



Modelo de predicción de precios de productos agropecuarios como base para la priorización de Proyectos Integrales de Desarrollo Agropecuario y Rural con enfoque territorial en Colombia en el marco de los impactos producidos por la COVID-19

Oscar Camilo Luna
Diciembre 2020

Director de proyecto
V́ctor Manuel Mondrag3n

Universidad de Bogot3 Jorge Tadeo Lozano
Facultad de Ciencias Naturales e Ingenier3a
Maestr3a en Ingenier3a y An3lítica de Datos
Bogot3 D.C.

La Agencia de Desarrollo Rural (ADR) del Ministerio de Agricultura de la República de Colombia tiene dentro de sus actividades misionales la realización de una convocatoria a nivel nacional para la presentación de perfiles de proyectos agropecuarios, con el fin de cofinanciar iniciativas productivas para beneficio de pequeños y medianos productores rurales. Con la información histórica de los proyectos cofinanciados desde el año 2017 hasta el año 2020 y con el objetivo de proporcionar la información requerida para la toma de decisiones por parte de la Presidencia de la ADR y el Ministerio de Agricultura, surge la necesidad de analizar el comportamiento del precio de los productos agropecuarios y determinar si la implementación de modelos de predicción puede hacer parte de las directrices que permitan la priorización de los Proyectos Integrales de Desarrollo Agropecuario y Rural con Enfoque Territorial (PIDAR) a ser cofinanciados en Colombia para el cuatrienio, en el marco del impacto socioeconómico generado por la COVID-19. Este proyecto pretende aportar técnicamente conocimientos relacionados con previsión de series de tiempo de precios, para fortalecer la capacidad instalada de la ADR y aportar herramientas que le permitan presentar a la Presidencia de la ADR las recomendaciones sobre cuáles deben ser los PIDAR a gestionar, promover y financiar por parte de la agencia.

Palabras clave: SARIMA, series de tiempo, precios agropecuarios, tendencia, periodicidad, estacionariedad, previsión.

The Ministry of Agriculture's Rural Development Agency (ADR) of the Republic of Colombia has within its missional activities the realization of a national level call for agricultural projects profiles, to co-finance productive initiatives for the benefit of small and medium rural producers. With historical information of co-financed projects from 2017 to 2020 and the aim of providing required information by the Presidency of the ADR and the Ministry of Agriculture for decision-making, it becomes necessary to analyze the behavior of the price of agricultural products and determine if the implementation of prediction models can be part of the guidelines that allow to prioritize the Comprehensive Agricultural and Rural Development Projects with a Territorial Approach (PIDAR) to be co-financed in Colombia for the four-year period, in the framework of the socioeconomic impact generated by COVID-19. This project intends to contribute technical knowledge about prices time series forecasting, strengthening the ADR's installed capacity and providing tools that allows to offer recommendations to the Presidency of the ADR on which should be the PIDARs to be managed, promoted, and financed by the agency.

Keywords: SARIMA, time series, agricultural prices, trends, seasonality, stationarity, forecasting.

Tabla de Contenidos

iv

1. Introducción	1
2. Objetivos	3
2.1. Objetivo General	3
2.2. Objetivos Específicos.....	3
3. Marco Teórico.....	4
3.1. PIDAR.....	4
Etapas de presentación de proyectos	5
3.2. COVID-19.....	6
3.3. Series de tiempo.....	8
Componentes de una serie de tiempo	9
Ejemplos de series de tiempo	10
3.4. Modelos de series de tiempo.....	11
Estacionariedad	12
Ruido blanco	12
Operador de retraso	13
Modelo AR	13
Modelo MA	13
Modelo ARMA	14
Modelo ARIMA	15
Modelo SARIMA	16
4. Estado del Arte.....	18
5. Desarrollo.....	23
5.1. Obtención de datos.....	25
5.2. Análisis previo de datos.....	29
5.3. Selección de modelo.....	32
6. Casos de estudio.....	41
6.1. Caso de estudio: Papa pastusa.....	41
6.2. Caso de estudio: Huevo blanco AA.....	44
7. Resultados y Discusión.....	50
Lista de referencias	52

Lista de Tablas

v

Tabla 1. Etapas de presentación de los PIDAR.	5
Tabla 2. Selección de modelo con respecto al comportamiento de la función ACF.	15
Tabla 3. Investigaciones marco.....	18
Tabla 4. Ejemplo datos tabla ADR_CORABASTOS_PRICES.	27

Figura 1. Cambio porcentual a nivel mundial en los precios de los productos básicos entre enero y abril de 2020..	7
Figura 2. Crecimiento de los precios reales de los alimentos en febrero y marzo (2020 vs 2017-19).	8
Figura 3. Ejemplo de ciclo económico de cuatro fases.	9
Figura 4. Población de EE. UU. en millones de habitantes y en intervalos de 10 años, 1790-1990.	10
Figura 5. Ventas de vino tinto australiano, enero 1980 - octubre 1991.	11
Figura 6. Proyectos cofinanciados por la ADR agrupados por departamento, 2017 - 2020.	24
Figura 7. Productos relacionados con los PIDAR, 2017 - 2020.	25
Figura 8. Precios de hortalizas publicados en la central mayorista de abastos de Bogotá para el 31 de julio de 2020.	26
Figura 9. Diseño tabla para precios de la central mayorista de abastos de Bogotá ADR_CORABASTOS_PRICES.	27
Figura 10. Consulta del valor histórico mensual del Índice de Precios al Consumidor – IPC.	28
Figura 11. Diseño tabla para valor histórico mensual del Índice de Precios al Consumidor en Colombia ADR_IPC.	29
Figura 12. Consulta en base de datos del histórico de precios del arroz corriente empleando Python.	30
Figura 13. Cálculo de precio actual para el histórico de precios del arroz corriente empleando Python.	31
Figura 14. Diagrama <i>Boxplot</i> del arroz corriente agrupado mensualmente entre 2012 y 2020.	32
Figura 15. Gráfico de serie de tiempo del precio del arroz corriente entre 2012 y 2020.	32
Figura 16. Gráfico de serie de tiempo de arroz corriente con valores agrupados por mes.	33
Figura 17. Gráfico de función de autocorrelación con 36 retrasos (<i>lags</i>) para el precio del arroz corriente agrupado por mes.	34
Figura 18. Gráfico de función de autocorrelación parcial con 36 retrasos (<i>lags</i>) para el precio del arroz corriente agrupado por mes.	34
Figura 19. Gráfico de serie de tiempo diferenciada para los precios del arroz corriente.	35
Figura 20. Gráfico de función de autocorrelación de la serie diferenciada para el precio del arroz corriente.	36
Figura 21. Gráfico de función de autocorrelación parcial de la serie diferenciada para el precio del arroz corriente	36
Figura 22. Código en Python para la evaluación de múltiples combinaciones de parámetros para el modelo SARIMA	37
Figura 23. Resumen de estimación en Python de un modelo SARIMA (0,1,1) x (1,1,1) ₁₂ .	38
Figura 24. Valor de la raíz del error cuadrático medio y comparación de los valores actuales y predichos para el precio del arroz corriente con un modelo SARIMA (0,1,1) x (1,1,1) ₁₂ .	39
Figura 25. Diagrama Bloxplot del comportamiento de precios mensuales de la papa pastusa, de 2012 a 2020.	41
Figura 26. Gráfico de serie de tiempo, función ACF y PACF para los precios mensuales de la papa pastusa, de 2012 a 2020.	42

Figura 27. Resumen modelo SARIMA(2,0,0) x (1,0,1) ₁₂ para los precios mensuales de la papa vii pastusa, de 2012 a 2020.	43
Figura 28. Valor del error RMSE y comparación de valores actuales y predichos de un modelo SARIMA(2,0,0) x (1,0,1) ₁₂ para los precios mensuales de la papa pastusa, de 2012 a 2020.	44
Figura 29. Diagrama Bloxplot del comportamiento de precios mensuales del huevo blanco AA, de 2012 a 2020.	45
Figura 30. Gráfico de serie de tiempo, función ACF y PACF para los precios mensuales del huevo blanco AA, de 2012 a 2020.	46
Figura 31. Gráfico comportamiento, función ACF y PACF para la serie de tiempo diferenciada de precios mensuales del huevo blanco AA, de 2012 a 2020.	47
Figura 32. Resumen modelo SARIMA(2,1,4) x (0,1,1) ₁₂ para los precios mensuales del huevo blanco AA, de 2012 a 2020.	47
Figura 33. Valor del error RMSE y comparación de valores actuales y predichos de un modelo SARIMA(2,1,4) x (0,1,1) ₁₂ para los precios mensuales del huevo blanco AA, de 2012 a 2020.	48

1. Introducción

El presente trabajo pretende analizar el comportamiento del valor de venta de varios productos agropecuarios y determinar si la implementación de modelos de previsión de esta clase de precios puede hacer parte de las directrices que permitan la priorización de los Proyectos Integrales de Desarrollo Agropecuario y Rural con Enfoque Territorial (PIDAR) que cofinancia la Agencia de Desarrollo Rural ADR, teniendo presente el impacto socioeconómico producido por la COVID-19. Para tal fin realiza la recolección y preparación de datos de precios de productos de origen agropecuario de Colombia y luego se selecciona y entrena un modelo de previsión de series de tiempo. Finalmente, se evalúan y ajustan los parámetros de dicho modelo para aplicarlo en dos casos de estudio.

Con todo lo anterior se espera lograr un aporte técnico de conocimientos que le permitan a la agencia contar con herramientas adicionales para la selección de proyectos. Cabe resaltar que las iniciativas impulsadas por la ADR benefician a pequeños y medianos productores rurales a lo largo y ancho del territorio colombiano y que instan a que los grupos poblacionales vulnerables participen en estos emprendimientos.

Bajo el interés académico el reto es hacer un entendimiento adecuado de los datos para definir posteriormente un método que permita seleccionar modelos previsivos de series de tiempo. Con la información arrojada por el modelo se tendrían criterios para elegir el momento idóneo para la ejecución de un proyecto productivo con respecto a un mayor margen de ventas del producto agropecuario según corresponda.

En el marco del desarrollo de este trabajo se aplica la metodología CRISP-DM para proyectos de minería de datos, en la cual se desarrollan actividades cíclicas de entendimiento del negocio y de los datos, la preparación de estos últimos y la posterior evaluación y despliegue. De manera conjunta, el modelamiento de las series de tiempo para la previsión de datos está soportado por análisis estadísticos que sugiere la adopción del modelo SARIMA (*Seasonal Autoregressive Integrated Moving Average*).

El trabajo contempla un marco teórico donde se abordan conceptos relacionados con los PIDAR, el impacto de la COVID-19, series de tiempo y sus componentes y la clasificación de modelos previsivos que se van a abarcar en el análisis de los datos. Después se presenta un estado del arte sobre trabajos de investigación relacionados con la predicción de precios de productos agrícolas. Continúa con la sección en donde muestra el proceso de obtención de los datos y la definición de la metodología de análisis y modelado de las series de tiempo. Posteriormente, se expone un apartado en el cual se presentan los casos de estudio para los productos papa pastusa y huevo blanco AA y se finaliza con los resultados y la discusión del trabajo realizado.

Para realizar una correcta lectura de este trabajo se recomienda contar con conceptos de estadística y procesos estocásticos.

2. Objetivos

2.1. Objetivo General

Desarrollar un modelo de predicción de precios de productos agropecuarios que establezcan las bases para la priorización de los Proyectos Integrales de Desarrollo Agropecuario y Rural con Enfoque Territorial (PIDAR) en Colombia por parte de la Agencia de Desarrollo Rural en el marco del impacto producido por la COVID-19.

2.2. Objetivos Específicos

- Recolectar datos de precios de productos de origen agropecuario en Colombia y aquellos relacionados con los efectos de la COVID-19 sobre estos mercados.
- Preparar los conjuntos de datos obtenidos, junto con los que tiene disponible la ADR sobre proyectos cofinanciados desde el año 2017 al año 2020, con el fin de generar el insumo requerido para el modelo de predicción.
- Seleccionar y entrenar un modelo de previsión de series de tiempo con los datos previamente preparados.
- Evaluar y realizar ajustes de parámetros sobre el modelo previamente seleccionado.
- Aplicar el modelo de previsión en dos casos de estudio para determinar el comportamiento a futuro de los precios de los productos analizados.

3. Marco Teórico

A continuación se presentan una serie de definiciones y conceptos con el propósito de establecer una adecuada contextualización sobre el desarrollo del proyecto.

3.1. PIDAR

Una de las funciones primordiales de la ADR es realizar el proceso de gestión y cofinanciación de proyectos, regidos por las políticas agropecuarias y los lineamientos sobre proyectos de inversión establecidos por el Gobierno de Colombia y dirigidos a organizaciones sociales, comunitarias y productivas, entidades territoriales o esquemas asociativos territoriales dedicados a actividades rurales (agropecuarias, pesqueras, acuícolas, forestales u otras comerciales o productivas enmarcadas dentro de este sector) y que se encuentren catalogadas como pequeños o medianos productores, sin distinción del grupo poblacional al que pertenezcan. Con respecto a esta clase de proyectos, Agencia de Desarrollo Rural [ADR] (2019) establece que:

Se consideran Proyectos Integrales de Desarrollo Agropecuario y Rural con Enfoque Territorial – PIDAR aquellas iniciativas que contemplan actividades limitadas en el tiempo, que utilizan parcialmente recursos públicos, con el fin de crear, ampliar, mejorar o recuperar la capacidad de producción de provisión de bienes o servicios por parte del Estado en el sector rural, enfocadas a incentivar la generación de ingresos o excedentes de producción, para contribuir a mejorar las condiciones de vida de los pobladores rurales y la competitividad del país. (p.9)

Etapas de presentación de proyectos. ADR (2009) indica que la presentación de un proyecto contempla 6 etapas, las cuales se describen en la Tabla 1.

Tabla 1. *Etapas de presentación de los PIDAR.*

Etapas	Descripción
Inscripción del perfil	Se debe relacionar la entidad u organización que el productor interesado representa, documentación acerca de proyectos similares, alcance de la propuesta, resultados esperados, estimación preliminar de la inversión, plazo de ejecución, localización del proyecto y número de potenciales beneficiarios (población objetivo).
Diagnóstico	Si se cumple con lo requerido en la etapa anterior, se hace un análisis integral de condiciones de los beneficiarios y del lugar en el que se realizará el proyecto, con el fin de definir la línea base de este.
Estructuración	Busca definir el alcance del proyecto y las condiciones de ejecución, con base en los principios de integralidad, orientación a mercados, sostenibilidad ambiental, aptitud productiva territorial, sostenibilidad financiera, inclusión social, fortalecimiento de capacidades organizativas, enfoque diferencial, articulación interinstitucional y con instrumentos de planificación.
Evaluación y calificación	La evaluación consiste en validar la viabilidad técnica, jurídica, ambiental, social, comercial y financiera a partir la estructuración del proyecto. La calificación consiste en aplicar criterios para cuantificar los proyectos y establecer una prioridad en su cofinanciación, la cual puede estar orientada a incrementar la competitividad de una cadena productiva agropecuaria o a una actividad específica de dicha cadena de valor.

Etapa	Descripción
Aprobación	Emisión de la decisión de cofinanciación con base en el concepto de evaluación y el puntaje de calificación realizados sobre el proyecto.
Ejecución	Ejecución física y financiera del proyecto posterior a la validación de requisitos de compromiso, corresponsabilidad, registro presupuestal, licenciamiento y constitución de garantía de cumplimiento.

Toda la documentación producto de estas fases debe reposar en el Banco de Proyectos de la Agencia.

3.2. COVID-19

De acuerdo con lo establecido por la Organización Mundial de la Salud [OMS] (2019), la COVID-19 es una enfermedad infecciosa causada por el coronavirus que se ha descubierto a finales del 2019. Actualmente es una pandemia que afecta a muchos países de todo el mundo. Su impacto se puede evidenciar en la caída mundial de algunos de los precios de los alimentos básicos tales como: el caucho, algodón, aceite, soja, maíz, cacao, café y trigo, siendo el arroz el único cereal que presentó una tendencia al alza, tal como lo muestra la Figura 1. A pesar de que se prevé estabilización de los precios durante el próximo año, existe el temor de padecer inseguridad alimentaria por causa de la recesión mundial y su influencia en el aumento del desempleo (Baffes & Nagle, 2020).

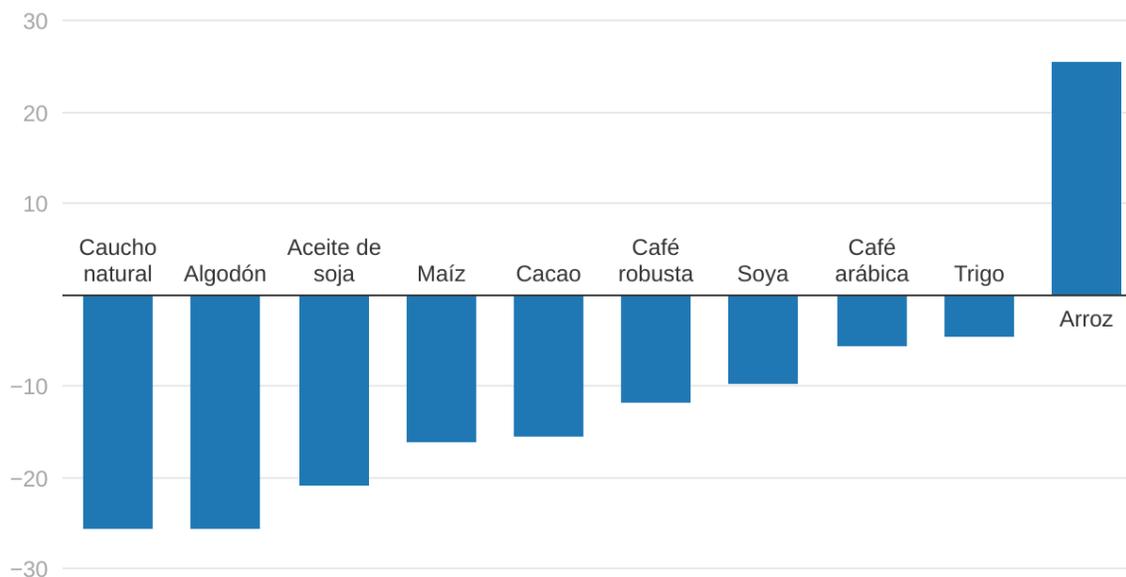


Figura 1. Cambio porcentual a nivel mundial en los precios de los productos básicos entre enero y abril de 2020. Baffes & Nagle (2020).

Por su parte la Organización de las Naciones para la Alimentación y la Agricultura [FAO] (2020) menciona que a nivel mundial el comportamiento de la inflación en los precios reales de los alimentos para los meses de febrero y marzo de 2020 vs 2017-19 tuvo un leve aumento, sin embargo la organización invita a prestar especial atención sobre los grupos más vulnerables de la población como se mencionó anteriormente, por los temas de acceso a la alimentación. En la Figura 2 se aprecian los resultados del análisis realizado por la FAO.

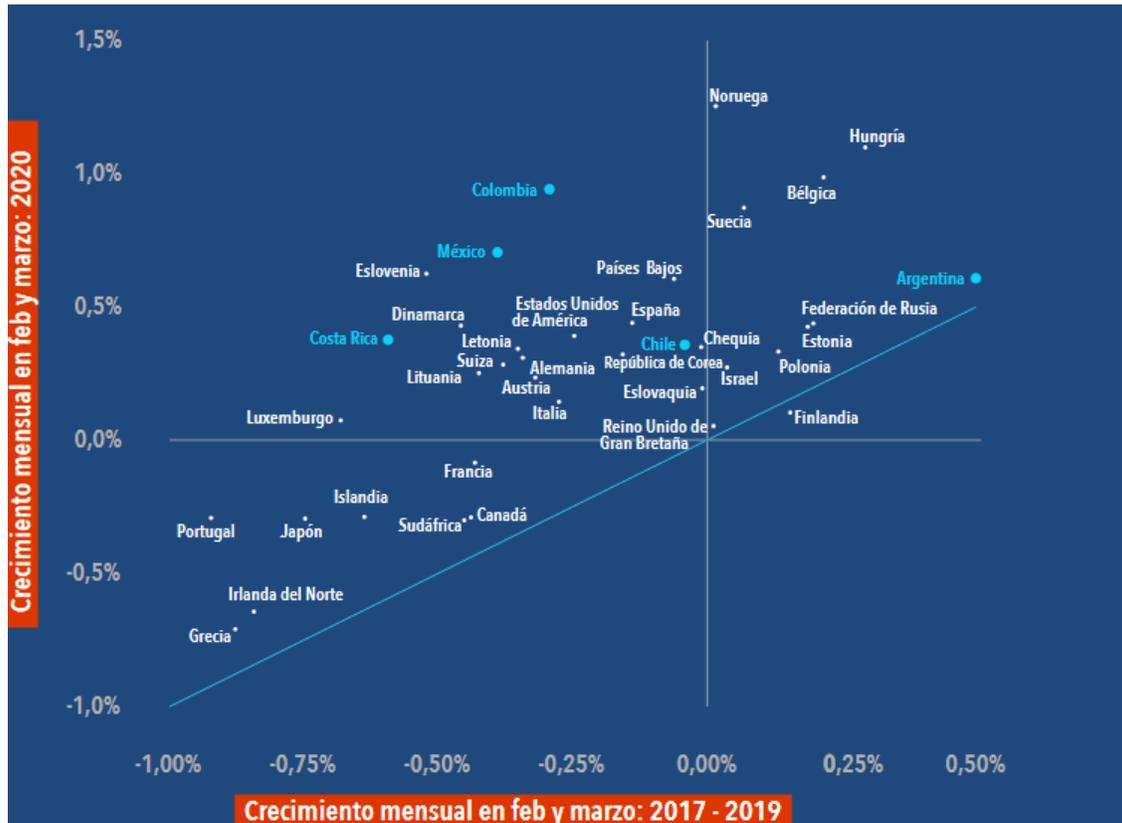


Figura 2. Crecimiento de los precios reales de los alimentos en febrero y marzo (2020 vs 2017-19). FAO (2020).

3.3. Series de tiempo

Una serie de tiempo es un conjunto de observaciones o puntos de datos secuenciales x_t medidos sobre un tiempo sucesivo t . Matemáticamente se define como un conjunto de vectores $x(t)$, $t = 0, 1, 2, \dots$ donde t representa el tiempo transcurrido. La variable x_t es considerada una variable aleatoria. Si la serie contiene datos de una sola variable se le denomina univariada, pero si se considera más de una variable por punto de observación, se cataloga como una serie multivariada (Adhikari & Agrawal, 2013).

Una serie de tiempo discreta es aquella en la cual el conjunto T_0 de tiempo sobre el que se realizan las observaciones es un conjunto discreto, como por ejemplo un

intervalo de tiempo fijo. Una serie de tiempo continua se obtiene cuando las observaciones son medidas continuamente sobre un intervalo de tiempo, por ejemplo cuando $T_0 = [0,1]$ (Brockwell & Davis, 2016).

Componentes de una serie de tiempo. De acuerdo con Adhikari & Agrawal (2013), una serie de tiempo en general se ve afectada por cuatro componentes principales, los cuales son: tendencia, periodicidad, irregularidad y ciclicidad. La tendencia hace referencia al movimiento a largo plazo en una serie de tiempo, es decir, la tendencia a crecer, decrecer o estancarse. La periodicidad es la fluctuación de la serie de tiempo de manera periódica por factores tales como el clima o las condiciones meteorológicas o hábitos de consumo dependiendo de la época del año. La irregularidad o variación aleatoria en una serie de tiempo se debe a influencias inesperadas que no se repiten en un patrón particular tales como un desastre natural. Finalmente, la ciclicidad hace referencia a los cambios a mediano plazo en la serie, provocados por circunstancias que se repiten en ciclos. Un ejemplo de ello se observa en la Figura 3, en la cual se muestran los componentes de un ciclo económico.

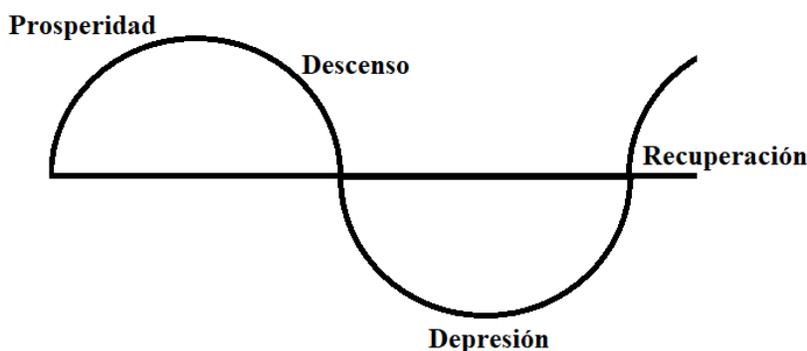


Figura 3. Ejemplo de ciclo económico de cuatro fases. Adhikari & Agrawal (2013).

Ejemplos de series de tiempo. La Figura 4 muestra la población de EE. UU. medida en intervalos de 10 años, sugiriendo una tendencia cuadrática o exponencial de los datos.

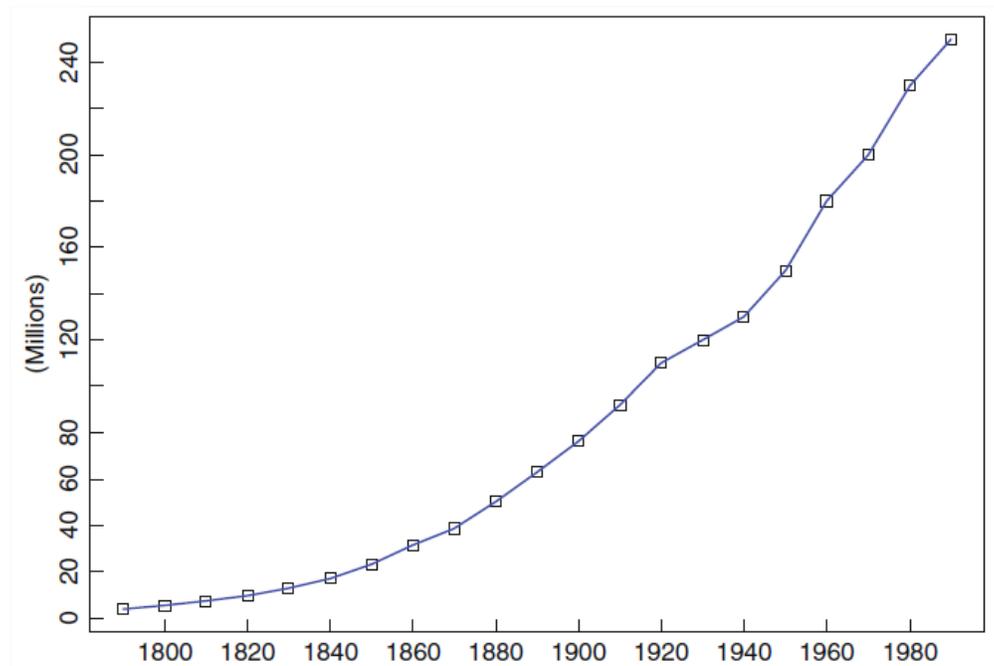


Figura 4. Población de EE. UU. en millones de habitantes y en intervalos de 10 años, 1790-1990. Brockwell & Davis (2016).

Por su parte, la Figura 5 presenta las ventas mensuales de vino tinto (en miles de litros) de productores australianos entre enero de 1980 a octubre de 1991. En este caso el conjunto T_0 de tiempo consiste en 142 entradas, correspondientes a la cantidad de meses que comprende el rango de tiempo considerado.

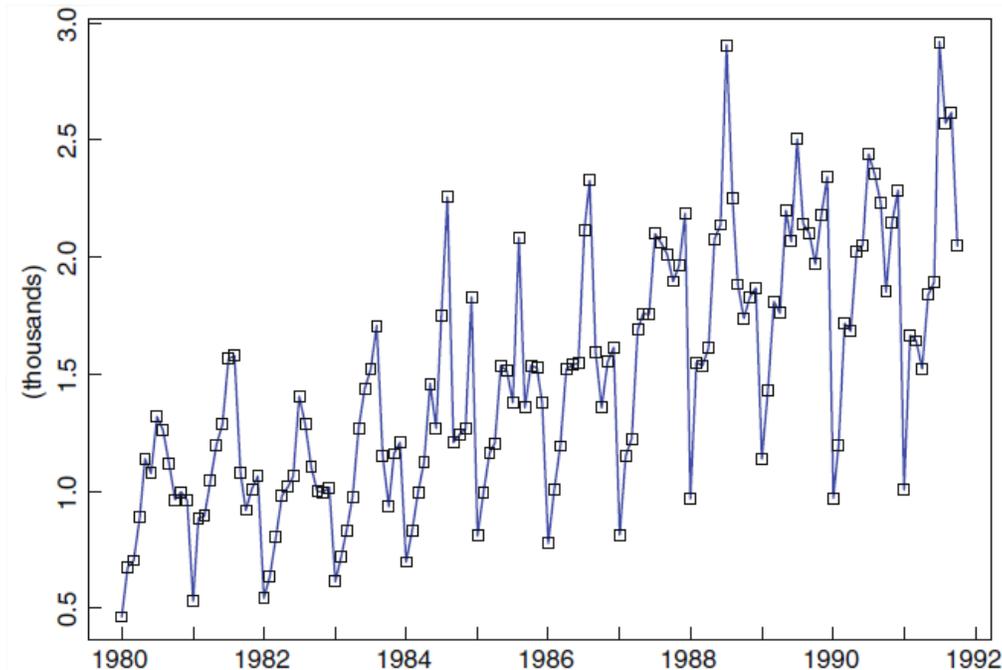


Figura 5. Ventas de vino tinto australiano, enero 1980 - octubre 1991. Brockwell & Davis (2016).

3.4. Modelos de series de tiempo

La intención de realizar análisis sobre una serie de tiempo es tratar de comprender el comportamiento, la naturaleza y los mecanismos que llevan a que dicha serie sea generada mediante la aplicación de modelos probabilísticos que representen los datos de la mejor manera posible y que de esta forma se puedan realizar predicciones. Esto tiene diversas aplicaciones sobre campos tales como la ingeniería, la ciencia, la sociología y la economía. Formalmente, un modelo de serie de tiempo para los datos observados $\{x_t\}$ es una especificación de las distribuciones conjuntas (o las medias y covarianzas) de una secuencia de variables aleatorias $\{X_t\}$ de la cual $\{x_t\}$ se postula como un valor observado (Brockwell & Davis, 2016).

Para poder comprender los múltiples modelos empleados en el análisis de series de tiempo, a continuación se definen varios conceptos.

Estacionariedad. El concepto de estacionariedad se puede visualizar como una forma de equilibrio estadístico (Adhikari & Agrawal, 2013). Una serie de tiempo $\{X_t, t = 0, \pm 1, \dots\}$ es estacionaria si tiene propiedades estadísticas similares a los de una serie diferida en el tiempo $\{X_{t+h}, t = 0, \pm 1, \dots\}$ para cada entero h .

Sea $\{X_t\}$ una serie de tiempo con $E(X_t^2) < \infty$. La media de $\{X_t\}$ es $\mu_x(t) = E(X_t)$. La covarianza de $\{X_t\}$ es $\gamma_x(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_x(r))(X_s - \mu_x(s))]$ para todo entero r y s . $\{X_t\}$ es estacionaria (débilmente) si $\mu_x(t)$ es independiente de t y si $\gamma_x(t+h, t)$ es independiente de t por cada h . Adicionalmente, se definen la función de autocovarianza (ACVF) de $\{X_t\}$ como $\gamma_x(h) = \text{Cov}(X_{t+h}, X_t)$ y la función de autocorrelación (ACF) de $\{X_t\}$ como $\rho_x(h) \equiv \frac{\gamma_x(h)}{\gamma_x(0)} = \text{Cor}(X_{t+h}, X_t)$. Esta última se emplea en la comparación de la correlación en el componente regular de una serie de tiempo (Brockwell & Davis, 2016). Otro concepto empleado es el de la función de autocorrelación parcial (PACF), el cual es usado para medir la correlación entre una observación k periodos atrás y la observación actual, es decir, para hacer una comparación periódica de la correlación (Adhikari & Agrawal, 2013).

Ruido blanco. Si $\{X_t\}$ es una secuencia de variables aleatorias no correlacionadas cada una con media cero y varianza σ^2 , se dice que dichas variables son independientes e idénticamente distribuidas y que la secuencia se denomina ruido blanco (Brockwell & Davis, 2016). El ruido blanco se usa para describir el error aleatorio dentro de las series de tiempo.

Operador de retraso. Dada una serie de tiempo $\{X_t, t = 0, \pm 1, \dots\}$ se define el operador de retraso (*lag operator*) como una acción que permite producir el valor anterior de una variable dentro de la secuencia (Adhikari & Agrawal, 2013). Es representado matemáticamente de la siguiente forma:

$$LX_t = X_{t-1}$$

Con base en este operador se puede hacer una representación simplificada de los modelos que se muestran a continuación, usando polinomios de retraso (*lag polynomials*).

Modelo AR. De acuerdo con Adhikari & Agrawal (2013) un modelo autorregresivo (*Autoregressive*) de orden p , $AR(p)$, es una representación de una serie de tiempo en la cual el valor futuro de una variable es la combinación lineal de p observaciones pasadas y un error aleatorio (ruido blanco) junto con un término constante. Matemáticamente el modelo AR es expresado de la siguiente manera:

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$$

Siendo X_t y ε_t el valor actual y el error aleatorio respectivamente, φ_i ($i = 1, 2, \dots, p$) los parámetros del modelo y c una constante.

Modelo MA. A diferencia del modelo autorregresivo, el modelo de medias móviles (*Moving Average*) de orden q , $MA(q)$, usa los errores pasados como variables explicativas, entendiendo esto como una regresión lineal de la observación actual contra los errores aleatorios de una o más observaciones previas. Se asume que los errores aleatorios siguen una distribución normal y adicionalmente se define que los modelos MA son siempre estacionarios (Adhikari & Agrawal, 2013).

Matemáticamente el modelo MA es expresado de la siguiente manera:

$$X_t = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Siendo μ la media de la serie y θ_j ($j = 1, 2, \dots, q$) los parámetros del modelo.

Modelo ARMA. El modelo autorregresivo de medias móviles (*Autoregressive Moving Average*) es la combinación de los modelos AR (orden p) y MA (orden q). Molina (2009a) indica que los modelos ARMA son útiles para representar una gran variedad de series estacionarias, debido a que los procesos que los componen tienen ciertas limitantes por separado. Para el caso de los modelos AR estos no pueden representar series de memoria muy corta donde el valor actual de la serie sólo esté correlacionado con un número pequeño de valores de la serie, mientras que los modelos MA son función de un número finito de errores pasados y no de todos.

El modelo ARMA(p, q) es expresado matemáticamente de la siguiente manera:

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

Donde p es el término autorregresivo y q el término de media móvil.

Otra forma de expresarlo es mediante polinomios del operador de retraso o polinomios de retraso, como se muestra a continuación (Adhikari & Agrawal, 2013):

$$\text{Modelo AR}(p): \varepsilon_t = \varphi(L)X_t.$$

$$\text{Modelo MA}(q): X_t = \theta(L)\varepsilon_t.$$

$$\text{Modelo ARMA}(p, q): \varphi(L)X_t = \theta(L)\varepsilon_t.$$

$$\text{Siendo } \varphi(L) = 1 - \sum_{i=1}^p \varphi_i L^i \text{ y } \theta(L) = 1 + \sum_{j=1}^q \theta_j L^j.$$

En esta clase de modelos se hace uso de las funciones de autocorrelación (ACF) y de autocorrelación parcial (PACF) con el fin de determinar el orden de los términos de MA y AR respectivamente, pues la ACF ayuda a revelar cuáles son los q primeros coeficientes no nulos y la PACF los p primeros coeficientes no nulos (Molina, 2009a). La Tabla 2, planteada por Campos (2018), describe el comportamiento de la función de autocorrelación y su incidencia en la selección del modelo más adecuado.

Tabla 2. Selección de modelo con respecto al comportamiento de la función ACF.

Comportamiento ACF	Modelo
Serie exponencial decayendo a cero.	Modelo autorregresivo AR. PACF se debe emplear para identificar el orden del modelo.
Picos alternativos positivos y negativos, decayendo a cero.	Modelo autorregresivo AR. PACF se debe emplear para identificar el orden del modelo.
Uno o más picos en serie, los restantes son cero.	Modelo de media móvil AM. El orden se identifica donde el gráfico se convierte en cero.
Después de algunos retrocesos en general la serie va decayendo.	Modelo autorregresivo de medias móviles ARMA.
La serie total es cero o casi cero.	Datos aleatorios.
Valores medios a intervalos fijos.	Se necesita incluir el término AR de periodicidad.
Picos visibles que no decaen a cero.	Serie no estacionaria

Modelo ARIMA. Dado que en la práctica muchas series de tiempo muestran un comportamiento no estacionario, los modelos ARMA son inadecuados para describir esta clase de series. Debido a lo anterior se propone el Modelo Autorregresivo Integrado de

Medias Móviles (*Autoregressive Integrated Moving Average*), el cual aplica diferenciación finita de puntos de datos en una serie no estacionaria para transformarla en estacionaria (Adhikari & Agrawal, 2013). Una serie es integrada de orden d , $d \geq 0$, si al diferenciarla d veces resulta una serie estacionaria (Molina, 2009b).

De acuerdo con Adhikari & Agrawal (2013) la formulación matemática del modelo ARIMA(p, d, q) usando polinomios de retraso se muestra a continuación:

$$\varphi(L)(1-L)^d X_t = \theta(L)\varepsilon_t$$

$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1-L)^d X_t = \left(1 + \sum_{j=1}^q \theta_j L^j\right)$$

Aquí p , d , y q son enteros mayores o iguales a cero y se refieren al orden de los componentes autorregresivo, integrado y de media móvil respectivamente. El entero d controla el nivel de diferenciación, siendo $d=1$ suficiente en la mayoría de los casos. Cuando $d=0$ se reduce a un modelo ARMA(p, q).

Modelo SARIMA. Si adicional al comportamiento no estacionario una serie de tiempo presenta un comportamiento periódico cada cierta cantidad de observaciones, existe una generalización del modelo ARIMA denominada Modelo Periódico Autorregresivo Integrado de Medias Móviles o SARIMA (*Seasonal Autoregressive Integrated Moving Average*). En este modelo se incorporan los órdenes autorregresivo P , integrado D y de media móvil Q para el componente periódico de la serie y un valor de periodo s que indica la frecuencia de observación del comportamiento, con el fin de realizar una diferenciación entre la observación actual y la del periodo anterior y de esta manera remover el componente no estacionario de la serie. El período s se puede expresar

por ejemplo en términos mensuales ($s = 12$), trimestrales ($s = 4$), semestrales ($s = 6$) o anuales ($s = 1$). Empleando los polinomios de retraso, el modelo SARIMA(p, d, q) x (P, D, Q) $_s$ se expresa de la siguiente manera:

$$\Phi_P(L^s)\varphi_p(L)(1-L)^d(1-L^s)^D X_t = \theta_Q(L^s)\theta_q(L)\varepsilon_t$$

$$\Phi_P(L^s)\varphi_p(L)z_t = \theta_Q(L^s)\theta_q(L)\varepsilon_t$$

Donde z_t es la serie diferenciada periódicamente (Adhikari & Agrawal, 2013).

4. Estado del Arte

Con el fin de establecer una guía de conocimientos aplicados sobre predicción de precios agropecuarios para la ADR, se decidió hacer una búsqueda de documentación relacionada con la producción en este sector. Como resultado se encontraron varios trabajos que muestran un análisis de precios de alimentos de la canasta básica familiar con el fin de pronosticar el comportamiento de dichos valores en el mercado. Esto adicionalmente permitió conocer conclusiones sobre la implementación de modelos de regresión en series de tiempo y así extender la búsqueda de trabajos de investigación en otras áreas. En la Tabla 3 se pueden apreciar las investigaciones marco sustento del proyecto.

Tabla 3. *Investigaciones marco.*

Referencia	Temática	Conceptos empleados
(Adanacioglu & Yercan, 2012)	Previsión de precios del tomate en la ciudad de Antalya (Turquía) con base en datos de precios desde el año 2000 al año 2010 y en la variabilidad de estos dada la periodicidad del producto.	Modelos AR, ARIMA y SARIMA
(Etuk & Igbudu, 2015)	Análisis de la variación de precios mensuales de limón autóctono con el fin de establecer herramientas para la gestión de información agrícola de siembra de cítricos en la región de Lattakia (Siria).	Modelos ARIMA y SARIMA, diagramas de correlación (correlogramas)
(Otu Archibong, Osuji, Opara, Ifeyinwa & Iheagwara, 2013)	Previsión de las tasas de inflación mensuales para el primer trimestre del año 2014 en Nigeria empleando datos de noviembre de 2003 a octubre de 2013.	Modelos ARMA, ARIMA y SARIMA
(Ruiz Hernández, Barrios Puente & Gómez Gómez, 2019)	Análisis de series de tiempo para generar una predicción de 12 meses sobre los precios promedio mensuales de la manzana red delicious en México, de enero de 1998 a julio de 2017.	Modelo SARIMA, correlogramas

Referencia	Temática	Conceptos empleados
(Wang & Wang, 2016)	Estudio de precios de productos agrícolas basado en información publicada en internet para construcción de modelo empírico de predicción de precios.	Rastreo de datos en páginas web (<i>web crawling</i>), modelo ARIMA
(Barrientos, Rondón & Melo, 2014)	Análisis del comportamiento de precios de las variedades de papa Parda Pastusa y Diacol Capiro en Colombia, desde 1995 a 2011.	Modelo SARIMA
(Botero & Cano, 2008)	Presentación de una metodología para la implementación de modelos de regresión sobre la serie de tiempo de precios de la energía en la bolsa de valores en Colombia.	Modelo AR con componente periódico

Para el caso de Adanacioglu & Yercan (2012) se muestra un análisis de la variabilidad y periodicidad de los precios del tomate con base en los datos recolectados. Se planteó un modelo SARIMA $(1, 0, 0) \times (1, 1, 1)_{12}$ para realizar la predicción de 12 meses sobre los precios del tomate, partiendo del análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) como elementos que conformaron los criterios para seleccionar los valores de los componentes del modelo.

En el trabajo de Etuk & Igbudu (2015) se usan gráficos de correlación (correlogramas) para posteriormente tabular los valores arrojados por estas gráficas, comparando la autocorrelación y la correlación parcial de la serie para determinar la naturaleza periódica de los datos y tomando los valores de error medio más pequeños, arrojados por el mismo análisis, para determinar que el modelo que más se ajusta a los datos es un SARIMA $(0,1,1) \times (1,0,1)_{12}$, con base en datos mensuales de precios de limón autóctono desde 2010 a 2013.

Mediante un análisis de la serie de tiempo de la tasa de inflación en Nigeria, Otu Archibong et al. (2013) encontraron una tendencia decreciente y un comportamiento no estacionario. A través de la diferenciación de la serie y la revisión de los valores de ACF

y PACF respectivos, se pudo comprobar que la serie presentaba un comportamiento periódico. Con la revisión de los coeficientes arrojados por el análisis de las funciones de autocorrelación y correlación parcial se logró determinar el orden de cada uno de los componentes del modelo, dando como resultado un modelo SARIMA(1, 1, 1) \times (0, 0, 1)₁₂.

De acuerdo con Ruiz Hernández et al. (2019), la serie de tiempo del precio de la manzana entre 1998 y 2017 mostró una varianza incremental, sobre la cual concluyen que la oscilación de los valores va aumentando a medida que pasa el tiempo. Para minimizar la varianza convirtieron la serie en logaritmos y luego la transformaron en primeras diferencias, para poder trabajar con una serie estacionaria en varianza y tendencia. Con los correlogramas para ACF y PACF de la serie en logaritmos, se planteó un modelo SARIMA(2,1,0) \times (1,1,0)₁₂ que posteriormente fue verificado por software especializado (Gretl). Finalmente se realizó la predicción sobre 12 meses y se verificó el error relativo de los valores para concluir que la previsión es válida.

En el artículo de Wang & Wang (2016) se menciona desde 1994 en China se encuentra una gran cantidad de información de precios disponibles en internet, pero que estos no han sido integrados como soporte en la toma de decisiones con respecto a políticas de producción o mercadeo agrícola. Por esta razón los autores implementaron un desarrollo para rastrear y extraer datos web (web crawling) de la plataforma de comercio de productos agrícolas denominada Xinxin. Con los datos obtenidos se hizo la formulación de un modelo ARIMA(3,1,2) y para probar la efectividad del modelo se tomaron los datos del precio del pepino entre el 16 de agosto y el 3 de septiembre de 2013 y se compararon contra la predicción realizada por el modelo. El resultado de la

comparación arrojó que más del 80% de los valores predichos se encontraban en un error relativo menor al 10%.

Por su parte, el estudio de Barrientos et al. (2014) busca abordar un modelo que permita hacer pronósticos del precio de la papa en las variedades Parda Pastusa y Diacol Capiro con base en los precios entre los años 1995 a 2011, con el fin de ayudar a una mejor planificación del cultivo de este alimento y de su mercadeo. La problemática surge debido a la caída del precio que desencadenó una crisis en mayo de 2013, afectando principalmente a los pequeños productores. Al realizar la inspección de los datos se encontró una tendencia al alza de los precios nominales y una tendencia a la baja de los precios reales, entendiendo estos últimos como los valores resultado de eliminar la inflación. Posteriormente se encontraron tendencias de periodicidad que corresponden a los momentos de siembra en la época de lluvias, después de ciclos de heladas y a lo largo del año. Como resultado de los análisis realizados se planteó un modelo $SARIMA(2,1,2) \times (2,1,2)_{12}$ y se concluyó que se pueden realizar predicciones de máximo 5 meses con un 95% de confianza.

Por último, la investigación de Botero & Cano (2008) trabaja la problemática del mercado mayorista de energía del sector eléctrico colombiano a raíz de la crisis durante los años ochenta y su incidencia en la formación de los precios de la energía en la bolsa en Colombia. Para la implementación del modelo de estimación de precios de la energía se tomó la serie de tiempo de costos con datos comprendidos entre enero de 1996 y junio de 2007 y se hizo un análisis para determinar ciclos, periodicidad, varianza de las observaciones, incrementos súbitos e irregularidades. Posteriormente se estudió la

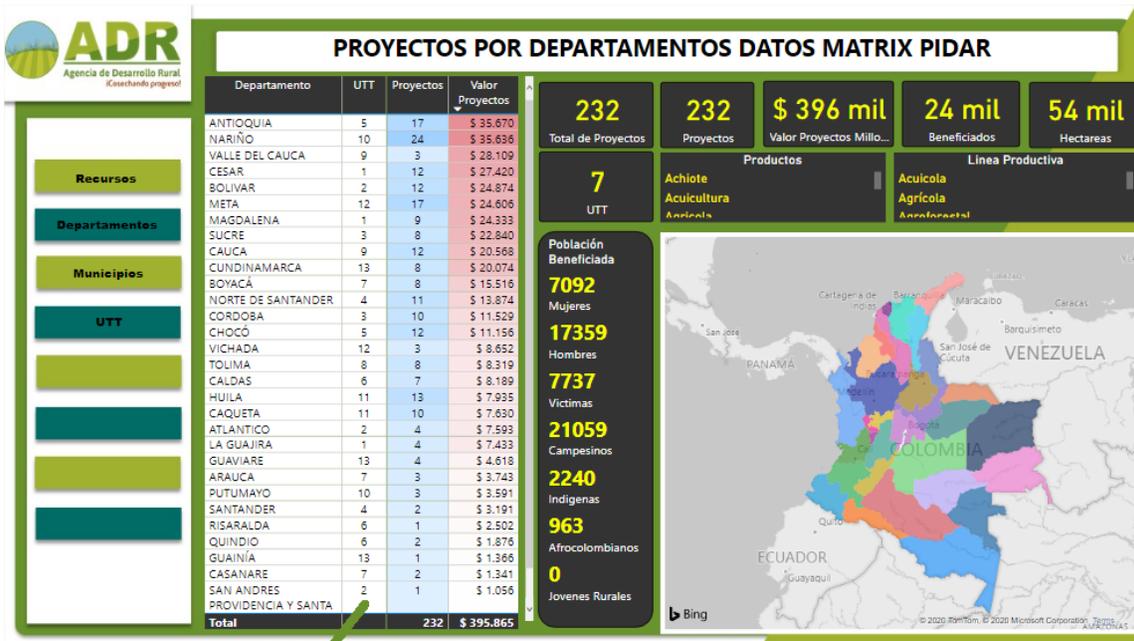
autocorrelación y la autocorrelación parcial de la variable de precios con respecto a valores anteriores, dando como resultado un modelo AR(2) con componente periódico. No obstante lo anterior, se tuvo que considerar solo una porción de la serie (junio 2001 a junio de 2007) debido a que en determinados momentos se presentaron intervenciones sobre el mercado que afectaron el comportamiento de los datos.

Dentro de este orden de ideas, las investigaciones consultadas revelan una metodología consistente en el análisis de las series de tiempo validando el componente periódico y estacionario, además de emplear las funciones de autocorrelación y autocorrelación parcial como punto de partida para determinar el orden del modelo que se plantee. También se destaca la inspección del error medio arrojado en las predicciones del modelo para determinar su exactitud.

5. Desarrollo

El punto de partida para el desarrollo del proyecto fue la revisión de la información de proyectos cofinanciados por la agencia desde el año 2017 a junio de 2020. En el informe de Agencia de Desarrollo Rural [ADR] (2020) se muestra la cantidad y el valor de los PIDAR por departamento, así como las líneas productivas y los productos relacionados con las iniciativas cofinanciadas. La Figura 6 revela que la mayor cantidad de proyectos se encuentran ubicados en el departamento de Nariño y que, a excepción de Amazonas, existen proyectos en todos los departamentos que conforman la república de Colombia. Por su parte, los productos relacionados con los PIDAR son presentados en la Figura 7, encontrando que la gran mayoría de ellos se pueden clasificar como productos de la canasta básica familiar.

Dado lo anterior, se planteó la hipótesis de que una forma que podría contribuir a la priorización de los PIDAR es analizar el precio de los alimentos de origen agropecuario mediante series de tiempo, con el fin de hallar un modelo que permita estimar en qué momento los precios pueden ser más favorables desde el punto de vista del productor, influenciando positivamente la rentabilidad de las iniciativas. Adicionalmente se estimó que el costo de esta clase de productos se vio afectado de cierta manera por las restricciones de movilidad y transporte producto de la emergencia sanitaria por COVID-19, en relación con la disponibilidad de mano de obra para los procesos de siembra, recolección y despacho.



Departamento	UTT	Proyectos	Valor Proyectos
ANTIOQUIA	5	17	\$ 35.670
NARIÑO	10	24	\$ 35.636
VALLE DEL CAUCA	9	3	\$ 28.109
CESAR	1	12	\$ 27.420
BOLIVAR	2	12	\$ 24.874
META	12	17	\$ 24.606
MAGDALENA	1	9	\$ 24.333
SUCRE	3	8	\$ 22.840
CAUCA	9	12	\$ 20.568
CUNDINAMARCA	13	8	\$ 20.074
BOYACÁ	7	8	\$ 15.516
NORTE DE SANTANDER	4	11	\$ 13.874
CORDOBA	3	10	\$ 11.529
CHOCÓ	5	12	\$ 11.156
VICHADA	12	3	\$ 8.652
TOLIMA	8	8	\$ 8.319
CALDAS	6	7	\$ 8.189
HUILA	11	13	\$ 7.935
CAQUETA	11	10	\$ 7.630
ATLANTICO	2	4	\$ 7.593
LA GUAJIRA	1	4	\$ 7.433
GUAVIARE	13	4	\$ 4.618
ARAUCA	7	3	\$ 3.743
PUTUMAYO	10	3	\$ 3.591
SANTANDER	4	2	\$ 3.191
RISARALDA	6	1	\$ 2.502
QUINDIO	6	2	\$ 1.876
GUAINÍA	13	1	\$ 1.366
CASANARE	7	2	\$ 1.341
SAN ANDRES	2	1	\$ 1.056
PROVIDENCIA Y SANTA CATALINA			
VAUPES	13	1	\$ 625
Total		232	\$ 395.865

Figura 6. Proyectos cofinanciados por la ADR agrupados por departamento, 2017 - 2020. ADR (2020).

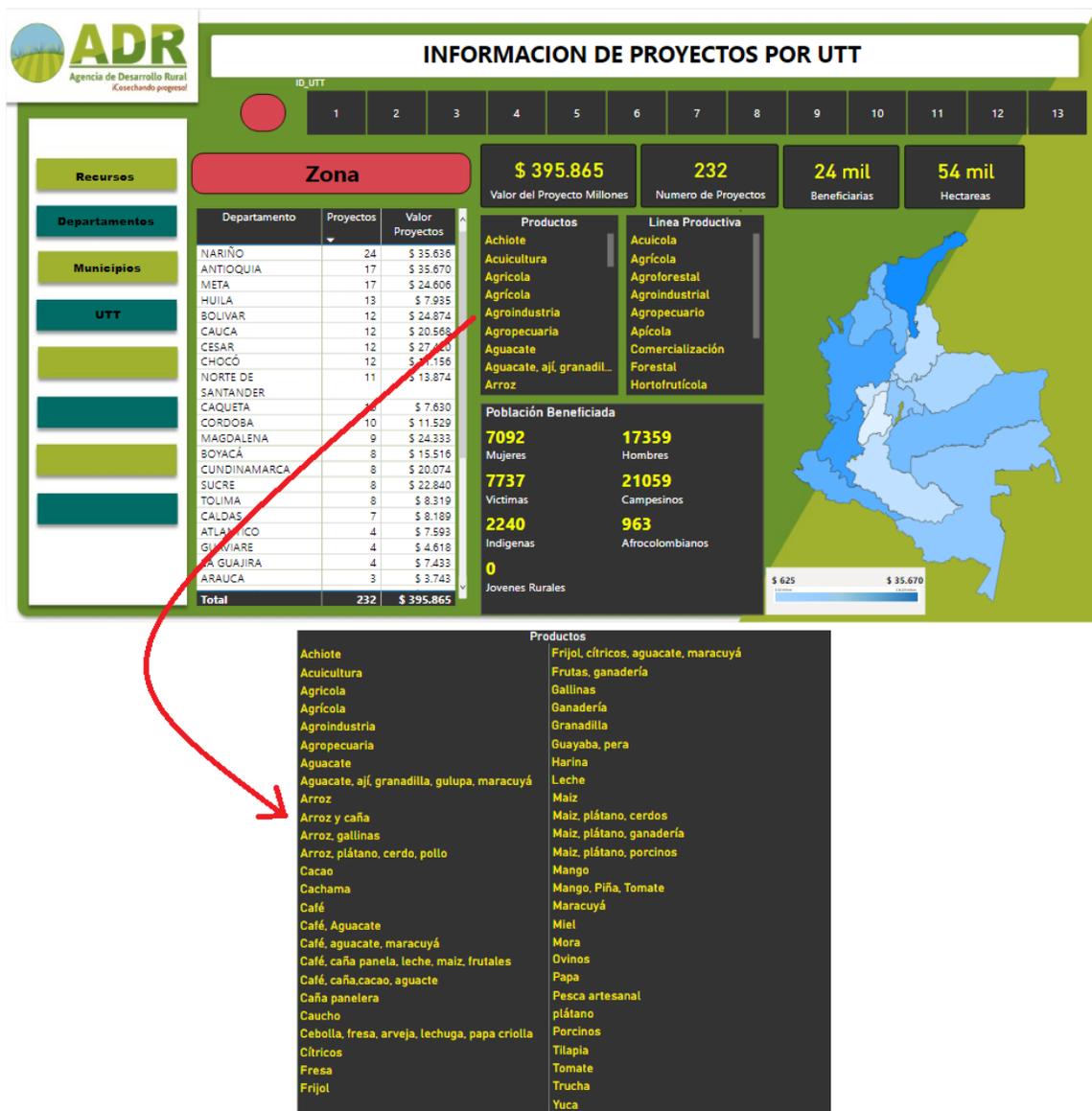


Figura 7. Productos relacionados con los PIDAR, 2017 - 2020. ADR (2020).

5.1. Obtención de datos

Teniendo en cuenta el rol que ocupa la central mayorista de abastos de Bogotá (CORABASTOS) en la fijación de precios de alimentos de origen agropecuario, se decidió tomar como base la información disponible de los valores de esta clase de productos en el rango de fechas de enero de 2012 a julio de 2020. Los datos que ofrece

CORABASTOS se encuentran distribuidos en ocho grandes grupos: hortalizas, frutas, tubérculos, plátanos, granos y procesados, lácteos, cárnicos y huevos. En la Figura 8 se muestra un ejemplo de visualización de precios de hortalizas para el día 31 de julio de 2020, discriminados por nombre, presentación, cantidad, unidad de medida y los valores del producto en las calidades extra, primera o *premium* y corriente.

Código	Producto	Presentación	Cantidad	Unidad	Extra	Primera	Corriente
100100	ACELGA	ATADO	10	KILO	100000	9000	8000
100200	AHUYAMA	KILO	1	KILO	1000	900	800
100301	AJO ROSADO	ATADO	5	KILO	40000	35000	30000
100400	ALCACHOFA	DOCENA	5	KILO	60000	55000	50000
100500	APIO	ATADO	10	KILO	15000	14000	12000

Figura 8. Precios de hortalizas publicados en la central mayorista de abastos de Bogotá para el 31 de julio de 2020. Corporación de Abastos de Bogotá [CORABASTOS] (2020).

Con la información de precios descargada, se creó una base de datos denominada ADR_PRICES en el motor PostgreSQL 12.4. En dicha base de datos se creó una tabla llamada ADR_CORABASTOS_PRICES para contener estos datos. El diseño de la tabla

es presentado en la Figura 9, mientras que en la Tabla 4 se observa un ejemplo de los datos contenidos.

adr_corabastos_prices	
123	cor_id int4
	cor_date date
ABC	cor_line varchar(150)
ABC	cor_chain varchar(250)
ABC	cor_unit varchar(50)
123	cor_unit_amount numeric(6,2)
ABC	cor_unit_label varchar(50)
123	cor_price_extra int4
123	cor_price_premium int4
123	cor_price_common int4

Figura 9. Diseño tabla para precios de la central mayorista de abastos de Bogotá ADR_CORABASTOS_PRICES.

Tabla 4. Ejemplo datos tabla ADR_CORABASTOS_PRICES.

Fecha precio	Línea	Producto	Unidad	Cantidad	Presentación	Precio calidad extra	Precio calidad primera	Precio calidad corriente
2020-07-01	CARNICOS	ALAS DE POLLO	KILO	1	KILO	7000	7000	7000
2020-07-01	FRUTAS	AGUACATE	KILO	1	KILO	2500	2000	1800
2020-07-01	GRANOS Y PROCESADOS	ACEITE (1000 c.c)	CAJA	12	KILO	5833	5833	5833
2020-07-01	HORTALIZAS	ACELGA	ATADO	10	KILO	1000	900	800
2020-07-01	HUEVOS	HUEVO BLANCO A	UNIDAD	1	KILO	250	250	250
2020-07-01	LACTEOS	CUAJADA	KILO	1	KILO	10000	10000	10000
2020-07-01	PLATANOS	PLATANO COLICERO	KILO	1	KILO	1400	1200	1000
2020-07-01	TUBERCULOS	ARRACACHA	BULTO	50	KILO	2200	2100	2000

Adicional a la fuente de datos descrita anteriormente y con el fin de aproximar los precios a valores actuales se decidió extraer la información del Índice de Precios al

Consumidor o IPC. De acuerdo con el Departamento Administrativo Nacional de Estadística [DANE] (2009) el IPC es una medida de la variación en el precio de bienes y servicios representativos del consumo de los hogares conocido como canasta. Se calcula registrando el precio de los artículos de la canasta en varios canales de distribución del país, tales como tiendas de barrio, supermercados, plazas de abastos y grandes superficies. Los datos históricos del valor total del IPC con periodicidad mensual se encuentran disponibles en el sitio web del Banco de la República de Colombia. La Figura 10 muestra la consulta realizada sobre este índice.

Inflación total al consumidor 1/
1.2.5. Serie histórica periodicidad mensual

Información disponible desde julio de 1954.

Nota: De acuerdo con el literal j) del artículo 2 del Decreto 3167 de 1968 le corresponde al DANE "Establecer índices de precios a nivel del productor, del distribuidor y del consumidor (...)", y el literal i) del numeral 1 del artículo 2 del Decreto 262 de 2004 según el cual el DANE debe "certificar la información estadística, siempre que se refiera a resultados generados, validados y aprobados por el Departamento".

Consulte la plataforma que ha dispuesto el DANE por medio de la cual expide y certifica los Índices de Precios al Consumidor.

Año(aaaa)- Mes(mm)	Índice	Inflación anual %	Inflación mensual %	Inflación año corrido %
2003-01	50,42	7,39	1,17	1,17
2003-02	50,98	7,24	1,11	2,30
2003-03	51,51	7,60	1,05	3,37
2003-04	52,10	7,85	1,15	4,56
2003-05	52,36	7,73	0,49	5,07
2003-06	52,33	7,21	-0,05	5,01
2003-07	52,26	7,04	-0,14	4,86
2003-08	52,42	7,26	0,31	5,18
2003-09	52,53	7,11	0,22	5,42
2003-10	52,56	6,58	0,06	5,48
2003-11	52,75	6,13	0,35	5,85

Figura 10. Consulta del valor histórico mensual del Índice de Precios al Consumidor – IPC. Banco de la República (2020).

De la misma forma en que se realizó para la información de precios, en el caso de los datos del IPC se creó una tabla en la base de datos, llamada ADR_IPC, en donde se

almacenó el valor del índice y el mes correspondiente. El diseño de esta tabla se puede apreciar en la Figura 11.

adr_ipc	
123  ipc_id	int4
 ipc_fecha	date
123  ipc_valor	numeric(6,2)

Figura 11. Diseño tabla para valor histórico mensual del Índice de Precios al Consumidor en Colombia ADR_IPC.

5.2. Análisis previo de datos

A partir de este punto se hizo uso del lenguaje de programación Python 3.0 (a través de la distribución libre Anaconda) y la aplicación web Jupyter Notebook para visualización de código y resultados. El código requerido para la generación de gráficos se basó en los ejemplos expuestos por Brownlee (2017).

Como paso inicial se seleccionó uno de los productos disponibles dentro de la tabla de precios cargadas, el cual fue “Arroz Corriente”. Mediante un comando de conexión a la base de datos se consultó la información y se cargó en una estructura para su posterior uso, lo cual puede verse en la Figura 12.

```

In [2]: #Creacion de conexion con base de datos ADR_PRICES
engine = create_engine('postgresql://postgres:admin@localhost/ADR_PRICES')

In [3]: #Lectura de la tabla con datos de IPC
ipc = pd.read_sql_table("adr_ipc", con=engine)

In [4]: #Lectura de la tabla con datos de precios
sqlPrices = "select * from adr_corabastos_prices where cor_chain = 'ARROZ CORRIENTE' ORDER BY cor_id ASC"
precios_df = pd.read_sql_query(sqlPrices, con=engine)
precios_df

```

Out[4]:

	cor_id	cor_date	cor_line	cor_chain	cor_unit	cor_unit_amount	cor_unit_label	cor_price_extra	cor_price_premium	cor_price_common
0	328445	2012-01-12	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2000	2000	2000
1	330231	2012-02-09	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2000	2000	2000
2	331591	2012-03-01	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2060	2060	2060
3	332947	2012-04-04	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2200	2200	2200
4	333265	2012-04-11	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2200	2200	2200
...
2118	703495	2020-07-16	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	3200	3200	3200

Figura 12. Consulta en base de datos del histórico de precios del arroz corriente empleando Python.

Tal como se mencionó al final del apartado anterior, para poder contar con el precio actualizado al momento de realizar el análisis se empleó la información cargada del IPC en dos vías: por un lado el valor del índice para cada uno de los meses y años que conforman las fechas de los precios y por el otro el valor para el mes anterior a la fecha actual en la que se realizó el análisis (esto debido a la disponibilidad de datos, que siempre son mes vencido). Con estos dos grupos de valores se implementó una fórmula dentro del código para calcular el precio actual del producto en cada fecha, la cual consistió en tomar el precio de calidad primera o *premium*, considerado como el valor intermedio existente en los datos y multiplicarlo por un factor resultante entre la división del valor actual del IPC sobre el valor del IPC del período del precio, es decir:

$$\text{precio actual} = \text{precio calidad primera} \times \left(\frac{\text{valor IPC actual}}{\text{valor IPC precio calidad primera}} \right)$$

Los resultados del cálculo del precio actual se muestran en la Figura 13.

```
In [7]: #Calculo del periodo (año-mes) para los datos de IPC
ipc['Periodo'] = ipc['ipc_fecha'].dt.to_period("M")

#Calculo del precio actual con base en el ajuste del IPC vigente / IPC del periodo del precio
df_actual = precios_df.merge(ipc[['ipc_fecha', 'ipc_valor', 'Periodo']], on="Periodo", how='left')
df_actual['Precio actual'] = df_actual['Precio Calidad Primera'] * (current_ipc/df_actual['ipc_valor'])
df_actual
```

Out[7]:

d	Fecha	Categorías	Producto	Presentacion	Cantidad	Unidad	Precio Calidad Extra	Precio Calidad Primera	Precio Calidad Corriente	Mes	Dia	Periodo	ipc_fecha	ipc_valor	Precio actual
5	2012-01-12	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2000	2000	2000	1	12	2012-01	2012-01-01	76.75	2742.149837
1	2012-02-09	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2000	2000	2000	2	9	2012-02	2012-02-01	77.22	2725.459725
1	2012-03-01	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2060	2060	2060	3	1	2012-03	2012-03-01	77.31	2803.955504
7	2012-04-04	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2200	2200	2200	4	4	2012-04	2012-04-01	77.42	2990.260914
5	2012-04-11	GRANOS Y PROCESADOS	ARROZ CORRIENTE	BULTO	50.0	KILO	2200	2200	2200	4	11	2012-04	2012-04-01	77.42	2990.260914

Figura 13. Cálculo de precio actual para el histórico de precios del arroz corriente empleando Python.

Lo siguiente a realizar, contando con la información ajustada, fue calcular el valor promedio por mes. Este dato es base para generar el diagrama *Boxplot*, la cual tiene como objetivo visualizar la distribución del precio con respecto los valores mínimo y máximo y a medidas de tendencia central como el promedio y la mediana. También permite identificar la existencia de valores atípicos.

En el caso del producto analizado, mediante la Figura 14, se puede apreciar que el precio a lo largo de los años 2012 a 2020 tiene un comportamiento creciente entre los meses de enero y abril pero con valores mínimos y máximos muy distantes. También se evidencia la presencia de valores atípicos en los meses de marzo, julio, agosto y diciembre.

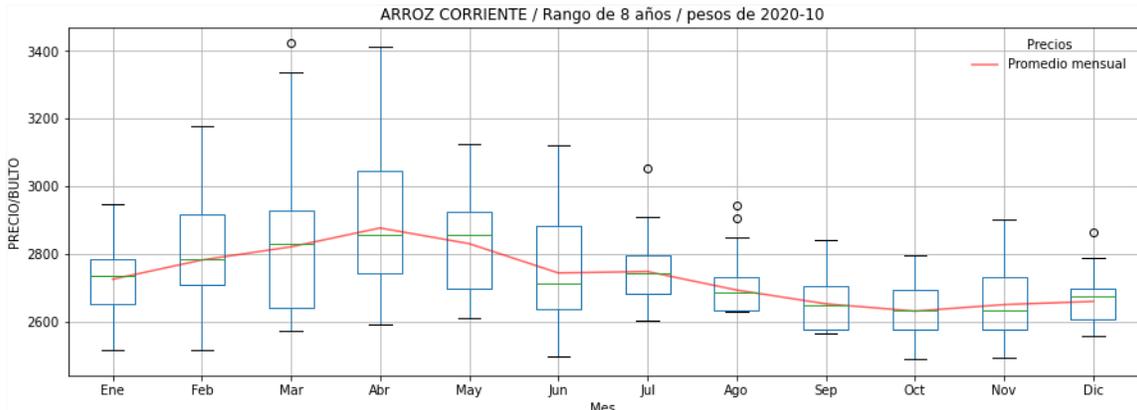


Figura 14. Diagrama *Boxplot* del arroz corriente agrupado mensualmente entre 2012 y 2020.

Lo anteriormente mencionado se evidencia con más claridad en la Figura 15, que corresponde al gráfico de la serie de tiempo y en el que se observan picos abruptos a lo largo de varios años.



Figura 15. Gráfico de serie de tiempo del precio del arroz corriente entre 2012 y 2020.

5.3. Selección de modelo

De acuerdo con Brownlee (2017) es posible realizar una prueba estadística, denominada prueba de Dickey-Fuller aumentada (ADF), que permite determinar la probabilidad de que una serie sea no estacionaria de acuerdo con unos niveles de

confianza establecidos (1%, 5% y 10%) y en el campo práctico sirve para validar si la serie debe ser diferenciada o no para transformarla en estacionaria. Para el caso analizado la prueba arrojó un estadístico mayor que el nivel más crítico (1%) y un valor p mayor al 5%, lo que llevó a inferir que la serie es no estacionaria. Los resultados se muestran a continuación:

```
ADF Statistic: -1.817728
p-value: 0.371697
Critical values:
  1%: -3.498
  5%: -2.891
 10%: -2.582
```

Luego de observar el comportamiento diario de la serie se procedió a agrupar los datos de manera mensual, calculando el promedio de los precios para cada período. Esto con el objetivo de suavizar la serie, tal como se puede apreciar en la Figura 16, además de facilitar la interpretación de las gráficas de las funciones ACF y PACF, presentadas en las Figuras 17 y 18 respectivamente.

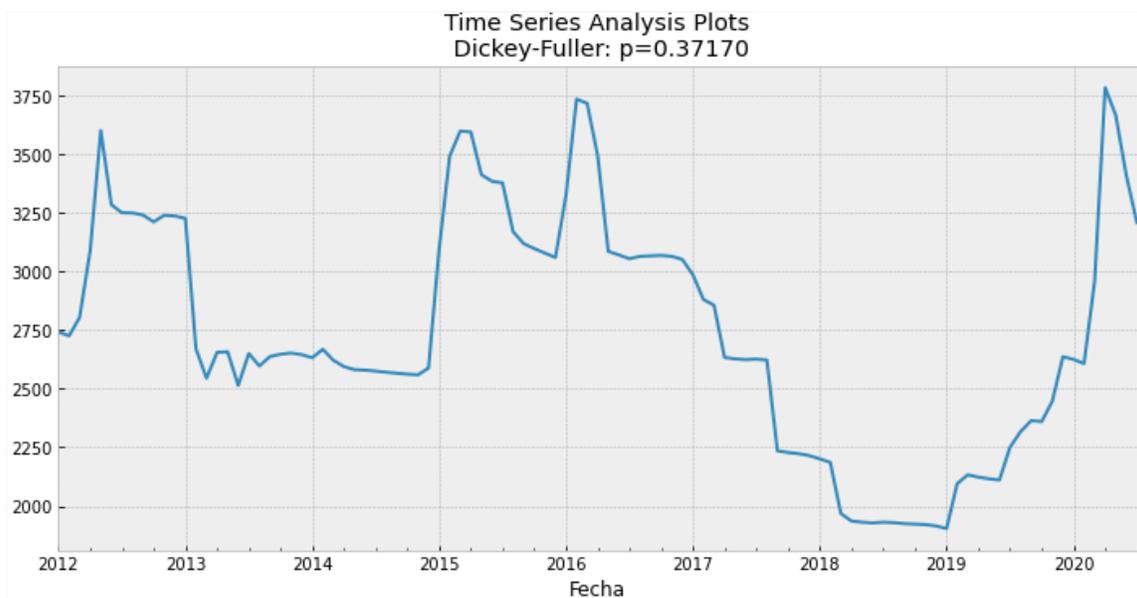


Figura 16. Gráfico de serie de tiempo de arroz corriente con valores agrupados por mes.

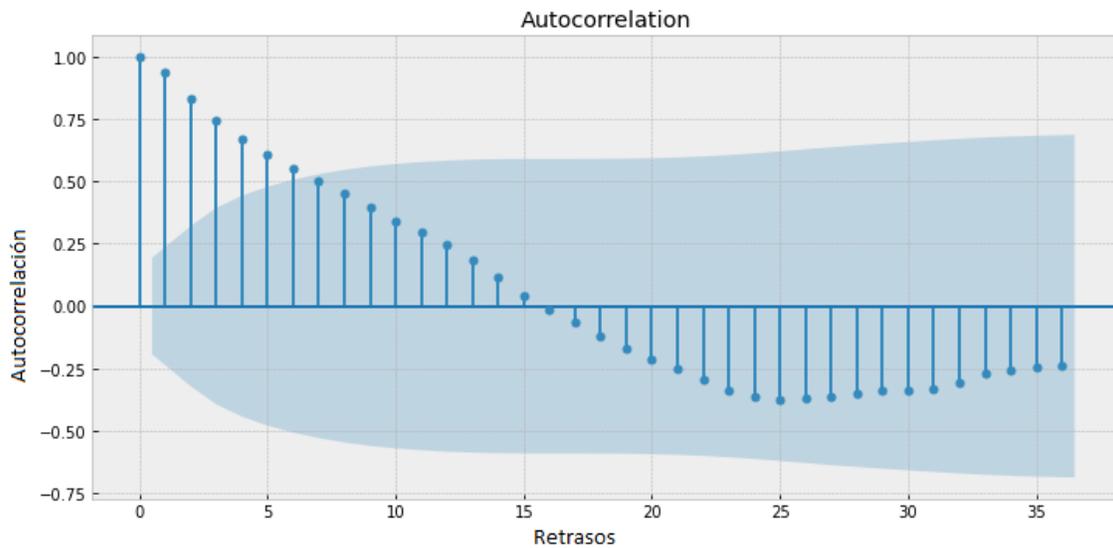


Figura 17. Gráfico de función de autocorrelación con 36 retrasos (*lags*) para el precio del arroz corriente agrupado por mes.

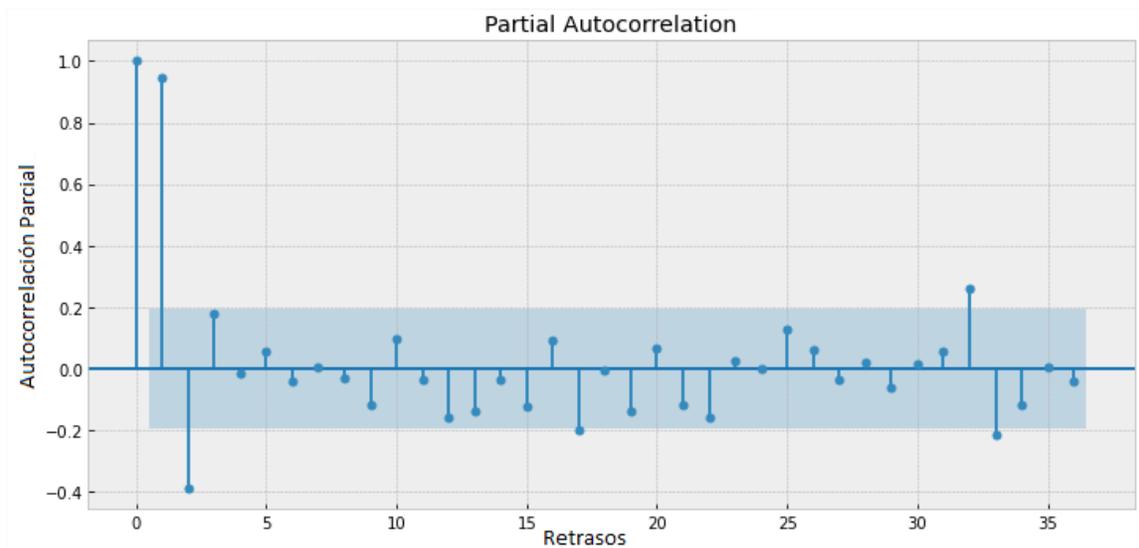


Figura 18. Gráfico de función de autocorrelación parcial con 36 retrasos (*lags*) para el precio del arroz corriente agrupado por mes.

El objetivo de analizar las gráficas de ACF y PACF es encontrar puntos en cada retraso que estén fuera del área de significancia. Estos puntos son la pauta para determinar el orden del término autorregresivo y de medias móviles. Con base en este

enunciado, la función de autocorrelación muestra 6 picos en los primeros *lags*, mientras que la función de autocorrelación muestra 3, lo que llevaría a plantear componentes AR(3) y MA(6).

Ya que se comprobó que la serie no es estacionaria, se procedió a hacer una primera diferenciación de la serie (con esto se tendría tentativamente un componente I(1)) y a realizar nuevamente la prueba ADF. En esta ocasión el estadístico arrojó un mayor valor que el punto más crítico y el valor p igual a cero, indicando la estacionariedad de la serie diferenciada:

```

ADF Statistic: -7.287442
p-value: 0.000000
Critical values:
  1%: -3.498
  5%: -2.891
 10%: -2.582

```

Posteriormente se graficó la serie diferenciada, presentándola en la Figura 19 y encontrando que los datos oscilan con respecto a cero. La función ACF de la Figura 20 solo mostró un punto levemente desbordado en el *lag* 1 y la función PACF de la Figura 21, por su parte, dos puntos en los *lags* 2 y 31.

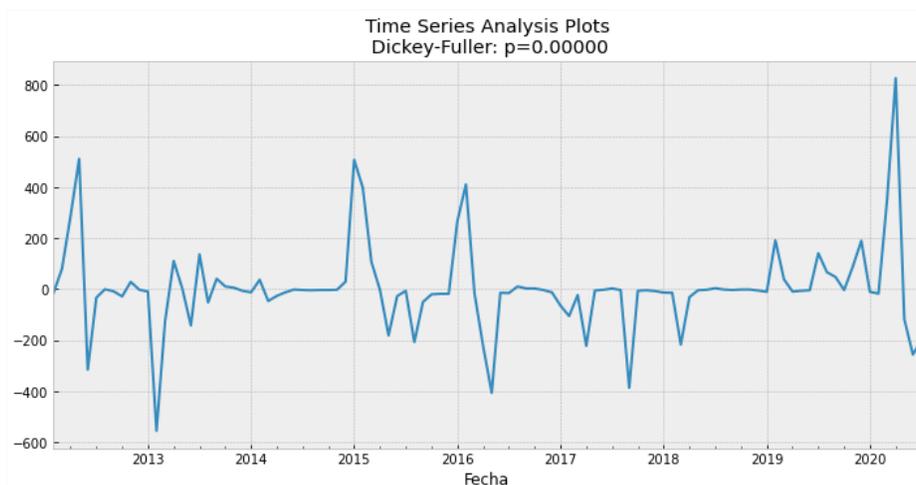


Figura 19. Gráfico de serie de tiempo diferenciada para los precios del arroz corriente.

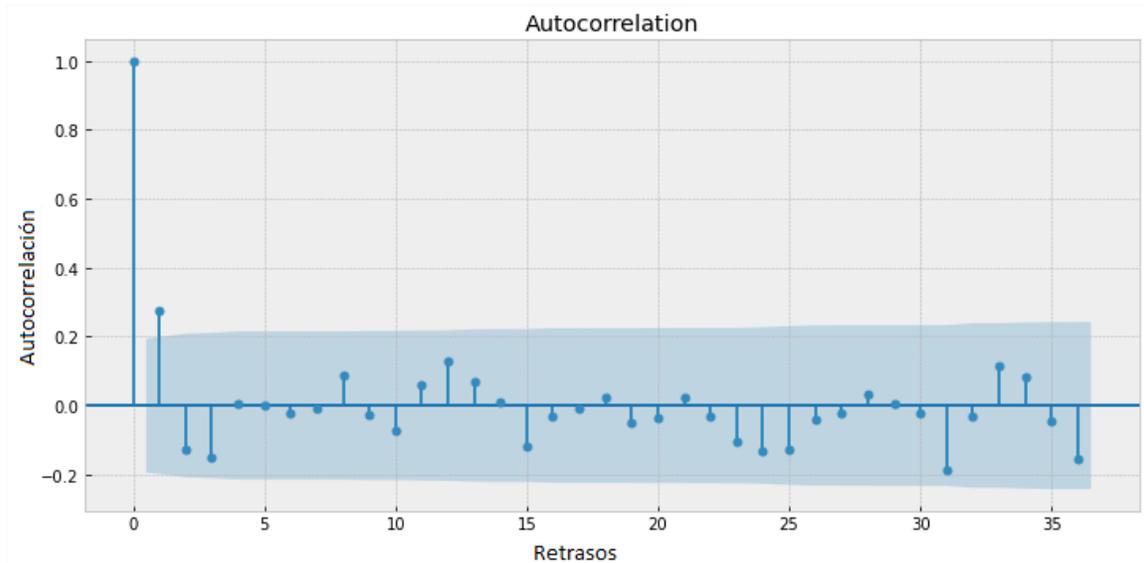


Figura 20. Gráfico de función de autocorrelación de la serie diferenciada para el precio del arroz corriente.

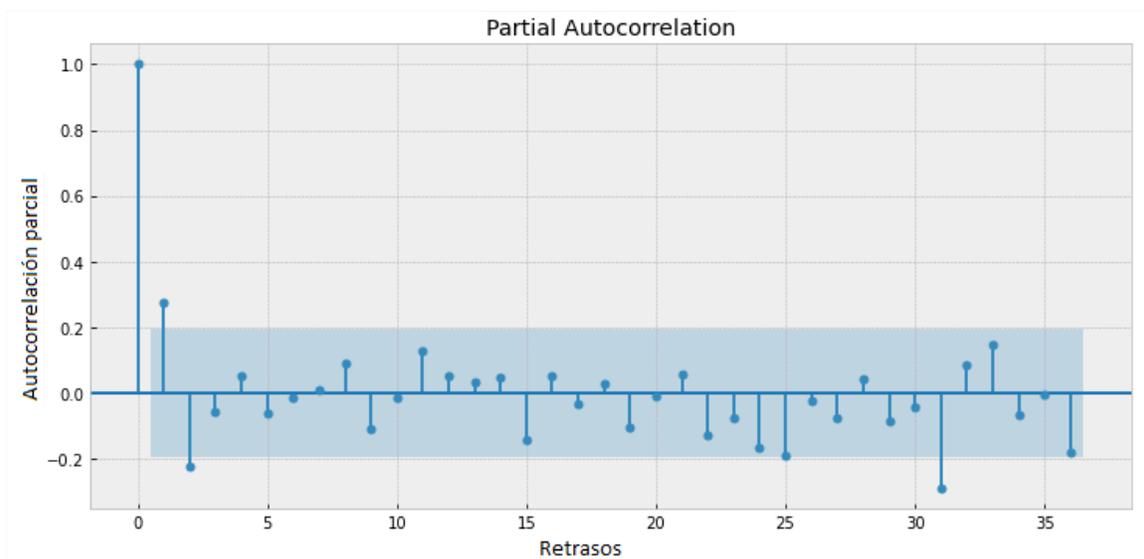


Figura 21. Gráfico de función de autocorrelación parcial de la serie diferenciada para el precio del arroz corriente

Con esta información se plantea tentativamente un modelo SARIMA(3,1,6) x (2,1,1)₁₂. Sin embargo es necesario hacer una evaluación del modelo con distintos órdenes y para ello se estableció un rango para los parámetros p (0 a 4), q (0 a 7), P (0 a

3) y Q (0 a 2). Todas las posibles combinaciones de estos rangos, junto con los valores de $d = 1$, $D = 1$ y $s = 12$ se cargaron en un modelo disponible en Python denominado SARIMAX para determinar cuál grupo de parámetros arroja el menor valor para el criterio de información de Akaike (AIC). El código que realizó este proceso se muestra en la Figura 22.

```
#Tentativamente es un modelo SARIMA(3,1,6)x(2,1,1)12
#Se hace una evaluación de varios modelos para encontrar el mas optimo
#por combinaciones de rangos de los parametros p, q, P y Q
ps = range(0, 4)
d=1
qs = range(0, 7)
Ps = range(0, 3)
D=1
Qs = range(0, 2)
s = 12
parameters = product(ps, qs, Ps, Qs)
parameters_list = list(parameters)
len(parameters_list)

#Definicion de funcion para encontrar los parametros adecuados de un modelo SARIMA
def optimizeSARIMA(parameters_list, d, D, s):
    """
    Return dataframe with parameters and corresponding AIC

    parameters_list - list with (p, q, P, Q) tuples
    d - integration order in ARIMA model
    D - seasonal integration order
    s - length of season
    """

    results = []
    best_aic = float("inf")

    for param in notebook.tqdm(parameters_list):
        # we need try-except because on some combinations model fails to converge
        try:
            model=sm.tsa.statespace.SARIMAX(serie_mensual['Precio actual'], order=(param[0], d, param[1]),
                                            seasonal_order=(param[2], D, param[3], s)).fit(dispatch=0)
        except:
            continue
        aic = model.aic
        # saving best model, AIC and parameters
        if aic < best_aic:
            best_model = model
            best_aic = aic
            best_param = param
            results.append([param, model.aic])

    result_table = pd.DataFrame(results)
    result_table.columns = ['parameters', 'aic']
    # sorting in ascending order, the lower AIC is - the better
    result_table = result_table.sort_values(by='aic', ascending=True).reset_index(drop=True)

    return result_table

%%time
result_table = optimizeSARIMA(parameters_list, d, D, s)
```

Figura 22. Código en Python para la evaluación de múltiples combinaciones de parámetros para el modelo SARIMA

Una vez la estimación de parámetros finalizó, el resultado obtenido fue que el conjunto de órdenes con menor valor AIC correspondió a $p=0$, $q=1$, $P=1$ y $Q=1$,

conformando así un modelo SARIMA (0,1,1) x (1,1,1)₁₂. El resumen del proceso de estimación del modelo se observa en la Figura 23, indicando un valor de AIC igual a 1205,683.

SARIMAX Results						
=====						
Dep. Variable:	Precio actual		No. Observations:	103		
Model:	SARIMAX(0, 1, 1)x(1, 1, 1, 12)		Log Likelihood	-598.842		
Date:	Mon, 30 Nov 2020		AIC	1205.683		
Time:	21:23:33		BIC	1215.683		
Sample:	01-31-2012		HQIC	1209.716		
	- 07-31-2020					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

ma.L1	0.3213	0.116	2.774	0.006	0.094	0.548
ar.S.L12	0.3870	0.261	1.484	0.138	-0.124	0.898
ma.S.L12	-0.9773	2.342	-0.417	0.676	-5.567	3.612
sigma2	2.988e+04	6.33e+04	0.472	0.637	-9.42e+04	1.54e+05
=====						
Ljung-Box (Q):	21.49		Jarque-Bera (JB):	26.06		
Prob(Q):	0.99		Prob(JB):	0.00		
Heteroskedasticity (H):	0.98		Skew:	0.63		
Prob(H) (two-sided):	0.95		Kurtosis:	5.32		
=====						

Figura 23. Resumen de estimación en Python de un modelo SARIMA (0,1,1) x (1,1,1)₁₂.

El valor del criterio AIC arrojado por el modelo a simple vista parece ser elevado, pero se desconoce si las predicciones tienen un alto grado de precisión. Para poder comprobar las posibles diferencias entre los precios de la serie y los valores que puede predecir el modelo, se realizó el cálculo de la raíz del error cuadrático medio o RMSE, empleando la información de la serie y contrastándola contra una predicción de precios desplazada 12 meses. Cabe aclarar que entre más pequeño sea el valor del error, más acertada es la predicción con respecto a los datos reales.

La Figura 24 muestra que el error es igual a 188.64 y que al comparar los valores actuales y los predichos se observa un ligero desfase de la serie de predicción, además de

que estos datos tienen más variabilidad cuando se están tratando zonas de descenso de la serie.

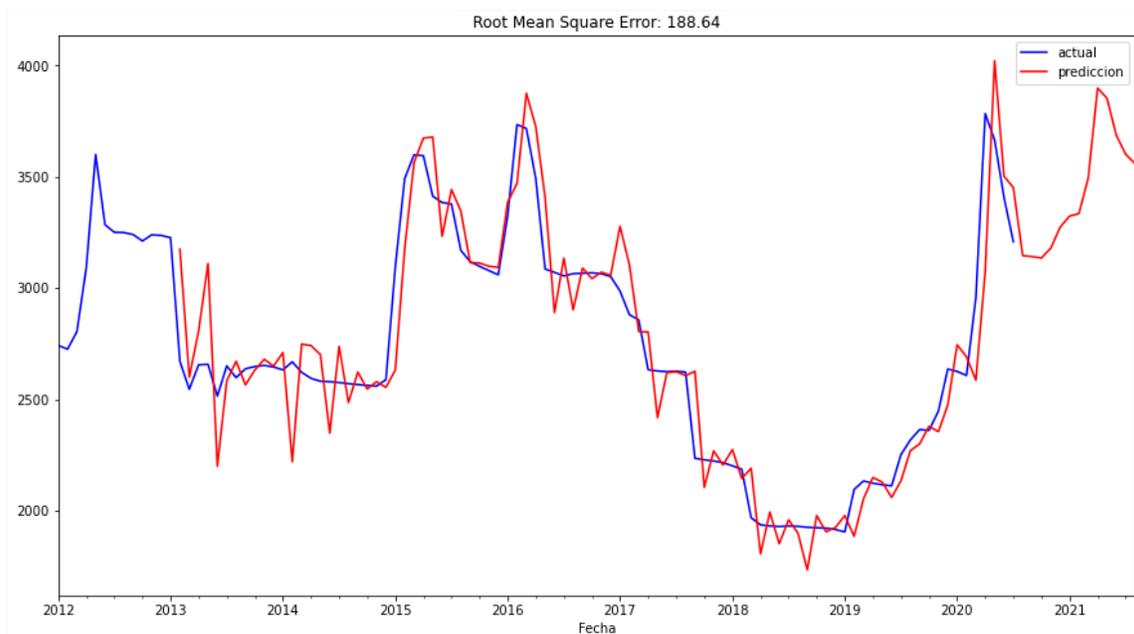


Figura 24. Valor de la raíz del error cuadrático medio y comparación de los valores actuales y predichos para el precio del arroz corriente con un modelo SARIMA $(0,1,1) \times (1,1,1)_{12}$.

Finalmente, se calculó el precio máximo, precio mínimo y el precio promedio de la serie con el objetivo de comparar estos valores con respecto al precio predicho del producto en un periodo (mes) particular. Si el precio estimado por el modelo se encuentra por encima del 70% del precio máximo, se habla de que el producto se vendería en un rango de precios altos. Si el precio estimado está ubicado por encima del 70% del precio promedio, pero, está por debajo del 70% del máximo, entonces el producto se vendería en un rango de precios promedio. Si no se cumple ninguna de las dos condiciones anteriores, entonces el producto se vendería en un rango de precios bajo.

Con todos los pasos descritos en este apartado se fijó una ruta para la evaluación de otros productos, teniendo en cuenta que es necesario el ajuste de los parámetros y su evaluación debido a que el comportamiento de los precios varía de un producto a otro.

6. Casos de estudio

Los casos de estudio para determinar el comportamiento futuro de precios se seleccionaron de la siguiente manera: el primero (papa pastusa) se consideró debido a que uno de los PIDAR más recientes tiene como producto a este alimento, mientras que el segundo (huevo blanco AA) se escogió debido a que es un producto de la canasta básica familiar y no se encontraron proyectos explícitos relacionados con este, razón por la cual y dependiendo del comportamiento hallado, podría llegar a ser considerado en un futuro por la ADR.

6.1. Caso de estudio: Papa pastusa

De acuerdo con los pasos establecidos para el análisis del producto, se generó el diagrama *Boxplot* para verificación de los datos mensualmente. La Figura 25 muestra que en los meses de marzo, junio, julio, y agosto hay presencia de datos anómalos, pero sin tener influencia significativa sobre el promedio, pues este no se encuentra por fuera de los percentiles Q1 y Q3 de datos.

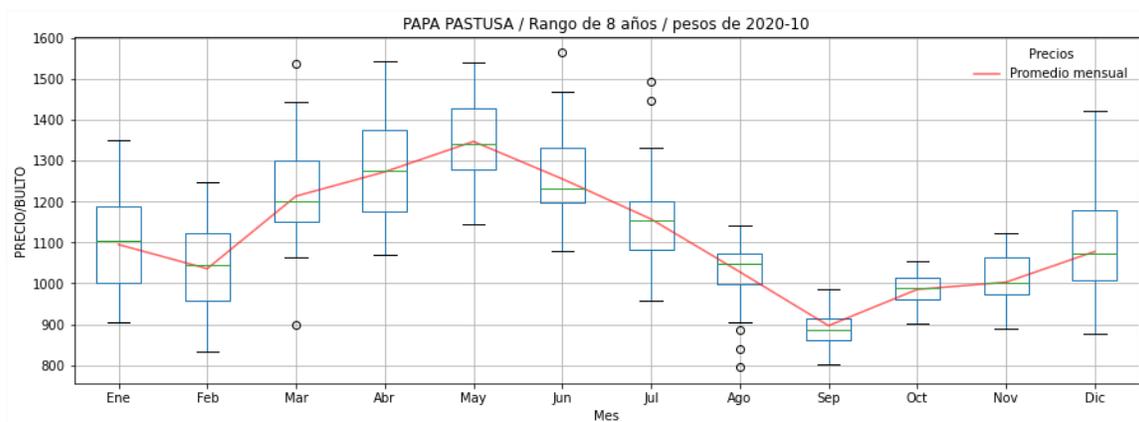


Figura 25. Diagrama Bloxplot del comportamiento de precios mensuales de la papa pastusa, de 2012 a 2020.

Por su parte, la prueba ADF reveló que la serie es estacionaria debido a que el estadístico es menor al valor crítico del 1%:

```

ADF Statistic: -4.866910
p-value: 0.000040
Critical values:
    1%: -3.498
    5%: -2.891
    10%: -2.582

```

La Figura 26 presenta el gráfico de la serie agrupado mensualmente y los diagramas de las funciones ACF y PACF. En el primero se aprecia que la serie presenta comportamientos cíclicos dos veces al año, mientras que las funciones sugieren un modelo tentativo SARIMA(3,0,2) x (3,0,2)₁₂. No se realizó diferenciación de la serie debido a los resultados antes descritos de la prueba ADF.

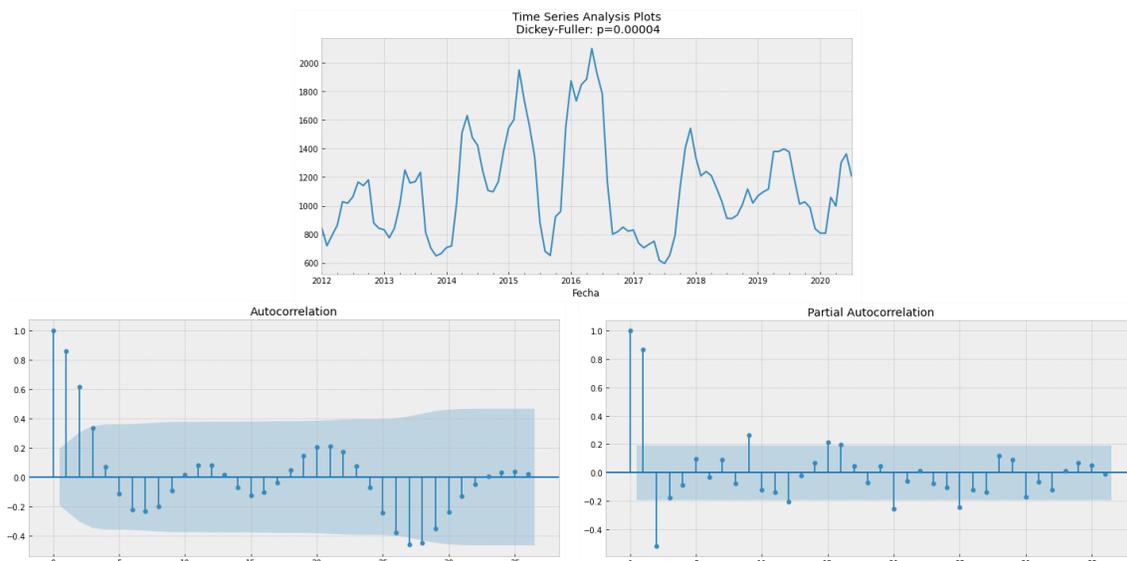


Figura 26. Gráfico de serie de tiempo, función ACF y PACF para los precios mensuales de la papa pastusa, de 2012 a 2020.

Después de la prueba de varios modelos con los rangos $p=[0,4]$, $q=[0,3]$, $P=[0,4]$, $Q=[0,3]$ se obtuvo un modelo SARIMA(2,0,0) x (1,0,1)₁₂. La Figura 27 muestra el resumen del modelo, con un valor AIC de 1352,052.

SARIMAX Results						
=====						
Dep. Variable:	Precio actual		No. Observations:	103		
Model:	SARIMAX(2, 0, 0)x(1, 0, [1], 12)		Log Likelihood	-671.026		
Date:	Mon, 30 Nov 2020		AIC	1352.052		
Time:	23:02:37		BIC	1365.226		
Sample:	01-31-2012		HQIC	1357.388		
	- 07-31-2020					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

ar.L1	1.3492	0.127	10.659	0.000	1.101	1.597
ar.L2	-0.3969	0.131	-3.035	0.002	-0.653	-0.141
ar.S.L12	0.9967	0.111	8.979	0.000	0.779	1.214
ma.S.L12	-0.9550	0.763	-1.251	0.211	-2.451	0.541
sigma2	2.307e+04	1.41e+04	1.631	0.103	-4656.744	5.08e+04
=====						
Ljung-Box (Q):	57.89		Jarque-Bera (JB):	14.62		
Prob(Q):	0.03		Prob(JB):	0.00		
Heteroskedasticity (H):	0.94		Skew:	-0.33		
Prob(H) (two-sided):	0.85		Kurtosis:	4.72		
=====						

Figura 27. Resumen modelo SARIMA(2,0,0) x (1,0,1)₁₂ para los precios mensuales de la papa pastusa, de 2012 a 2020.

La Figura 28 muestra que el valor del RMSE es de 180,14 y que el comportamiento de los valores predichos con respecto a los observados es similar, con algunos desfases en los descensos de la serie. Si bien el error es alto, los resultados pueden dar una idea de cómo se comportaría el precio de la papa a futuro.

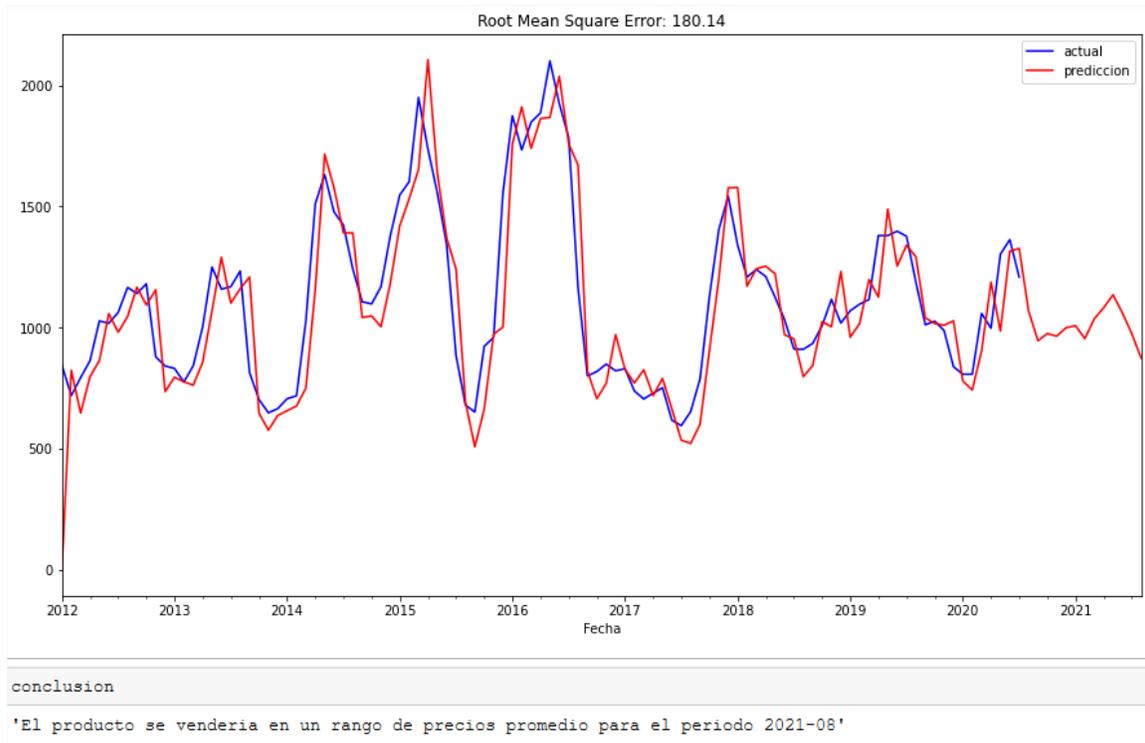


Figura 28. Valor del error RMSE y comparación de valores actuales y predichos de un modelo SARIMA(2,0,0) x (1,0,1)₁₂ para los precios mensuales de la papa pastusa, de 2012 a 2020.

Finalmente, partiendo de la predicción de 12 meses hacia adelante (agosto 2020 a agosto 2021), se evaluó el comportamiento de los precios para el período 2021-08, arrojando como resultado que el producto se vendería en un rango de precios promedio.

6.2. Caso de estudio: Huevo blanco AA

El segundo caso de estudio, enfocado en el huevo blanco AA, reveló que la serie tiene menos anomalías que el caso anterior, presentándose en los meses de abril, octubre y noviembre. También se observa que los precios tienden al alza en los meses de mayo y noviembre. Estas observaciones se pueden apreciar en el diagrama presentado en la Figura 29.

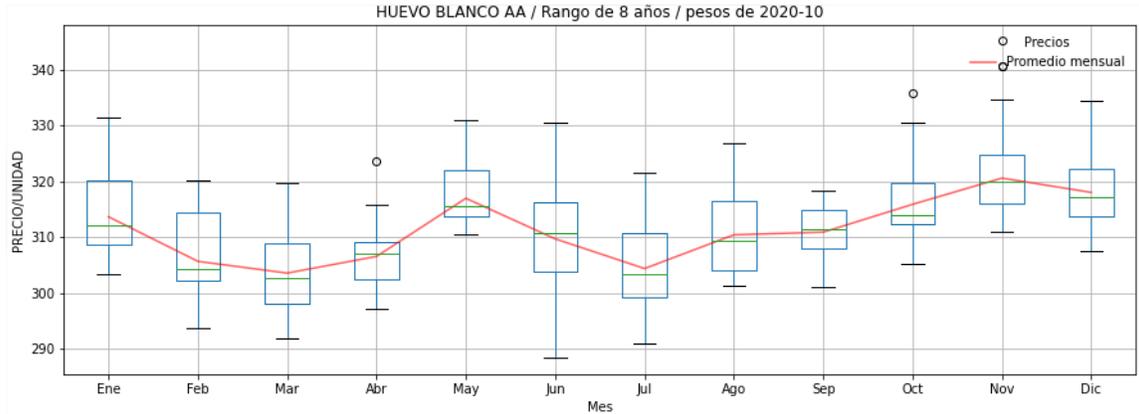


Figura 29. Diagrama Bloxplot del comportamiento de precios mensuales del huevo blanco AA, de 2012 a 2020.

Con respecto a la prueba ADF, esta mostró que serie no es estacionaria debido al valor del estadístico mayor que el valor crítico del 10% y además el valor p es mayor al 5%:

```

ADF Statistic: -2.263937
p-value: 0.183863
Critical values:
  1%: -3.498
  5%: -2.891
 10%: -2.582

```

Por su parte, la Figura 30 muestra que el gráfico de la serie mensual muestra que aparentemente no hay comportamientos periódicos y que existe un descenso evidente de los precios entre 2016 y 2019. Así mismo, las funciones ACF y PACF sugieren un componente autorregresivo AR(6) y un componente de medias móviles MA(4).

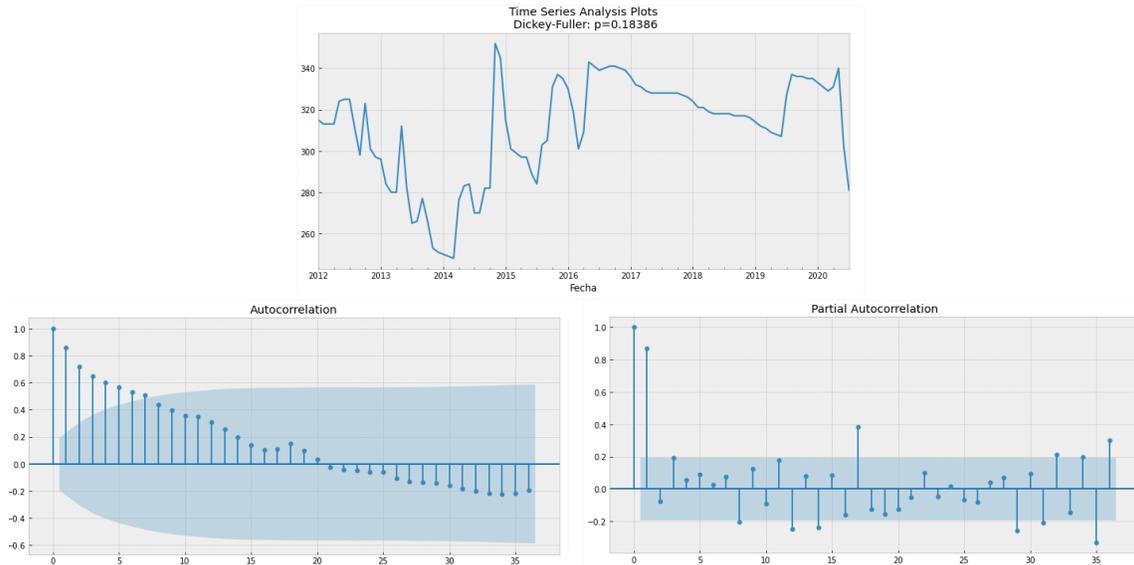


Figura 30. Gráfico de serie de tiempo, función ACF y PACF para los precios mensuales del huevo blanco AA, de 2012 a 2020.

El paso siguiente consistió en diferenciar la serie para transformarla en estacionaria, validando esto con la aplicación de la prueba ADF, que arrojó los resultados esperados:

```

ADF Statistic: -7.987500
p-value: 0.000000
Critical values:
    1%: -3.498
    5%: -2.891
    10%: -2.582

```

Al graficar la serie diferenciada se pudo observar el comportamiento de la variación precios de 2016 a 2019, en donde los datos son muy cercanos a cero, indicando una baja volatilidad (ocasionada por el descenso gradual del precio). Los valores de ACF y PACF permitieron establecer *a priori* un componente autorregresivo AR(2) y uno de medias móviles MA(1). Teniendo en cuenta lo anterior, se planteó un modelo SARIMA(4,1,6) x (2,1,1)₁₂. Las gráficas analizadas se aprecian en la Figura 31.

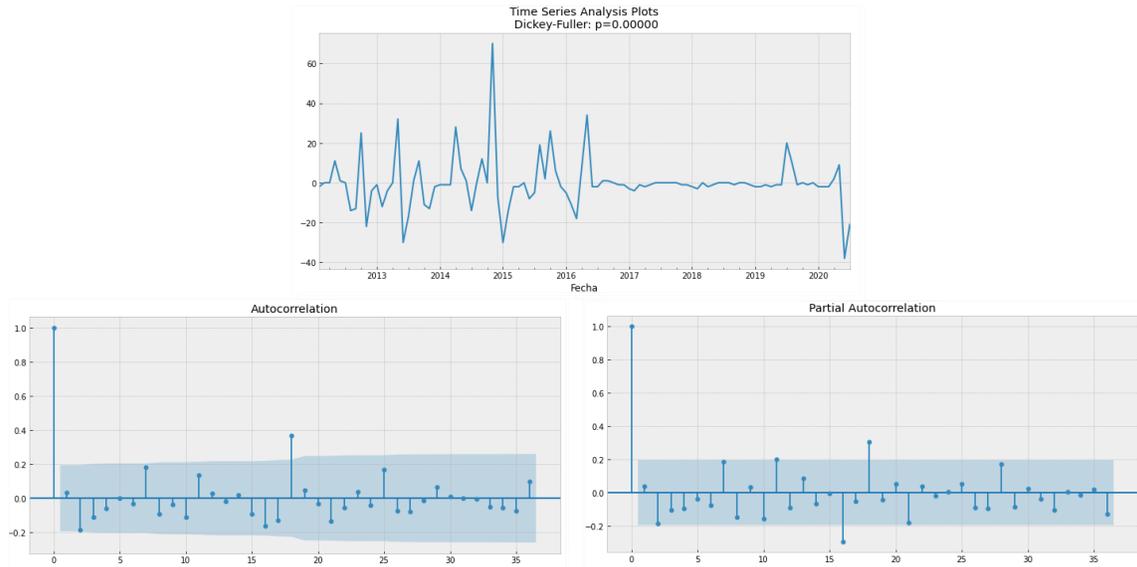


Figura 31. Gráfico comportamiento, función ACF y PACF para la serie de tiempo diferenciada de precios mensuales del huevo blanco AA, de 2012 a 2020.

Posteriormente se probaron varios modelos con los rangos $p=[0,5]$, $q=[0,7]$, $P=[0,3]$, $Q=[0,2]$, dando como resultado un modelo SARIMA(2,1,4) x (0,1,1)₁₂. El resumen del modelo, con un valor AIC de 1735,759, se encuentra en la Figura 32.

SARIMAX Results						
Dep. Variable:	Precio actual		No. Observations:	103		
Model:	SARIMAX(2, 1, 4)x(0, 1, [1], 12)		Log Likelihood	-359.880		
Date:	Mon, 30 Nov 2020		AIC	735.759		
Time:	22:41:10		BIC	755.758		
Sample:	01-31-2012		HQIC	743.824		
	- 07-31-2020					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-1.8453	0.047	-39.379	0.000	-1.937	-1.753
ar.L2	-0.9654	0.044	-21.773	0.000	-1.052	-0.879
ma.L1	1.9831	0.179	11.065	0.000	1.632	2.334
ma.L2	0.8861	0.377	2.353	0.019	0.148	1.624
ma.L3	-0.4646	0.325	-1.430	0.153	-1.101	0.172
ma.L4	-0.3318	0.145	-2.291	0.022	-0.616	-0.048
ma.S.L12	-0.9806	2.866	-0.342	0.732	-6.597	4.636
sigma2	129.0590	364.535	0.354	0.723	-585.417	843.535
Ljung-Box (Q):	23.19	Jarque-Bera (JB):	439.96			
Prob(Q):	0.98	Prob(JB):	0.00			
Heteroskedasticity (H):	0.33	Skew:	1.62			
Prob(H) (two-sided):	0.00	Kurtosis:	13.34			

Figura 32. Resumen modelo SARIMA(2,1,4) x (0,1,1)₁₂ para los precios mensuales del huevo blanco AA, de 2012 a 2020.

Al realizar la gráfica del comportamiento de los valores predichos y su contraste con respecto a los valores observados, disponible en la Figura 33, se encontró que la serie de predicciones tiene un comportamiento similar cuando los precios están al alza. El RMSE del modelo es de 13,87, dando a entender que podría ser confiable la predicción, pero se generan dudas con respecto a la serie en los descensos debido a la variación de esta. Se considera que para el modelo el error es bajo debido a que la fluctuación tiene como eje el promedio de valores reales y sería necesaria la vigilancia de los precios reales contra la predicción para determinar si es confiable. En cuanto a la predicción de 12 meses, el modelo sugiere que el producto se vendería en un rango de precios alto para el período 2021-08.

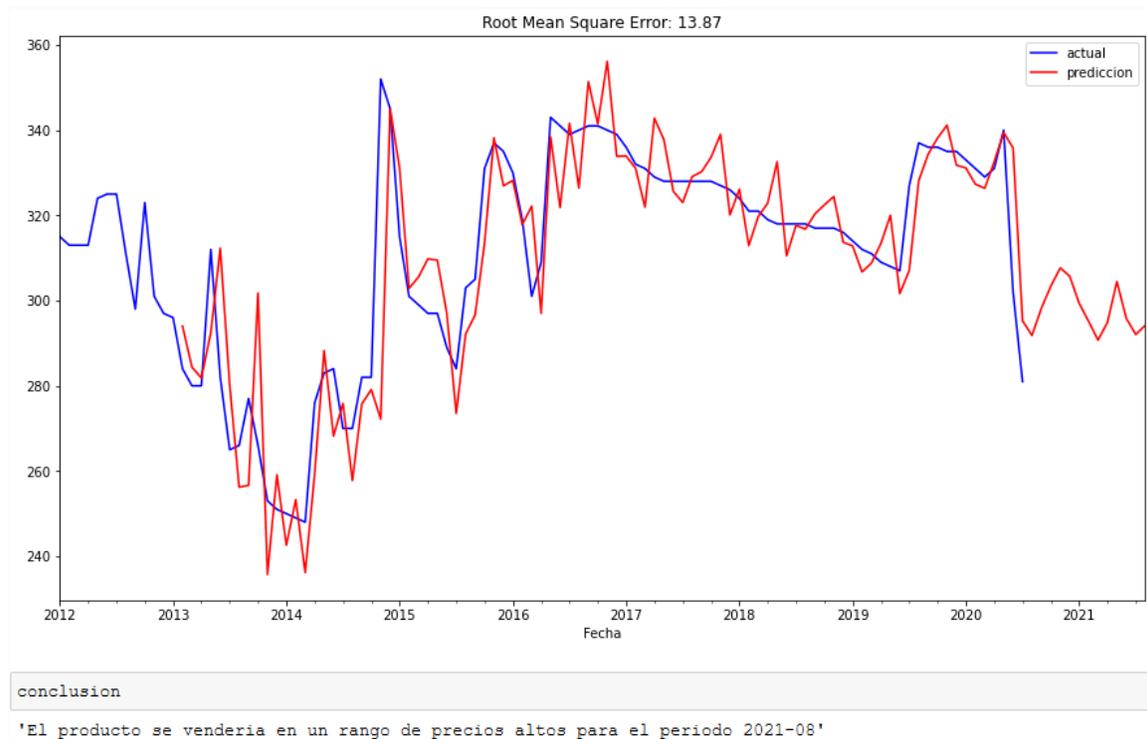


Figura 33. Valor del error RMSE y comparación de valores actuales y predichos de un modelo SARIMA(2,1,4) x (0,1,1)₁₂ para los precios mensuales del huevo blanco AA, de 2012 a 2020.

Como anotación final de ese capítulo, los datos de precios extraídos de la central de abastos de Bogotá, los valores de IPC totales por mes y el código de Python base para los análisis realizados se encuentran disponibles en el siguiente enlace:

<https://drive.google.com/drive/folders/1KAipVmDf3Zz58dBSLwOnBLuT7dgxvTpi?usp=sharing>.

7. Resultados y Discusión

Después de haber realizado el análisis de los productos agropecuarios seleccionados se encontró que la estimación de las órdenes o parámetros de cada modelo es un proceso de observación y de prueba iterativa. Conviene no solamente que los precios tengan un valor aproximado al presente sino que se realice una adecuación o estandarización de estos para tratar de reducir las anomalías existentes.

Con respecto a los datos extremos se evidencia que la existencia de altos niveles de incertidumbre dificulta el proceso de modelado de las series. La volatilidad que pueden sufrir los datos debido a crisis económicas, las decisiones gubernamentales, los efectos ambientales adversos u otras externalidades tiene incidencia en el manejo de las ventanas de tiempo para la predicción y al margen de tolerancia del error generado por el modelo.

Otro aspecto para tener en cuenta es que los modelos de predicción deben ser ajustados y evaluados cada vez que las series se alimenten con nuevos datos, que pueden afectar los patrones de tendencia, periodicidad o dispersión. Es posible un modelo ajustado para un producto que históricamente ha presentado cierta clase de comportamiento deje de ser adecuado debido que la serie no es estacionaria y a que los datos reales cambien la tendencia que se venía presentando.

En cuanto al papel de la estimación de precios en la priorización de los PIDAR cofinanciados por la ADR se puede inferir que, con un adecuado entendimiento de los datos y de las recomendaciones enunciadas en los párrafos anteriores, se pueden sentar

las bases para considerar los mejores momentos para ejecutar proyectos dependiendo del producto involucrado, para tener márgenes de venta más altos que puedan beneficiar a los productores. El impacto de la COVID-19, tal como se pudo ver en los documentos del Banco Mundial y de la Organización de las Naciones Unidas para la Alimentación y la Agricultura, se ve reflejado en la posibilidad de acceso a los productos de la canasta básica familiar por parte de las personas que se han visto afectadas a nivel económico por la pérdida de empleo, más que por la tendencia al alza en los precios de los alimentos.

Finalmente, uno de los pasos que podría realizarse a futuro es el almacenamiento de los parámetros y condiciones de ejecución de los modelos para cada uno de los productos para tener en cuenta en la priorización de proyectos, de tal manera que puedan integrarse dinámicamente al proceso de toma de decisiones.

Lista de referencias

- Adanacioglu, H., & Yercan, M. (2012). An analysis of tomato prices at wholesale level in Turkey: An application of SARIMA model. Researchgate.
https://www.researchgate.net/publication/268807959_An_analysis_of_tomato_prices_at_wholesale_level_in_Turkey_An_application_of_SARIMA_model.
- Adhikari R., & Agrawal, R. K. (2013). *An Introductory Study on Time Series Modeling and Forecasting*. Saarbrücken: LAP Lambert Academic Publishing.
- Agencia de Desarrollo Rural. (2019). *Reglamento para estructuración, aprobación y ejecución de los proyectos integrales de desarrollo agropecuario y rural con enfoque territorial*. ADR.
https://perfildeproyecto.adr.gov.co/public/files/documentos/REGLAMENTO_PIDAR.pdf.
- Agencia de Desarrollo Rural. (2020). *Informe Matriz PIDAR* [Informe de PowerBI]. Agencia de Desarrollo Rural, Bogotá, Colombia.
- Baffes, J. & Nagle, P. (2020). *Las perspectivas de los mercados de productos básicos y los efectos del coronavirus en seis gráficos*.
<https://blogs.worldbank.org/es/voices/las-perspectivas-de-los-mercados-de-productos-basicos-y-los-efectos-del-coronavirus-en-seis>
- Banco de la República. (2020). *Índice de precios al consumidor (IPC)*. IPC Total Base 2018. <https://www.banrep.gov.co/es/estadisticas/indice-precios-consumidor-ipc>
- Barrientos, J. C., Rondón D., C., & Melo, S. E. (2014). Comportamiento de precios de las variedades de papa Parida Pastusa y Diacol Capiro en Colombia (1995-2011). *Revista Colombiana de Ciencia Hortícolas*, 8(2), 272-286.
https://revistas.uptc.edu.co/index.php/ciencias_hortícolas/article/view/3469/pdf_11.
- Botero, S., & Cano, J. A. (2008). ANÁLISIS DE SERIES DE TIEMPO PARA LA PREDICCIÓN DE LOS PRECIOS DE LA ENERGÍA EN LA BOLSA DE COLOMBIA. *SciELO*.
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-47722008000100007.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (3 ed.). Cham: Springer.
- Brownlee, J. (2017). *Introduction to Time Series Forecasting with Python*. Vermont: Machine Learning Mastery.
- Campos, F. (2018). *Análisis de series temporales en R. ARIMA*.
<https://www.diegocalvo.es/analisis-de-series-temporales-en-r-arima/>.
- Corporación de Abastos de Bogotá. (2020). *Boletín precios*. CORABASTOS.
<https://precios.corabastosonline.co/#/boletin/grupos>.
- Departamento Administrativo Nacional de Estadística. (2009). *Índice de Precios al Consumidor (IPC)*. DANE. <https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc>.
- Etuk, E. H., & Igbudu, R. C. (2015). Appropriate marketing information system tools for citrus plantation in Lattakia, Syria: a revisitacion. *Journal of Multidisciplinary*

- Engineering Science and Technology (JMEST). <https://www.jmest.org/wp-content/uploads/JMESTN42350363.pdf>.
- Molina, I. (2009a). *Procesos de Media Móvil y ARMA* [PDF]. Series Temporales. <http://halweb.uc3m.es/esp/Personal/personas/imolina/MiDocencia/SeriesTemporales/Tema4SeriesEstud.pdf>.
- Molina, I. (2009b). *Procesos Integrados* [PDF]. Series Temporales. <http://halweb.uc3m.es/esp/Personal/personas/imolina/MiDocencia/SeriesTemporales/Tema5SeriesEstud.pdf>.
- Organización de las Naciones para la Alimentación y la Agricultura. (2020). *COVID-19: Inflación en los precios reales de los alimentos*. FAO. <http://www.fao.org/3/CA9235ES/CA9235ES.pdf>
- Organización Mundial de la Salud. (2019). *Preguntas y respuestas sobre la enfermedad por coronavirus (COVID-19)*. OMS. [OMShttps://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses](https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses)
- Otu Archibong, O., Osuji, G. A., Opara, J., Ifeyinwa, M. H., & Iheagwara, A. I. (2013). Application of Sarima Models in Modelling and Forecasting Nigeria's Inflation Rates. Researchgate. https://www.researchgate.net/profile/George_Osuji/publication/314467439_Application_of_Sarima_Models_in_Modelling_and_Forecasting_Nigeria's_Inflation_Rates/links/5da45d7045851553ff8f6533/Application-of-Sarima-Models-in-Modelling-and-Forecasting-Nigerias-Inflation-Rates.pdf.
- Ruiz Hernández, J. A., Barrios Puente, G., & Gómez Gómez, A. A. (2019). Análisis del precio de la manzana mediante un modelo SARIMA. SciELO. <http://www.scielo.org.mx/pdf/remexca/v10n2/2007-0934-remexca-10-02-225.pdf>.
- Wang, X., & Wang, C. (2016). Empirical Study on Agricultural Products Price Forecasting based on Internet-based Timely Price Information. DOI. <http://article.nadiapub.com/IJAST/vol87/4.pdf>.