
**Pronóstico del precio promedio por
tonelada del arroz paddy desde julio
hasta diciembre de 2018 en los
Llanos Orientales**



Presentado por

César Humberto Parada Vergara

LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Fundación Universitaria Los Libertadores

Facultad de Ingeniería y Ciencias Básicas

Especialización en Estadística Aplicada

Bogotá D.C, Colombia

2018

**Pronóstico del precio promedio por
tonelada del arroz paddy desde julio
hasta diciembre de 2018 en los
Llanos Orientales**



LOS LIBERTADORES

Presentado por
César Humberto Parada Vergara

en cumplimiento parcial de los requerimientos para optar al título

de

Especialista en Estadística Aplicada

Dirigida por

Sébastien Lozano Forero

Profesor

Fundación Universitaria Los Libertadores

Facultad de Ingeniería y Ciencias Básicas

Especialización en Estadística Aplicada

Bogotá D.C, Colombia

2018

Notas de aceptación



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Bogotá DC, Agosto de 2018.



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Las directivas de la Fundación Universitaria Los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores y a los resultados de su trabajo.

Dedicatoria



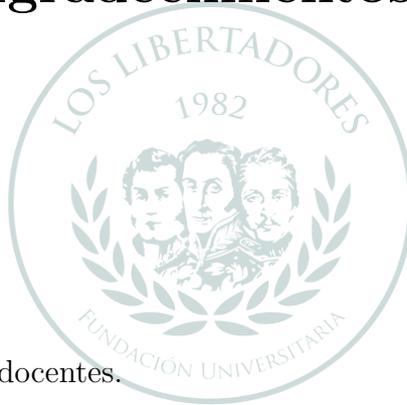
Dedico este trabajo a todas las personas que me han acompañado y apoyado durante el tiempo de estudio.

A mi madre, Marleny Vergara Bocanegra.

A mis hermanos: Jimmy, Félix y William.

A mis amigos.

Agradecimientos



A Dios, familia, amigos y docentes.

LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Índice general

1	Introducción	3
2	Planteamiento del Problema	5
2.1	Objetivos	8
2.1.1	Objetivo General	8
2.1.2	Objetivos Específicos	8
2.2	Justificación	9
3	Marco Teórico / conceptual	11
3.1	Sector primario en Colombia	11
3.2	El arroz en Colombia	12
3.3	Fijación de los precios del arroz en Colombia	14
3.4	Modelos estadísticos para series de tiempo	16
3.4.1	Modelo ARMA	17
3.4.2	Metodología Box-Jenkins	18
3.4.3	Método Holt-Winters	19
4	Marco Metodológico	21
4.1	Bases de datos	21
4.1.1	Precio promedio mensual en pesos por tonelada del arroz paddy zona llanos	21
4.1.2	Valoración nominal por rubros de los costos por hectárea del arroz en Colombia	22
4.1.3	Rendimiento de toneladas por hectárea de arroz paddy	23
4.1.4	Promedio mensual de las precipitaciones en Villavicencio (mm)	24
4.2	Tratamiento a las bases de datos	25
5	Análisis y Resultados	27
5.1	Construcción de los modelos de predicción	27
5.1.1	Modelo ARMA	28
5.1.2	Modelo ARIMA	31

5.1.3	Modelo SARIMAX	34
5.1.4	Modelo Holt-Winters	37
5.2	Elección del mejor modelo construido y pronóstico del precio promedio por tonelada del arroz paddy para la región de los Llanos Orientales	40
5.3	Depuración de la base de datos de costos por hectárea para el arroz paddy en la zona de los Llanos	43
6	Conclusiones y Recomendaciones	45
	Apéndice A Script de los modelos construidos	49
	Apéndice B Script de pronósticos y sistemas de regresión múltiple	57

Índice de figuras

3.1	Pasos de la metodología Box-Jenkins	19
4.1	Precio promedio mensual en pesos colombianos por tonelada del arroz Paddy zona llanos(Villavicencio)	22
4.2	Valoración nominal por rubros de los costos por hectárea del arroz en Colombia	23
4.3	Rendimiento de toneladas por hectárea de arroz paddy	24
4.4	Promedio mensual de las precipitaciones en Villavicencio (mm)	25
5.1	Precio promedio mensual por tonelada del arroz paddy en los Llano Orientales	27
5.2	ACF y PACF de la serie de tiempo	28
5.3	Comportamiento de los residuales	29
5.4	Datos observados VS datos ajustados ARMA(1,1)	30
5.5	Serie original y serie diferenciada	31
5.6	ACF Y PACF de serie diferenciada	31
5.7	Serie original y serie transformada con Box-Cox	32
5.8	ACF Y PACF de serie transformada con Box-Cox	32
5.9	Comportamiento de los residuales ARIMA	33
5.10	Datos observados VS datos ajustados ARIMA(1,1,2)x(0,0,1)[12]	34
5.11	Promedio de precipitaciones en Villavicencio	35
5.12	ACF Y PACF de precipitaciones	35
5.13	Comportamiento de los residuales SARIMAX	36
5.14	Datos observados VS datos ajustados SARIMAX	37
5.15	Holt-Winters aditivo y multiplicativo	38
5.16	Comportamiento de los residuales en Holt-Winters aditivo	38
5.17	Datos observados VS datos ajustados Holt-Winters	39
5.18	Precios promedio de la tonelada del Arroz Paddy para la región de los Llanos Orientales	42

Índice de cuadros

3.1	Holt-Winters aditivo	20
3.2	Holt-Winters multiplicativo	20
5.1	Tabla de estadísticos	28
5.2	AIC y BIC de los modelos ARMA(p,q)	29
5.3	Datos observados y datos ajustados ARMA(1,1)	30
5.4	Valores de AIC, BIC	33
5.5	Datos observados y datos ajustados ARIMA(1,1,2)x(0,0,1)[12]	34
5.6	Valores de AIC, BIC de Modelo SARIMAX	36
5.7	Datos observados y datos ajustados SARIMAX	37
5.8	Datos observados y datos Holt-Winters	39
5.9	Parámetros ARMA(1,1)	40
5.10	Parámetros ARIMA(1,1,2)x(0,0,1)[12]	40
5.11	Parámetros SARIMAX(1,1,2)x(0,0,1)[12]	41
5.12	Parámetros Holt-Winters	41
5.13	RMSE, AIC y BIC de cada modelo	41
5.14	Precio promedio por tonelada del arroz paddy en los Llanos Orientales desde julio hasta diciembre de 2018	42
5.15	Valor del $\Pr(> t)$ de cada costo	43
5.16	Valor del $\Pr(> t)$ de cada costo	44

Pronóstico del precio promedio por tonelada del arroz paddy desde julio hasta diciembre de 2018 en los Llanos Orientales

Resumen

La Federación Nacional de Arroceros (FEDEARROZ) no cuenta con un modelo de predicción o de pronóstico para el precio promedio del arroz paddy por toneladas para los Llanos Orientales. Esta federación cuenta con una base de datos robusta donde registra mes a mes los precios promedios del arroz paddy que cumple con las características de una serie de tiempo en donde se pueden aplicar herramientas estadísticas para realizar predicciones que contribuyen a la toma de decisiones acertadas. Con esa base de datos se construyen y comparan modelos como Holt-Winters, ARMA, SARIMAX y ARIMA para elegir o determinar el más adecuado y que se ajuste a las observaciones registradas cronológicamente. Asimismo, se realiza una depuración de la base de datos de costos por toneladas del arroz paddy para determinar los costos que más influyen en la fijación del precio promedio. En conclusión, de los modelos construidos se escoge el SARIMAX debido a que el criterio de información bayesiano (BIC) es menor a otros modelos y, se depura la base de datos de los rubros, donde se identifica los fertilizantes, la protección al cultivo y la recolección como los costos que más influyen en la fijación del precio promedio del arroz paddy.

Palabras claves: arroz paddy, precio promedio, modelo ARMA, modelo ARIMA, modelo SARIMAX, modelo Holt-Winters, ACF, PACF, AIC, BIC y RMSE.

Capítulo 1

Introducción

Este trabajo consta de seis capítulos, en el primero se encuentran la introducción; en el segundo está el planteamiento del problema, el objetivo general y los objetivos específicos, y la justificación; en el tercero, el marco teórico; en el cuarto, el marco metodológico; en el quinto, el análisis y los resultados; en el sexto, las conclusiones y recomendaciones; y por último, los apéndices.

En el segundo capítulo se contextualizará el problema de estudio, que esta relacionado con la inestabilidad del mercado arrocero y los incrementos desmedidos en los precios [1] que afectan a las regiones donde se cultiva este cereal. Asimismo, este capítulo, tendrá el objetivo general y los objetivos específicos del proyecto que serán la carta de navegación para predecir el precio promedio del arroz paddy desde Julio hasta Diciembre de 2018 en los Llanos Orientales y por último, la justificación de este trabajo.

En el tercer capítulo se realizará una introducción económica del sector primario en Colombia, haciendo énfasis en el cultivo de arroz que anualmente tiene dos cosechas en el año, es decir que es un bien que se produce semestralmente, lo que repercute en los precios mes a mes [2] y se evidencia en la base de datos que será objeto de estudio en este trabajo. También, se describirá la situación del arroz en el país, con una definición formal del cereal y las diferentes orientaciones que se tienen para cultivar el grano. Asimismo, se explicará la fijación del precio del arroz paddy en Colombia, definiendo la oferta, la demanda, el costo e interacción entre estos. Por último, se expondrá los modelos que se usaran en el trabajo para el pronóstico de los precios del arroz paddy en la zona de los Llanos Orientales.

En el cuarto capítulo se encuentran las bases de datos donde está la información de los precios mensuales por tonelada del arroz paddy para los Llanos Orientales, la valoración nominal por rubros de los costos del arroz paddy por hectárea, el rendimiento del arroz

paddy toneladas por hectáreas, y el tratamiento que se realizará de las bases de datos.

En el quinto capítulo se construirán los modelos ARMA, ARIMA, SARIMAX y la metodología Holt-Winters que permiten predecir el precio promedio del arroz paddy en los Llanos Orientales desde julio hasta diciembre de 2018. Después, se hará la comparación de los modelos y se escogerá el mejor teniendo en cuenta el criterio de información bayesiano (BIC) para luego realizar el pronóstico y la gráfica de los precios futuros. Además, se depurará la base de datos de los costos para elegir las variables que más influye en la fijación del precio promedio del arroz paddy para esta región del país.

En el sexto capítulo se expondrán las conclusiones y recomendaciones con base en los análisis y resultados obtenidos. Por ultimo, están los apéndices donde se encontrarán los script más relevantes en el análisis de los datos.

Capítulo 2

Planteamiento del Problema

El arroz es un cereal que hace parte de la canasta familiar en los hogares colombianos, además la producción de arroz en Colombia genera alrededor de 500 mil empleos directos e indirectos en más de 215 municipios, los cuales dependen en un 90% de ésta actividad [3]. Detrás de este alimento hay una industria que históricamente se ha posicionado en el país. Los primeros arroceros se ubicaron en la región central a orillas del río Magdalena o de afluentes que se represaron para construir sistemas de riego y beneficiarse de estos. Asimismo, el gobierno central incentivó la producción arrocerera en otras regiones para disminuir las importaciones venezolanas y ecuatorianas. Una de las regiones beneficiada fue los Llanos Orientales, en especial el departamento del Meta, en la zona del Ariari y, el departamento del Casanare en los municipios de Paz de Ariporo, Trinidad, Aguazul y San Luis de Palenque.

FEDEARROZ arrojó que para el 2016 el consumo per capital del grano en la zona rural fue de 46 kilogramos y en la urbana de 39 kilogramos, consumo que está por encima de otros productos como la papa, el café, el maíz, etc. Según el Ministerio de Agricultura y Desarrollo Rural, para ese mismo año se lograron más de 185.000 nuevas hectáreas sembradas, de las cuales se destaca la producción de arroz con 84.056 hectáreas [4], es decir, representa más del 50% del área nueva sembrada por encima de la palma de aceite, café, maíz tecnificado, cacao, frutales, soya, entre otros. Pero el optimismo del 2016 en la siembra del arroz se ve con preocupación para el 2018 porque el gremio arrocerero quiere evitar una sobreoferta del producto, y en efecto la disminución en los precios, ganancias nulas y altos inventarios en el negocio.

Por otro lado, el Estado en su tiempo otorgó beneficios crediticios con intereses bajos y subsidios para los nuevos cultivadores del cereal. Estas políticas se mantuvieron vigentes hasta la apertura económica del país con los tratados de libre comercio con otras naciones, aumentado el contrabando de arroz a un menor precio que entra por la frontera

terrestre con Venezuela y Ecuador [5], teniendo consecuencias en la disminución del precio interno del grano que afecta la cadena productiva y deja pérdidas para los agricultores.

Los productores del arroz ven que el precio del cereal cada año varia a causa del contrabando, de las tasas de interés para los créditos conferidos a esta agroindustria y el incumplimiento de los subsidios por parte del Estado, sumándole el comportamiento atípico del clima en algunos años en esta región del planeta por su proximidad a la línea ecuatorial.

Lo anterior influye en el precio interno del arroz pero varía según la región productora, debido a que en algunos centros de cultivo se cuenta con sistema de riego que suple el déficit de agua en los ciclos de siembra de secano, y además permite sembrar en otra época y aprovechar, no solo la oferta ambiental disponible, sino las mejores condiciones del mercado en precios y costos [6], en cambio en las regiones donde son deficientes los canales hídricos artificiales, como los Llanos Orientales, la siembra del arroz depende en su mayoría de la temporada de lluvia en la fase vegetativa, reproductiva y madurativa.

La falta de planificación escalonada en la región Oriental para la siembra del arroz conlleva a que se recurra a los molinos que compran el arroz en cascara, lo tratan y lo reservan. Estas reservas del cereal permiten que haya una aparente homogeneidad en el precio del arroz. Sin embargo, factores como la interacción oferta-demanda determinan el crecimiento o decrecimiento del costo del grano, cuestión que afecta tanto al sector productor como industrial. Asimismo, hay un debate que se está llevando a cabo en el Consejo Nacional del Arroz, donde se intenta poner de acuerdo a todas las partes para fijar un precio mínimo de venta entre el productor y el industrial [7], cerrando la posibilidad de generar pronósticos de los precios del cereal a corto plazo.

Tanta es la tensión entre cultivadores e industriales que en la región Oriental los molinos determinan el precio de compra del arroz paddy, situación incómoda para los agricultores porque deben negociar el precio de venta y en algunos casos acceder a las pretensiones de los industriales. En ocasiones, los agricultores llegan a un acuerdo de compra con un precio fijo mínimo con los molinos siempre y cuando este se encargue de adquirir toda la producción. Esa tensión genera que se fije implícitamente un precio mínimo del grano.

Tales acciones de fijar el precio mínimo del cereal son para evitar que haya volatilidad en el costo y producción del producto, en otras palabras, para soslayar la incertidumbre y especulación en el sector arrocerero. Además, la inflación del país es un factor importante para dar cabida a dicha especulación en los precios, situación que beneficia o afecta a la

cadena productiva y empresarios del grano. En esta línea, la especulación acentúa la condiciones de explotación del trabajo y lo precariza y esto tiene significativas consecuencias de orden social [8] en las regiones donde se vive de la siembra, recolección y venta de arroz paddy.

FEDEARROZ registra mes a mes el precio promedio del arroz paddy, construyendo una serie de tiempo en términos estadísticos. Pero no elabora pronósticos del precio promedio para los meses venideros, debido a las razones que se mencionaron anteriormente. Las predicciones del precio promedio del cereal contribuyen a la toma acertada de decisiones en la cadena productiva, recurriendo a la aplicación de métodos estadísticos como base teórica y práctica. En efecto, no realizar estos pronósticos evidencia riesgos en las diferentes variables que intervienen en la siembra, recolección y venta del cereal.

Con base a lo anterior, las herramientas estadísticas que se aplican para las series de tiempo determinan modelos de pronóstico confiables, por lo tanto, el precio promedio mensual por tonelada del arroz paddy desde 1996 hasta el 2018 en la zona de los Llanos Orientales presentan las características de una serie de tiempo. Así, con estas observaciones registradas cronológicamente, el presente trabajo resolverá la siguiente pregunta: ¿cuál será el precio promedio mensual del arroz paddy por tonelada para la zona de los Llanos Orientales desde el mes de julio hasta diciembre de 2018?

2.1 Objetivos

2.1.1 Objetivo General

Proponer un modelo que permita generar pronósticos del precio promedio mensual del arroz paddy por tonelada en la región de los Llanos Orientales desde julio a diciembre de 2018.

2.1.2 Objetivos Específicos

- Consolidar la base de datos sobre los costos fijos, variables, directos e indirectos que intervienen en la fijación del precio promedio del arroz paddy por tonelada en la región de los Llanos Orientales.
- Estimar el valor de la pluviosidad como variable exógena en el precio promedio del arroz paddy por tonelada en la región de los Llanos Orientales para un modelo SARIMAX.
- Validar los supuestos estadísticos del modelo propuesto para el pronóstico del precio promedio del arroz paddy por tonelada en la región de los Llanos Orientales.
- Generar pronósticos para el precio promedio del arroz paddy en la región de los Llanos Orientales.

2.2 Justificación

El presente trabajo de investigación documental se enfocará en pronosticar el precio promedio mensual del arroz paddy por tonelada en la región de los Llanos Orientales, ya que no existen registros históricos de pronósticos del precio del cereal para la toma acertada de decisiones a la hora de planificar la siembra del grano. Así, el presente trabajo permitirá construir varios modelos estadísticos que permitan pronosticar dichos precios, y se elegirá el mejor, además, profundizar en los conocimientos teóricos de la siembra, precios, costos, oferta, elaboración, prueba y elección de modelos estadísticos de predicción.

Capítulo 3

Marco Teórico / conceptual

En este capítulo se realizará una contextualización breve de la economía del sector primario en Colombia. Luego se explicará el papel del arroz en el país y se definirá los conceptos de oferta, demanda, costo y, la interacción que existe entre ellos para la fijación del precio de un producto agrícola. Por último se expondrá los modelos estadísticos para realizar pronósticos en una serie de tiempo.

3.1 Sector primario en Colombia

En términos generales la economía colombiana ha tenido un desarrollo histórico que se ha visto marcado por periodos de crecimiento, de desaceleración y, como resultado de estos cambios no deja de presentarse una constante consolidación, mostrando así una evolución dinámica, beneficiando diferentes sectores del país, tales como el social, el político, empresarial, comercial, agrícola, industrial, financiero, tecnológico, entre otros, que dieron un impulso y reconocimiento del desarrollo económico en el continente, posicionándose como una economía emergente y aceptada por las organizaciones internacionales, tal como la Organización para la Cooperación y el Desarrollo Económico (OCDE), que ha confirmado la confianza de la comunidad internacional en Colombia [9].

En el comportamiento de la economía del país intervienen variables que permiten el dinamismo en esta materia, entre ellas está la oferta, la demanda, los impuestos, las normas reguladoras, etc., que se ve reflejado en el Producto Interno Bruto (PIB) o en el total de bienes y servicios producidos por el país durante un periodo de tiempo determinado.

El país cuenta con recursos diversificados que contribuyen a fortalecer el PIB y son agrupados en los distintos sectores productivos o económicos, es decir, estos sectores son considerados como la división de la economía de un país o región, que abarca todas las etapas de la exploración de los recursos naturales, su industrialización, su preparación para el consumo y su utilización. Esta división es dada en el siguiente orden: sector pri-

mario, sector secundario, sector terciario y sector cuaternario.

El sector primario es el conjunto de actividades económicas relacionadas con la obtención de productos procedentes del mar, de la tierra o el aire, es decir, obtenidos directamente de la naturaleza sin ningún proceso de transformación [10], por ejemplo la minería, la pesca de mar o de cuerpos de aguas continentales, la ganadería y la agricultura. La función del sector primario es producir alimentos y materia prima, para así, satisfacer la demanda tanto interna como externa [11] con el propósito de abastecer la región o el país. Cuando este sector cumple con su función se dinamiza la economía, se fortalece la interacción entre la oferta y la demanda.

En Colombia el sector agropecuario agrupa a todas las organizaciones gremiales, pequeños y grandes empresarios, cultivadores, entre otros, que se benefician del crecimiento de este sector. Las perspectivas del crecimiento agropecuario se ven reflejadas en tres factores:

- 1) el incremento en la demanda mundial de alimentos que se conjuga con las condiciones geoclimáticas colombianas y permiten el aumento sostenido de la oferta agropecuaria;
- 2) la evolución reciente del sector agropecuario ha mostrado un crecimiento superior al de otros renglones de la economía y 3) el incremento del gasto del Gobierno Nacional dirigido al sector agropecuario [12].

Para terminar, Colombia no ha presentado desabastecimiento de los productos primarios en los últimos años porque este sector de la economía ha cumplido con su función y su oferta alcanza la geografía nacional. El año pasado el sector lideró el crecimiento del Producto Interno Bruto (PIB) o crecimiento económico del país [9], se espera que para 2018 el sector mantenga su ritmo y no se vea afectado por fenómenos climáticos o protestas de los transportadores, campesinos, empresarios, etc.

3.2 El arroz en Colombia

Según el Diccionario de la Real Academia, el arroz es una planta anual propia de terrenos muy húmedos, cuyo fruto es un grano oval rico en almidón [13], siendo considerado un cereal y a su vez de la familia gramíneas. Las regiones donde se cultiva el arroz son terrenos que cuentan con dispensas de agua o temporada de lluvias para mantener la humedad indicada en la siembra y crecimiento del cereal.

La cadena productiva de arroz en Colombia está organizada en tres partes, la primera es el cultivo y producción de arroz paddy verde, la segunda es el procesamiento y molienda del arroz paddy seco y la producción de arroz blanco y otros derivados y, a tercera es la comercialización mayorista y minorista de arroz blanco y sus derivados [14]. Cabe aclarar

que hay una relación vertical intrínseca entre el arroz paddy verde y el arroz paddy seco, debido a que su diferencia no es tan marcada, sólo corresponde al cambio que se da con base en el eslabón, por lo general cuando se hace referencia de arroz paddy contiene al verde y al seco.

El arroz se cultiva en tres fases, la primera, es la vegetativa que por lo general dura entre 55 a 60 días, en este tiempo se da la germinación de la semilla, emergencia, macollamiento y elongación del tallo. La segunda, es la reproductiva que se desarrolla entre 35 a 40 días y, va desde la iniciación y desarrollo de la panícula, hasta la floración. La tercera fase, maduración, se demora entre 30 a 40 días, comenzando con el llenado y desarrollo de los granos o estado lechoso y pastoso hasta la madurez del grano o cosecha. El tiempo que duran las tres fases oscila entre 110 a 140 días, sin estar por debajo de los 20 °C y por encima de los 32 °C con un registro pluvial de 1200 mm bien distribuidos durante todo el ciclo, asimismo, el terreno donde se cultive el cereal requiere tener una formación arcillosa para conservar por más tiempo la humedad.

FEDEARROZ ofrece para los cultivadores una guía de trabajo para la siembra del cereal, teniendo en cuenta la planificación económica y financiera y, la planificación agronómica. La primera consiste en

elaborar el flujo de caja proyectado y establecer un nivel de costos de producción; determinar la necesidad de capital y créditos necesarios; presupuestar de acuerdo a las actividades; planificar los gastos teniendo en cuenta los limitantes tecnológicos y oferta ambiental; Provisionar los insumos y recursos necesarios; aprovechar los descuentos por volumen o compra de contado; comprar lo necesario; capacitar al personal [15].

La segunda, planificación agronómica, se ejecuta en su totalidad y a tiempo para obtener buenos resultados, teniendo en cuenta: la elaboración de un cronograma de actividades; establecer y planificar el inventario de la finca, recursos, implementos, maquinaria, mano de obra; establecer la mejor época de siembra [15].

Según FEDEARROZ, el manejo agronómico que se da al arroz en Colombia presenta nueve factores:

la época de siembra y selección de la variedad; el manejo agronómico por ambiente; los componentes del suelo (físicos, químicos, biológicos); la preparación y adecuación de suelos; el riego, drenaje y uso racional del agua; la siembra y densidad de plantas; una nutrición oportuna y balanceada; el monitoreo y manejo fitosanitario; las alternativas de control biológico y natural [15].

Pero esos factores varía según la región, en el caso de los Llanos Orientales los agricultores han ido adecuando los suelos para el desarrollo de la actividad arrocera, asimismo la riqueza hídrica de esta región ha sido subutilizada por los agricultores del cereal porque no hay sistemas de almacenamientos y de riego. En consecuencia, los arroceros prevén que la siembra del grano coincida con el comienzo de la temporada de lluvias que se concentra principalmente entre los meses de abril a octubre, hace que las siembras se escalonen geográficamente, iniciando en la región del río Ariari (Meta) y finalizando en el departamento de Arauca [16], generando subregiones homogéneas debido a sus condiciones productivas, geográficas y climáticas.

Por último, para el 2018 se espera que el arroz siga aumentando su participación en el sector agropecuario y alcance un techo para regular la oferta en el país, tal es la preocupación que en los Llanos Orientales se aconsejó a los agricultores que siembren con responsabilidad, es decir, sin expandirse más las áreas [17]. A pesar de ese consejo, el arroz es protagonista en el sector agrícola colombiano por sus cifras y aporte al PIB y, porque durante el proceso de producción es fuente de empleo en las regiones donde se cultiva.

3.3 Fijación de los precios del arroz en Colombia

El precio de un producto se define como la cantidad de dinero dada a cambio de una mercancía o servicio. El precio es el valor de un bien expresado en términos monetarios [18]. Aunque el uso de la palabra precio varía de acuerdo al contexto, por ejemplo: el trabajador recibe un pago por su servicio y se le llama salario; se llama renta al alquiler de la tierra o de un inmueble, entre otros. En el caso colombiano, algunos precios son fijados por el Estado, como la gasolina, pero según Sabino hay precios que surgen de la interacción entre la oferta y la demanda del mercado [18], siendo este el caso del arroz.

La demanda es la cantidad de bienes y servicios que un consumidor o los consumidores están dispuestos a comprar a un determinado precio en un tiempo dado [10]. La cuantificación de la demanda está trazada por la dimensión temporal y relacionada con el precio de un producto o servicio a través de la ley de demanda, es decir, cuando el precio de un bien o servicio aumenta, la cantidad de ese bien o servicio que las personas están dispuestas o pueden comprar durante un periodo particular de tiempo disminuye [10], en caso contrario, si el precio decae de un bien o servicio la demanda crecerá.

La demanda de un producto varía por diferentes factores: precios de bienes sustitutos; precios de bienes complementarios; precios de bienes esperados en el futuro; e ingreso de los consumidores. En cuanto a bienes se entiende como la adquisición de un objeto o servicio a través de un pago o contraprestación y, por consumidor como una persona u

organización que demanda bienes o servicios a cambio de un pago. Los factores, los bienes y el consumidor producen elasticidad en la demanda que es la responsable de medir la magnitud en que disminuye la demanda de un producto ante un aumento de su precio [10], pero existen bienes con demanda inelástica, como el arroz, que es un bien que no es fácil de sustituir en la dieta alimentaria y las alteraciones del precio no modifica la demanda.

La oferta se define como el monto que los productores están dispuestos a vender en el mercado durante un periodo dado a un precio específico [10], asimismo el número de la demanda y la cantidad ofrecida se mide a través de un periodo o intervalo en el tiempo. La oferta se relaciona con el proceder de los productores o vendedores de un bien y está vinculada con la ley de la oferta que establece que si el precio de un bien o servicio aumenta, la cantidad de ese bien o servicio que los productores están dispuestos o pueden vender durante un tiempo determinado aumentará [10], en caso contrario si disminuye el precio de un bien, la cantidad ofrecida también, estableciendo una relación directa entre el precio y la oferta.

Igual que la demanda, también se presenta elasticidad en la oferta, debido a que es una magnitud, la respuesta de los productores ante un cambio en el precio no será siempre la misma [10] y mide la sensibilidad de la oferta ante el cambio de un precio. Un factor que influye en la estabilidad de la oferta es la probabilidad de sustituir los recursos con los que se producen los bienes y servicios, otro factor, es el tiempo que transcurre entre el aumento del precio y la respuesta de los vendedores [10], por ejemplo, el arroz tiene una oferta inelástica a corto plazo porque las hectáreas sembradas se determinan con anticipación.

Los bienes ofrecidos y la cantidad de demandante tienden a alcanzar el punto de equilibrio o un precio equilibrado, permitiendo una interacción que no afecta ni a la oferta ni a la demanda. Este equilibrio es el valor donde la cantidad ofrecida iguala a la cantidad demandada [10], por ejemplo, en los productos agrícolas lo ideal es no estar por debajo del punto de equilibrio porque generaría un desabastecimiento en la región o el país. Los cambios en el punto de equilibrio se dan por la variación en la demanda o en la oferta. En el caso de la demanda, si se pierde el equilibrio no hay incentivo para producir o consumir [10], en caso de la oferta, si aumenta los productos ofrecidos los precios tienden a disminuir, afectando a los oferentes, por ejemplo, el caso del arroz en Colombia que FEDEARROZ pide regular las hectáreas cultivadas del cereal para no presentar una sobreoferta.

El costo es el gastos necesario para llevar a cabo el proceso productivo [10], en otras palabras, el valor monetario de cada rubro necesario para la producción de un bien o ser-

vicio, en el caso del arroz, el costo por hectárea tiene en cuenta los valores de la asistencia técnica, del arriendo de predios, la preparación del terreno, del riego, de los fertilizantes, de la protección del cultivo, de la recolección, entre otros.

Dentro de los costos se identifican costos fijos, variables, directos, indirectos, entre otros. En el sector agropecuario intervienen el costo fijo que no varían con la producción, como el pago de arriendo. Otro es el costo variable que cambia según el volumen de producción, como la compra de herbicidas. Además, intervienen el costo directo e indirecto, el primero es indispensable en el proceso productivo, por ejemplo, los fertilizantes; el segundo, son costos que intervienen indirectamente en el proceso productivo y no son tan indispensables como la mano de obra indirecta.

La sumatoria de todos los costos determina el costo total y este a su vez es necesario para establecer el costo unitario, debido a que es el valor de la razón entre el costo total y la cantidad de unidad producida. Los precios del arroz paddy en el país están correlacionados con la evolución de los precios internacionales y con la estacionaridad de la producción nacional de arroz [14]. La estacionaridad se da por la relación entre el agricultor de arroz que fija el precio de venta y el molino que también tiene un precio de compra, pero en algunos casos el agricultor realiza un acuerdo de compra con el molino que almacena, descascara y limpia el cereal para que sea apto para el consumo humano.

3.4 Modelos estadísticos para series de tiempo

FEDEARROZ desde 1996 hasta la actualidad lleva registro por mes del precio promedio del arroz paddy, consolidando así una base de datos robusta que se ajusta a las características de una serie de tiempo, es decir, de un conjunto de observaciones de una variable medida en puntos sucesivos en el tiempo o en periodos de tiempo sucesivos [19].

Cada serie de tiempo tiene un patrón que es consecuencia del comportamiento de los componentes que intervienen en ella, en algunos casos ese patrón es causado por las variables, que son de dos tipos: endógena o exógena. La primera, es una variable con un valor determinado por las interacciones establecidas dentro del modelo, por ejemplo, el precio y la demanda; la segunda, no depende de valores actuales, pasados y futuros del termino de error [20] y que afectan el modelo, por ejemplo, el clima de una región.

En una serie de tiempo se presentan cuatro componentes: de tendencia, ciclicidad, estacionaridad y aleatoriedad. El primero es un desplazamiento gradual hacia valores relativamente altos o bajos a través de un lapso largo; el segundo, es toda sucesión recurrente de puntos que caiga abajo y arriba de la línea de tendencia y que dure más de un año; el

tercero, se usa para representar cualquier variación que se presente con regularidad en un lapso menor a un año; y el cuarto, es el factor residual que da cuenta de las desviaciones de los valores reales de la serie de tiempo, de los valores que se esperan al considerar los defectos de los componentes anteriores [19].

Para realizar pronósticos en una serie de tiempo existen varias metodologías, entre ellas están: Box-Jekins y Holt-Winters. La primera tiene que ver con el montaje de un modelo en la clase de modelos SARIMA (modelo estacional auto-regresivo integrado de medias móviles); la segunda, es un modelo de alisamiento o suavizamiento exponencial que se puede generalizar fácilmente para tratar las series temporales que contienen tendencias y variaciones estacionales [21]. Estas metodologías son usadas en Econometría, por tanto, los precios del arroz constituyen un problema econométrico porque se aplican metodologías estadísticas como las enunciadas anteriormente.

3.4.1 Modelo ARMA

El modelo ARMA está compuesto por procesos autoregresivos (AR) de orden p y procesos de medias móviles (MA) de orden q . El AR(1), matemáticamente se expresa así:

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t \quad (3.1)$$

Donde δ es la media de Y y u_t es un término de error aleatorio. Si la media es cero y la varianza constante, ruido blanco, se afirma que Y_t sigue un proceso estocástico autorregresivo de primer orden o AR(1), en otras palabras, el valor de Y en el tiempo t depende de su valor en el periodo anterior y de un término aleatorio [20]. Si los valores de Y están expresados como desviaciones de su valor medio, se dice que el valor de pronóstico de Y en el periodo t es alguna ($=\alpha_1$) de su valor en el periodo $(t-1)$ más una perturbación aleatoria en el tiempo t , asimismo, los valores de Y están expresados alrededor del valor de su media.

Un modelo AR(2) esta expresado por la ecuación:

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + u_t \quad (3.2)$$

En este modelo, el valor de Y en el tiempo de t depende de sus valores en los que dos periodos anteriores, los valores de Y expresados alrededor del valor de su media σ [20]. El modelo AR(p), en términos generales se expresa matemáticamente:

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \dots + \alpha_p(Y_{t-p} - \delta) + y_t \quad (3.3)$$

Por otro lado, un MA(1) es un modelo de Y que se describe por medio de la siguiente ecuación:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} \quad (3.4)$$

En la anterior expresión matemática, la μ es una constante y u es el termino del error estocástico (ruido blanco) y, debido a que Y en el periodo t es igual a una constante más un promedio móvil de los términos de error presente y pasado [20]. En términos generales un MA(q) es tan sólo una combinación lineal de términos de error de ruido blanco y su ecuación es:

$$Y_t = \mu + \beta_0 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (3.5)$$

Con base a lo anterior, un AR(p) y MA(q) juntos forman un ARMA(p,q), es decir, donde p es el orden de la parte autorregresiva y q es el orden de la parte de media móvil. El modelo ARMA(1,1) es el más simple, donde θ es una constante y se escribe:

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \quad (3.6)$$

En conclusión, un modelo ARMA(p,q) tendrá p términos autorregresivos y q términos de promedios móviles.

3.4.2 Metodología Box-Jenkins

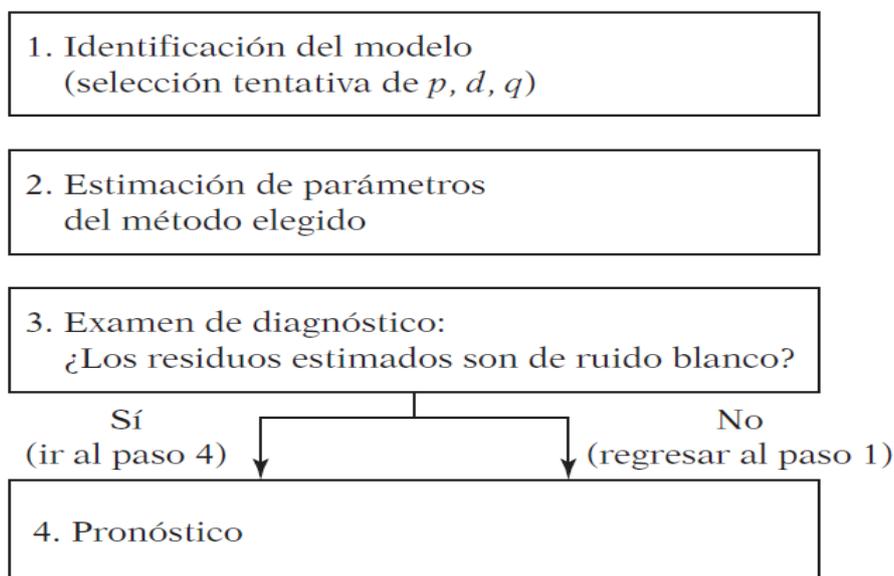
La metodología Box-Jenkins es conocida en el medio estadístico como ARIMA, es decir, proceso autorregresivo integrado de promedios móviles. Otra forma de expresarlo es ARIMA(p, d, q), donde p denota el número de términos autorregresivos, d el número de veces que la serie se diferencia para hacerse estacionaria y q el número de términos de promedios móviles. El objetivo de Box-Jenkins es

identificar y estimar un modelo estadístico que se interprete como generador de los datos muestrales. Entonces, si se va a pronosticar con este modelo estimado, debe suponerse que sus características son constantes a través del tiempo y, en particular, en periodos futuros. Así, la sencilla razón para requerir datos estacionarios es que todo modelo que se infiera a partir de estos datos pueda interpretarse como estacionario o estable en sí mismo, y proporcione, por consiguiente, una base válida para pronosticar [22].

La metodología estima cuatro pasos para elaborar un modelo ARIMA. El primer paso es la identificación, el segundo, es la estimación, el tercero, es el examen de diagnóstico y el cuarto paso es el pronóstico.

En la identificación se encuentran los valores adecuados de p , d y q con ayuda del correlograma y el correlograma parcial. La estimación establece los parámetros de los términos autorregresivos y de promedios móviles incluidos en el modelo, es decir, se selecciona el modelo ARIMA. El examen de diagnóstico consiste en evaluar si el modelo ARIMA se ajusta a los datos de forma razonablemente buena, en tal caso que no suceda se realiza el procedimiento de nuevo, por eso, se efectúa una prueba a los residuales del modelo para observar que son ruido blanco y determinar que el modelo ARIMA es el adecuado. El pronóstico se realiza con base a modelo ARIMA, que en términos generales es confiable para periodos a corto plazo. La imagen presenta una síntesis de la metodología Box-Jenkins:

Figura 3.1: Pasos de la metodología Box-Jenkins



Fuente: Guajarati

Cabe recordar que para utilizar la metodología Box-Jenkins, se debe tener una serie de tiempo estacionaria o una serie de tiempo que sea estacionaria después de una o más diferenciaciones [20], además, se recomienda que no se diferencie la serie más de dos veces.

3.4.3 Método Holt-Winters

El alisamiento o suavizamiento exponencial se puede generalizar fácilmente para tratar series de tiempo que contiene tendencia y variaciones estacionales [21]. Existen varios métodos de suavizamiento, tales como el suavizamiento exponencial simple, el método lineal de Holt y el método de Holt-Winters, así como sus variaciones. Para este trabajo, es de interés sólo el último de los tres nombrados anteriormente.

En el método Holt-Winters se presentan dos efectos: aditivo y multiplicativo. El primero, es más confiable cuando el patrón estacional en los datos no cambia conforme la serie incrementa o disminuye de valor; el segundo, se presenta cuando el patrón estacional en los datos depende del tamaño de los datos. Las ecuaciones para el Holt-Winters aditivo son:

Cuadro 3.1: Holt-Winters aditivo

Componente	Aditivo
Nivel	$S_t = \alpha(D_t - C_{t-T}) + (1 - \alpha)(S_{t-1} + G_{t-1})$
Tendencia	$G_t = \beta(S_t - S_{t-1}) + (1 - \beta)G_{t-1}$
Factor estacional	$C_t = \gamma(D_t - S_t) + (1 - \gamma)C_{t-T}$

De la tabla anterior, las constantes de alisado α , β , γ deben satisfacer únicamente valores comprendidos entre 0 y 1. A continuación se presentan las ecuaciones del Holt-Winters multiplicativo:

Cuadro 3.2: Holt-Winters multiplicativo

Componente	Multiplicativo
Nivel	$S_t = \alpha\left(\frac{D_t}{C_{t-T}}\right) + (1 - \alpha)(S_{t-1} + G_{t-1})$
Tendencia	$G_t = \beta(S_t - S_{t-1}) + (1 - \beta)G_{t-1}$
Factor estacional	$C_t = \gamma\left(\frac{D_t}{S_t}\right) + (1 - \gamma)C_{t-T}$

Se cumple la misma condición del efecto aditivo para las constantes de alisado α , β , γ en el multiplicativo. El método Holt-Winters está basado en un algoritmo iterativo que a cada tiempo (mes o semana) realiza un pronóstico sobre el comportamiento de la serie de tiempo. Dicho algoritmo consta de tres componentes: nivel, tendencia y estacionaridad. En el algoritmo cada uno de los componentes se ajusta, comparando la serie real con la pronosticada para el mismo lapso de tiempo. Este modelo se utiliza con frecuencia en empresas para predecir comportamientos a corto plazo.

Capítulo 4

Marco Metodológico

En este capítulo se explicará de donde se extraerá los datos para realizar el pronóstico del precio promedio del arroz paddy para los Llanos Orientales. Cabe aclarar que este trabajo es un estudio de orden documental, es decir, es un trabajo sobre bases de datos disponibles al público establecidas por otras instituciones.

4.1 Bases de datos

Para pronosticar el precio promedio del arroz paddy se tendrán en cuenta cuatro bases de datos que están disponibles al público, tres son tomadas de la Federación Nacional de Arroceros (FEDEARROZ) y la otra del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM).

4.1.1 Precio promedio mensual en pesos por tonelada del arroz paddy zona llanos

Esta base de datos se encuentra en la página web de FEDEARROZ. Al ingresar a www.fedearroz.com.co, se despliega la ventana principal de la página, en la parte superior se encuentran diferentes opciones tales como: La Federación, Servicios al Arroceros, Insumos Fedearroz, Publicaciones, Estadísticas Arroceras, Fondo Nacional del Arroz, El Arroz, Información Jurídica. Al ingresar a la opción Estadísticas Arroceras se despliega una pestaña con dos títulos, uno de series históricas y otro de estudios económicos. Se selecciona el primero y se elige precios. Después se despliega una ventana con los siguientes criterios: Producto, Desde, Hasta. En el primero se selecciona Arroz Paddy Zona Llanos (Villavicencio), en el segundo se escoge 1996 y en el último se elige 2018.

En la primera columna de la base de datos están los meses y en la primera fila los años desde 1996 hasta 2018. Los valores representan el precio promedio para el arroz paddy por tonelada y están en pesos colombianos (COP). Para esta base de datos se tiene la salvedad de que no es el precio solo para Villavicencio, sino que la seccional de

FEDEARROZ Villavicencio promedia el precio de las seccionales en el Llano, es decir, el precio del promedio del arroz paddy es la media de los precios en las seccionales de: Aguazul (Casanare), Yopal (Casanare), Acacias (Meta), Granada (Meta), y Villavicencio (Meta). Por esa razón en la base de datos aparece Zona Llanos (Villavicencio).

Figura 4.1: Precio promedio mensual en pesos colombianos por tonelada del arroz Paddy zona llanos(Villavicencio)



Fuente: FEDEARROZ

4.1.2 Valoración nominal por rubros de los costos por hectárea del arroz en Colombia

Esta base de datos se encuentra en la página web de FEDEARROZ. Al ingresar a www.fedearroz.com.co, se despliega la ventana principal de la página, en la parte superior se encuentran diferentes opciones tales como: La Federación, Servicios al Arrocero, Insumos Fedearroz, Publicaciones, Estadísticas Arroceras, Fondo Nacional del Arroz, El Arroz, Información Jurídica. Al ingresar a la opción Estadística Arroceras se despliega una pestaña con dos títulos, uno de series históricas y otro de estudios económicos. Se elige la primera y se selecciona costos. En esta última aparecerá las opciones: Zona, Sistema,

Semestre, Desde, Hasta. En la primera opción se escoge Llanos; en la segunda, secano o riego; en la tercera, semestre 1 o 2; en la cuarta, 1996; en la quinta, 2018.

La base de datos consta de 10 columnas, el nombre de cada columna, de izquierda a derecha, es: Años, Asistencia Técnica, Arriendos, Preparación Terreno, Riego, Fertilizantes, Protección al Cultivo, Recolección, Otros, Total. Los valores están en pesos colombianos (COP).

Figura 4.2: Valoración nominal por rubros de los costos por hectárea del arroz en Colombia



Fuente: FEDEARROZ

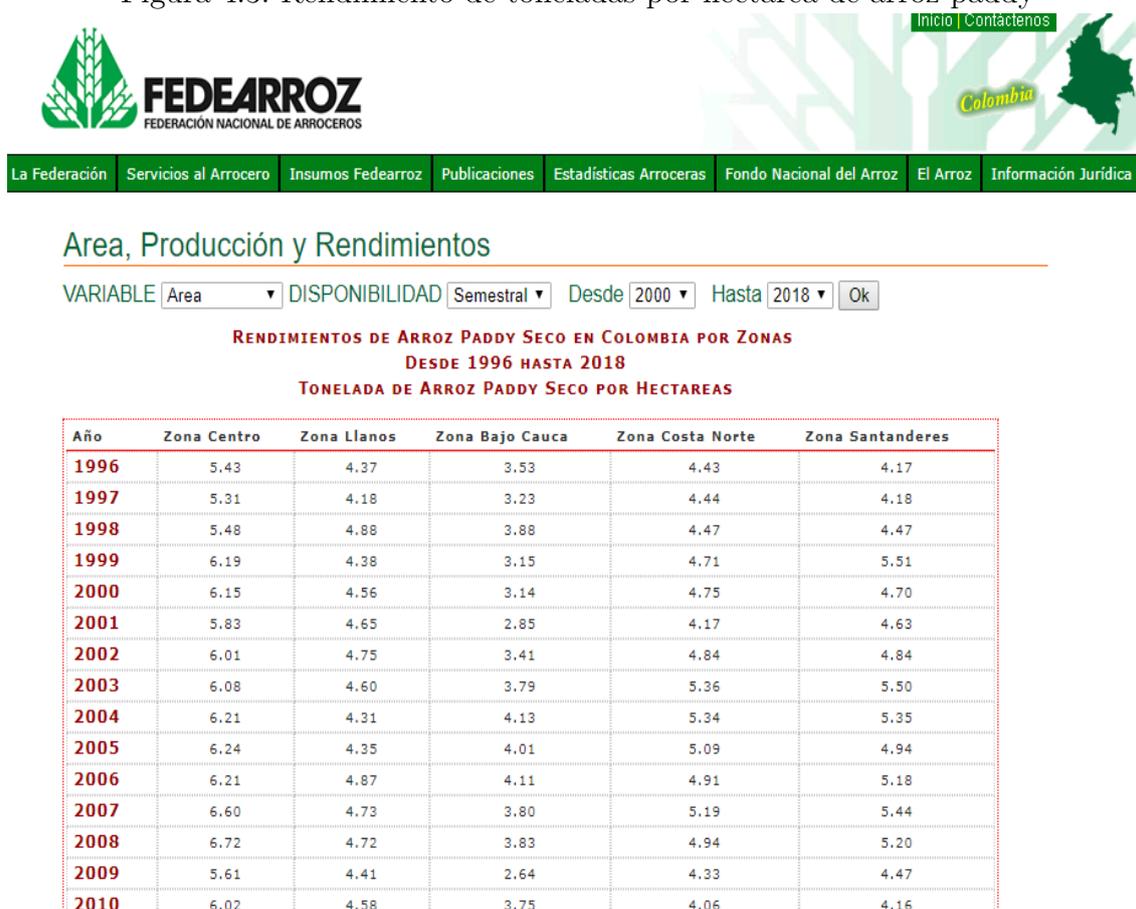
4.1.3 Rendimiento de toneladas por hectárea de arroz paddy

Esta base de datos se encuentra en la página web de FEDEARROZ. Al ingresar a www.fedearroz.com.co, se despliega la ventana principal de la página, en la parte superior se encuentran diferentes opciones tales como: La Federación, Servicios al Arroceros, Insumos Fedearroz, Publicaciones, Estadísticas Arroceras, Fondo Nacional del Arroz, El Arroz, Información Jurídica. Al ingresar a la opción Estadística Arroceras se despliega una pestaña con dos títulos, uno de series históricas y otro de estudios económicos. Se

elige la primera y se selecciona Área, Producción y Rendimientos. En esta última aparecerá cuatro criterios: Variable, Disponibilidad, Desde, Hasta. En el primero se selecciona rendimiento, en el segundo zona, en el tercero desde 1996 y en el último hasta 2017.

La base de datos tiene seis columnas organizadas en orden de izquierda a derecha así: Año, Zona Centro, Zona Llanos, Zona Bajo Cauca, Zona Costa Norte, Zona Santanderes. Los valores representan las toneladas de arroz paddy por hectáreas.

Figura 4.3: Rendimiento de toneladas por hectárea de arroz paddy



Fuente: FEDEARROZ

4.1.4 Promedio mensual de las precipitaciones en Villavicencio (mm)

La base de datos fue suministrada por el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) a través del servicio de solicitud de información por medio de la página web de la entidad, <http://www.ideam.gov.co>, con el radicado N° 20180713103405 de 12 de julio de 2018.

Esta base de datos tiene un encabezado principal que suministra información de los datos. Asimismo, la base de datos está estructurada por 16 columnas, organizada de izquierda a derecha de la siguiente manera: Año, Estación, Entidad, Enero, Febrero, Marzo, Abril, Mayo, Junio, Julio, Agosto, Septiembre, Octubre, Noviembre, Diciembre y, Valor Anual. La serie de la base de datos comienza en 1924, registrando datos hasta la fecha.

Figura 4.4: Promedio mensual de las precipitaciones en Villavicencio (mm)

VALORES TOTALES MENSUALES DE PRECIPITACION (mms)													SISTEMA DE INFORMACION NACIONAL AMBIENTAL			
FECHA DE PROCESO : 2018/07/25								ESTACION : 35035020					APTO VANGUARDIA			
LATITUD 0409 N		TIPO EST		SP		DEPTO		META			FECHA-INSTALACION		1924-OCT			
LONGITUD 7337 W		ENTIDAD		01 IDEAM		MUNICIPIO		VILLAVICENCIO			FECHA-SUSPENSION					
ELEVACION 0423 m.s.n.m		REGIONAL		03 META-GUAVIAR		CORRIENTE		GUATIQUIA								
A#O	EST	ENT	ENERO *	FEBRE *	MARZO *	ABRIL *	MAYO *	JUNIO *	JULIO *	AGOST *	SEPTI *	OCTUB *	NOVIE *	DICIE *	VR ANUAL *	
1980	2	01	125.8	39.0	225.9	447.2	467.8	708.9	511.8	336.7	352.5	440.7	249.1	198.0	4103.4	
1981	2	01	41.5	224.4	117.5	415.6	885.6	561.5	409.6	341.4	396.9	621.3	457.1	233.7	4706.1	
1982	2	01	19.5	66.4	251.0	964.1	576.0	359.2	507.7	454.8	398.4	624.2	397.4	182.7	4801.4	
1983	2	01	93.8	259.6	207.5	768.5	852.7	430.1	377.6	401.5	356.5	501.6	506.2	334.6	5090.2	
1984	2	01	255.3	294.3	157.1	520.6	416.9	557.5	341.1	520.6	486.5	501.6	303.2	75.6	4430.3	
1985	2	01	1.0	23.1	122.5	294.9	686.2	596.9	415.3	387.8	410.3	475.9	443.2	61.1	3918.2	
1986	2	01	35.6	156.6	178.6	445.8	747.4	541.6	583.3	386.8	376.2	584.5	530.3	93.5	4660.2	
1987	2	01	88.3	111.3	265.7	427.1	461.9	354.5	422.8	353.7	323.3	565.9	452.6	121.1	3948.2	
1988	1	01	1.1	16.4	167.0	390.7	525.2	537.6	370.0	484.7	246.6	392.9	495.8	160.7	3788.7	
1989	2	01	64.9	161.3	100.4	277.8	615.2	474.8	265.1	377.2	308.6	507.8	505.5	205.3	3863.9	
1990	1	01	82.6	325.2	304.4	630.7	640.6	424.8	413.8	234.3	240.5	308.3	519.7	340.3	4465.2	
1991	1	01	24.9	116.7	361.2	669.4	640.2	557.0	480.8	347.5	370.1	257.2	322.7	65.3	4213.0	
1992	1	01	54.0	36.5	212.1	338.2	670.8	518.1	401.5	473.7	538.1	385.7	329.3	215.2	4173.2	
1993	1	01	145.3	95.3	373.9	784.3	409.5	408.2	615.0	389.5	543.1	642.4	386.5	125.7	4918.7	
1994	1	01	99.8	100.8	333.3	596.2	674.9	367.0	519.5	365.9	597.6	756.1	450.1	74.3	4935.5	
1995	2	01	81.9	14.4	220.2	359.7	687.9	621.7	389.4	514.9	511.0	600.6	349.3	133.5	4484.5	
1996	2	01	24.3	209.4	198.3	542.6	877.1	499.5	487.0	368.9	526.6	466.7	711.1	280.2	5191.7	
1997	2	01	140.3	209.4	105.8	469.2	701.6	628.7	540.8	413.7	584.6	240.4	302.4	13.2	4350.1	
1998	1	01	36.3	215.8	402.0	606.8	706.2	599.1	518.6	502.3	347.5	482.3	365.7	234.0	5016.6	
1999	1	01	91.0	248.5	192.0	735.9	524.3	669.9	431.6	333.7	396.1	594.2	434.4	202.5	4854.1	
2000	1	01	64.0	163.9	104.2	379.9	923.6	351.8	333.6	354.6	200.6	288.3	371.0	145.8	3681.3	
2001	1	01	12.8	38.5	158.3	330.4	765.1	567.8	515.6	368.0	393.3	572.8	372.9	417.4	4512.9	
2002	1	01	9.8	7.9	365.4	611.1	897.0	635.2	614.3	341.7	477.5	606.3	201.2	222.6	4990.0	
2003	1	01	.4	21.5	146.5	621.6	400.8	694.0	307.2	417.3	632.8	399.2	547.6	214.4	4403.3	
2004	1	01	46.6	324.3	161.1	424.8	736.6	655.8	574.7	374.7	428.1	383.6	391.0	336.5	4837.8	
2005	1	01	95.9	278.9	246.8	594.1	780.3	588.8	561.5	394.5	350.7	531.3	609.9	139.0	5171.7	

Fuente: IDEAM

4.2 Tratamiento a las bases de datos

Con la base de datos principal se realizará un diagnóstico descriptivo de los datos y su representación gráfica. Con esos resultados se procederá a construir los modelos de pronóstico, empezando por la metodología Box-Jenkins y por último Holt-Winters para escoger el mejor modelo. Luego, se depurará la base de datos de la valoración nominal por rubros de los costos por hectárea del arroz para evaluar las variables que más inciden en los precios del cereal. Por último se estimará el valor de la pluviosidad en el cultivo de arroz con base al promedio mensual de las precipitaciones en la región de Villavicencio.

Después de haber realizado lo anterior se analizarán los resultados para pronosticar el precio promedio por tonelada del arroz paddy en los Llanos Orientales y los costos que más influyen a la hora de establecer el precio del cereal.

Capítulo 5

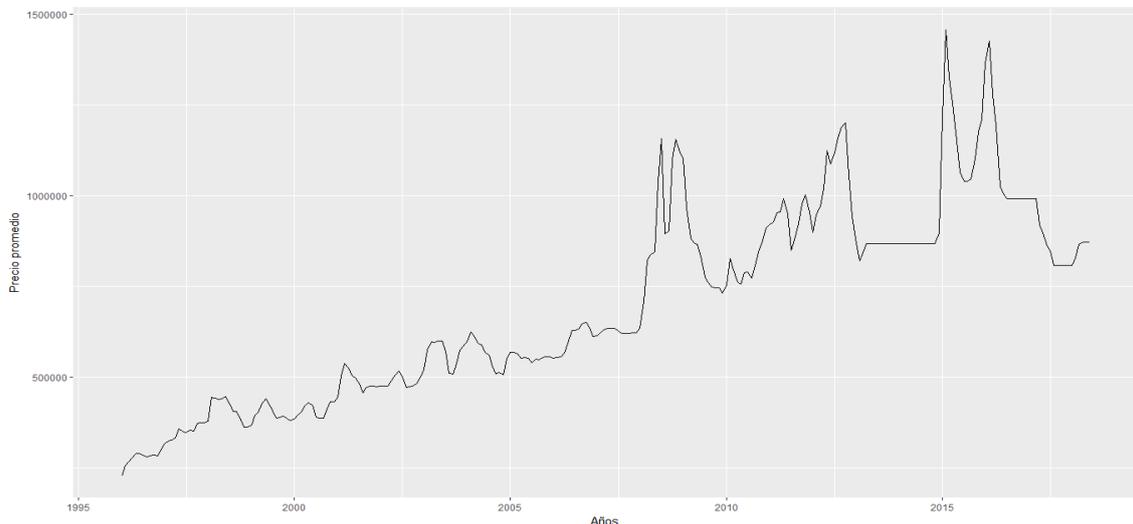
Análisis y Resultados

En este capítulo se realizará la construcción de los modelos de pronóstico, luego se depurará la base de costos que dejará en evidencia las variables que influyen a la hora de fijar el precio promedio por tonelada del arroz paddy en la región de los Llanos Orientales.

5.1 Construcción de los modelos de predicción

La serie de tiempo del precio promedio mensual por tonelada del arroz paddy en la región de los Llanos Orientales tiene 270 observaciones, comenzando desde Enero de 1996 hasta Junio de 2018. A continuación se presentan la gráfica de las observaciones:

Figura 5.1: Precio promedio mensual por tonelada del arroz paddy en los Llanos Orientales



Fuente: Autor

De la gráfica anterior, entre el año 1996 a 2007 hay un comportamiento cíclico de los datos, a partir del 2007 se evidencia un aumento considerable en el precio promedio, que luego decae y aumenta nuevamente, generando un cambio brusco en la varianza de las observaciones. Dicho comportamiento se repite en los meses siguiente a excepto de algunos

donde el precio es constante, dando indicios que la serie no es estacionaria.

Para constatar lo anterior, se realiza los cálculos descriptivos de las observaciones. En la siguiente tabla se muestran los resultados:

Cuadro 5.1: Tabla de estadísticos

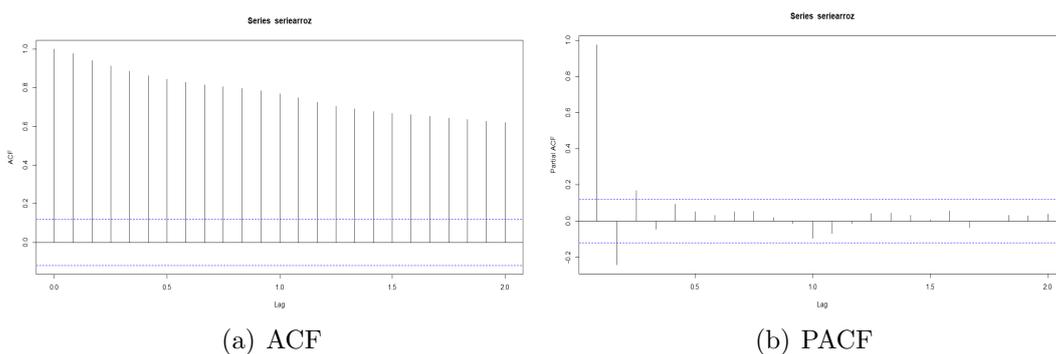
Estadístico	Valor
Media	693998
Varianza	70903661566
Desviación Estándar	266277.4
Valor Mínimo	227800
Valor Máximo	1458000

En el cuadro anterior se observa que la desviación estándar es el 34 % de la media aritmética, demostrando que los datos están dispersos como se evidencia en la gráfica y no son estacionarios. Con la prueba de raíz unitaria Dickey-Fuller se confirma que la serie carece de estacionaridad debido a que el p-valor es de 0.0941, siendo mayor a 0.05.

5.1.1 Modelo ARMA

Para comenzar a construir este modelo, se graficará la Función de Autocorrelación (ACF) y la Función Parcial de Autocorrelación (PACF) para evidenciar los rezagos con los que se relacionan los datos con el pasado. Las dos gráficas muestran esos retardos:

Figura 5.2: ACF y PACF de la serie de tiempo



Fuente: Autor

Con base a las gráficas anteriores, en la función ACF se observa que la serie tiene un decrecimiento exponencial debido al comportamiento de las observaciones, asimismo, la gráfica PACF arroja que los datos alcanzan una estacionaridad en el primer rezago.

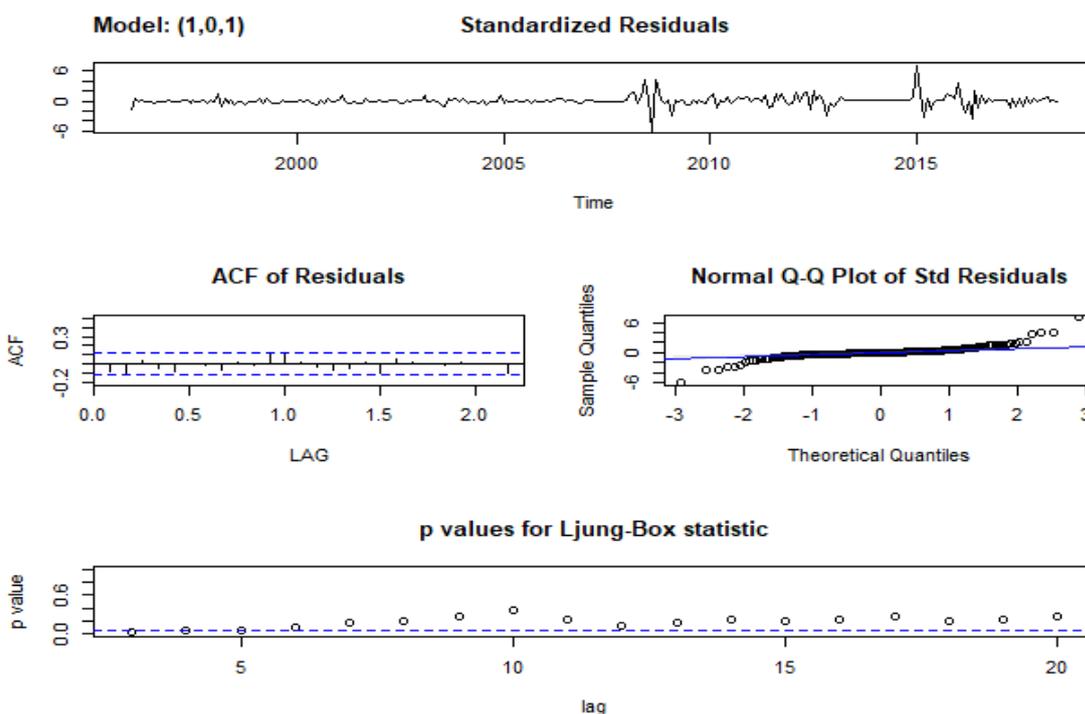
Con esa información, se proponen dos modelos: ARMA(1,1) y ARMA (2,2) y se seleccionan los valores del criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC) para determinar el modelo confiable para el pronóstico del precio promedio por tonelada del arroz paddy en la región de los Llanos Orientales:

Cuadro 5.2: AIC y BIC de los modelos ARMA(p,q)

Modelo	AIC	BIC
ARMA(1,1)	6548.713	6563.106
ARMA(2,2)	6548.552	6570.142

El AIC y BIC dan aviso sobre el modelo que se usará para el pronóstico de datos, en ese orden se selecciona el BIC que tenga el valor más pequeño, siendo el ARMA(1,1). A continuación las gráficas muestran el comportamiento de los residuales del modelo seleccionado:

Figura 5.3: Comportamiento de los residuales



Fuente: Autor

De la figura anterior se observa que los residuales evidencian una estacionaridad, comprobando lo evaluado por el test de Dickey-Fuller. Además, el gráfico ACF muestra que las observaciones están dentro de los límites establecidos. Asimismo, los residuales tienen una tendencia a la normalidad y el test de Ljung-Box refleja que la mayoría de los p-value

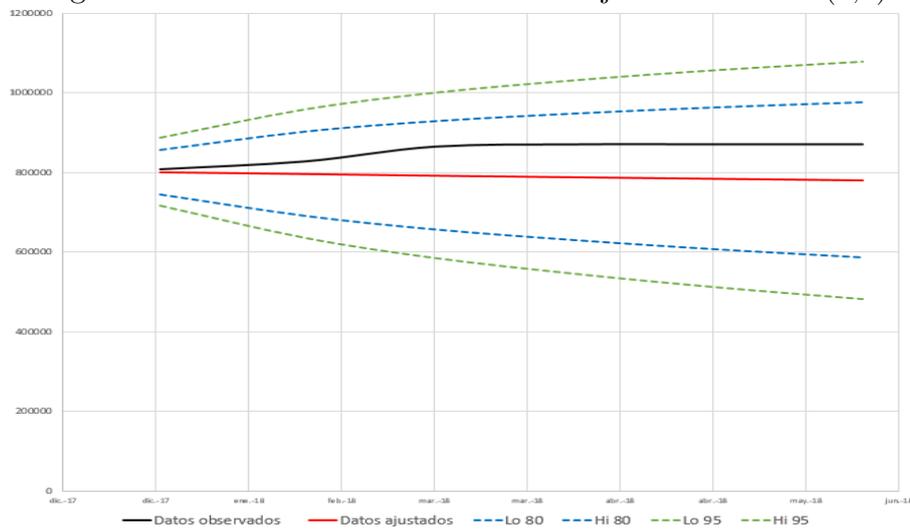
están por encima del límite de rechazo, es decir, existe evidencia estadísticamente significativa que los residuos del modelo son ruido blanco gaussiano, por lo cual, se procede a realizar el pronóstico de datos. Por tanto, se elimina parcialmente los últimos seis datos de la serie original y se aplica el modelo, luego se comparan las observaciones iniciales con las pronosticadas. La siguiente tabla muestra los resultados:

Cuadro 5.3: Datos observados y datos ajustados ARMA(1,1)

Tiempo	Datos observados	Datos ajustados
Enero 2018	808000	800013.1
Febrero 2018	828000	795313.6
Marzo 2018	866000	790774.7
Abril 2018	872000	786391.0
Mayo 2018	872000	782157.2
Junio 2018	872000	778068.1

A continuación se presenta la gráfica de los datos observados con los datos ajustados para visualizar los resultados:

Figura 5.4: Datos observados VS datos ajustados ARMA(1,1)



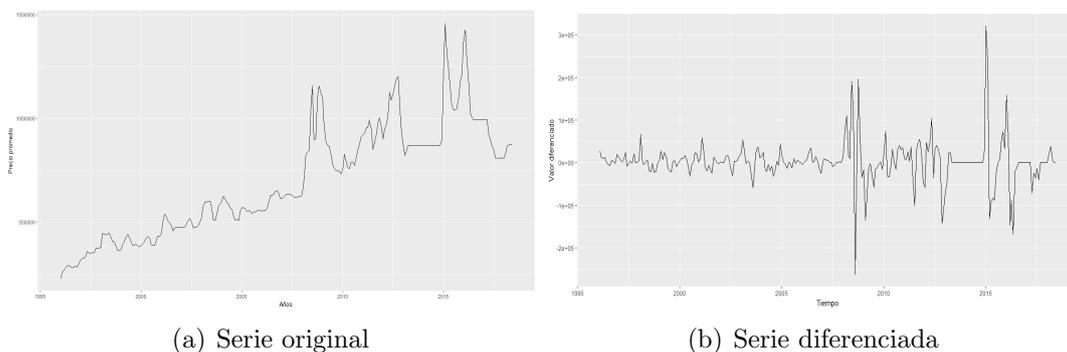
Fuente: Autor

Los datos ajustados por el modelo ARMA(1,1) mantienen una tendencia decreciente a comparación de los observaciones iniciales ya que su variación es creciente. Pero los precios ajustados por el modelo y los observados se encuentran dentro de los intervalos de confianza que determinan el modelo.

5.1.2 Modelo ARIMA

Para volver la serie estacionaria se aplica el test de Kwiatkowski–Phillips–Schmidt–Shin (KPSS) y Phillips-Perron (PP) para determinar sólo una diferenciación de los datos, siendo consecuente con lo observado en la gráfica PACF que se usó para el modelo anterior. En seguida se presenta un paralelo entre la serie original y la serie diferenciada una vez:

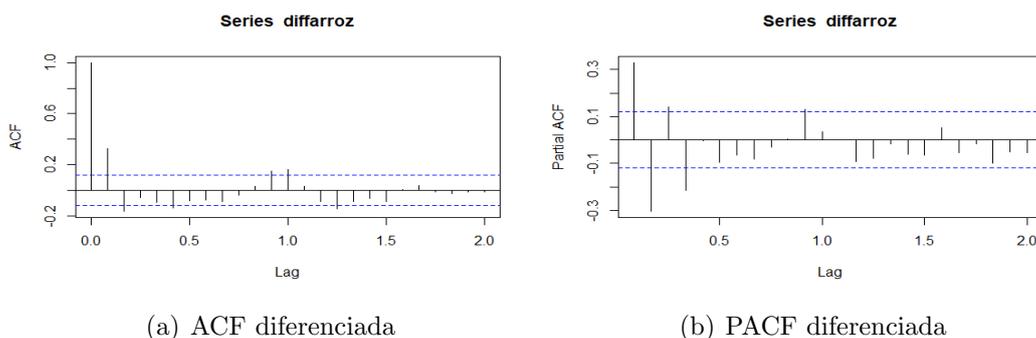
Figura 5.5: Serie original y serie diferenciada



Fuente: Autor

Con la serie diferenciada se observa un comportamiento de Ruido Blanco a comparación de la serie inicial, lo que implica graficar el ACF y PACF para observar el comportamiento de las observaciones diferenciadas.

Figura 5.6: ACF Y PACF de serie diferenciada



Fuente: Autor

La función ACF y PACF muestran que hay un rezago significativo y una visible estacionaridad de la serie, por tanto, se aplica del tes de Dickey-Fuller para comprobar lo anterior. Las dos hipótesis a considerar son:

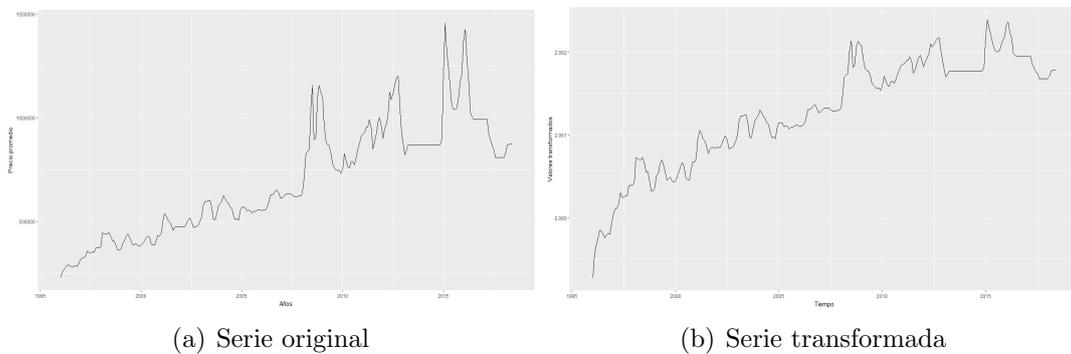
H_0 : Los datos no son estacionarios.

Ha: Los datos son estacionarios.

El test arroja que el valor de p-value es de 0.01, siendo menor que 0.05, por tanto se acepta la hipótesis alterna, es decir, los datos son estacionarios.

Además, de la versión diferenciada de la serie, se realizará la transformación de Box-Cox, $1/Y^{0.5}$, debido a que el valor de lambda es -0.4843741, para ver el comportamiento de la varianza. Las dos gráficas muestran la serie original y la serie de las observaciones transformadas con Box-Cox.

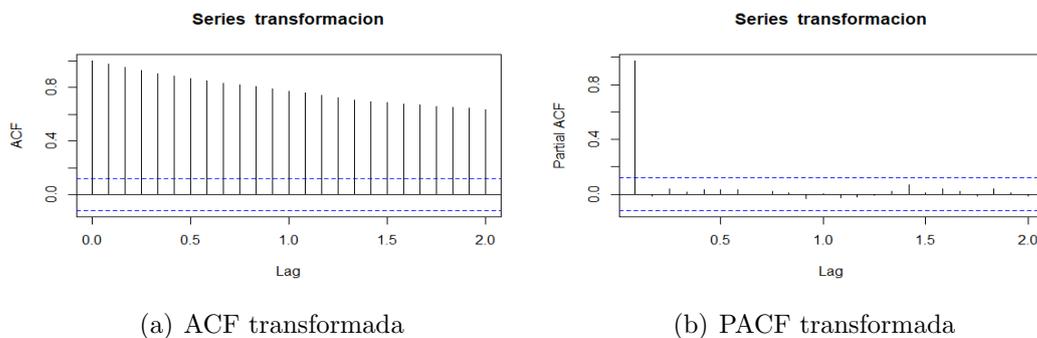
Figura 5.7: Serie original y serie transformada con Box-Cox



Fuente: Autor

Como lambda en la transformación Box-Cox es negativa, la gráfica de la serie transformada tiene esa forma y al aplicar el test Dickey-Fuller, se tiene que el valor de p-value es de 0.2939, mayor al 0.05 y por tanto se concluye que la serie no es estacional. La función ACF y PACF para los datos transformados con Box-Cox corroboran lo evaluado en test de Dickey-Fuller:

Figura 5.8: ACF Y PACF de serie transformada con Box-Cox



Fuente: Autor

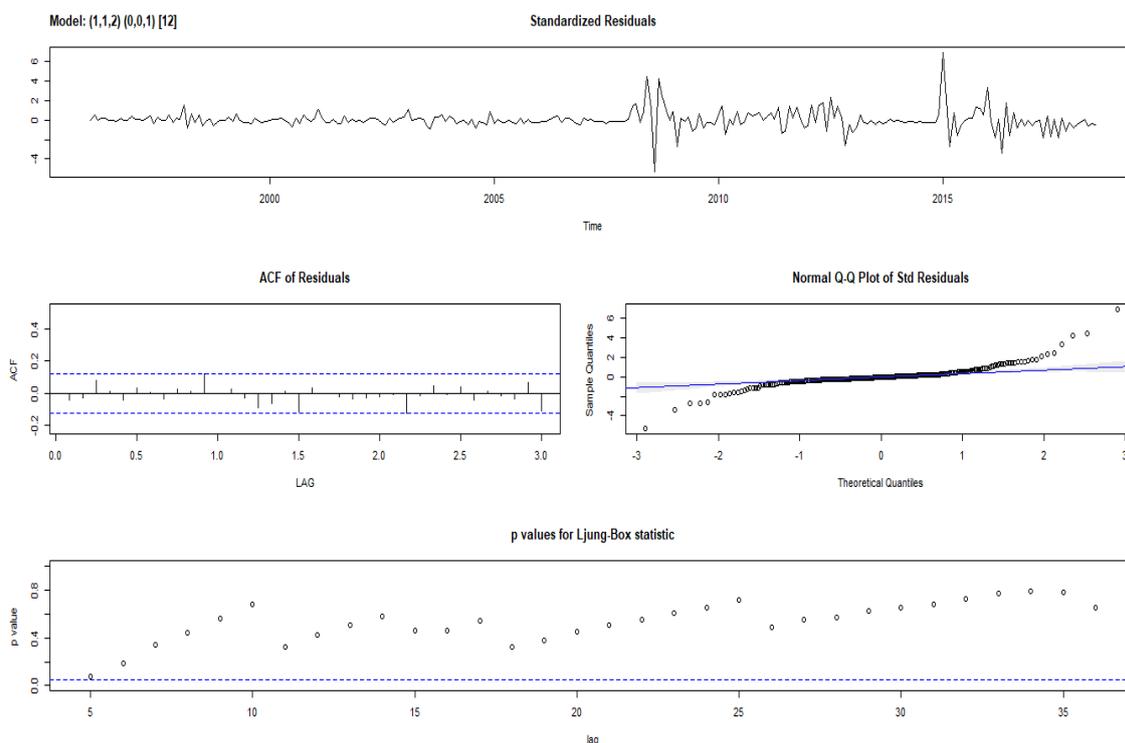
Con base a la información, y optando por el método de diferenciación se plantean tres modelos y uno de ellos se elegirá teniendo en cuenta el valor del AIC y BIC. Los modelos son: a) ARIMA(1,1,2)x(0,0,1)[12]. b) ARIMA(1,1,1)x(0,0,1)[12]. c) ARIMA(2,1,2)x(0,0,1)[12]. Los valores de cada modelo del AIC y BIC son:

Cuadro 5.4: Valores de AIC, BIC

Modelo	AIC	BIC
ARIMA(1,1,2)x(0,0,1)[12]	6511.918	6529.892
ARIMA(1,1,1)x(0,0,1)[12]	6516.831	6531.21
ARIMA(1,1,1)x(0,0,1)[12]	6512.587	6534.155

De acuerdo con la información de la tabla se escoge el modelo ARIMA(1,1,2)x(0,0,1)[12], debido a que los valores del AIC y BIC son menores con respecto a los otros dos modelos. Para confirmar el modelo se observa el comportamiento de los residuales:

Figura 5.9: Comportamiento de los residuales ARIMA



Fuente: Autor

La imagen anterior nos exhibe cuatro gráficas, la primera muestra los valores estandarizados de los residuales, la segunda, confirman que los residuales están dentro de los límites establecidos, la tercera, la distribución normal de los datos y, la última el test de

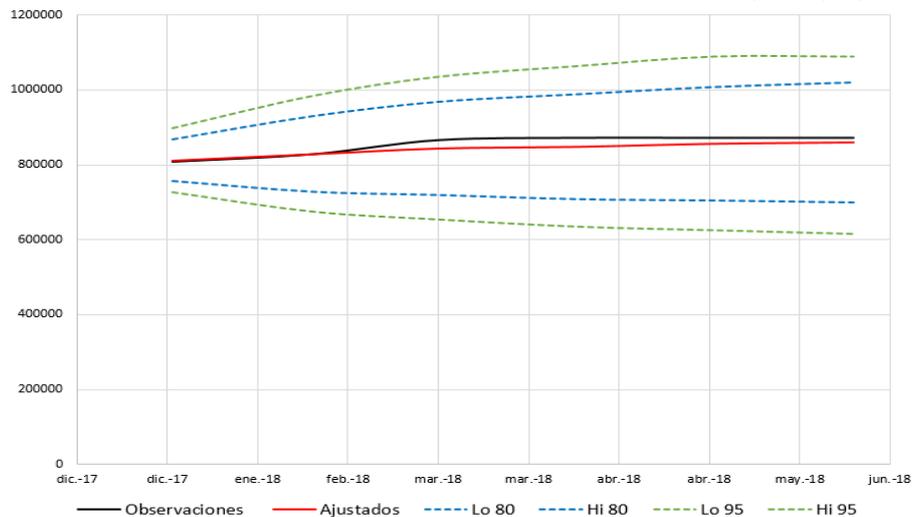
Ljung-Box para los p-value reflejando que los residuales están por encima de la zona de rechazo. Con base en lo anterior, se eliminaran 6 valores de la serie original para realizar un diagnóstico con el modelo escogido y observar el comportamiento de la predicción con respecto a las observaciones iniciales. Dichos resultados son:

Cuadro 5.5: Datos observados y datos ajustados ARIMA(1,1,2)x(0,0,1)[12]

Tiempo	Datos observados	Datos ajustados
Enero 2018	808000	812002.8
Febrero 2018	828000	828730.2
Marzo 2018	866000	843624.9
Abril 2018	872000	848211.5
Mayo 2018	872000	856061.8
Junio 2018	872000	859639.2

Los resultados de la tabla se observan mejor en la siguiente gráfica donde se muestra una tendencia decreciente de los valores diagnosticados por el modelo ARIMA a diferencia de las observaciones iniciales.

Figura 5.10: Datos observados VS datos ajustados ARIMA(1,1,2)x(0,0,1)[12]



Fuente: Autor

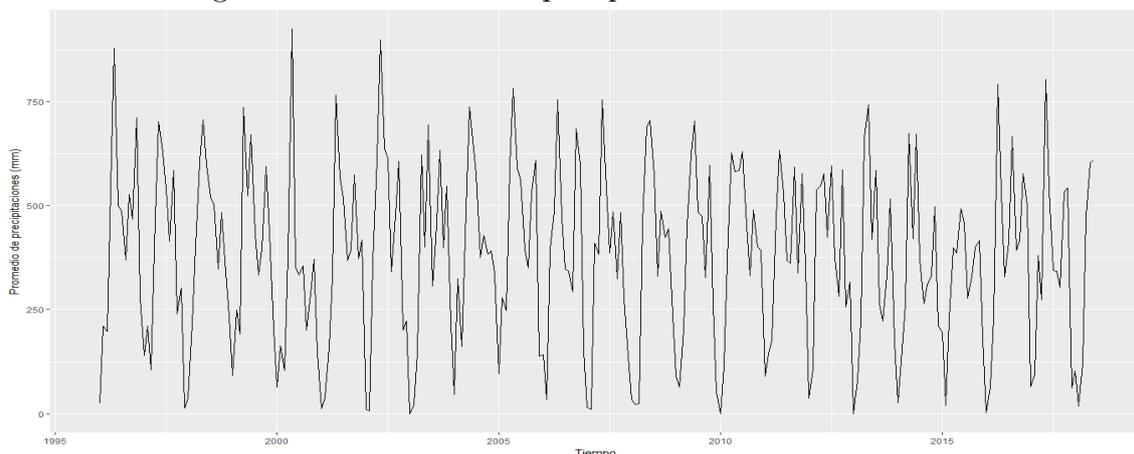
Según la gráfica anterior, los precios ajustados y los precios observados están dentro de los intervalos de confianza que el modelo determina.

5.1.3 Modelo SARIMAX

Para construir este modelo se utilizan los modelos de prueba de la Metodología Box-Jenkins, es decir, ARIMA(1,1,2)x(0,0,1)[12], ARIMA(1,1,1)x(0,0,1)[12] y ARIMA(2,1,2)x(0,0,1)[12]. Además, se tiene en cuenta una variable exógena para que realice la regresión

en la serie de tiempo del precio promedio del arroz paddy. Básicamente estos componentes, modelos ARIMA y variable exógena, conforman el modelo SARIMAX. La variable exógena escogida para el modelo es el promedio mensual de la precipitación medida en milímetros. Dicho promedio muestra el siguiente comportamiento:

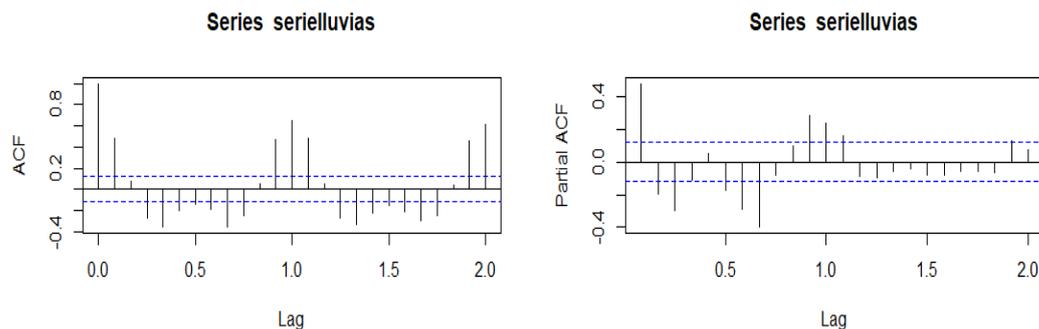
Figura 5.11: Promedio de precipitaciones en Villavicencio



Fuente: Autor

Al aplicar el test de Dickey-Fuller a las observaciones de la serie de tiempo del promedio de las precipitaciones, el p-value es menor a 0.05, dando a entender que la serie es estacionaria, información que se confirmará con el ACF y PACF:

Figura 5.12: ACF Y PACF de precipitaciones



(a) ACF de precipitaciones

(b) PACF de precipitaciones

Fuente: Autor

En las dos gráficas se observa que el ACF continua con la tendencia de la serie, pero el PACF muestra que hasta el segundo rezago las observaciones están dentro de los límites permitidos.

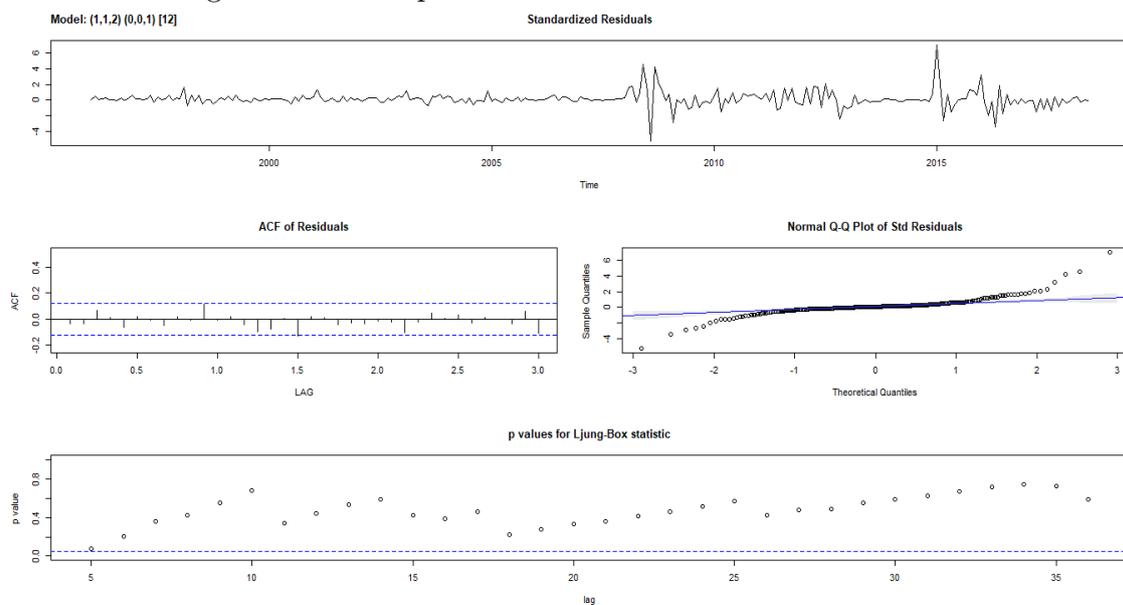
Con respecto a los Modelos ARIMA se aplica la variable exógena sin diferenciar, arrojando los siguientes resultados en AIC y en el BIC:

Cuadro 5.6: Valores de AIC, BIC de Modelo SARIMAX

Modelo	AIC	BIC
SARIMAX(1,1,2)x(0,0,1)[12]	6489.46	6511.006
SARIMAX(1,1,1)x(0,0,1)[12]	6494.701	6512.656
SARIMAX(1,1,1)x(0,0,1)[12]	6490.338	6515.475

Con base en la información de la tabla, el modelo SARIMAX que más se ajusta a las observaciones es el ARIMA(1,1,2)x(0,0,1)[12], debido a que los resultados del AIC y BIC son más pequeños a comparación de los otros dos modelos. A continuación se presenta el comportamiento de los residuales:

Figura 5.13: Comportamiento de los residuales SARIMAX



Fuente: Autor

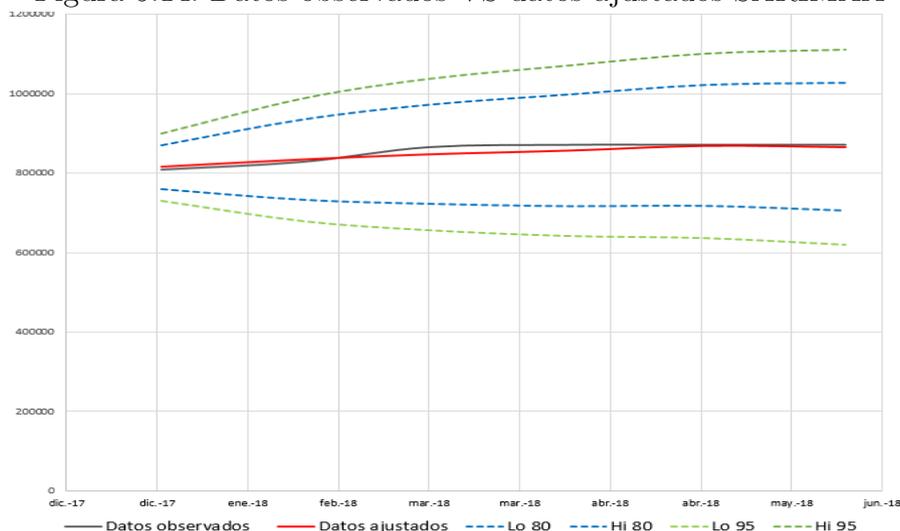
De las gráficas anteriores, se deduce que el comportamiento de los residuales está dentro de los parámetros permitidos, es decir, los residuales muestran un comportamiento de ruido blanco, asimismo, la ACF muestra que todos los valores están dentro de los límites permitidos, además, algunos valores están dispersos con respecto a la línea de normalidad y el test de Ljung-Box muestra que los valores no están en zona de rechazo. Con esa información se realizará el pronóstico de los datos, utilizando el algoritmo que se empleó para los modelos anteriores:

Cuadro 5.7: Datos observados y datos ajustados SARIMAX

Tiempo	Datos observados	Datos ajustados
Enero 2018	808000	815248.4
Febrero 2018	828000	833907.3
Marzo 2018	866000	847390.5
Abril 2018	872000	857136.7
Mayo 2018	872000	869371.6
Junio 2018	872000	866063.3

Los valores ajustados del modelo SARIMAX son cercanos a los valores de la serie original. La siguiente gráfica refleja mejor el comportamiento de los datos ajustados con respecto a las observaciones iniciales:

Figura 5.14: Datos observados VS datos ajustados SARIMAX



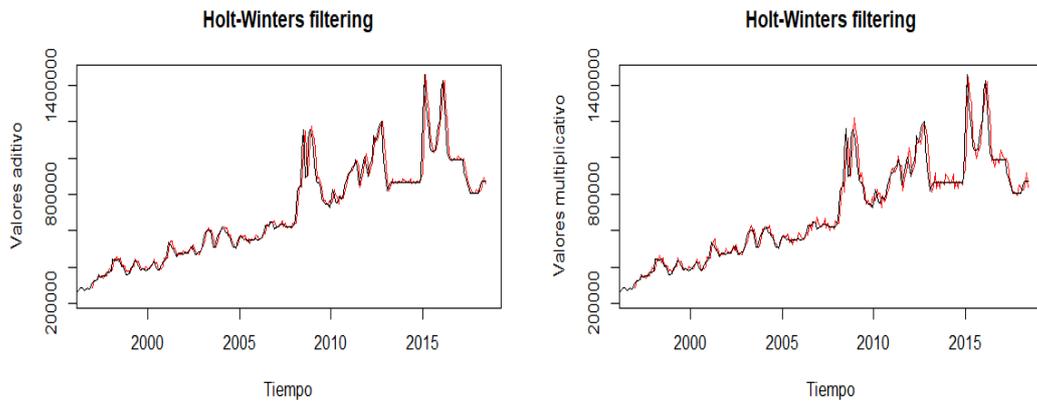
Fuente: Autor

La gráfica refleja que los valores pronosticados por el modelo se acercan de los resultados iniciales, además, el valor del precio promedio esta dentro de los intervalos de confianza.

5.1.4 Modelo Holt-Winters

El método de suavizamiento exponencial, Holt-Winters, es aditivo o multiplicativo. Para ello se gráfica la serie de observaciones iniciales y el componente estacional aditivo o multiplicativo y, con base a la suma de los cuadrados de error (ESS) se deduce el tipo de componente. Las gráficas siguientes reflejan cuando es aditivo o multiplicativo:

Figura 5.15: Holt-Winters aditivo y multiplicativo



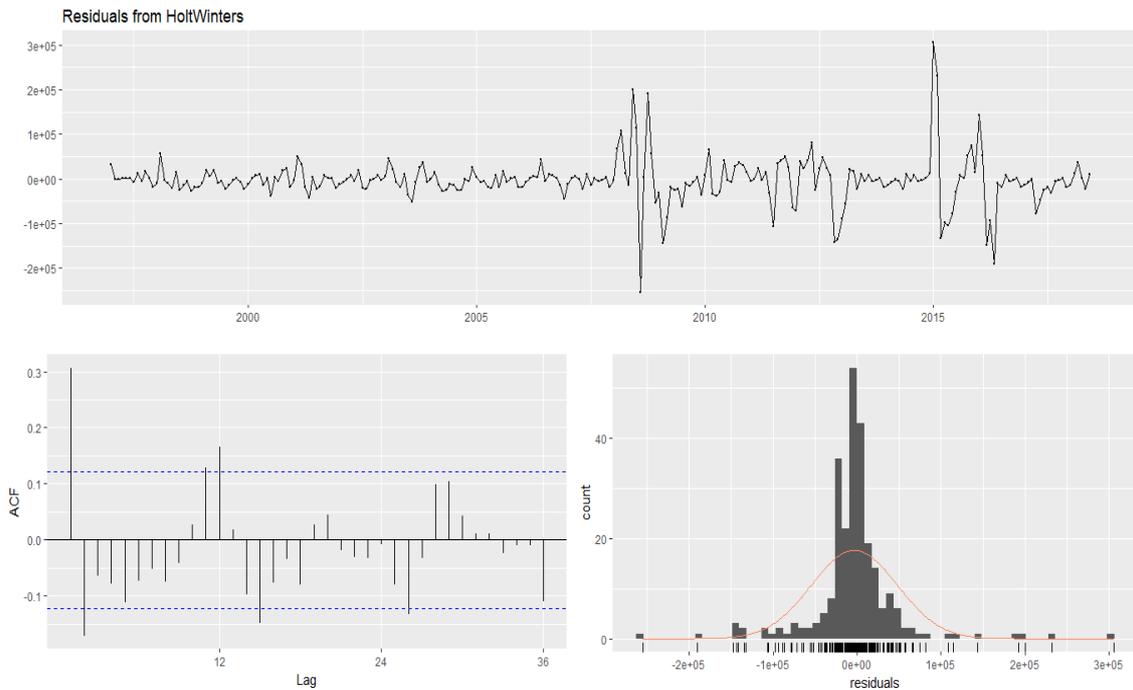
(a) Holt-Winters aditivo

(b) Holt-Winters multiplicativo

Fuente: Autor

Para afirmar si el modelo es aditivo o multiplicativo se recurre a la suma de los cuadrados de error (ESS), donde el valor menor indica el componente estacional. El ESS aditivo tiene un valor de $6.71855e+11$ y el ESS multiplicativo de $7,478579e+11$, por tanto, el componente es estacional aditivo. A continuación se evidencia el comportamiento de los residuales:

Figura 5.16: Comportamiento de los residuales en Holt-Winters aditivo



Fuente: Autor

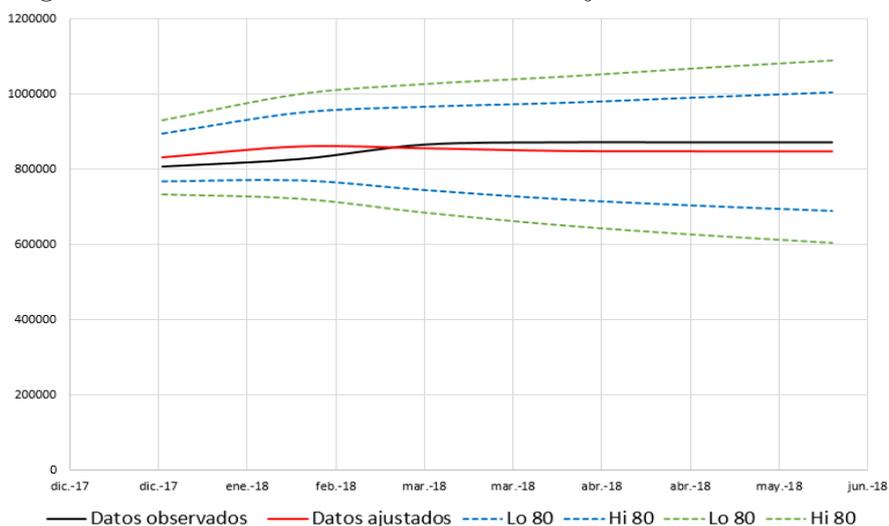
Los residuales de este modelo, después de aplicar el suavizamiento exponencial, muestran que en la gráfica ACF hay una aparente autocorrelación, esta información se corrobora con el test de Ljung-Box que arroja un resultado del p-value de 2.636e-08, evidenciando la autocorrelación. A continuación se realizará el pronóstico siguiendo el mismo procedimiento que se llevó para los modelos anteriores. Dichos valores pronosticados se contrastan con las observaciones iniciales, como muestra el siguiente cuadro:

Cuadro 5.8: Datos observados y datos Holt-Winters

Tiempo	Datos observados	Datos ajustados
Enero 2018	808000	831793.0
Febrero 2018	828000	861032.0
Marzo 2018	866000	856163.9
Abril 2018	872000	849104.3
Mayo 2018	872000	847933.1
Junio 2018	872000	847826.6

De acuerdo con los datos arrojados en la tabla, se procede a graficar para ver el comportamiento de los datos y la tendencia de estos:

Figura 5.17: Datos observados VS datos ajustados Holt-Winters



Fuente: Autor

La gráfica anterior muestra que los datos ajustados por el modelo Holt-Winters siguen de cerca el comportamiento de los datos observados y están dentro de los intervalos de confianza que el mismo modelo arroja.

5.2 Elección del mejor modelo construido y pronóstico del precio promedio por tonelada del arroz paddy para la región de los Llanos Orientales

Los modelos tienen unos parámetros que indican el comportamiento de este con respecto a las observaciones históricas. Asimismo, un modelo con muchos parámetros puede llegar a ser fiable pero que no sirve como modelo predictivo, debido a que es demasiado específico. En cambio, un modelo con pocos parámetros es más fácil de controlar y puede ser aplicable en otras situaciones.

En el caso del modelo ARMA(1,1), los parámetros son:

Cuadro 5.9: Parámetros ARMA(1,1)

Variable	Coefficiente	Error estándar
ϕ	0.9666	0.0155
θ	0.6287	0.0532
intercept	671775.8	117617.0
σ^2	1.912e+09	NULL

En el cuadro anterior, ϕ representa los procesos autorregresivos y θ los de medias móviles.

El otro modelo candidato a pronosticar el precio promedio del arroz paddy para los Llanos Orientales es el ARIMA(1,1,2)x(0,0,1)[12], cuyos parámetros son:

Cuadro 5.10: Parámetros ARIMA(1,1,2)x(0,0,1)[12]

Variable	Coefficiente	Error estándar
ϕ_1	0.7983	0.0677
θ_1	-0.2839	0.0707
θ_2	-0.6066	0.0511
Θ_1	0.1003	0.0611
σ^2	1.834e+09	NULL

Con respecto a la información anterior, ϕ y θ representan, respectivamente, los procesos autorregresivos y los de medias móviles, además, Θ representa los procesos de medias móviles estacionales. Los subíndices que acompañan a cada letra son el número de procesos.

Por otro lado, el modelo SARIMAX(1,1,2)x(0,0,1)[12] incluye una variable externa al modelo, denominada exógena. Con esa consideración, los parámetros de este modelo son:

Cuadro 5.11: Parámetros SARIMAX(1,1,2)x(0,0,1)[12]

Variable	Coefficiente	Error estándar
ϕ_1	0.8004	0.0664
θ_1	-0.2820	0.0691
θ_2	-0.6101	0.0506
Θ_1	0.1010	0.0607
xreg	-10.0832	8.8615
σ^2	1.832e+09	NULL

La misma aclaración que se ha hecho para los modelos anteriores de ϕ , θ , Θ , se cumple para este modelo, igual que los subíndices. Además, *xreg* representa la variable exógena.

Por último, para el Holt-Winters, los parámetros de suavizado y los coeficientes son:

Cuadro 5.12: Parámetros Holt-Winters

Variable	Coefficiente	Variable	Coefficiente
α	1	s4	-10574.306
β	0.0002802993	s5	-17745.139
γ	0	s6	-4107.639
a	867145.139	s7	4929.861
b	5486.435	s8	7396.528
s1	4967.361	s9	3438.194
s2	-8032.639	s10	3346.528
s3	-8336.806	s11	19863.194

Con los parámetros expuestos de cada modelo y con base a la información del error de la raíz cuadrada media (RMSE) que calcula el error cuadrático medio entre dos vectores numéricos; el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC), se escoge el mejor modelo. El siguiente cuadro contiene los RMSE, AIC, y BIC de cada modelo:

Cuadro 5.13: RMSE, AIC y BIC de cada modelo

Modelo	RMSE	AIC	BIC
ARMA(1,1)	72762.56	6548.713	6563.106
ARIMA(1,1,2)x(0,0,1)[12]	15758.11	6511.918	6529.892
SARIMAX(1,1,2)x(0,0,1)[12]	10776.67	6489.46	6511.006
Holt-Winters	23950.47	7163.291	7228.063

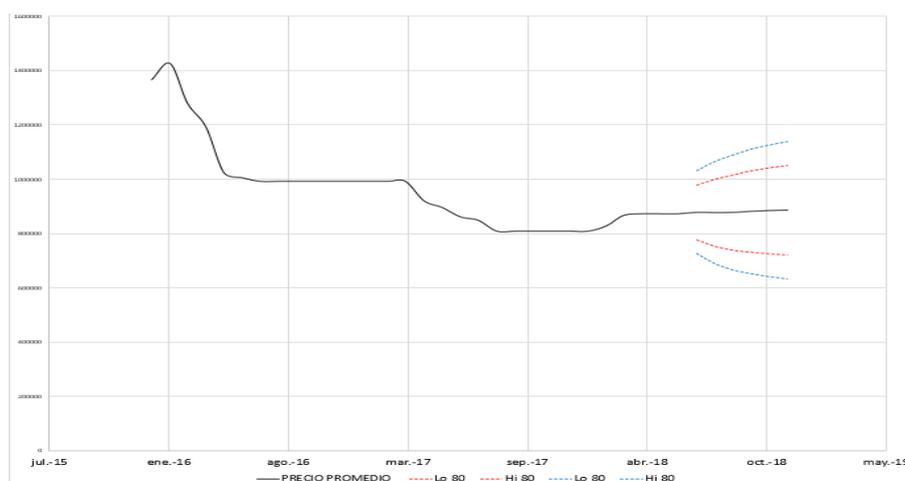
El modelo de pronóstico que mejor predice el precio promedio por toneladas del arroz paddy para la región de los Llanos Orientales es el SARIMAX(1,1,2)x(0,0,1)[12], debido a que el AIC, el BIC y el RMSE, son menores con respecto a los otros modelos. Para tener más certeza del modelo escogido, se realiza la prueba de Ljung-Box, donde el p-value es 0.682, por lo que no se rechaza la hipótesis nula de ruido blanco. Como ya se determinó el modelo, se predecirá el precio del cereal desde julio hasta diciembre de 2018 y luego se mostrará la gráfica de predicción.

Cuadro 5.14: Precio promedio por tonelada del arroz paddy en los Llanos Orientales desde julio hasta diciembre de 2018

Tiempo	Precio promedio
Julio 2018	877609.8
Agosto 2018	876676.8
Septiembre 2018	877387.2
Octubre 2018	881683.2
Noviembre 2018	884405.0
Diciembre 2018	886100.6

La siguiente gráfica visualiza el comportamiento del precio promedio por tonelada para el arroz paddy en la zona de los Llanos Orientales entre enero de 2016 hasta diciembre de 2018, incluyendo los intervalos de confianza que propone el modelo SARIMAX. La franja roja muestra que el precio está en un intervalo de confianza del 80 % y la franja azul del 95 %. Es decir, aunque el precio promedio real en ese tiempo varié y no sea igual al pronosticado, se espera que no exceda los límites de los intervalos de confianza.

Figura 5.18: Precios promedio de la tonelada del Arroz Paddy para la región de los Llanos Orientales



Fuente: Autor

Para finalizar, la gráfica muestra el comportamiento de los datos pronosticados con una tendencia similar a la que tienen los datos históricos de 2018, es decir, que hay un aumento el precio promedio del cereal para esta región del país. También, visualiza el comportamiento histórico del promedio de los precios para 2018 con los precios pronosticados por el modelo SARIMAX(1,1,2)x(0,0,1)[12].

5.3 Depuración de la base de datos de costos por hectárea para el arroz paddy en la zona de los Llanos

La base de datos de costos tiene las siguientes variables: asistencia técnica, arriendos, preparación terreno, riego, fertilizantes, protección al cultivo, recolección, otros y total. Los costos tienen su influencia en el momento de fijar el precio por tonelada del arroz paddy en los Llanos Orientales, esos costos están por hectáreas, entonces, con ayuda del rendimiento del promedio anual de toneladas por hectárea se calcula el costo de una tonelada de arroz paddy. Con este procedimiento, se escogen los costos con mayor influencia en el precio del arroz paddy.

Al correr el análisis del sistema de regresión múltiple, se obtiene que el costo fertilizante es altamente significativa al 100 %, el costo protección del cultivo es significativo al 99 % y el costo otros es significativo al 95 %, asimismo, el coeficiente de determinación del modelo es de 96,16 % y el coeficiente de determinación ajustado de 93.6 %.

Cuadro 5.15: Valor del $\Pr(> |t|)$ de cada costo

Costo	$\Pr(> t)$
Asistencia técnica	0.498304
Arriendos	0.923452
Preparación terreno	0.227391
Riego	0.131643
Fertilizantes	0.000489 ***
Protección al cultivo	0.027344 *
Recolección	0.460832
Otros	0.067503 .

Con base al cuadro anterior, se depura el modelo de regresión múltiple quitando uno a uno los costos que no son estadísticamente significativos para disminuir la distancia entre el coeficiente de determinación y el coeficiente de determinación ajustado. En esa purga al modelo, se identificó que los costos estadísticamente significativos son:

Cuadro 5.16: Valor del $\Pr(> |t|)$ de cada costo

Costo	$\Pr(> t)$
Fertilizantes	8.63e-05 ***
Protección al cultivo	0.00728 **
Recolección	0.00141 **

El cuadro muestra que el costo fertilizantes es estadísticamente significativo 100 % y los otros dos costos, protección al cultivo y recolección, son al 99.9 %. Además el coeficiente de determinación del modelo es de 93.27 % y el coeficiente de determinación ajustado de 92.09 %, indicando la confianza del modelo. Asimismo, la prueba de normalidad de Shapiro-Wilk arroja que el valor del p-value es de 0.2532, siendo mayor a 0.05, por tanto las observaciones provienen de una distribución normal. En conclusión, de los ocho costos que intervienen en la fijación del precio del arroz paddy, los costos que tiene mayor influencia son: fertilizante, protección al cultivo y recolección.

Capítulo 6

Conclusiones y Recomendaciones

- EL mejor modelo construido de predicción para el precio promedio por tonelada del arroz paddy para la zona de los Llanos Orientales es el SARIMAX(1,1,2)x(0,0,1)[12], debido a que el criterio de información de Akaike (AIC), el criterio de información bayesiano (BIC) y la raíz del error cuadrático medio (RMSE) del modelo son menores a comparación de los otros modelos. Asimismo, la ecuación del modelo SARIMAX(1,1,2)x(0,0,1)[12] es:

$$(1 - 0,8004\beta)(1 - \beta)Y_t = (1 + 1010)(1 - 0,2820\beta)(1 - 0,6101\beta^2)a_t - 10,0832_t\beta$$

- Los costos fueron sometidos a una depuración, teniendo en cuenta las características de los modelos de regresión múltiple, que evidenció la significancia estadística en cada uno de ellos. Después de probar varios modelos de regresión donde se quitaron las variables con poco valor estadístico y de seleccionar aquellas que estaban con una significancia alta, se concluyó que los costos más influyentes, en la fijación del precio promedio por tonelada del arroz paddy para la zona de los Llanos Orientales, son: los fertilizantes, la protección al cultivo y la recolección.
- Con base en la información que proporcionó el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), se halló que la media de precipitaciones por mes es una variable exógena que influye en la determinación del precio promedio del arroz paddy por tonelada para la región de los Llanos Orientales y contribuye a que el modelo SARIMAX(1,1,2)x(0,0,1)[12] sea mas robusto para pronosticar que el ARMA(1,1), ARIMA(1,1,2)x(0,0,1)[12] y Holt-Winters.
- El precio promedio por tonelada del arroz paddy que se pronostico desde julio hasta diciembre de 2018 para los Llanos Orientales está dentro de los intervalos de confianza que el modelo determina, y muestran una tendencia de crecimiento en este periodo.

Bibliografía

- [1] A. Garces. “Diagnostico de gestión tecnológica del cultivo de arroz en la región de Venadillo, Tolima: Siembra tecnificada versus siembra tradicional”. Trabajo de grado. Facultad de Ciencias Económicas y Administrativas. Pontificia Universidad Javeriana, feb. de 2009.
- [2] T. David y N. Hernández. “Efectos del tratado de libre comercio con Estados Unidos en el subsector arrocero colombiano”. Trabajo de grado. Escuela de Administración, Universidad del Rosario, 2015.
- [3] FEDERACIÓN NACIONAL DE ARROCEROS (FEDEARROZ). *Discurso instalación del XXXIV congreso nacional arrocero del doctor Rafael Hernández*. 2014. URL: <http://www.fedearroz.com.co/revistanew/arroz508.pdf>; consulta: noviembre, 2015.
- [4] MINISTERIO DE AGRICULTURA Y DESARROLLO RURAL. *El 83.5% de los alimentos que consumen los colombianos son producidos por nuestros campesinos*. 2016. URL: <https://www.minagricultura.gov.co/noticias/Paginas/El-83-de-los-alimentos-que-consumen-los-colombianos-son-producidos-por-nuestros-campesinos.aspx>.
- [5] Chica J. Tirado Y. y Barreto J. “Indicadores de competitividad del cultivo del arroz en Colombia y Estados Unidos.” En: *Revista de Ciencias Agrícolas*. 32.2 (2016), págs. 16-31.
- [6] Vargas P. “Índice de disponibilidad hídrica (IDH) y su importancia en el desarrollo del cultivo de arroz.” En: *Revista Arroz* 63.517 (2015), págs. 4-11.
- [7] EL TIEMPO. *Gobierno alista norma para fijar un precio mínimo del arroz*. Citado el 07 de mayo 2018. URL: <http://www.eltiempo.com/economia/sectores/gobierno-alista-norma-para-fijar-un-precio-minimo-del-arroz-214236>.
- [8] R. Zapata. “La especulación financiera y las crisis de la acumulación.” En: *Revista Ensayos de Economía*. 13.23 (2003), págs. 94-118.
- [9] F. Arias. *Crecimiento del sector agropecuario estará este año entre 1,3% y 2,4%*. Citados el 02 de marzo de 2018. URL: <http://www.elcolombiano.com/negocios/campo-crecera-en-2018-pese-a-retos-que-afronta-EI8280969>.

- [10] M. Austillo. *Fundamentos de economía*. Primera Edición. Editorial Instituto de investigaciones económicas, 2012. URL: <http://ru.iiec.unam.mx/2462/1/FundamentosDeEconomiaSecuenciaCorrecta.pdf>.
- [11] W. Mina y R. Padilla. “El sector agropecuario como respuesta a la paz.” En: *Revista de Estudiantes de Administración de Empresas*. 8 (2015), págs. 218-227.
- [12] M. Linero. “Sector agropecuario en Colombia.” En: *Documento de docencia N 11 3* (2017), págs. 6-20.
- [13] DICCIONARIO DE LA REAL ACADEMIA ESPAÑOLA. 2014. URL: <http://www.elcolombiano.com/negocios/campo-crecera-en-2018-pese-a-retos-que-afronta-EI8280969>.
- [14] ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA (ANDI). *Política comercial para el arroz*. 2013. URL: <http://www.andi.com.co/Uploads/Archivo%20final%20ANDI.pdf>.
- [15] FEDERACIÓN NACIONAL DE ARROCEROS (FEDEARROZ). *Guía de trabajo*. 2015. URL: http://www.fedearroz.com.co/docs/Guia%5C_de%5C_trabajo%5C_baja.pdf.
- [16] FEDERACIÓN NACIONAL DE ARROCEROS (FEDEARROZ). *Dinámica del sector arrocerro en los Llanos Orientales de Colombia*. 2011. URL: http://www.fedearroz.com.co/%5C%5Cdoc%5C_economia/Dinamica%5C_del%5C_sector%5C_arrocero%5C_en%5C_los%5C_Llanos%5C_orientales.pdf.
- [17] EL TIEMPO. *Fedearroz pide que las siembras no superen 340.000 hectáreas*. Citado el 11 de abril 2018. URL: <http://www.eltiempo.com/economia/sectores/areas-sembradas-con-arroz-en-colombia-203380>.
- [18] C. Sabino. *Diccionario de economía y finanzas*. Primera Edición. Editorial Penapo, 1991. URL: <http://www.eumed.net/cursecon/dic/P.htm#precio>.
- [19] D. Anderson. *Estadística para administración y economía*. Decima Edición. Cengage Learning, 2008.
- [20] D. Gujarati. *Econometría*. Quinta Edición. Editoria McGrawHill, 2009.
- [21] C. Chatfield. *The analysis of time Series and introduction*. fifth. Chapman y Hall/CRC, 1995.
- [22] M. Pokorny. *An Introduction to Econometrics*. Basil Blackwell, 1987.

Apéndice A

Script de los modelos construidos

Serie de tiempo de Precio Promedio Mensual Arroz Paddy Zona Llanos (Villavicencio)
Pesos / Tonelada

	Jan	Feb	Mar	Apr	May	Jun	Jul
1996	227800	255000	266700	277000	289000	288000	284000
1997	316500	324700	326700	333100	357000	348700	347200
1998	378100	444000	442000	437300	439300	445300	426100
1999	368000	395300	401900	425700	439300	425300	407300
2000	384000	394700	405200	421700	430400	422000	390300
2001	443200	502000	536000	524000	504000	498000	481333
2002	475200	475200	475200	488800	505600	516000	501440
2003	520000	574000	597000	595200	600000	600000	569400
2004	600000	624000	612800	591882	588000	568000	560000
2005	568000	568000	564800	552000	554000	552000	540000
2006	552000	552750	556600	567000	593400	627000	628000
2007	613200	622705	630000	635000	635000	635000	625750
2008	634400	709250	819500	836800	846000	1036750	1157600
2009	1102750	967250	882000	869000	866500	833500	776000
2010	753250	827500	796000	763000	756000	787200	789000
2011	919500	927000	952600	955000	992000	950800	849750
2012	898000	946000	972000	1019250	1123200	1088000	1118000
2013	868000	820000	843500	867000	867000	867000	867000
2014	867000	867000	867000	867000	867000	867000	867000
2015	1218000	1458000	1327000	1235200	1152000	1064000	1040000
2016	1366000	1426000	1280000	1192000	1024000	1004800	992000
2017	992000	992000	992000	920000	896000	860800	848000
2018	808000	828000	866000	872000	872000	872000	
	Aug	Sep	Oct	Nov	Dec		

Pronóstico del precio promedio por tonelada del arroz paddy desde julio hasta diciembre de 2018 en los Llanos Orientales

1996	277600	282700	285300	283300	302300
1997	352000	351300	372500	373300	373300
1998	405100	404500	381600	361300	362100
1999	385300	388500	392700	383300	380000
2000	386700	387000	410000	432000	432000
2001	456800	470000	475520	475200	473200
2002	471200	473600	476000	482000	499000
2003	510500	508000	537900	574000	585600
2004	528000	508800	512000	506000	550000
2005	549600	548000	553000	556000	556000
2006	631800	646250	650250	635400	610750
2007	620000	620000	620400	622000	622000
2008	895500	904000	1099800	1154750	1120000
2009	760000	748600	745000	746750	730600
2010	773750	806000	845750	875500	910600
2011	877800	923500	977000	1002000	955500
2012	1158400	1188000	1200000	1056750	941500
2013	867000	867000	867000	867000	867000
2014	867000	867000	867000	867000	897800
2015	1040000	1046400	1102000	1175000	1208000
2016	992000	992000	992000	992000	992000
2017	808000	808000	808000	808000	808000
2018					

Prueba de Dickey-Fuller a la serie de tiempo.

```
> adf.test(seriearroz)
```

Augmented Dickey-Fuller Test

```
data: seriearroz
```

```
Dickey-Fuller = -3.1653, Lag order = 6, p-value = 0.0941
```

```
alternative hypothesis: stationary
```

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) y Phillips-Perron (PP)

```
> ndiffs(seriearroz, test=c("kpss"))
```

```
[1] 1
```

```
> ndiffs(seriearroz, test=c("pp"))
```

```
[1] 1
```

Valores de la serie del precio promedio del arroz diferenciada:

```
> diff(tsa1)
```

	Jan	Feb	Mar	Apr	May	Jun
1996		27200	11700	10300	12000	-1000
1997	14200	8200	2000	6400	23900	-8300
1998	4800	65900	-2000	-4700	2000	6000
1999	5900	27300	6600	23800	13600	-14000
2000	4000	10700	10500	16500	8700	-8400
2001	11200	58800	34000	-12000	-20000	-6000
2002	2000	0	0	13600	16800	10400
2003	21000	54000	23000	-1800	4800	0
2004	14400	24000	-11200	-20918	-3882	-20000
2005	18000	0	-3200	-12800	2000	-2000
2006	-4000	750	3850	10400	26400	33600
2007	2450	9505	7295	5000	0	0
2008	12400	74850	110250	17300	9200	190750
2009	-17250	-135500	-85250	-13000	-2500	-33000
2010	22650	74250	-31500	-33000	-7000	31200
2011	8900	7500	25600	2400	37000	-41200
2012	-57500	48000	26000	47250	103950	-35200
2013	-73500	-48000	23500	23500	0	0
2014	0	0	0	0	0	0
2015	320200	240000	-131000	-91800	-83200	-88000
2016	158000	60000	-146000	-88000	-168000	-19200
2017	0	0	0	-72000	-24000	-35200
	Jul	Aug	Sep	Oct	Nov	Dec
1996	-4000	-6400	5100	2600	-2000	19000
1997	-1500	4800	-700	21200	800	0
1998	-19200	-21000	-600	-22900	-20300	800
1999	-18000	-22000	3200	4200	-9400	-3300
2000	-31700	-3600	300	23000	22000	0
2001	-16667	-24533	13200	5520	-320	-2000
2002	-14560	-30240	2400	2400	6000	17000
2003	-30600	-58900	-2500	29900	36100	11600
2004	-8000	-32000	-19200	3200	-6000	44000
2005	-12000	9600	-1600	5000	3000	0
2006	1000	3800	14450	4000	-14850	-24650
2007	-9250	-5750	0	400	1600	0

2008	120850	-262100	8500	195800	54950	-34750
2009	-57500	-16000	-11400	-3600	1750	-16150
2010	1800	-15250	32250	39750	29750	35100
2011	-101050	28050	45700	53500	25000	-46500
2012	30000	40400	29600	12000	-143250	-115250
2013	0	0	0	0	0	0
2014	0	0	0	0	0	30800
2015	-24000	0	6400	55600	73000	33000
2016	-12800	0	0	0	0	0
2017	-12800	-40000	0	0	0	0

Modelos ARMA(p,q)

```
> summary(arma1.1)
```

Call:

```
arima(x = diffarroz, order = c(1, 0, 1), seasonal = list(order = c(0, 0, 0)))
```

Coefficients:

	ar1	ma1	intercept
	-0.2094	0.7424	2406.572
s.e.	0.0929	0.0632	3828.564

sigma² estimated as 1.903e+09: log likelihood = -3255.79, aic = 6519.58

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE
Training set	-22.66529	43626.72	23504.66	NaN	Inf
	MASE	ACF1			
Training set	0.8299811	-0.01047679			

```
> summary(arma2.2)
```

Call:

```
arima(x = diffarroz, order = c(2, 0, 2), seasonal = list(order = c(0, 0, 0)))
```

Coefficients:

	ar1	ar2	ma1	ma2	intercept
	0.7463	0.1021	-0.2860	-0.7140	2953.6502

s.e. 0.0932 0.0924 0.0681 0.0676 347.4864

sigma² estimated as 1.785e+09: log likelihood = -3248.61, aic = 6509.23

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	854.7923	42243.62	23456	NaN	Inf	0.8282628

	ACF1
Training set	-4.892221e-05

Test Dickey-Fuller para datos diferenciados

```
> adf.test(diffarroz)
```

Augmented Dickey-Fuller Test

data: diffarroz

Dickey-Fuller = -7.6422, Lag order = 6, p-value = 0.01

alternative hypothesis: stationary

Resumen de los modelos escogidos bajo la Metodología Box-Jenkins

```
> summary(model1)
```

Call:

```
arima(x = tsa1, order = c(1, 1, 2), seasonal = c(0, 0, 1), transform.pars = 12)
```

Coefficients:

	ar1	ma1	ma2	sma1
	0.7983	-0.2839	-0.6066	0.1003
s.e.	0.0677	0.0707	0.0511	0.0611

sigma² estimated as 1.834e+09: log likelihood = -3250.96, aic = 6511.92

Training set error measures:

	ME	RMSE	MAE	MPE
Training set	4082.913	42746.45	23276.34	0.6394959

	MAPE	MASE	ACF1
Training set	3.060531	0.9072505	-0.04337444

```
> summary(model2)
```

Call:

```
arima(x = tsal, order = c(1, 1, 1), seasonal = c(0, 0, 1), transform.pars = 12)
```

Coefficients:

	ar1	ma1	sma1
	-0.2252	0.7432	0.1078
s.e.	0.0965	0.0663	0.0597

sigma² estimated as 1.883e+09: log likelihood = -3254.42, aic = 6516.83

Training set error measures:

	ME	RMSE	MAE	MPE
Training set	1512.969	43314.66	23773.09	0.2589966

	MAPE	MASE	ACF1
Training set	3.093071	0.9266126	-0.01250945

```
> summary(model3)
```

Call:

```
arima(x = tsal, order = c(2, 1, 2), seasonal = c(0, 0, 1), transform.pars = 12)
```

Coefficients:

	ar1	ar2	ma1	ma2	sma1
	0.7093	0.1119	-0.2313	-0.6738	0.1092
s.e.	0.0991	0.0980	0.0738	0.0718	0.0609

sigma² estimated as 1.825e+09: log likelihood = -3250.29, aic = 6512.59

Training set error measures:

	ME	RMSE	MAE	MPE
Training set	4183.113	42640.88	23480.05	0.65231

	MAPE	MASE	ACF1
Training set	3.079167	0.9151908	-0.009515228

Resumen de los modelos escogidos para el modelo SARIMAX

```
> summary(model8)
```

Call:

```
arima(x = tsal, order = c(1, 1, 2), seasonal = c(0, 0, 1), xreg = sta2, transform.pars
```

Coefficients:

	ar1	ma1	ma2	sma1	sta2
	0.8016	-0.2829	-0.6098	0.1043	23.7024
s.e.	0.0658	0.0687	0.0503	0.0605	15.4972

sigma² estimated as 1.818e+09: log likelihood = -3249.8, aic = 6511.59

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE
Training set	4035.127	42560.96	23405.72	0.6307152	3.071499

	MASE	ACF1
Training set	0.9122936	-0.0386224

```
> summary(model9)
```

Call:

```
arima(x = tsal, order = c(1, 1, 1), seasonal = c(0, 0, 1), xreg = sta2, transform.pars
```

Coefficients:

	ar1	ma1	sma1	sta2
	-0.2113	0.7376	0.1110	20.7034
s.e.	0.0965	0.0664	0.0592	15.4142

sigma² estimated as 1.871e+09: log likelihood = -3253.52, aic = 6517.04

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE
Training set	1466.97	43169.89	23666.83	0.2542186	3.07171

	MASE	ACF1
Training set	0.9224708	-0.01228531

```
> summary(model10)
```

Call:

```
arima(x = tsal, order = c(2, 1, 2), seasonal = c(0, 0, 1), xreg = sta2, transform.pars
```

Coefficients:

	ar1	ar2	ma1	ma2	sma1	sta2
	0.7243	0.0969	-0.2377	-0.6675	0.1119	22.0116

Pronóstico del precio promedio por tonelada del arroz paddy desde julio hasta diciembre de 2018 en los Llanos Orientales

s.e. 0.0989 0.0979 0.0739 0.0719 0.0604 15.5609

sigma² estimated as 1.811e+09: log likelihood = -3249.3, aic = 6512.59

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE
Training set	4124.923	42482.13	23530.2	0.6424811	3.081262
	MASE	ACF1			
Training set	0.9171454	-0.009396537			

Apéndice B

Script de pronósticos y sistemas de regresión múltiple

Predicciones de los tres modelos construidos.

ARMA(1,1)

```
> pronostico=arima(seriearroz[1:264], order=c(1,0,1))
```

```
> forecast(pronostico, h=6)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
265	800013.1	743424.6	856601.6	713468.4	886557.7
266	795313.6	688773.2	901853.9	632374.1	958253.0
267	790774.7	653109.1	928440.3	580233.3	1001316.1
268	786391.0	625015.7	947766.3	539588.8	1033193.2
269	782157.2	601448.5	962865.9	505787.0	1058527.3
270	778068.1	581027.9	975108.3	476721.1	1079415.1

ARIMA(1,1,2)x(0,0,1) [12]

```
> forecast(model11, h=6)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2018	812002.8	756523.8	867481.7	727155.1	
Feb 2018	828730.2	728013.4	929447.1	674697.1	
Mar 2018	843624.9	719106.7	968143.0	653190.8	
Apr 2018	848211.5	707947.5	988475.6	633696.2	
May 2018	856061.8	704404.6	1007719.0	624122.2	
Jun 2018	859639.2	699260.6	1020017.8	614361.3	
					Hi 95
Jan 2018	896850.5				
Feb 2018	982763.4				
Mar 2018	1034059.0				

Apr 2018 1062726.9

May 2018 1088001.5

SARIMAX(1,1,2)x(0,0,1) [12]

```
> modelo10=arima(sta11, order=c(1,1,2), seasonal = c(0,0,1), t=12, xreg = sta21)
> forecast(modelo10, xreg = sta21)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2018	815248.4	760141.2	870355.6	730969.15	899527.6
Feb 2018	833907.3	733512.7	934301.9	680366.96	987447.6
Mar 2018	847390.5	723143.6	971637.4	657371.24	1037409.7
Apr 2018	857136.7	717093.9	997179.5	642959.67	1071313.7
May 2018	869371.6	717896.0	1020847.2	637709.63	1101033.5
Jun 2018	866063.3	705839.8	1026286.8	621022.62	1111104.0

Modelo Holt-Winters

```
> holtserie
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2018	831793.0	767714.2	895871.9	733792.9	929793.2
Feb 2018	861032.0	770271.8	951792.1	722226.3	999837.7
Mar 2018	856163.9	744831.8	967496.0	685896.2	1026431.7
Apr 2018	849104.3	720346.9	977861.7	652186.9	1046021.7
May 2018	847933.1	703751.4	992114.9	627426.2	1068440.1
Jun 2018	847826.6	689634.8	1006018.4	605893.2	1089760.1

Pronósticos del precio promedio

```
> forecast(pronos, xreg = tsa2)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jul 2018	877609.8	777883.8	977335.8	725092.04	1030128
Aug 2018	876676.8	753335.6	1000018.0	688042.73	1065311
Sep 2018	877387.2	738425.4	1016349.0	664863.52	1089911
Oct 2018	881683.2	731427.8	1031938.6	651887.38	1111479
Nov 2018	884405.0	725516.2	1043293.9	641405.49	1127405
Dec 2018	886100.6	720329.8	1051871.4	632576.02	1139625

Modelo de regresión de los costos.

```
> summary(r1c5)
```

Call:

```
lm(formula = 'Promedio precio anual' ~ Fertilizantes + 'Proteccion al cultivo' +  
  Recoleccion, data = costos_anual_por_tonelada)
```

Residuals:

Min	1Q	Median	3Q	Max
-152782	-36971	-4924	31801	150122

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.131e+05	6.577e+04	3.241	0.00480	**
Fertilizantes	2.826e+00	5.526e-01	5.114	8.63e-05	***
'Proteccion al cultivo'	-1.547e+00	5.077e-01	-3.048	0.00728	**
Recoleccion	3.861e+00	1.015e+00	3.806	0.00141	**

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

Residual standard error: 75270 on 17 degrees of freedom

Multiple R-squared: 0.9327, Adjusted R-squared: 0.9209

F-statistic: 78.57 on 3 and 17 DF, p-value: 3.619e-10

Test de Shapiro-Wilk de modelo de los costos

```
> shapiro.test(rstandard(rlc5))
```

Shapiro-Wilk normality test

data: rstandard(rlc5)

W = 0.94331, p-value = 0.2532