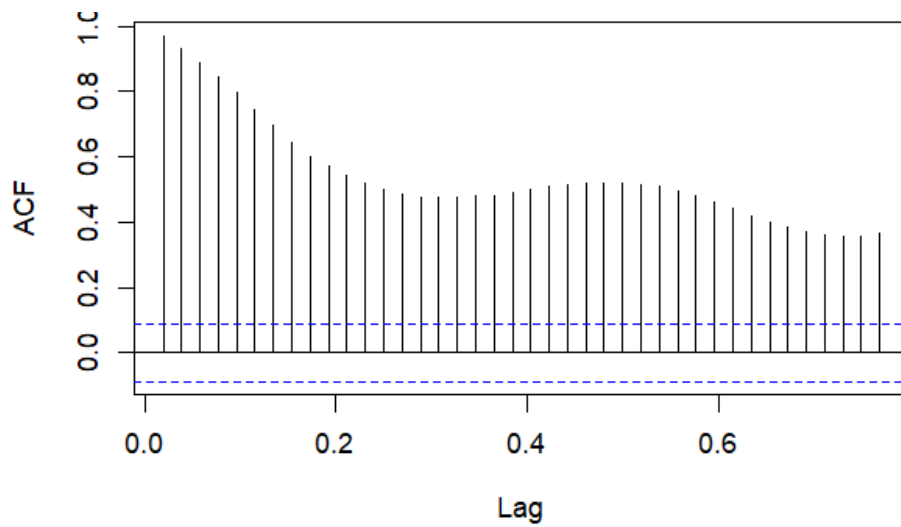
Gráfico 16 Monthplot para los precios del aguacate hass en la central mayorista de Antioquia



Nota: Elaboración propia a partir de los datos de SIPSA-DANE.

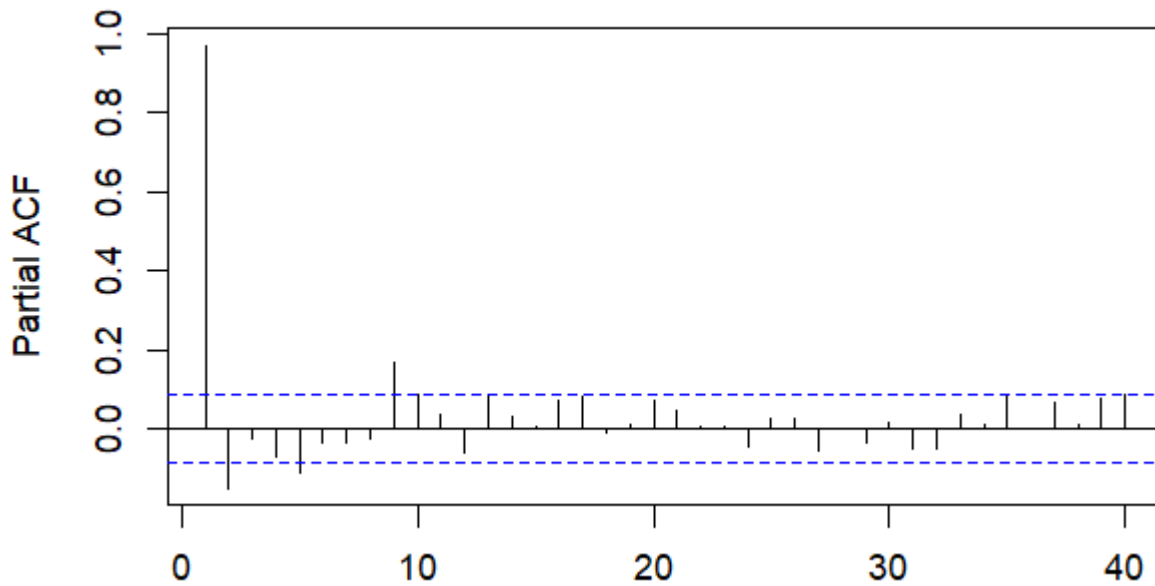
Al hacerlos gráficos ACF y PACF de la serie, los cuales se encuentran representados en los gráficos 17 y 18, se ve que la ACF decae lentamente y se observa un posible patrón estacional en ella. Por tanto, se deberán realizar las pruebas de raíces unitarias tanto en el componente estacionario como en el componente estacional.

Gráfico 17 ACF para los precios del aguacate hass en la central mayorista de Antioquia



Nota: Elaboración propia.

Gráfico 18 PACF para los precios del aguacate hass en la central mayorista de Antioquia



Nota: Elaboración propia.

Al realizar las pruebas de raíces unitarias Dickey y Fuller aumentado en el componente estacionario que se encuentra en la

Figura 9, se concluye que al menos existe una raíz unitaria en el componente estacionario. Por tal razón, para efecto de la aplicación del algoritmo descrito en el numeral 6.1.1.1.3.1. la diferenciación representada con la letra d, tomará el valor de 1.

Figura 9 Pruebas de raíces unitarias estacionarias

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression drift

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-574.73  -99.97   -7.69   98.33 1201.17

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  65.8460966  33.8600793    1.945 0.052410 .
z.lag.1      -0.0223174   0.0124588   -1.791 0.073887 .
z.diff.lag1   0.0922641   0.0464166    1.988 0.047418 *
z.diff.lag2  -0.0260428   0.0464615   -0.561 0.575387
z.diff.lag3   0.0424761   0.0464261    0.915 0.360701
z.diff.lag4   0.0946022   0.0462896    2.044 0.041538 *
z.diff.lag5   0.0113184   0.0464243    0.244 0.807489
z.diff.lag6  -0.0007324   0.0465309   -0.016 0.987448
z.diff.lag7   0.0094433   0.0463328    0.204 0.838587
z.diff.lag8  -0.1681794   0.0462556   -3.636 0.000308 ***
z.diff.lag9  -0.0714839   0.0461381   -1.549 0.121969
z.diff.lag10 -0.0713531   0.0463558   -1.539 0.124414
z.diff.lag11  0.0469833   0.0467977    1.004 0.315910
z.diff.lag12 -0.0612945   0.0468759   -1.308 0.191649
z.diff.lag13 -0.0595480   0.0469467   -1.268 0.205273
z.diff.lag14 -0.0344590   0.0472303   -0.730 0.466000
z.diff.lag15 -0.0669271   0.0472301   -1.417 0.157130
z.diff.lag16 -0.0957929   0.0472267   -2.028 0.043084 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 194.7 on 472 degrees of freedom
Multiple R-squared:  0.1168,    Adjusted R-squared:  0.08501
F-statistic: 3.672 on 17 and 472 DF,  p-value: 1.245e-06

Value of test-statistic is: -1.7913 1.9451

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```



Nota: Elaboración propia.

Con respecto a las pruebas de raíces unitarias para el componente estacional, se utilizó el test de Hegy, el cual está representado en la Figura 10. De estas pruebas se deduce que la serie tiene al menos una raíz unitaria del componente estacional. Por esta razón en el momento de aplicar el algoritmo para poder determinar el mejor orden del modelo, la diferenciación en el componente estacional, representado por la letra D, tomará el valor de 1.

Figura 10 Pruebas de raíces unitarias estacionales

```
HEGY test for unit roots
data: db_serie

      statistic p-value
t_1      -2.1629  0.3848
t_2      -2.7888  0.0028 **
F_3:4      4.2925  0.0114 *
F_5:6       1.874  0.1481
F_7:8       6.0837  0.0017 **
F_9:10      6.4375  0.0012 **
F_11:12     5.2778  0.004  **
F_13:14    10.4149  0  ***
F_15:16    11.632  0  ***
F_17:18    13.5821  0  ***
F_19:20     6.1343  0.0016 **
F_21:22     4.9877  0.0054 **
F_23:24     7.661  3e-04 ***
F_25:26     5.4865  0.0032 **
F_27:28    10.7398  0  ***
F_29:30    10.9989  0  ***
F_31:32    10.2576  0  ***
F_33:34    15.602  0  ***
F_35:36     8.7035  1e-04 ***
F_37:38     6.8814  7e-04 ***
F_39:40    11.4902  0  ***
F_41:42     8.3165  2e-04 ***
F_43:44     9.2187  0  ***
F_45:46     8.2649  2e-04 ***
F_47:48    10.2833  0  ***
F_49:50     8.1779  2e-04 ***
F_51:52     7.5604  4e-04 ***
F_2:52     75.7755  0  ***
F_1:52     74.3313  0  ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Deterministic terms: constant + trend
Lag selection criterion and order: AIC, 0
P-values: based on response surface regressions
```

Nota: Elaboración propia.

Ya realizadas las pruebas de raíces unitarias para ambos componentes (estacionario y estacional), se procede a determinar el mejor orden del modelo y a estimar los coeficientes del mismo. El

algoritmo presentado en el numeral 6.1.1.1.3.1 tiene en cuenta los diferentes datos atípicos que se pueden llegar a presentar en una serie. Sobre esta consideración, describimos en la Figura 11 los resultados obtenidos y allí se observa que el orden del modelo es un SARIMA (0,1,1) (0,1,1) 52 con los coeficientes ya calculados. Adicionalmente, se obtuvo que la serie de tiempo analizada tenía 3 tipos de datos atípicos: uno aditivo en la semana 58, un cambio temporal en la semana 460 y un cambio de nivel en la semana 437. En estos datos atípicos, ya el modelo tiene en cuenta el coeficiente al momento de utilizarlo para la predicción.

Figura 11 Modelo resumido para los precios del aguacate hass de la central mayorista Antioquia

```
Series: train
Regression with ARIMA(0,1,1)(0,1,1)[52] errors

Coefficients:
      ma1      sma1      LS 437      AO 58      TC 460
      0.0159 -0.9484 1250.0613 -542.5907 395.0810
s.e. 0.0540 0.2336 184.1436 128.5592 129.7501

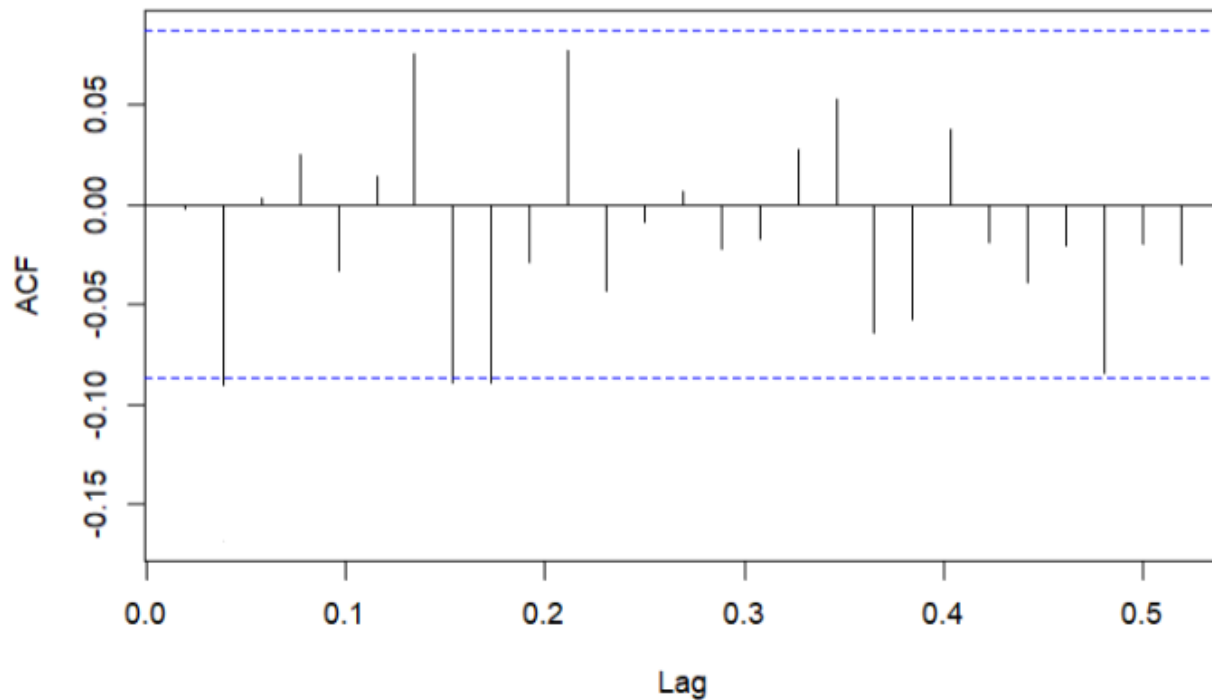
sigma^2 = 32034: log likelihood = -2978.58
AIC=5969.16 AICc=5969.35 BIC=5993.73

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE
Training set -0.08130028 168.2132 117.9179 -0.08617672 4.453341
              MASE      ACF1
Training set 0.241039 -0.00201035
```

Nota: Elaboración propia.

Después de obtener el modelo se procede con la comprobación de los errores. Es decir, que éstos sean un ruido blanco y sigan una distribución normal. Para el primero, se realizó la prueba de Ljung y Box y se contrastó gráficamente con la ACF que se encuentra representada en la Figura 12, donde se muestra que los errores son un ruido blanco; respecto al segundo, se realizaron las pruebas de bondad de ajuste, obteniendo como resultado que los errores siguen una distribución normal.

Figura 12 ACF residuales del modelo

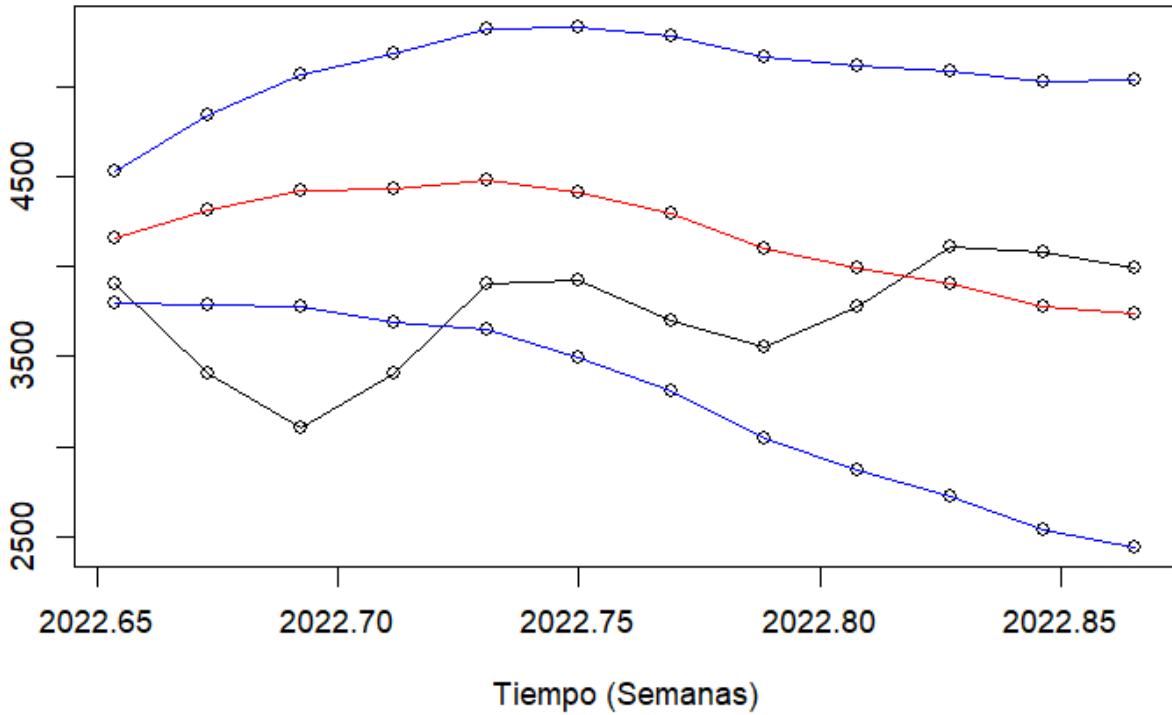


Nota: Elaboración propia.

Ya comprobados los supuestos de los modelos Box y Jenkins, se procede a medir la capacidad de predicción del modelo. Para esto se tomaron como muestra de test las 12 últimas semanas. Los resultados de la predicción se encuentran en el Gráfico 19, y muestran que las cuatro primeras semanas no se obtuvo una buena predicción dado que se salió de los intervalos de confianza. No obstante, esta predicción se fue ajustando en las semanas siguientes.

Para poder determinar la capacidad de predicción del modelo, se utilizaron las métricas MAPE, RMSE y MAE, cuyos resultados se encuentran en la Tabla 8. Esta tabla muestra que el modelo utilizado tiene una buena capacidad de predicción, en vista de que el error porcentual (MAPE) es del 5% y el MAE y RMSE que son de 203 y 237 respectivamente, son bajos en comparación con la escala de los precios del aguacate, que es de \$4.500 por kilo. Es decir, un error de \$203 en promedio con respecto a un precio promedio de \$4.500, muestra un error bajo.

Gráfico 19 Comparativo de la predicción de los precios del aguacate hass de la central mayorista de Antioquia con respecto a los precios reales con modelo Box-Jenkins



Nota: Elaboración propia.

Tabla 8 Resultados del modelo de predicción modelo Box y Jenkins

Indicador de desempeño	Resultado
MAPE	5.619%
RMSE	237.7815
MAE	203.0366

Nota: Elaboración propia.

Antes de realizar el gráfico del comportamiento de las próximas semanas entre los meses de octubre y diciembre (12 semanas), se aplicará la metodología de redes neuronales LSTM, y luego se compararán las métricas MAPE, RSME y MAE, con las métricas obtenidas en el modelo Box y Jenkins. El que presente mejores métricas, será el modelo seleccionado para realizar la predicción final.

### 6.2.2.3.2 Redes Neuronales LSTM

Como punto de partida se transformó la serie y se aplicó la metodología de ventanas de tiempo para poder aplicar la red neuronal LSTM. La transformación realizada corresponde a la transformación min-max y las ventanas de tiempo se realizaron a 12 semanas. El resultado son las series que se observan en la Tabla 9.

Tabla 9 Aplicación del método de ventanas para el entrenamiento de la red neuronal LSTM

	Scaled	shift_1	shift_2	shift_3	shift_4	shift_5	shift_6	shift_7	shift_8	shift_9	shift_10	shift_11	shift_12
0	0.322030	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	0.243243	0.322030	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	0.243243	0.243243	0.322030	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	0.196583	0.243243	0.243243	0.322030	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	0.148394	0.196583	0.243243	0.243243	0.322030	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...
481	0.674401	0.668281	0.652473	0.647629	0.640999	0.635900	0.628761	0.603009	0.597909	0.596634	0.641254	0.618307	0.578786
482	0.681285	0.674401	0.668281	0.652473	0.647629	0.640999	0.635900	0.628761	0.603009	0.597909	0.596634	0.641254	0.618307
483	0.679500	0.681285	0.674401	0.668281	0.652473	0.647629	0.640999	0.635900	0.628761	0.603009	0.597909	0.596634	0.641254
484	0.704997	0.679500	0.681285	0.674401	0.668281	0.652473	0.647629	0.640999	0.635900	0.628761	0.603009	0.597909	0.596634
485	0.734319	0.704997	0.679500	0.681285	0.674401	0.668281	0.652473	0.647629	0.640999	0.635900	0.628761	0.603009	0.597909

486 rows x 13 columns

Nota: Elaboración propia.

Luego de tener los datos preparados, se define la arquitectura de la red neuronal, a partir de definir la estructura inicial que se observa en la Figura 13. Allí se visualiza que la red solo cuenta con 3 capas, la función de pérdida definida fue el MSE y el optimizador utilizado fue el adam. Con esta estructura se obtuvo entonces que la red neuronal tendrá un total de 2.661 parámetros a estimar.

Sobre esta arquitectura inicial se compiló el modelo, el cual se utilizará para entrenarlo con los datos de entrenamiento y determinar el poder de predicción con los datos de evaluación, que, para poder comparar los resultados con el modelo de Box y Jenkins, se mantuvo las 12 últimas semanas.

Figura 13 Arquitectura modelo de redes neuronales LSTM para la serie de precios de aguacate hass de la central mayorista de Antioquia

```
[ ] K.clear_session()
    model = Sequential()

    model.add(LSTM(20, input_shape=(1, a)))
    model.add(Dropout(0.2))
    model.add(Dense(1))
    model.compile(loss='mean_squared_error', optimizer='adam')
```

```
[ ] model.summary()
```

Model: "sequential"

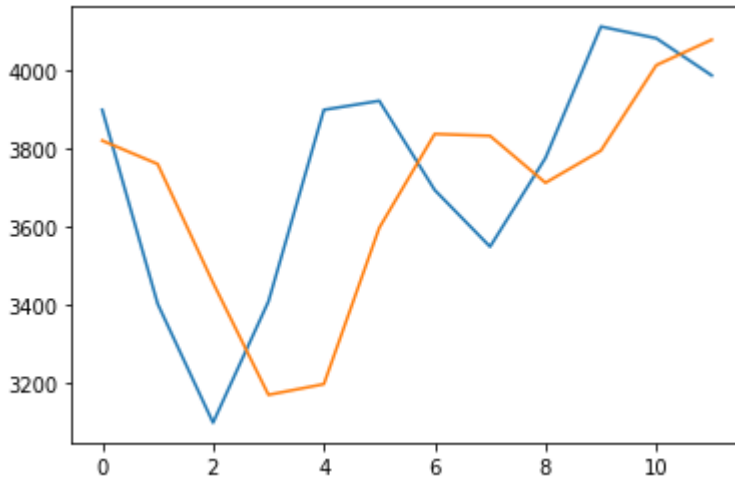
Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 20)	2640
dropout (Dropout)	(None, 20)	0
dense (Dense)	(None, 1)	21

=====  
Total params: 2,661  
Trainable params: 2,661  
Non-trainable params: 0  
=====

Nota: Elaboración propia.

Al ajustar el modelo con los datos de entrenamiento con un número de épocas igual a cien y un número de baches igual a 1, se obtuvieron los resultados que se observan en el Gráfico 20. En él se aprecia un ajuste aceptable de la predicción con respecto a los datos reales. Es de resaltar que este primer resultado, no es el resultado definitivo, debido a que no se ha calibrado la red neuronal. El objetivo es comparar el modelo no calibrado y el calibrado, para determinar cómo mejoran los resultados cuando se tunean los hiperparámetros.

Gráfico 20 Comparativo de la predicción de los precios del aguacate hass de la central mayorista de Antioquia con respecto a los precios reales con modelo Redes neuronales LSTM



Nota: Elaboración propia.

Si bien es cierto que hay un ajuste aceptable en el Gráfico 20, es importante calcular los indicadores para evaluar el poder predictivo que obtuvo este primer modelo. Dichos resultados se encuentran en la Tabla 10 y muestran buenos resultados, y que el modelo inicial tiene un poder predictivo aceptable. No obstante, estos resultados se pueden mejorar si se calibran los hiperparámetros, es decir, si se calibra el número de capas, neuronas por capa, función de activación, función de pérdida, optimizador, batches y épocas.

En este sentido, se realizará el mismo ejercicio pero con el tuneo de los hiperparámetros, para comparar los resultados y validar si hubo un efecto de mejora. Después, se comparan los resultados obtenidos de la red neuronal LSTM con los resultados del modelo box y Jenkins.

Tabla 10 Resultados del modelo de predicción modelo Redes Neuronales LSTM

Indicador de desempeño	Resultado
MAPE	6.879%
RMSE	307.5667
MAE	252.212

Nota: Elaboración propia.

Aplicando el algoritmo descrito en el numeral 6.1.1.1.3.2 para el tuneo de los hiperparámetros, se obtuvo como resultado la arquitectura de red neuronal LSTM que se describe en la Figura 14. El dicha figura, se aprecia un total de 7 capas donde la primera es de 25 neuronas, las 4 siguientes de 16, la 5 capa no tiene neuronas y la capa final de una neurona que es el resultado de usar un optimizador adam, funciones de activación lineal y un he normal como inicializador de los pesos iniciales. Sobre esta estructura se tiene entonces un total de parámetro de 5.049.

Figura 14 Arquitectura final modelo de redes neuronales LSTM para la serie de precios de aguacate hass de la central mayorista de Antioquia con calibración en los hiperparámetros

```
[ ] K.clear_session()
model = Sequential()
model.add(LSTM(25, input_shape=(1, a)))
model.add(Dense(16, activation = 'linear', kernel_initializer = 'he_normal'))
model.add(Dense(16, activation = 'linear', kernel_initializer = 'he_normal'))
model.add(Dense(16, activation = 'linear', kernel_initializer = 'he_normal'))
model.add(Dense(16, activation = 'linear', kernel_initializer = 'he_normal'))
model.add(Dropout(0.0))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
```

```
[ ] model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 25)	3800
dense (Dense)	(None, 16)	416
dense_1 (Dense)	(None, 16)	272
dense_2 (Dense)	(None, 16)	272
dense_3 (Dense)	(None, 16)	272
dropout (Dropout)	(None, 16)	0
dense_4 (Dense)	(None, 1)	17

```
=====
Total params: 5,049
Trainable params: 5,049
Non-trainable params: 0
```

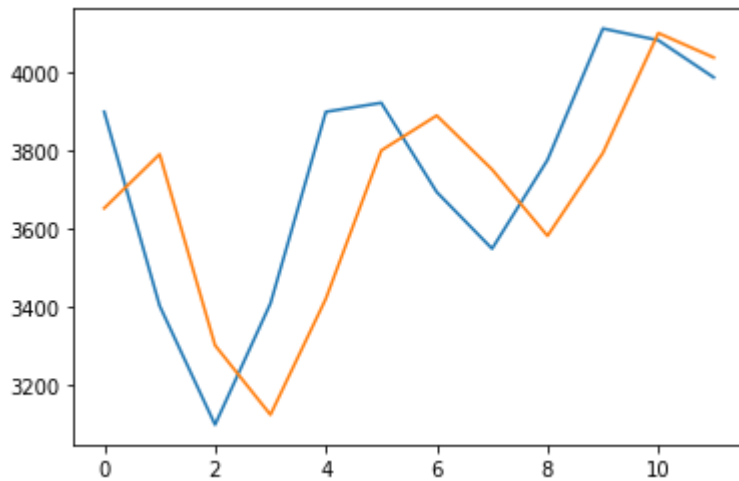
Nota: Elaboración propia.

Al ajustar el modelo con los datos de entrenamiento con un número de épocas y baches ya calibrados con valores de 100 y 10 respectivamente, se obtuvieron los resultados que se observan



en el Gráfico 21. Aquí observamos un mejor ajuste en comparación del Gráfico 20. Ahora bien, para poder corroborar esta afirmación se calcularon las métricas de evaluación de modelos que presentamos en la Tabla 11.

Gráfico 21 Comparativo de la predicción de los precios del aguacate hass de la central mayorista de Antioquia con respecto a los precios reales con modelo Redes neuronales LSTM calibrado



Nota: Elaboración propia.

Los resultados de la Tabla 11 muestran que el modelo calibrado tiene un poder de predicción aceptable para la serie de precios analizado. Ahora bien, se compararán estos resultados con los obtenidos en el modelo inicial para ver qué tanto mejoró el modelo luego de su calibración.

Tabla 11 Resultados del modelo de predicción modelo Redes Neuronales LSTM con tuneo en los hiperparámetros

Indicador de desempeño	Resultado
MAPE	6.129%
RMSE	257.807
MAE	225.089

Nota: Elaboración propia.

En la Tabla 12 se hace la comparación del modelo calibrado con respecto al modelo inicial, y se puede observar que el modelo calibrado tiene un poder de predicción mejor que el modelo no

calibrado, mostrando mejoras importantes en las 3 métricas de evaluación. Entonces se selecciona la red neuronal calibrada para compararla con los resultados obtenidos en el modelo de Box y Jenkins.

Tabla 12 Comparación de los modelos de redes neuronales LSTM con y sin tuneo en los hiperparámetros

Indicador de desempeño	Resultado modelo inicial (sin tuneo hiperparámetros)	Resultado modelo Tuneado
MAPE	6.879%	6.129%
RMSE	307.5667	257.807
MAE	252.212	225.089

Nota: Elaboración propia.

En la Tabla 13, se comparan los modelos utilizados para la predicción de los precios del aguacate hass en la central mayorista de Antioquia. Los resultados muestran que el modelo Box y Jenkins tuvo un mejor desempeño que la red neuronal. Así las cosas, se utilizará el modelo Box y Jenkins para hacer la predicción final para los meses de octubre, noviembre y diciembre (12 semanas), para determinar el comportamiento del precio del aguacate hass.

Es de resaltar que para efectos del caso de estudio, se seleccionó el modelo Box y JENKINS. No obstante, para otros sistemas productivos puede suceder que el modelo de redes neuronales LSTM sea el seleccionado debido a su desempeño.

Tabla 13 Comparación de los modelos de Box-Jenkins y redes neuronales LSTM

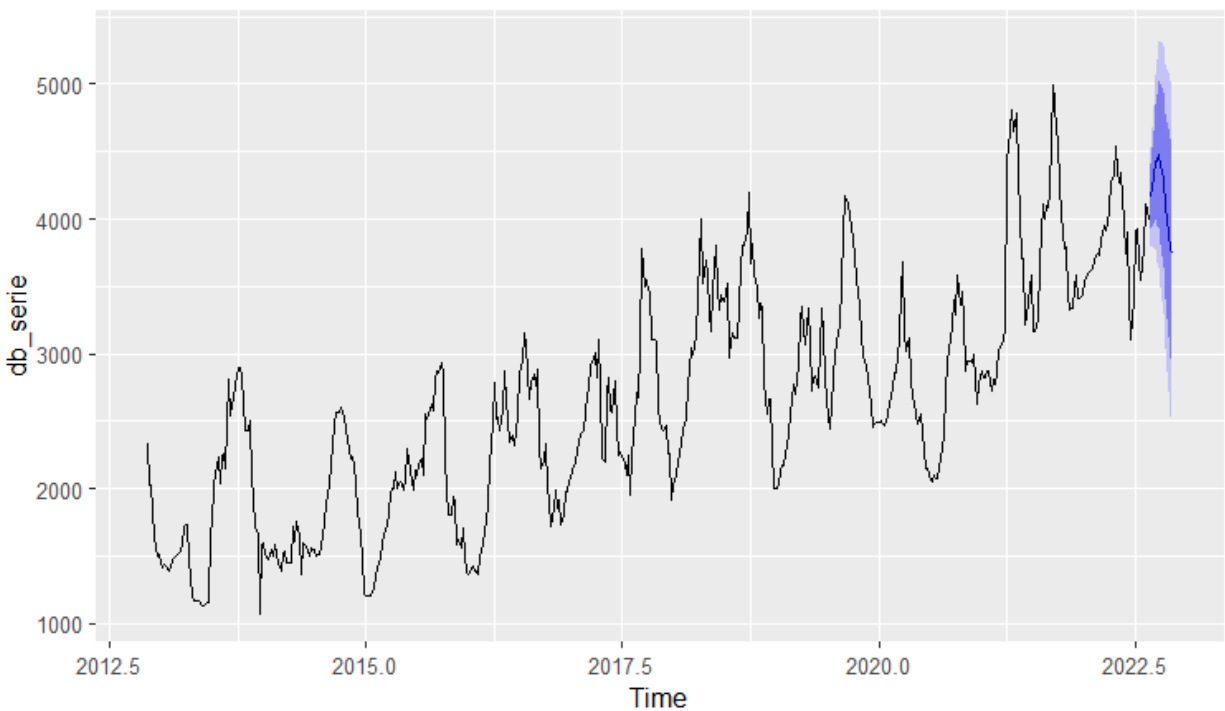
Indicador de desempeño	Modelo Box-Jenkins	Redes Neuronales LSTM
MAPE	5.619%	6.129%
RMSE	237.7815	257.807
MAE	203.0366	225.089

Nota: Elaboración propia.

Al realizar la predicción final, se observa que los precios tienden a la baja, lo cual está de acuerdo con el análisis realizado en el contexto del monitoreo. Entonces, si el productor de aguacate hass va a vender su producto en estas fechas, debe de tener en cuenta que percibirá menos ingresos, y si una entidad financiera decide prestarle a un productor de aguacate en estas fechas, debe medir el riesgo financiero teniendo en cuenta este aspecto para determinar la viabilidad de otorgar un crédito.

Adicionalmente, obsérvese que los resultados de la predicción tienen un sentido lógico si los comparamos con el Gráfico 16 debido a que, según este gráfico, los precios del aguacate hass tienden a disminuir. Es decir, que la predicción realizada si está considerando el componente de estacionalidad de la serie de precios.

Gráfico 22 Predicción del precio del aguacate hass en la central mayorista de Antioquia para los periodos octubre – diciembre de 2022

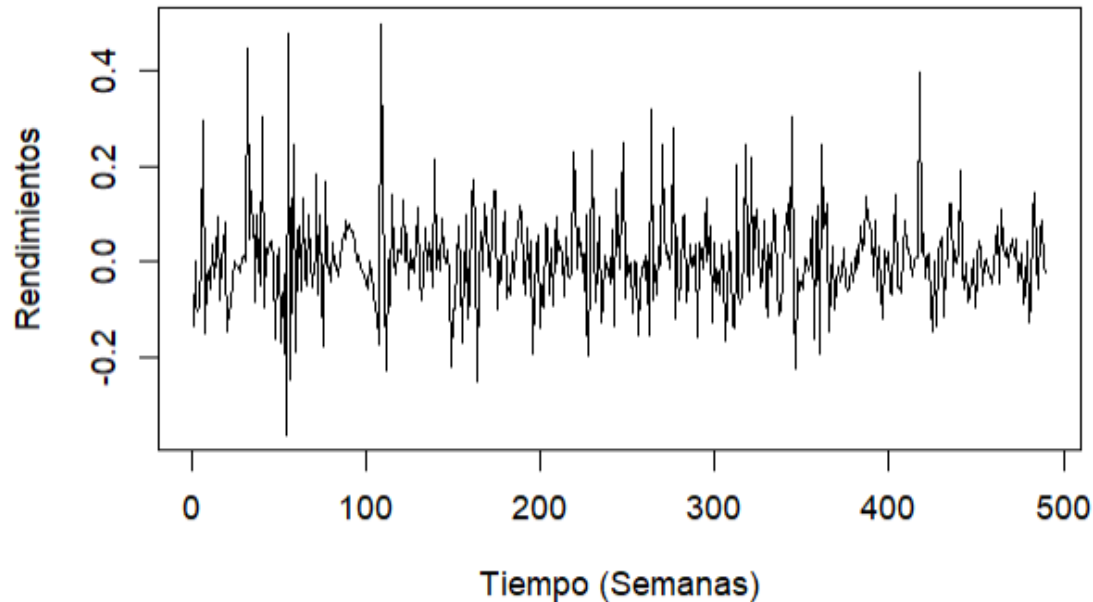


Nota: Elaboración propia.

### 6.2.2.3.3 *Modelo ARCH-GARCH*

Habiendo pronosticado la media de los precios del aguacate hass de la central mayorista de Antioquia utilizando los métodos Box-Jenkins y redes neuronales LSTM, se pronostica la volatilidad de los precios utilizando los modelos ARCH-GARCH con la volatilidad histórica que se encuentra representada en la Figura 15.

Figura 15 Rendimientos de los precios de aguacate hass para la central mayorista de Antioquia



Nota: Elaboración propia.

Para iniciar, se procede a estimar un modelo ARIMA para la serie de rendimientos asumiendo que dicha serie es ruido blanco y sobre los errores de este modelo, se determina si existe efecto ARCH, al elevar al cuadrado los errores y aplicar la prueba de Ljung y Box.

Al realizar este procedimiento, se observa en la Figura 16 que los rendimientos de los precios del aguacate hass tienen efecto ARCH. Sobre esta base, se procede con la determinación del modelo de varianza, para la cual se tomó como punto de referencia un modelo GARCH (1,1)

Figura 16 Prueba de efecto ARCH sobre los rendimientos de los precios

Box-Ljung test

data: resMM1^2  
 X-squared = 20.223, df = 12, p-value = 0.06298

Nota: Elaboración propia.

En la Figura 17 se observa la aplicación del modelo de varianza con un GARCH (1,1), sobre el cual se aplicarán los diagnósticos en el término de error para determinar si es un modelo adecuado para pronosticar la volatilidad de los precios del aguacate hass.

Figura 17 Modelo de varianza para los rendimientos de los precios

```
Title:
  GARCH Modelling
Call:
  garchFit(formula = ~garch(1, 1), data = Rendimiento_Precio, trace = FALSE)
Mean and Variance Equation:
  data ~ garch(1, 1)
<environment: 0x00000202b6cef278>
 [data = Rendimiento_Precio]
Conditional Distribution:
  norm
Coefficient(s):
      mu      omega      alpha1      beta1
0.0043148  0.0048433  0.1277341  0.0751633
Std. Errors:
  based on Hessian
Error Analysis:
      Estimate Std. Error t value Pr(>|t|)
mu      0.004315  0.003462  1.246  0.2127
omega   0.004843  0.000885  5.473 4.43e-08 ***
alpha1  0.127734  0.049797  2.565  0.0103 *
beta1   0.075163  0.137522  0.547  0.5847
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Log Likelihood:
  582.4526      normalized:  1.14656
Description:
  sun oct  2 13:28:59 2022 by user: Analítica

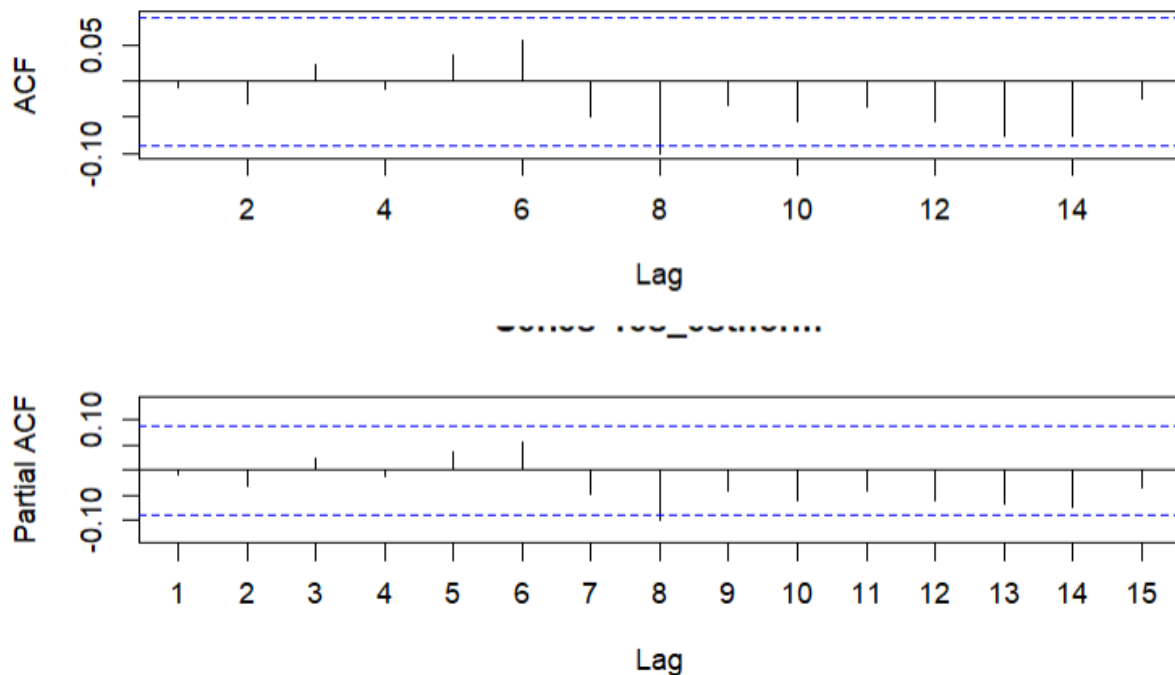
Standardised Residuals Tests:
      Jarque-Bera Test  R      Chi^2  Statistic p-Value
Shapiro-wilk Test    R      W      0.9371858 8.33815e-14
Ljung-Box Test       R      Q(10)  46.37443  1.226699e-06
Ljung-Box Test       R      Q(15)  78.09872  1.55248e-10
Ljung-Box Test       R      Q(20)  93.78579  1.602762e-11
Ljung-Box Test       R^2    Q(10)  13.22872  0.2111636
Ljung-Box Test       R^2    Q(15)  15.13431  0.4417865
Ljung-Box Test       R^2    Q(20)  20.84246  0.4064572
LM Arch Test         R      TR^2   13.17763  0.356263

Information Criterion Statistics:
      AIC      BIC      SIC      HQIC
-2.277372 -2.244062 -2.277495 -2.264310
```

Nota: Elaboración propia.

Al tomar los residuales del modelo, se realizaron las pruebas de Ljung y Box y se graficaron las ACF y PACF de dichos residuales para determinar si son ruido blanco. Los resultados obtenidos al aplicar este ejercicio se pueden observar en las Figuras 18 y 19, donde se muestra que los errores son un ruido blanco. Por tanto, se puede tomar el modelo GARCH (1,1) para pronosticar la volatilidad de los precios del aguacate hass.

Figura 18 Residuales modelo de varianza para los rendimientos de los precios



Nota: Elaboración propia.

Figura 19 Pruebas sobre los residuales del modelo de varianza para los rendimientos de los precios

```

Box-Ljung test
data: res_estnorm
X-squared = 19.844, df = 15, p-value = 0.178

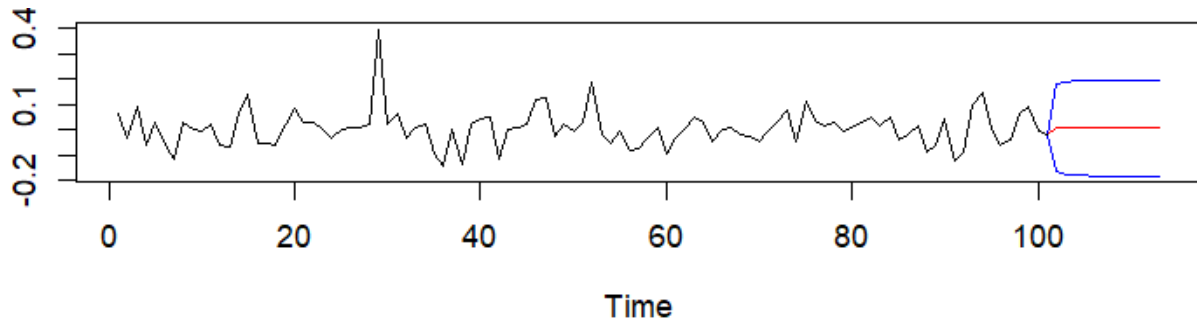
Box-Ljung test
data: res_estnorm^2
X-squared = 7.0189, df = 15, p-value = 0.9571

```

Nota: Elaboración propia.

Finalmente, al realizar la predicción con un modelo GARCH (1,1), se obtiene que la volatilidad de los precios del aguacate hass para las próximas 12 semanas será de un 9,36% con unos intervalos de confianza de -16,71% y 18,10%.

Figura 20 Pronostico de los rendimientos de los precios del aguacate hass para la central mayorista de Antioquia



Nota: Elaboración propia.

#### 6.2.2.4 Análisis de vulnerabilidad

Para el análisis de la vulnerabilidad del precio del aguacate hass, se realizó el modelo de regresión lineal, donde las variables explicativas fueron la tasa representativa del mercado y la cantidad de importaciones y exportaciones del aguacate en Colombia. Las bases de datos tomadas para el cálculo fueron:

- Tasa representativa del mercado: Información reportada por el Banco de la República.
- Importaciones y exportaciones: Información reportada en el Trade Map del Centro de Comercio Internacional (ITC).

Las variables mencionadas se ingresaron en el algoritmo para la estimación del modelo de regresión lineal. Es de aclarar que se utilizaron los logaritmos para interpretar las relaciones a nivel porcentual. Sobre esta base, se obtuvo como resultado el modelo de regresión que se encuentra en la Figura 21.

Entonces, se observa que las importaciones y la tasa representativa del mercado explican significativamente el comportamiento del precio del aguacate hass. Igualmente, se obtiene un R2 de 0.47, lo cual lo hace un modelo aceptable.

Figura 21 Modelo de regresión lineal para los precios de la central mayorista de Antioquia con respecto a la cantidad de importaciones, exportaciones y tasa representativa del mercado

```
Call:
lm(formula = log(Precio_interno) ~ log(Importaciones) + log(Exportaciones) +
    log(TRM), data = Base_General)

Residuals:
    Min       1Q   Median       3Q      Max
-0.63077 -0.13004  0.01218  0.13743  0.45748

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.16818    1.43192    0.816  0.4165
log(Importaciones) -0.05068    0.02392   -2.118  0.0366 *
log(Exportaciones) -0.01994    0.01647   -1.211  0.2289
log(TRM)         0.93662    0.18641    5.024 2.19e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2293 on 101 degrees of freedom
(3 observations deleted due to missingness)
Multiple R-squared:  0.4758,    Adjusted R-squared:  0.4602
F-statistic: 30.56 on 3 and 101 DF,  p-value: 3.856e-14
```

Nota: Elaboración propia.

En el marco de las ideas desarrolladas, se proceden a validar los supuestos del modelo de regresión con el fin de interpretar correctamente los coeficientes obtenidos en la modelación.

La primera validación realizada es la heterocedasticidad en el término de error. Para hacer la validación se aplicó la prueba de Breusch-Pagan. Al observar los resultados en la Figura 22, se deduce que los errores se deben corregir con la matriz de Wald para poder realizar una correcta interpretación de los coeficientes del modelo de regresión.



Figura 22 Prueba de Breusch-Pagan para la heterocedasticidad en el término de error

```
studentized Breusch-Pagan test  
  
data: Modelo_Econometrico_1  
BP = 7.8658, df = 3, p-value = 0.04887
```

Nota: Elaboración propia.

Al aplicar la corrección con la matriz de Wald, se observa en la Figura 23 que las variables de importación y TRM siguen siendo estadísticamente significativas con un nivel de confianza del 95%. A partir de este resultado, se tiene entonces que por el aumento en 1% en la cantidad de aguacate importada, el precio del aguacate hass en la central mayorista de Antioquia disminuye en un 0.05%, con respecto a la TRM y se observa que por el aumento en un 1% de la TRM, el precio del aguacate hass aumentaría en un 0.93%.

Figura 23 Modelo de regresión lineal con la corrección de los errores con la matriz de wald

```
t test of coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.168181	1.355195	0.8620	0.39073	
log(Importaciones)	-0.050679	0.023901	-2.1203	0.03643	*
log(Exportaciones)	-0.019938	0.015290	-1.3040	0.19519	
log(TRM)	0.936618	0.171197	5.4710	3.261e-07	***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Nota: Elaboración propia.

Luego de hacer la corrección por la matriz de Wald, se valida el supuesto de la multicolinealidad. Para lograrlo se calculó el valor inflado de la varianza que se observa en la Figura 24, en la que se aprecia que ninguna de las variables tomadas para el análisis del precio del aguacate presenta multicolinealidad.

Figura 24 Valor inflado de la varianza del modelo

log(Importaciones)	log(Exportaciones)	log(TRM)
1.359755	3.325073	3.818215

Nota: Elaboración propia.

Finalmente, se ajustaron los modelos de regresión lineal para datos atípicos utilizando las regresiones robustas de Yohai y Huber, obteniendo como resultados las regresiones de la Figura 25 y Figura 26. En ellas se observa un aumento en el coeficiente de la TRM, es decir, por el aumento de un 1% en la TRM, el precio del aguacate aumentaría en un 0.97%.

Figura 25 Modelo de regresión con datos atípicos Yohai

[1] "Estimador Yohai"

```
Robust Regression with Bisquare Function
Convergence achieved after: 8 iterations
source      SS      df      MS      F
model       4.87    3      1.62    29.86
error       3.91    101
tot         8.79    104
rsquared    0.554
mse         0.03875796

Coefficients:
              estimate  std error  t value  p value
(Intercept)   0.92411205  1.51690991   0.61  0.54376
log(Importaciones) -0.05763571  0.02551416  -2.26  0.02603
log(Exportaciones) -0.02360918  0.01732525  -1.36  0.17601
log(TRM)       0.98391821  0.19678502   5.00  0.00000
```

Nota: Elaboración propia.

Figura 26 Modelo de regresión con datos atípicos Huber

[1] "Estimador Huber"

```
Robust Regression with Huber Function
Convergence achieved after: 9 iterations
source      SS      df      MS      F
model       5.05    3      1.68    29.45
error       4.26   101
tot         9.31   104
rsquared    0.543
mse         0.04214844
```

Coefficients:

	estimate	std error	t value	p value
(Intercept)	1.01721396	1.51973093	0.67	0.50481
log(Importaciones)	-0.05955045	0.02551618	-2.33	0.02159
log(Exportaciones)	-0.02366853	0.01743926	-1.36	0.17774
log(TRM)	0.97507754	0.19760319	4.93	0.00000

Nota: Elaboración propia.

### 6.2.3 Riesgo financiero

A continuación, se presenta la aplicación de la metodología propuesta para la medición del riesgo financiero. El propósito es medir la sensibilidad de la rentabilidad de un productor de aguacate hass y calcular el Valor en Riesgo sobre la simulación del valor presente neto (VPN).

#### 6.2.3.1 Simulación Montecarlo

Como punto inicial, se construyó el flujo de caja libre para un productor de 10 hectáreas de aguacate hass en Antioquia, tal y como se observa en la Tabla 14. Posteriormente, se desarrolló la Simulación Montecarlo, para lo cual se tomaron como variables los precios de la central mayorista de Antioquia y los rendimientos del aguacate hass. Para el primero, se tomó la distribución normal, mientras que, para el segundo, se tomó una distribución triangular. Por otro lado, se realizaron 20.000 simulaciones en un solo ciclo. Finalmente, se tomó como variable de análisis el valor presente neto.

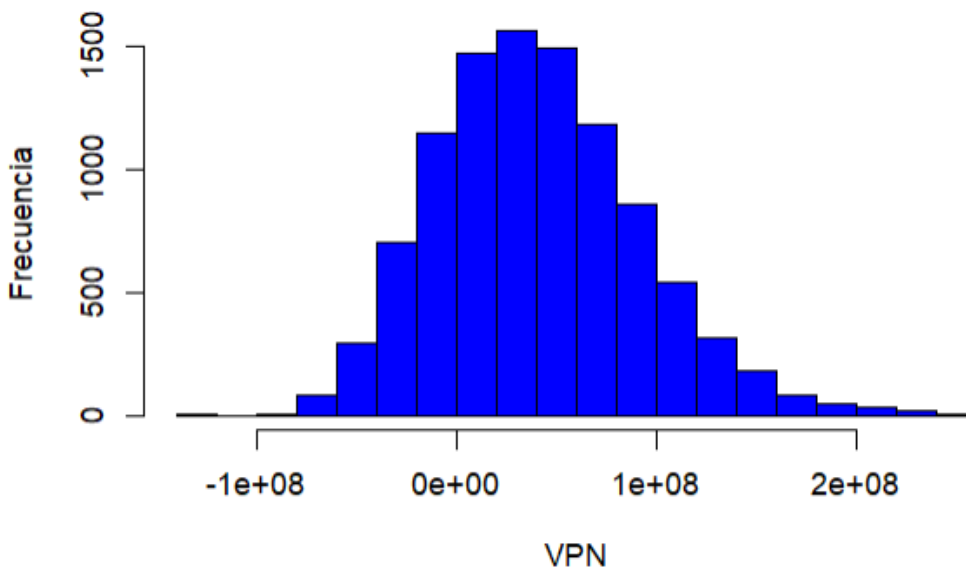
Tabla 14 Flujo de caja libre para el sistema productivo de aguacate

	Año 1	Año 2	Año 3	Año 4
<b>UTILIDAD NETA</b>	19378617	23254340	27130064	12378617
<b>(+) Gastos de depreciación</b>	15000000	15000000	15000000	15000000
<b>(=) GIF (Generación interna de fondos</b>	34378617	38254340	42130064	27378617
<b>(+) Intereses</b>	2000000	1800000	1600000	1200000
<b>(=) Flujo de caja bruto</b>	36378617	40054340	43730064	28578617
<b>(-) Variación KTNO</b>	6000000	7000000	2000000	2000000
<b>(=) EGO (Efectivo generado por la operación)</b>	30378617	33054340	41730064	26578617
<b>(-) CAPEX</b>	90000000	0	0	0
<b>(=) FLUJO DE CAJA LIBRE</b>	-59621383	33054340	41730064	26578617

Nota: Elaboración propia.

Sobre las bases descritas en el párrafo anterior, se obtuvo como resultados el Gráfico 23 el cual muestra la función de densidad de probabilidad de los VPN simulados. Sobre este gráfico, se calcularon los intervalos de confianza en los que podía oscilar el indicador, obteniendo como resultado la Figura 27, donde se observa con una probabilidad del 95% que las ganancias y pérdidas de este productor de aguacate hass pueden oscilar entre - 58 millones y 139 millones.

Gráfico 23 Histograma de los valores presente netos simulados



Nota: Elaboración propia.

Figura 27 Intervalo de confianza para el Valor Presente Neto

	Intervalo 2,5%	Intervalo 97,5%
95% de confianza	-58908567	139921854

Nota: Elaboración propia.

También se calculó el VaR sobre el VPN, obteniendo que con una probabilidad del 95% las pérdidas del proyecto no superarán los 42 millones de pesos. Para medir si esto es riesgoso o no, se calculó la probabilidad de que el VPN fuera inferior a cero ( $P_{\text{VPN}} \leq 0$ ), obteniendo como resultado un 21,22%. Es decir, existe una probabilidad del 21% de que el negocio del productor agropecuario no sea viable. En este sentido, se tiene que el negocio, bajo las simulaciones realizadas, se clasifica en un riesgo medio-alto, debido a que, para este tipo de proyectos, se esperan probabilidades inferiores al 10%, para considerarlo riesgo bajo.

Finalmente, se midió la relación porcentual entre las variables críticas con el VPN del negocio y se analizaron el impacto y la significancia de las variables utilizadas, obteniendo como resultado la Figura 28 en la cual se puede analizar que por el aumento de un 1% en el precio de venta, el VPN aumentaría en un 4.3%, mientras que, por el aumento en un 1% en los rendimientos, el VPN aumentaría en un 3.9%. Se observa igualmente que ambas variables son estadísticamente significativas.

Figura 28 Modelo de regresión lineal para medir el impacto de las variables sobre el VPN

```
Call:
lm(formula = log(Matriz_VPN) ~ log(Precio_Simulado) +
log(Produccion_Simulada))

Residuals:
    Min       1Q   Median       3Q      Max
-6.8413 -0.0948  0.1531  0.2705  0.3259

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -30.73238    0.32825  -93.62  <2e-16 ***
log(Precio_Simulado)  4.30719    0.03490  123.42  <2e-16 ***
log(Produccion_Simulada)  3.90859    0.02824  138.42  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4802 on 7770 degrees of freedom
(2227 observations deleted due to missingness)
Multiple R-squared:  0.778,    Adjusted R-squared:  0.7779
F-statistic: 1.362e+04 on 2 and 7770 DF,  p-value: < 2.2e-16
```

Nota: Elaboración propia.

## **Conclusiones y recomendaciones**

Como conclusión principal de esta consultoría, se tiene que las metodologías desarrolladas dan cumplimiento al objetivo general, en la medida en que dichos indicadores permiten soportar la toma de decisiones de los productores agropecuarios y de las entidades financieras. En esta dirección, la inclusión de los indicadores y metodologías de medición de riesgo de mercado y del riesgo financiero propuestas para la Plataforma Agro de SFA Cebár, generarán un valor agregado a las diferentes partes interesadas de la empresa. Para las compañías de financiamiento, debido a que tendrán una visión completa del producto sujeto a financiación y tendrán más elementos de juicio para determinar el monto del préstamo y para reconocer cuáles riesgos corren al momento de otorgar el crédito. En lo que hace a los productores, tendrán un panorama amplio de las amenazas asociadas al riesgo de mercado y al riesgo financiero, y podrán conocer el impacto que puede tener la materialización de éstas sobre su rentabilidad, permitiéndoles prepararse ante las situaciones adversas en términos de precios, y de este modo mitigar el impacto sobre su rentabilidad, como también, les permitirá tomar decisiones para reducir o transferir los riesgos ante la materialización de alguna amenaza.

Con respecto al caso aplicado, se concluye que el productor de aguacate hass que comercializa en la central mayorista de Antioquia presenta un riesgo de mercado medio-alto, según los indicadores históricos de los precios del aguacate; además los indicadores de monitoreo y predicción muestran que los precios en esta central para el trimestre posterior al período analizado tenderán a la baja. Por tanto, el productor agropecuario deberá buscar mecanismos para mitigar este riesgo. Con respecto a la entidad financiera que entregue crédito a este productor, debe tener en cuenta la posibilidad de otorgar periodos de gracias para este último trimestre del año, con el fin de que el cliente agropecuario tenga el flujo de caja necesario para amortizar la deuda. Para el caso del riesgo financiero, se evidenció en el caso aplicado que se encuentra en un nivel medio, debido a la alta correlación de los precios con los flujos de caja y por el comportamientos de dichos precios que tenderán a la baja.

De otro lado, los resultados obtenidos en la presente consultoría, pueden convertir a SFA CEBAR en un referente para la evaluación del riesgo de mercado y del riesgo financiero del sector

agropecuario, y le abren camino para futuras investigaciones y aportes al sector agropecuario. Adicionalmente, le otorga una ventaja importante con respecto a la competencia, dado que ésta solo se limita a la evaluación financiera.

Así mismo, los indicadores y metodologías propuestas para la Plataforma Agro de SFA CEBAR, son aplicables al contexto del sector agropecuario colombiano, dado que, dentro de las metodologías halladas, existen algunas donde las variables requeridas para la aplicación, tales como precios en bolsa de valores, no se disponen en Colombia. Sobre la consideración anterior, las metodologías seleccionadas se adaptan a las fuentes de información disponibles en la Plataforma Agro y en las fuentes públicas de información como DANE, FINAGRO, entre otras.

Por otro lado, el desarrollo de la presente consultoría dejó en evidencia que los avances en materia de evaluación de riesgos de mercado y financiero en el contexto nacional han sido precarios porque si bien es cierto que las entidades financieras que prestan para el sector agropecuarios realizan estudios de riesgo crediticio a través de scoring de riesgo, estos análisis solo se limitan a la información de cada cliente y no a la información del entorno en el que éste desempeña su actividad productiva. Por tal razón, se espera que esta consultoría motive a la compañía a continuar investigando metodologías de evaluación de riesgos de mercado y riesgo financiero para el agro colombiano.

En cuanto a las limitaciones que se evidenciaron a lo largo del desarrollo de la consultoría, se concluye que la disponibilidad de información para el sector agropecuario es precaria, en comparación de la información reportada en otros países. Lo cual limitó, el desarrollo de metodologías adicionales para la medición del riesgo de mercado y financiero.

Por otro lado, con respecto al aspecto de programación, es preciso mencionar que dependiendo del ambiente de desarrollo de la Plataforma Agro, se recomienda al equipo de sistemas de información de la empresa buscar paquetes que lean códigos de R para ejecutar los algoritmos propuestos en cada una de las metodologías seleccionadas, para que el entorno de desarrollo no presente problema.



Finalmente, se recomienda al equipo financiero y de riesgos, seguir profundizando sobre nuevas metodologías para la medición de riesgos de mercado y financiero, dado que esto, marcará un factor diferenciador frente a las entidades bancarias y los productores del sector agropecuario.

## Referencias bibliográficas

- Assouto, A., Houensou, D., & Semedo, G. (2020). Price risk and farmers' decisions: A case study from Benin. *Scientific African*, 8. <https://doi.org/10.1016/j.sciaf.2020.e00311>
- Barten, A. P., & Vanloot, C. (1996). Price dynamics in agriculture: An exercise in historical econometrics. *Economic Modelling*, 13(3), 315–331. [https://doi.org/10.1016/0264-9993\(96\)01011-5](https://doi.org/10.1016/0264-9993(96)01011-5)
- Bayramoglu, B., & Chakir, R. (2016). The impact of high crop prices on the use of agro-chemical inputs in France: A structural econometric analysis. *Land Use Policy*, 55, 204–211. <https://doi.org/10.1016/j.landusepol.2016.03.027>
- Benavides, G. (2004). *Price volatility forecasts for agricultural commodities: an application of historical volatility models, option implieds and composite approaches for futures prices of corn and wheat*.
- Berger, A., Pena, S., & Melón, M. (2015). Cría ganadera y riesgo. Evaluación mediante simulación Montecarlo. *Revista de La Escuela de Prefecionamiento En Investigación Operativa*, 37, 119–136.
- Casas, M., & Cepeda, E. (2008). Modelos ARCH, GARCH y EGARCH: aplicaciones a series financieras. *Cuadernos de Economía*, 27(1), 287–319.
- Chuluunsaikhan, T., Ryu, G. A., Yoo, K. H., Rah, H., & Nasridinov, A. (2020). Incorporating deep learning and news topic modeling for forecasting pork prices: The case of South Korea. *Agriculture (Switzerland)*, 10(11), 1–22. <https://doi.org/10.3390/agriculture10110513>
- Cowpertwait, P., & Metcalfe, A. (2011a). Introductory time series with R. In *Journal of Applied Statistics* (Vol. 38, Issue 10). Springer. <https://doi.org/10.1080/02664763.2010.517940>
- Cowpertwait, P., & Metcalfe, A. (2011b). Introductory time series with R. In *Journal of Applied Statistics* (Vol. 38, Issue 10). Springer. <https://doi.org/10.1080/02664763.2010.517940>
- Eissa, M. A., & Al Refai, H. (2019). Modelling the symmetric and asymmetric relationships between oil prices and those of corn, barley, and rapeseed oil. *Resources Policy*, 64(September), 101511. <https://doi.org/10.1016/j.resourpol.2019.101511>
- Gómez, M., Danglot, C., & Vega, L. (2003). Sinopsis de pruebas estadísticas no paramétricas. Cuándo usarlas. *Revista Mexicana de Pediatría*, 70(2), 91–99. <https://doi.org/10.1227/00006123-198203000-00001>

- Grewal, D., & Daneshyari, M. D. (2022). Machine learning prediction of agricultural produces for Indian Farmers using LSTM. *International Journal of Multidisciplinary Research and Growth Evaluation*, 113–119. <https://doi.org/10.54660/anfo.2022.3.5.5>
- Herrera, G. P., Lourival, R., da Costa, R. B., Mendes, D. R. F., Moreira, T. B. S., de Abreu, U. G. P., & Constantino, M. (2018). Econometric analysis of income, productivity and diversification among smallholders in Brazil. *Land Use Policy*, 76(February), 455–459. <https://doi.org/10.1016/j.landusepol.2018.02.025>
- Heymans, A. (2008). *Managing an agricultural commodities portfolio in South Africa with pairs trading*. THE CENTRE FOR BUSINESS MATHEMATICS AND INFORMATICS OF THE NORTH-WEST UNIVERSITY.
- Insana, A. (2022). Does systematic risk change when markets close? An analysis using stocks' beta. *Economic Modelling*, 109. <https://doi.org/10.1016/j.econmod.2022.105782>
- Jarque, C., & Bera, A. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review*, 55(2), 163–172.
- Kadigi, I. L., Richardson, J. W., Mutabazi, K. D., Philip, D., Bizimana, J. C., Mourice, S. K., & Waized, B. (2020). Forecasting yields, prices and net returns for main cereal crops in Tanzania as probability distributions: A multivariate empirical (MVE) approach. *Agricultural Systems*, 180(October 2018), 102693. <https://doi.org/10.1016/j.agsy.2019.102693>
- Kahn, M. N. (2010). *Technical Analysis Plain and Simple* (I. Pearson Education, Ed.; 3rd ed.). 9780137042012.
- Kilgallen, T. (2012). Testing the Simple Moving Average across Commodities, Global Stock Indices, and Currencies. *The Journal of Wealth Management*. [www.ijjournals.com](http://www.ijjournals.com)
- León-Castro, E., Espinoza-Audelo, L. F., Merigó, J. M., Herrera-Viedma, E., & Herrera, F. (2021). Measuring volatility based on ordered weighted average operators: The case of agricultural product prices. *Fuzzy Sets and Systems*, 422, 161–176. <https://doi.org/10.1016/j.fss.2020.08.006>
- Li, G. Q., Xu, S. W., & Li, Z. M. (2010). Short-term price forecasting for agro-products using artificial neural networks. *Agriculture and Agricultural Science Procedia*, 1, 278–287. <https://doi.org/10.1016/j.aaspro.2010.09.035>
- López, P. (2019). *Rentabilidad y riesgos en la producción de papa blanca comercial. Los casos de Ayacucho y Lima*. Universidad Naiconal Agraria La molina.

- Macijauskas, L. (2012). *Simple Moving Average as a Risk Management Method in Main Asset Classes*. 114–121. <https://doi.org/10.3846/bm.2012.016>
- Méndez, C. E. (1988). *Metodología. Guía para elaborar diseño de investigación en ciencias económicas, contables, administrativas*. McGraw-Hill.
- Miguez, D. F. (2014). Análisis de riesgos en emprendimientos agropecuarios. evaluación de resultados económicos esperados en proyectos productivos en el oeste de la provincia de Buenos Aires. *Revista de Investigación En Modelos Financiero*, 1, 69–92.
- Mohd, N., & Bee, Y. (2011a). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Mohd, N., & Bee, Y. (2011b). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Montgomery, D., & Runger, G. (2008). *Probabilidad y Estadística aplicada a la Ingeniería*. Limusa Wiley.
- Murugesan, R., Mishra, E., & Krishnan, A. H. (2021). *Deep Learning Based Models: Basic LSTM, Bi LSTM, Stacked LSTM, CNN LSTM and Conv LSTM to Forecast Agricultural Commodities Prices*. <https://doi.org/10.21203/rs.3.rs-740568/v1>
- Onour, I., & Sergi, B. (2011). Modeling and forecasting volatility in the global food commodity prices. *Agricultural Economics*, 57(3), 132–139. <https://doi.org/https://doi.org/10.17221/28/2010-agricecon>
- Perez, F. (2006). Modelación de la volatilidad y pronóstico del precio del café. *Revista Ingeniería Universidad de Medellin*, 5(9), 45–58.
- Piot-Lepetit, I., & M'Barek, R. (2011). *Methods to Analyse Agricultural Commodity Price Volatility*. Springer.
- Rehman, A., Jingdong, L., Chandio, A. A., & Hussain, I. (2017). Livestock production and population census in Pakistan: Determining their relationship with agricultural GDP using econometric analysis. *Information Processing in Agriculture*, 4(2), 168–177. <https://doi.org/10.1016/j.inpa.2017.03.002>
- Rezitis, A. N., & Tsionas, M. (2019). Modeling asymmetric price transmission in the European food market. *Economic Modelling*, 76(July 2018), 216–230. <https://doi.org/10.1016/j.econmod.2018.08.004>

- Rezitis, A., & Stavropoulos, K. (2010). Modeling beef supply response and price volatility under CAP reforms: The case of Greece. *Food Policy*, 35(2), 163–174. <https://doi.org/https://doi.org/10.1016/j.foodpol.2009.10.005>
- Sekhar, C. (2004). Agricultural Price Volatility in International and Indian Markets. *Economic and Political Weekly*, 39(43), 4729–4736. <https://doi.org/http://www.jstor.org/stable/4415712>
- Senapati, A. K. (2020). Evaluation of risk preferences and coping strategies to manage with various agricultural risks: evidence from India. *Heliyon*, 6(3), e03503. <https://doi.org/10.1016/j.heliyon.2020.e03503>
- Siddig, K., & Grethe, H. (2014). International price transmission in CGE models: How to reconcile econometric evidence and endogenous model response? *Economic Modelling*, 38, 12–22. <https://doi.org/10.1016/j.econmod.2013.11.038>
- Silveira, R., Maciel, L., Mattos, F., & Ballini, R. (2017). Volatility persistence and inventory effect in grain futures markets: evidence from a recursive model. *Revista de Administração*, 52(4), 403–418. <https://doi.org/https://doi.org/10.1016/j.rausp.2017.08.003>
- Tadesse, G., Algieri, B., Kalkuhl, M., & von Braun, J. (2014). Drivers and triggers of international food price spikes and volatility. *Food Policy*, 47, 117–128. <https://doi.org/10.1016/j.foodpol.2013.08.014>
- Thi Nhung, N., Thi Nguyet Nuong, N., & Ha Uyen, N. (2021). Efficiency of Bollinger Bands in Forward Operations for Agricultural Products in Vietnam. *VNU Journal of Economics and Business*, 1(2), 56–68. <https://doi.org/10.25073/2588-1108/vnueab.4604>
- Wang, C., Zhao, J., & Huang, M. (2010). Measurement of the fluctuation risk of the China fruit market price based on VaR. *Agriculture and Agricultural Science Procedia*, 1, 212–218. <https://doi.org/10.1016/j.aaspro.2010.09.026>
- Xiong, T., Li, C., & Bao, Y. (2018). Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China. *Neurocomputing*, 275, 2831–2844. <https://doi.org/10.1016/j.neucom.2017.11.053>
- Zou, H. F., Xia, G. P., Yang, F. T., & Wang, H. Y. (2007). An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing*, 70(16–18), 2913–2923. <https://doi.org/10.1016/j.neucom.2007.01.009>