



**Pronóstico de la inflación mexicana 2019-2020, a través del
Índice Nacional de Precios al Consumidor (INPC) general.
Modelo ARIMA estacional; programado en lenguaje R**

Soriano Urbán Jorge Stephano Matrícula: 210373538

Asesor: Padilla Hermida Ricardo

Área de Concentración: Economía Aplicada

Febrero 2020

TRABAJO TERMINAL DE LICENCIATURA
LICENCIATURA EN ECONOMÍA
UAM-AZCAPOTZALCO

Índice

Introducción	3
Definición del tema.....	4
Objetivo general	5
Objetivos específicos.....	5
Capítulo 1. El Índice Nacional de Precios al Consumidor (INPC)	6
1.1 Los índices de precios	6
1.2 El INPC	6
1.3 La inflación	7
1.4 La inflación en México	8
1.5 Evidencia empírica	9
Capítulo 2. Series de Tiempo	11
2.1 Componentes	11
2.1.1 Tendencia	11
2.1.2 Ciclo	13
2.1.3 Estacionalidad	14
2.1.4 Aleatoriedad	15
2.2 Pronóstico	16
2.2.1 Pronóstico financiero	16
2.2.2 Pronóstico de la demanda	16
2.2.3 Pronóstico de la oferta	17
Capítulo 3. Modelo ARIMA y SARIMA	18
3.1 Modelo ARIMA	18
3.2 Estacionariedad	18
3.2.1 Raíz unitaria	19
3.2.2 Transformaciones de las series de tiempo	19
3.2.2.1 Transformaciones de estabilización de la varianza	19
3.2.2.2 Transformaciones de estabilización del nivel	20
3.3 Proceso Autorregresivo (AR)	20
3.4 Proceso de Promedios Móviles (MA)	21
3.5 Proceso Autorregresivo Integrado de Promedios Móviles (ARIMA)	22
3.6 Órdenes ARIMA	22
3.7 Modelo SARIMA	23

3.7.1 Proceso Estacional Autorregresivo (SAR)	23
3.7.2 Proceso de Promedios Móviles Estacionales (SMA)	24
3.7.3 Órdenes estacionales	25
Capítulo 4. Metodología Box-Jenkins.....	26
4.1 Identificación	26
4.1.1 Función de Autocorrelación (FAC) y Función de Autocorrelación Parcial (FACP)	26
4.2 Estimación	27
4.3 Examen diagnóstico	27
4.3.1 Criterios de selección del modelo	27
4.3.1.1 Criterio Akaike	27
4.3.1.2 Criterio Schwarz	28
4.3.1.3 Medidas de precisión	28
4.3.2 Análisis de los residuos	30
4.3.2.1 Distribución normal	30
4.3.2.2 Sesgo	31
4.3.2.3 Curtosis	32
4.3.2.4 Estadístico Cuantil-Cuantil (Q-Q <i>plot</i>)	32
4.4 Pronóstico	32
Capítulo 5. Aplicación de la metodología Box-Jenkins al INPC con el lenguaje «R»	33
5.1 Identificación de los órdenes no estacionales y estacionales	36
5.2 Estimación de los modelos de ajuste	39
5.3 Examen de diagnóstico de los residuos de los modelos de ajuste ..	45
5.4 Pronóstico del INPC general, 2019-2020	53
5.5 Pronóstico de la inflación anualizada, 2019-2020	54
Conclusiones.....	55
Anexos	57
Bibliografía	77

Introducción

Este proyecto terminal tiene por objetivo pronosticar el comportamiento de la inflación general de México, 2019-2020, utilizando la metodología Box-Jenkins en el lenguaje de programación estadística R. Está formado por cinco capítulos: el primero señala algunas características del INPC, y de la inflación; cómo medirla y cuál es el nivel objetivo de inflación fijado por el Banco de México. El segundo es una introducción a los componentes que tienen las series de tiempo económicas y que deben de eliminarse antes de hacer el pronóstico, de no hacerlo, las conclusiones del análisis predictivo serán erróneas; este segundo capítulo también señala algunos casos donde el pronóstico resulta útil. El tercer y cuarto capítulos definen el marco teórico sobre el que se apoya la metodología que se utilizó para realizar el modelo predictivo.

Finalmente, en el capítulo cinco se utilizó la metodología Box-Jenkins y se eligió, entre varios modelos de ajuste, el modelo $SARIMA(3,1,1) (1,1,2)_{24}$ como aquel que mejor se ajusta a los valores reales del INPC general, con un nivel de confianza de 84.83 %; posteriormente, con los valores pronosticados del INPC se realizó el pronóstico de la inflación general. Los resultados del pronóstico prevén que la inflación en México, 2019, se acercará gradualmente a 3 % (objetivo del Banco de México), pero repuntará casi al iniciar el año 2020, ese patrón de comportamiento al alza hará que la inflación tenga una trayectoria alrededor de 4 % (umbral superior de la inflación objetivo del Banco de México), a partir de la primera quincena de febrero del año 2020, hasta terminar el periodo pronosticado (primera quincena de julio del mismo año).

Definición del tema

En el trabajo terminal se analizará el Índice Nacional de Precios al Consumidor (INPC) general,¹ a partir de la primera quincena de enero de 1988 a la primera quincena de julio de 2019, para predecir su comportamiento comenzando la segunda quincena de julio de 2019 y terminando en la primera quincena de julio 2020 (un año), con el objetivo de pronosticar la inflación quincenal anualizada, esto calculando las variaciones del INPC (cociente entre el valor quincenal del INPC pronosticado respecto al correspondiente valor quincenal observado del INPC del año anterior). La investigación predice el comportamiento de la inflación analizando el comportamiento pasado del INPC mediante la metodología Box-Jenkins, es decir, es un análisis ARIMA univariante debido a que el valor del INPC tiene un componente estacional (el valor actual del INPC está compuesto por una proporción de su valor del año anterior), el análisis se hará mediante un modelo ARIMA estacional (SARIMA).

¹ Es un indicador económico que contempla tanto el INPC subyacente como el INPC no subyacente; el primero se divide en mercancías (bienes comerciables) y servicios (bienes no comerciables), el segundo se compone de agropecuarios, energéticos y tarifas autorizadas por el gobierno. Esta distinción es igual de importante, ya que los precios de los productos agropecuarios tienen una alta volatilidad y sus variaciones obedecen sobre todo a cuestiones climatológicas. En cambio, los energéticos y tarifas autorizadas por el gobierno son precios administrados, o bien, fijados en función de criterios políticos (Heath, 2012, pág. 287).

Objetivo general

Pronosticar el comportamiento de la inflación general para el período de la segunda quincena de julio de 2019 a la primera quincena de julio de 2020.

Objetivos específicos

- 1) Eliminar el problema de raíz unitaria
- 2) Eliminar el componente tendencial
- 3) Eliminar el componente estacional
- 4) Obtener los patrones AR y MA, y patrones AR y MA estacionales
- 5) Generar diversos modelos de ajuste
- 6) Realizar pruebas estadísticas para detectar proceso de ruido blanco y falta de autocorrelación en los residuos
- 7) Elegir aquellos modelos de ajuste que tengan el menor valor de los criterios Akaike, Schwarz y de las medidas de exactitud
- 8) Elegir el mejor modelo de ajuste
- 9) Realizar simulación de pronóstico del modelo de ajuste seleccionado
- 10) Pronosticar los valores del INPC general para el período de la segunda quincena de julio de 2019 a la primera quincena de julio de 2020
- 11) Pronosticar la inflación general en México para el mismo período

Capítulo 1. El Índice Nacional de Precios al Consumidor (INPC)

1.1 Los índices de precios

El Banco de México define a los índices de precios como estadísticos que miden los cambios porcentuales o proporcionales de un conjunto de precios a lo largo del tiempo. Señala como uno de sus usos principales la medición de la inflación de una economía y que para calcular un índice de precios se requiere definir un grupo de productos o canasta para dar seguimiento a sus precios, determinar la importancia relativa de cada componente en el total de la canasta, así como la manera más apropiada de promediar las variaciones de sus precios (Banco de México, s.f.).

1.2 El INPC

El INPC es un indicador cuyo valor refleja la variación de los precios en el tiempo de un conjunto de bienes y servicios que consume la población de México. Este indicador se utiliza para conocer el ritmo al que aumentan los precios y así el Banco de México pueda aplicar medidas para frenar o controlar la inflación, es medido con una frecuencia quincenal, sin embargo, pueden ser medidos mensual y anualmente por el Instituto Nacional de Estadística y Geografía (Inegi), el cual levanta periódicamente una Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), para conocer qué es lo que más consume la gente en México. En esta encuesta se identifican los productos de mayor consumo, después se recopilan los precios de dichos productos en tiendas de todo el país. Además, el Inegi da seguimiento cada mes, a cerca de 235 000 precios de productos en 46 ciudades del país para medir la inflación, la información de esos miles de precios se procesa para identificar aquellos rubros en los cuales hay mayor importancia en el consumo de las familias tomando en cuenta qué tanto se gastó en ellos (Banco de México, s.f.).

Para conocer cuál es la inflación anualizada, por ejemplo, de la primera quincena de enero del 2017 y, teniendo en cuenta que el valor del INPC de la primera

quincena de enero de 2016 fue 89.22861951, y el de la primera quincena de enero de 2017 fue 93.49194663, se calcula el siguiente cociente.²

$$\text{Inflación} = \left(\frac{INPC_{1Q\ m\ t}}{INPC_{1Q\ m\ t-1}} - 1 \right) \times 100 ;$$

$$\text{Inflación anualizada}_{1Q\ Enero\ 2017} = \left(\frac{93.49194663_{1Q\ Enero\ 2017}}{89.22861951_{1Q\ Enero\ 2016}} - 1 \right) \times 100 = \mathbf{4.78\ \%}$$

1.3 La inflación

La inflación es un fenómeno económico presente en todos los países que consiste en un aumento de los precios de la mayoría de los bienes y servicios que se comercian. Este fenómeno se debe a varios factores, uno de ellos es un desajuste entre el nivel de demanda y el nivel de producción (Lexico, 2019) debido, sobre todo, a la excesiva circulación de dinero en la economía de un país (el Banco de México se encarga de regular la cantidad de dinero que circula en el país), la población tiene mayor capacidad de compra por lo que aumenta su gasto, ese aumento genera un incremento en la demanda de bienes y servicios, la capacidad productiva del país se mantiene pero aumenta la demanda de productos, provocando una escasez de bienes y servicios. Para estabilizar el nivel de demanda y el nivel productivo aumentan los precios, así se logra racionar la demanda, hay quienes disminuyen la cantidad de consumo de ciertos bienes y servicios, algunos, debido a su bajo nivel de ingreso, dejan de comprarlos, otros, comprarán bienes sustitutos.

A los agentes económicos les afecta el aumento de precios porque debilita su capacidad de compra; además, les resulta más complicado administrar y distribuir sus ingresos porque los precios son tomados como referencia para comprar, ahorrar, invertir, pagar deudas, etcétera.

El impacto de la inflación es mayor para aquel sector de la población que posee menos ingreso.

² El subíndice *1Q* hace referencia a la primera quincena, el subíndice *m* al mes, el subíndice *t* al año actual y el subíndice *t-1* al año anterior (Instituto Nacional de Estadística y Geografía).

1.4 La inflación en México

México ha mantenido una inflación sana, por lo menos en la última década, según el Banco de México; la inflación ha sido baja y estable debido a las bases institucionales del banco central mexicano que han hecho un objetivo prioritario la estabilidad inflacionaria, la mayor transparencia y la autonomía, añade que llegar a la estabilidad de los precios significa tener una condición favorable para el bienestar social y el crecimiento económico. El Banco de México ha aplicado a partir de 2001 una estrategia que se sustenta en mantener objetivos de inflación, ya en 2003 se fijó como objetivo una inflación anual de 3 %.

La política monetaria³ en México tiene como objetivo controlar la inflación. El Banco de México utiliza un esquema de objetivos de inflación que funciona de la siguiente manera:

1. Define el objetivo,⁴ rango y plazo para obtener la inflación, esto debe ser del dominio público.
2. Observa la inflación subyacente.⁵
3. El banco central analiza todas las causas que provocan el fenómeno inflacionario para tener contemplado cómo se comportará en el futuro.⁶
4. La información económica relevante y las decisiones de política monetaria se comunican de manera transparente.
5. Publica el calendario con las fechas en las que dará a conocer sus decisiones de política monetaria, y diez días hábiles después de cada anuncio, da a conocer la minuta correspondiente a la sesión de la Junta de Gobierno durante la cual se deliberó y tomó la decisión correspondiente (Banco de México, s.f.).

³ Se conoce como política monetaria lo que hace el banco central para administrar la cantidad de dinero y crédito en la economía.

⁴ La meta de inflación objetivo es de 3 % con un rango de variación de \pm un punto porcentual.

⁵ La inflación subyacente es aquella medida de la inflación que elimina todos los precios de los bienes y servicios que influyen en el comportamiento de esta y con ello identificar la tendencia.

⁶ El informe de estos análisis se publica trimestralmente y son la base para la toma de decisiones o acciones necesarias por los agentes económicos.

1.5 Evidencia empírica

Una breve inspección del estudio del análisis predictivo de la inflación en México muestra que diversos investigadores utilizan modelos autorregresivos y de promedios móviles integrados, principalmente estacionales. Algunos investigadores utilizan modelos ARIMA estacionales univariantes, como Rodríguez (2006), quien implementó un modelo estacional autorregresivo integrado de promedio móvil (SARIMA) por sus siglas en inglés Seasonal Autoregressive Integrated Moving Average, para pronosticar la inflación acumulada de los dos últimos meses del año 2005, hay investigadores como (Capistrán, Constandse, & Ramos-Francia), quienes se centran en el análisis del componente tendencial de la inflación para pronosticarla a corto y mediano plazos.

En los estudios que tienen como propósito pronosticar, es común observar que los analistas comparan diversas metodologías predictivas con el objetivo de seleccionar aquella que mejor se ajuste a las observaciones reales de la serie de tiempo analizada. Durán *et al.* (2012) compararon los métodos SARIMA, modelos de datos de panel, modelos de corrección del equilibrio y modelos de factores dinámicos para pronosticar el INPC agregado, concluyendo que los modelos autorregresivos son eficientes para predecir la inflación agregada un periodo hacia adelante; no obstante, al aumentar el horizonte de proyección esta capacidad se ve disminuida frente a otras estrategias y, que en el mediano plazo, modelos econométricos más complejos que consideran relaciones de cointegración, modelos con efectos fijos o con factores dinámicos aplicados a series de tiempo desagregadas se muestran como los procedimientos más eficientes.

Algunas metodologías predictivas se fundamentan en la inteligencia artificial (IA), como las redes neuronales artificiales (RNA), que utilizan diferentes algoritmos de aprendizaje y sus neuronas artificiales identifican las relaciones lineales y no lineales de las series de tiempo García (2015), pudiendo ajustar diferentes tendencias tanto lineales como no lineales a las series de tiempo económicas, las RNA tienen la capacidad de «aprender» de manera iterativa el comportamiento de los datos históricos o en tiempo real y proyectar ese comportamiento al futuro; sin embargo, a veces, utilizar metodologías tan sofisticadas como esta no siempre genera mejores resultados; Flores (2007) implementó un análisis comparativo entre un modelo SARIMA, Holt-Winters y RNA para pronosticar el INPC, y

concluyó que la metodología de RNA presenta menor capacidad predictiva para pronosticarlo.

Al inspeccionar los estudios analíticos predictivos de los índices de precios en Colombia, se observa que para pronosticar el Índice de IPC al Consumidor (IPC) también se implementan modelos autorregresivos de promedios móviles integrados estacionales y no estacionales; Alonso y Rivera (2017) implementaron un modelo SARIMA para pronosticar el IPC total de Colombia, y concluyeron que a partir de los índices de precios desagregados del nivel de Grupo 12 por ciudad y por nivel de ingresos, se puede construir un mejor pronóstico para la inflación mensual durante el 2014 y 2015, en lugar de emplear un modelo univariado para el IPC nacional total.

De igual forma, Ramírez y Huérfano (2014) implementaron un modelo SARIMA para pronosticar el IPC de Colombia y concluyeron que la coyuntura mundial puede tener efectos inesperados sobre esta variable. Investigadores como Espinal *et al.* (2016) (Espinal Rodríguez, Giraldo Botero, Peña Córdoba, & Lotero, 2016) implementaron un modelo ARIMA y de descomposición mediante el filtro Holt-Winters para pronosticar el IPC de Medellín, concluyendo que los hallazgos obtenidos a partir del modelo ARIMA fueron satisfactorios; sin embargo, es posible que se logre un modelo más preciso si se lleva a cabo un análisis del IPC desagregado. Por otra parte, Rodríguez y Siado (2003) implementaron un modelo SARIMA, modelo no paramétrico Kernel y un modelo STAR para pronosticar la inflación colombiana, y concluyeron que para cada horizonte de predicción el modelo Kernel permite mejorar los pronósticos respecto a los modelos paramétricos ARIMA y no lineal STAR.

Capítulo 2. Series de Tiempo

Una serie de tiempo es un registro de datos (de cualquier índole) que están relacionados entre sí y a un punto en el tiempo y el espacio, y se ordenan de manera cronológica, se almacenan en una base de datos relacional⁷ o no relacional,⁸ una hoja de cálculo e incluso una hoja de papel. Existen datos históricos y datos en tiempo real (datos en flujo constante o en *streaming*);⁹ por ejemplo, los datos intradías de los activos financieros de las bolsas de valores, los datos que generan los sensores embebidos en sistemas electrónicos, por nombrar algunos. Estos registros, multivariantes, se almacenan con el propósito de procesarlos y analizarlos a fin de realizar una predicción y anticiparse a los hechos a fin de tomar mejores decisiones.

2.1 Componentes

Los registros de las series de tiempo contienen patrones o componentes tendenciales y estacionales, factores exógenos y cualquier otro dato pertinente disponible (Woody, y otros, 2016, pág. 66), estos afectan el comportamiento real de la variable y, por ende, repercute en el análisis de la serie de tiempo, llevando a tomar malas decisiones e interpretaciones erróneas de los resultados obtenidos en el proceso del análisis predictivo.

2.1.1 Tendencia

El análisis de las series de tiempo puede revelar si hay tendencia a largo plazo, ya sea decreciente o creciente. Una gráfica de la población de los Estados Unidos respecto al tiempo podría mostrar una tendencia creciente a largo plazo, mientras una gráfica del porcentaje de adultos en esa misma población que fuma muestra una tendencia decreciente sobre el tiempo (Woody, y otros, 2016, pág. 68).

⁷ Es un conjunto de información relacionada sobre un tema particular, organizada de una forma tal que suministra una base o fundamento para procedimientos, como la recuperación o consulta de la información, la elaboración de conclusiones con base en los datos y la toma de decisiones de una organización (Torres Remon, 2012, pág. 47).

⁸ Una base de datos no relacional es aquella que no usa el esquema tabular de filas y columnas que se encuentra en la mayoría de los sistemas de base de datos más tradicionales (Microsoft, 2018).

⁹ Los datos de *streaming* son datos que se generan constantemente a partir de miles de fuentes de datos, que normalmente envían los registros de datos simultáneamente en conjuntos de tamaño pequeño (varios kilobytes) (Amazon, 2019).

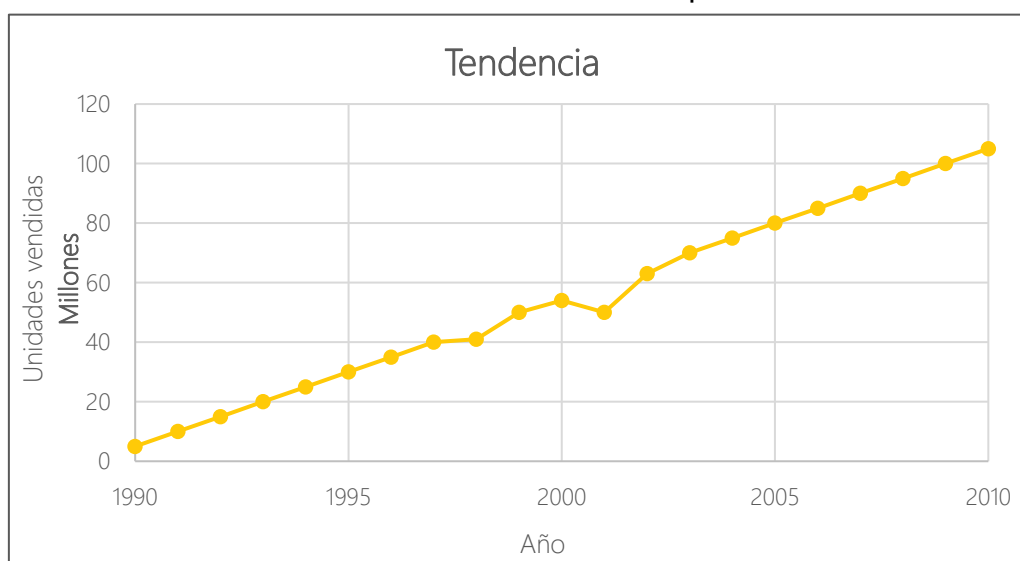
Cuadro 1. Tendencia lineal y tendencias no lineales de una serie de tiempo

Tendencia lineal y tendencias no lineales de una serie de tiempo	
Tendencia lineal	$T = \beta_0 + \beta_1 \text{Tiempo}$
Tendencia cuadrática	$T = \beta_0 + \beta_1 \text{Tiempo} + \beta_2 \text{Tiempo}^2$
Tendencia cúbica	$T = \beta_0 + \beta_1 \text{Tiempo} + \beta_2 \text{Tiempo}^2 + \beta_3 \text{Tiempo}^3$
Tendencia exponencial	$T = \beta_0 e^{\beta_1 \text{Tiempo}}$
Tendencia exponencial modificada simple	$T = a + \beta r^{\text{Tiempo}}$
Tendencia logística	$T = 1/(a + \beta r^{\text{Tiempo}})$

Fuente: Elaboración propia.

El proceso autorregresivo del análisis predictivo de series de tiempo, es decir, la autorregresión, detecta las fluctuaciones lineales, no lineales y estacionales en los datos históricos y proyecta estas tendencias al futuro (International Business Machines, s.f.).

Gráfica 1. Tendencia lineal en una serie de tiempo



Fuente: Elaboración propia en Excel con base en la tabla 6-1 del libro *Data Science with Microsoft SQL Server 2016*, página 69.

Los datos de la serie de tiempo muestran una tendencia de largo plazo incrementándose año con año. Woody y otros, mencionan, por ejemplo, las ventas de *smart phones*.

2.1.2 Ciclo

En diversos sectores de la economía ostentan los ciclos económicos los cuales tienen un período de expansión y contracción, estos están conformados por diferentes etapas que afectan las decisiones que los agentes económicos toman. El subgobernador del Banco de México Jonathan Heath, en su libro *Lo que indican los indicadores*, señala: «Los ciclos económicos son altibajos que experimenta la actividad económica través del tiempo.» (Heath, 2012, pág. 31).

El ciclo económico está compuesto por dos etapas principales:

1. Disminución: La actividad económica se encuentra en recesión.
2. Crecimiento: Se divide en dos fases más: recuperación y expansión.

En la práctica se designa al punto máximo del ciclo económico como el inicio, por lo que la primera etapa es la recesión, en la siguiente etapa la economía comienza a crecer, se conoce como recuperación y comienza en el momento que concluye la recesión, esta etapa concluye cuando se regresa a punto máximo anterior, a partir de ese punto comienza la tercera etapa llamada expansión. Jonathan Heath señala que algunos analistas e instituciones contemplan a la etapa de crecimiento como una sola (llamada expansión) sin tomar en cuenta la etapa de recuperación.

Grafica 2. Etapas del ciclo económico típico



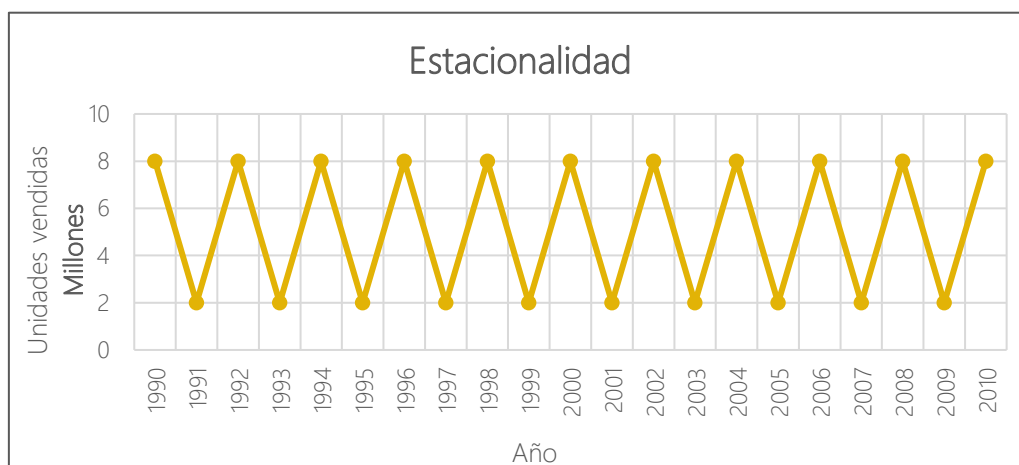
Fuente: Elaboración propia en Excel con base en la gráfica 3.1 del libro *Lo que indican los indicadores*, página 33.

2.1.3 Estacionalidad

El análisis de las series de tiempo puede revelar si existen patrones estacionales en los datos. Por ejemplo, para ciertos productos de consumo, los picos de venta del *Black Friday* podrían siempre sugerir que las ventas de la cuarta semana de noviembre de ese año se correlacionan con las ventas del año pasado y con las ventas de la cuarta semana de noviembre de los años anteriores (Woody, y otros, 2016, pág. 69).

Para identificar la tendencia real en una serie de datos se necesita tener en cuenta la variación de los datos de la serie, debido a los efectos estacionales. Para la descomposición estacional las variables no deben contener datos perdidos o incrustados (International Business Machines, s.f.).

Gráfica 3. Estacionalidad en una serie de tiempo

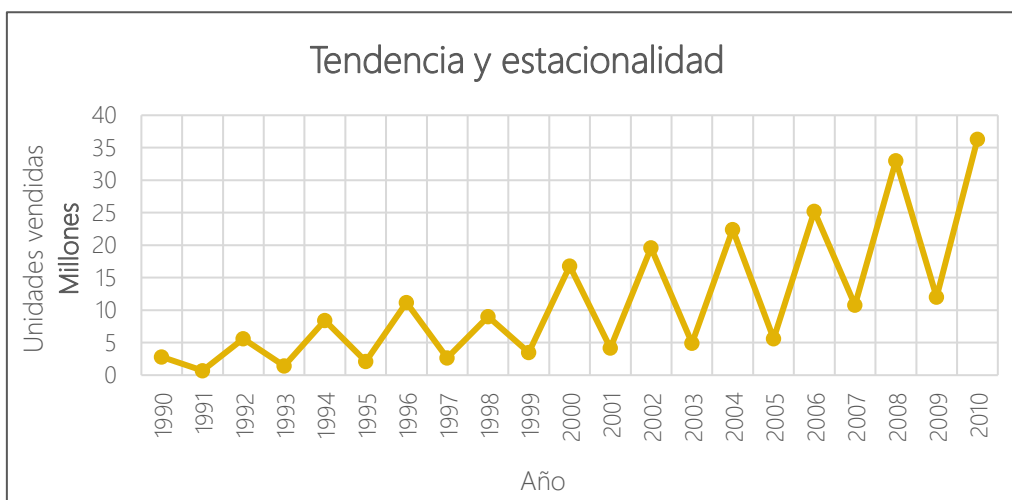


Fuente: Elaboración propia en Excel con base en la tabla 6-1 del libro *Data Science with Microsoft SQL Server 2016*, página 69.

Los datos de la serie de tiempo muestran un patrón estacional sin tendencia que se repite en el tiempo.

Por ejemplo: Las ventas de una empresa que alcanzan su máximo cada año (Woody, y otros, 2016, pág. 69).

Gráfica 4. Tendencia y estacionalidad en una serie de tiempo



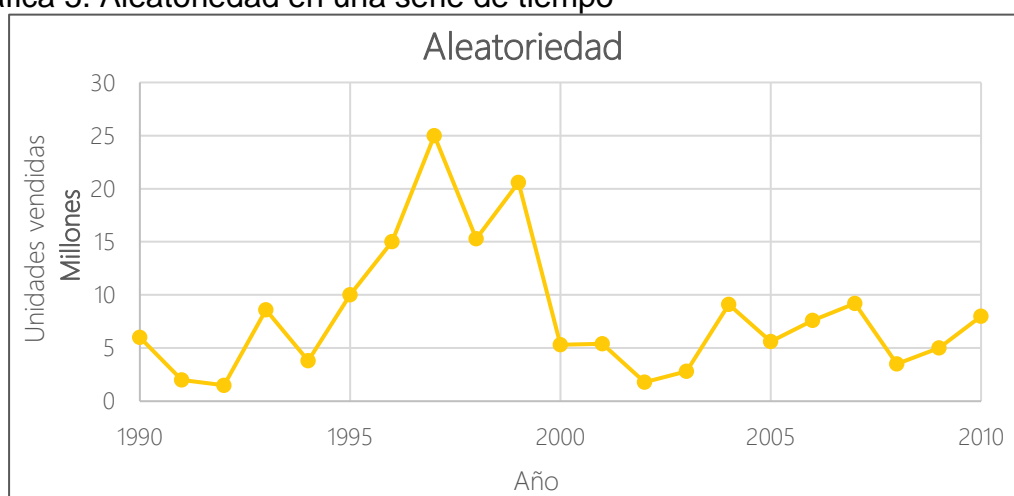
Fuente: Elaboración propia en Excel con base en la tabla 6-1 del libro *Data Science with Microsoft SQL Server 2016*, página 69.

Los datos de la serie de tiempo muestran ambos componentes; una tendencia que se incrementa a lo largo del tiempo y un patrón estacional. Por ejemplo: Las ventas de helado (Woody, y otros, 2016, pág. 69).

2.1.4 Aleatoriedad

El comportamiento irregular de una serie de tiempo está compuesto por fluctuaciones causadas por sucesos impredecibles o no periódicos, como el clima poco usual, huelgas, guerras, rumores, elecciones y cambio de leyes (Unidad Académica de Matemáticas, s.f.).

Gráfica 5. Aleatoriedad en una serie de tiempo



Fuente: Elaboración propia en Excel con base en la tabla 6-1 del libro *Data Science with Microsoft SQL Server 2016*, página 69.

Los datos de la serie de tiempo no muestran una tendencia clara ni patrones estacionales. Nota: los métodos de series de tiempo no son adecuados para este conjunto de datos (Woody, y otros, 2016, pág. 69).

2.2 Pronóstico

El pronóstico se usa ampliamente en muchas aplicaciones y decisiones comerciales críticas que dependen de tener una imagen lo más precisa posible del futuro (Woody, y otros, 2016, pág. 66).

2.2.1 Pronóstico financiero

El pronóstico financiero es un proceso en el que se prevén las métricas financieras clave, como los ingresos, costos y beneficios, esto puede mostrar el comportamiento de indicadores clave de rendimiento (KPI) de ingresos, costos y beneficios, así como de nuevos clientes o precios promedios de venta. El propósito de este tipo de pronóstico es administrar el desempeño financiero de una organización respecto a su plan financiero. Las empresas deben monitorear sus ventas y costos para garantizar que la organización entregue los datos financieros requeridos en el plan para financiar operaciones en curso, como la fabricación, las inversiones en nuevos productos y la contratación. El pronóstico es una función esencial en el proceso de gestión del desempeño financiero (Woody, y otros, 2016, pág. 67).

2.2.2 Pronóstico de la demanda

El pronóstico de la demanda es un paso crítico en la administración de la cadena de suministro de una organización. El propósito de pronosticar la demanda es ayudar a informar cuántas unidades de un producto (a veces llamado *SKU*) será demandado por los clientes en el futuro, así la empresa puede planear la cantidad de materia prima necesaria para comprar, así como dónde y cuándo se fabricará. El pronóstico es un elemento crítico que garantiza que se fabrique la cantidad correcta, en aras de minimizar los costos de inventario y desabastecimientos (Woody, y otros, 2016, pág. 67).

La previsión de la demanda suele ser exclusiva de un sector industrial. Por ejemplo, pronosticar la demanda en la industria energética implica estimar la cantidad de energía que los clientes probablemente usarán en el futuro. La previsión de la demanda en la industria manufacturera se centra en la cadena de suministro y el mejoramiento de la fabricación (Woody, y otros, 2016, pág. 67).

En la industria del cuidado de la salud, podría implicar pronosticar el número de pacientes que necesitan servicios de atención médica para planificar cuántas camas de hospital hay disponibles (Woody, y otros, 2016, pág. 67).

2.2.3 Pronóstico de la oferta

El pronóstico de la oferta a menudo tiene estrecha colaboración con la previsión de la demanda. Por ejemplo, una organización municipal de agua quiere pronosticar no solo la demanda de agua, sino también cuánta agua habrá disponible para abastecer la demanda del cliente. Una organización de atención médica podría pronosticar el suministro de atención médica calificada y de profesionales que estarán disponibles para satisfacer la demanda de futuros servicios de atención médica (Woody, y otros, 2016, pág. 67).

Capítulo 3. Modelo ARIMA y SARIMA

3.1 Modelo ARIMA

El modelo autorregresivo integrado de promedios móviles (*autoregressive integrated moving average*, *ARIMA*) es un método utilizado para pronosticar, centra su atención en el análisis de la propiedad probabilística o estocástica de las series de tiempo económicas, este método se aplica a partir de la no estacionariedad de una serie de tiempo.

ARIMA es un acrónimo formado por los siguientes aspectos clave del modelo (Jason, 2019):

- *AR*: Autorregresión. Utiliza la relación dependiente entre una observación y algunos números de observaciones rezagadas.
- *I*: Integración. Se usa para la diferenciación de observaciones, con el fin de volver una serie de tiempo a estacionaria.
- *MA*: Promedio Móvil. Utiliza la relación de dependencia entre una observación y los residuos.

Cada uno de esos aspectos son especificados en el modelo como un parámetro, se usa la siguiente notación estándar: $ARIMA(p, d, q)$, donde los parámetros p , d y q (órdenes del modelo *ARIMA*) son sustituidos con números enteros para indicar el modelo *ARIMA* específico que se va a utilizar.

Un problema con *ARIMA* es que no soporta datos estacionales, es decir, series de tiempo con un ciclo repetitivo. *ARIMA* asume que los datos no son estacionales o que se les haya eliminado el componente estacional, mediante algún método de ajuste, como una diferenciación estacional (Brownlee, 2018).

3.2 Estacionariedad

Una serie de tiempo es estacionaria cuando su media, varianza y autocovarianza no varían en el tiempo (Villavicencio), la mayoría de las series de tiempo económicas son no estacionarias, pero aplicando sobre ellas un logaritmo natural o calculando sus primeras diferencias¹⁰ la serie se convierte en estacionaria, una

¹⁰ Diferenciar es restar al valor presente (t) su valor pasado ($t - 1$), las primeras diferencias restan un valor pasado inmediato al valor presente.

vez que la serie de tiempo es estacionaria ya se puede aplicar la metodología *ARIMA*.

3.2.1 Raíz unitaria

Raíz unitaria es una situación de no estacionariedad, una serie de tiempo es estacionaria cuando su media, varianza y autocovarianza no cambian en el tiempo; a saber, la media es una medida de tendencia central que resulta de sumar todos los valores de un conjunto de datos y luego dividirlo entre el número total de valores del conjunto de datos, la varianza es una medida de dispersión de los datos y representa la variabilidad de una serie de datos respecto a su valor medio, cabe mencionar que una serie de tiempo puede contener más de una raíz unitaria (Gujarati, 2010, pág. 744).

Distinguir entre series de tiempo estacionarias y no estacionarias es fundamentalmente importante para saber si la tendencia (lenta evolución de largo plazo de la serie de tiempo) es determinista¹¹ o estocástica¹² (Gujarati, 2010, pág. 745).

3.2.2 Transformaciones de la serie de tiempo

ARIMA es eficaz siempre y cuando la serie se pueda convertir en estacionaria mediante la aplicación de transformaciones tales como el logaritmo natural, la diferenciación simple o diferenciación estacional (International Business Machines, s.f.).

3.2.2.1 Transformaciones de estabilización de la varianza

Las series en las que la varianza cambia a lo largo del tiempo con frecuencia se pueden estabilizar con transformaciones funcionales, tales como transformación logarítmica natural o transformación de raíz cuadrada. No se pueden usar transformaciones para estabilizar la varianza de series con valores negativos (International Business Machines, s.f.).

¹¹ Se llama tendencia determinista si la tendencia de una serie de tiempo es del todo predecible y no variable (Gujarati, 2010, pág. 745).

¹² Se llama tendencia estocástica si la tendencia de una serie no es predecible (Gujarati, 2010, pág. 745).

3.2.2.2 Transformaciones de estabilización del nivel

Un suave descenso de las funciones de autocorrelación indica que todos los valores de las series están estrechamente correlacionados con el valor anterior (International Business Machines, s.f.).

Para estabilizar el nivel se utiliza la diferenciación de la serie de tiempo:

- Diferenciación simple: Se calculan las diferencias existentes entre cada valor y el valor anterior.
- Diferenciación estacional: Se calculan las diferencias existentes entre cada valor y el valor estacional anterior

Si se utiliza la diferenciación simple o estacional de manera simultánea con la transformación logarítmica o de raíz cuadrada siempre se aplicará primero la transformación que estabiliza la varianza. Si se utiliza la diferenciación simple y estacional, los valores de la serie resultante son iguales independientemente de si se aplica primero una diferenciación u otra, según un documento publicado por International Business Machines (IBM).

3.3 Proceso Autorregresivo (AR)

Un proceso autorregresivo de orden p , $AR(p)$, es un proceso estocástico

$(Y_t) = t \in \mathbb{Z}$ que sigue el siguiente modelo:

$$Y_t = c + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z}$$

donde: $c, \alpha_1, \dots, \alpha_p \in \mathbb{R}$ son constantes.

El proceso (ε_t) es un ruido blanco con varianza σ^2 , los α_t se llaman innovaciones, porque representan la nueva información que aparece en cada instante (Universidad Carlos III de Madrid).

Teniendo la siguiente expresión:

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + u_i$$

donde δ es la media de Y y u_i es un término de error aleatorio no correlacionado con media cero y varianza constante σ^2 (ruido blanco), se dice que Y_t sigue un proceso estocástico autorregresivo de primer orden, $AR(1)$, en este caso el valor

de Y en el tiempo t depende de su valor en el periodo anterior y de un término aleatorio, aunque en este caso los valores de Y están expresados como desviaciones de su valor promedio. El modelo señala que el valor pronosticado de Y en el periodo t es alguna proporción ($= \alpha_1$) de su valor en el periodo $(t - 1)$ más un choque o perturbación aleatoria en el periodo t . Si se considera el siguiente modelo:

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

se dice que Y_t sigue un proceso autorregresivo de segundo orden, o sea, que el valor de Y en el tiempo depende de sus valores en los dos periodos anteriores.

De la siguiente expresión:

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

se tiene que, de manera general, Y_t es un proceso autorregresivo de orden p , $AR(p)$.

Definición y ecuaciones obtenidas de (Gujarati, 2010, págs. 775-776)

3.4 Proceso de Promedios Móviles (MA)

Un proceso de promedios móviles es una combinación lineal de términos de error de ruido blanco (Gujarati, 2010, pág. 776).

Este proceso se puede representar en el siguiente modelo de Y_t de la siguiente forma:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1}$$

donde μ es una constante y u es el término de error estocástico de ruido blanco. Y en el periodo t es igual a una constante más un promedio móvil de los términos de error presente y pasado. Así, en este caso decimos que Y sigue un proceso de promedios móviles de primer orden, es decir, $MA(1)$.

Si Y sigue la siguiente forma:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2}$$

Y sigue un proceso $MA(2)$. Generalizando,

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \cdots + \beta_q u_{t-q}$$

es un proceso $MA(q)$.

Definición y ecuaciones obtenidas de (Gujarati, 2010, pág. 776)

3.5 Proceso autorregresivo integrado de promedios móviles (*ARIMA*)

Gujarati señala que la mayoría de las series de tiempo económicas no son estacionarias, recordemos que *ARIMA* supone que la serie de tiempo a analizar es estacionaria, para volver una serie no estacionaria a estacionaria se debe diferenciar d veces y luego aplicar los procesos $AR(p)$ y $MA(q)$, se dice que una serie de tiempo es $ARIMA(p, d, q)$ si se ha diferenciado, y se le llama serie de tiempo *autorregresiva integrada de promedios móviles*, en donde p corresponde al número de términos autorregresivos, d es el número de veces que la serie debe diferenciarse para hacerse estacionaria y q es el número de términos de promedios móviles; por ejemplo, una serie de tiempo de la forma $ARIMA(2,1,2)$ tiene que diferenciarse una vez ($d = 1$) antes de hacerse estacionaria (en primeras diferencias) y se puede modelar como un proceso $ARMA(2,2)$, ya que tiene dos términos *AR* y dos términos *MA*.

Si sucediera que $d = 0$, es decir, la serie fuera estacionaria desde un inicio, y si fuera un proceso $ARIMA(p, 0, 0)$, significa que hay un proceso estacionario $AR(p)$ puro; y si fuera un proceso $ARIMA(0, 0, q)$ significa que es un proceso estacionario $MA(q)$ puro (Gujarati, 2010, pág. 777).

3.6 Órdenes *ARIMA*

Orden autorregresivo (p): Especifica los valores previos de la serie utilizados para predecir los valores actuales. Por ejemplo: un orden $AR(2)$ especifica que se van a utilizar los valores de la serie correspondiente a dos periodos de tiempo del pasado para predecir el valor actual.

Diferencia (d): Especifica el orden de diferenciación aplicado a la serie antes de estimar los modelos.

La diferenciación es necesaria si hay tendencias (las series con tendencias suelen ser no estacionarias y el modelo *ARIMA* asume la estacionariedad) y se utiliza para eliminar su efecto. El orden de la diferenciación representa el grado de

la tendencia, de modo que la diferenciación de primer orden representa tendencias lineales, la diferenciación de segundo orden representa tendencias cuadráticas, etcétera (International Business Machines, s.f.).

Media móvil (q): Los órdenes de la media móvil especifican el modo en que se utilizan las desviaciones respecto a la media de la serie para los valores previos, con el fin de predecir los valores actuales. Por ejemplo: los órdenes de media móvil 1 y 2 especifican que las desviaciones del valor medio de la serie de cada uno de los periodos de tiempo se tienen en cuenta al predecir los valores actuales de la serie (International Business Machines, s.f.).

3.7 Modelo *SARIMA*

El método *ARIMA* ayuda a analizar series de tiempo con componente tendencial; sin embargo, este método no funciona para series de tiempo con componente estacional (Brownlee, 2018), el modelo *SARIMA* (*Seasonal Autoregressive Integrated Moving Average*), es una extensión del modelo *ARIMA* que permite trabajar con series de tiempo con componente tendencial y estacional.

Brownlee menciona que el modelo *SARIMA* está conformado por los siguientes elementos estacionales P , D , Q y S , donde P es el orden estacional autorregresivo, D es el orden de las diferencias estacionales, Q es el orden de promedios móviles estacionales y S es el número de frecuencias estacionales.

3.7.1 Proceso Estacional Autorregresivo (*SAR*)

Se considera que una serie de tiempo Y_t admite un proceso autorregresivo de primer orden, que se denota por *SAR*(1) y se modela a través de la ecuación:

$$Y_t = \phi_1 Y_{t-s} + \varepsilon_t$$

Donde:

$Y_t, Y_{t-1}, \dots, Y_{t-s}, \dots$ son variables aleatorias concebidas como realizaciones de un proceso estocástico en los momentos del tiempo $t, t-1, t-2, \dots$, que se caracterizan por $E(Y_t) = E(Y_{t-1}) = E(Y_{t-2}) = \dots$ valor finito

ϕ_1 , junto con la varianza del proceso σ_ε^2 , son los parámetros que definen el modelo (estos parámetros se deben estimar)

ε_t es un proceso constituido por variables aleatorias independientes e igualmente distribuidas que deben cumplir:

- La esperanza matemática de ε_t es nula; $E(\varepsilon_t) = 0$
- La varianza de ε_t es constante; $E(\varepsilon_t \varepsilon_{t+s}) = \sigma_\varepsilon^2 \quad \forall s = 0$
- Las autocovarianzas de ε_t son nulas; $E(\varepsilon_t \varepsilon_{t+s}) = 0 \quad \forall s \neq 0$
- La variable ε_t se distribuye de manera normal

La variable aleatoria que reúne estas características se denomina variable aleatoria ruido blanco.

Definición y ecuaciones obtenidas de (Cabrer Borrás, 2014)

3.7.2 Proceso de Promedios Móviles (*SMA*)

Se considera que una serie de tiempo Y_t admite un proceso de promedios móviles de primer orden y se denota por *SMA*(1), y se modela a través de la ecuación:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Donde:

Y_t , es una variable aleatoria concebida como realización de un proceso estocástico en los momentos del tiempo $t, t - 1, t - 2, \dots$

θ_1 , junto con la varianza del proceso σ_ε^2 , son los parámetros que definen el modelo (estos parámetros se deben estimar)

ε_t es un proceso constituido por variables aleatorias independientes e igualmente distribuidas que deben cumplir:

- La esperanza matemática de ε_t es nula; $E(\varepsilon_t) = 0$
- La varianza de ε_t es constante; $E(\varepsilon_t \varepsilon_{t+s}) = \sigma_\varepsilon^2 \quad \forall s = 0$
- Las autocovarianzas de ε_t son nulas; $E(\varepsilon_t \varepsilon_{t+s}) = 0 \quad \forall s \neq 0$
- La variable ε_t se distribuye de acuerdo con una función normal

La variable aleatoria que reúne estas características se denomina variable aleatoria ruido blanco.

Definición y ecuaciones obtenidas de (Cabrer Borrás, 2014)

3.7.3 Órdenes estacionales

Los componentes estacionales autorregresivos de media móvil y diferenciación tienen los mismos roles que los componentes no estacionales correspondientes. Sin embargo, en el caso de los órdenes estacionales los valores de la serie actual se ven afectados por uno o más periodos estacionales (International Business Machines, s.f.).

IBM describe el siguiente ejemplo: para los datos mensuales (periodo estacional 12), un orden estacional de 1 significa que el valor de la serie actual se ve afectado por el valor de la serie 12 periodos antes del actual, un orden estacional de 1 para los datos mensuales equivale a la especificación de un orden no estacional de 12.

Capítulo 4. Metodología Box-Jenkins

Es un proceso iterativo que obtiene los patrones adecuados de *AR* y *MA*; a fin de utilizar esta técnica, se debe tener una serie de tiempo estacionaria o una serie de tiempo que sea estacionaria después de una o varias diferenciaciones (Gujarati, 2010, pág. 777).

4.1 Identificación

Gujarati señala que en esta fase se detectan los órdenes p , q y d del proceso autorregresivo, promedios móviles y diferenciación, respectivamente, mediante el análisis de las funciones de autocorrección (*FAC*) y de autocorrelación parcial (*FACP*), el valor de d se detecta aplicando la prueba *Augmented Dickey-Fuller* (*test ADF*) a la serie de tiempo con el objetivo de rechazar la hipótesis nula de no estacionariedad de esta si el valor $p < 0.05$.

4.1.1 Función de autocorrelación (*FAC*) y función de autocorrelación Parcial (*FACP*)

Las funciones *FAC* y *FACP* son medidas de asociación entre valores de series actuales y pasadas e indican cuáles son los valores de series pasados más útiles para predecir valores futuros (International Business Machines, s.f.).

La *FAC* en el retardo k es la autocorrelación entre los valores de las series que se encuentran a k intervalos de distancia. Si el primer rezago de un gráfico de *FAC* es estadísticamente significativo, indica que existe una fuerte correlación entre el valor de cada serie y el valor anterior; si el segundo rezago es estadísticamente significativo, indica que existe una fuerte correlación entre el valor de cada serie y el valor que aparece dos puntos anteriores. La *FACP* en el retardo k es la autocorrelación entre los valores de las series que se encuentran a k intervalos de distancia, teniendo en cuenta los valores de los intervalos intermedios. En los gráficos *FAC* y *FACP* se indica hasta qué punto la variable depende de ella misma; en el eje X se indica el retardo en el que se calcula la autocorrección y en el eje Y se indica el valor de la correlación, la cual debe estar entre -1 y 1 (International Business Machines, s.f.).

4.2 Estimación

Al haber identificado los valores apropiados de p y q , se estiman los parámetros de los términos autorregresivos y de promedios móviles incluidos en el modelo. Ese cálculo se realiza mediante mínimos cuadrados simples, pero, frecuentemente, se calculan mediante métodos de estimación no lineal (modelos paramétricos) (Gujarati, 2010, pág. 777).

4.3 Examen diagnóstico

Después de haber seleccionado un modelo *ARIMA* en particular y de haber estimado sus parámetros, se analiza si el modelo seleccionado se ajusta a los datos de manera razonable ya que puede existir otro modelo *ARIMA* que también se ajuste. Una prueba para ver si el modelo seleccionado se acepta o no, es graficar el residual estimado del modelo seleccionado, si estos se comportan como ruido blanco,¹³ se acepta el modelo de ajuste, si no, se repite el proceso a partir del paso 1 (Gujarati, 2010, pág. 777).

4.3.1 Criterios de selección del modelo

4.3.1.1 Criterio Akaike

Este criterio se utiliza para seleccionar un modelo de ajuste¹⁴ entre todos aquellos modelos de ajuste detectados, mide la bondad de ajuste de un modelo estadístico, describe la relación entre sesgo y varianza en la construcción del modelo, se selecciona aquel modelo de ajuste donde el criterio Akaike (*AIC*) tenga el valor más bajo, se calcula de la siguiente manera:

$$AIC = \log\left(\frac{\sum_{t=1}^T e_t^2}{T}\right) + \left(\frac{2k}{T}\right)$$

donde

¹³ Es un tipo especial de proceso estocástico, es puramente aleatorio si tiene una media igual a cero, una varianza constante y no está correlacionado; si también es independiente, se conoce como estrictamente ruido blanco (Gujarati, 2010, pág. 741).

¹⁴ Los valores medidos en el mundo real nunca se ajustan de forma perfecta a un modelo, debido en primer lugar a errores de medida, pero también a que cualquier modelo matemático es una simplificación del mundo real, y si tuviera en cuenta todos los factores que influyen en un conjunto de variables, sería inmanejable. Por tanto, no tiene sentido aspirar a encontrar un modelo que prediga exactamente los valores medidos, y debemos admitir que el modelo cometerá un cierto error (Angulo, 2011).

k = Número de parámetros en el modelo estadístico

T = Tamaño de la muestra

4.3.1.2 Criterio Schwarz

Este criterio, también conocido como criterio de información bayesiano, es utilizado para seleccionar un modelo de ajuste entre todos aquellos modelos de ajuste detectados, y, al igual que el criterio Akaike, mide la bondad de ajuste de un modelo estadístico, se selecciona aquel modelo de ajuste donde el criterio Schwarz (BIC) tenga el valor más bajo; se calcula de la siguiente manera:

$$BIC = \log\left(\frac{\sum_{t=1}^T e_t^2}{T}\right) + \left(\frac{k \log(T)}{T}\right)$$

donde

k = Número de parámetros en el modelo estadístico

T = Tamaño de la muestra

4.3.1.3 Medidas de precisión

El error del pronóstico o precisión es la diferencia entre el valor real y el valor pronosticado de un período específico (PRICING revenue management , s.f.):

$$u_t = Y_t - \hat{Y}_t$$

donde u_t es el error del pronóstico en el período t , Y_t es el valor real en ese período y \hat{Y}_t es el valor del pronóstico del mismo periodo.

Algunas medidas de precisión miden el error del pronóstico y son las siguientes:¹⁵

ME (error medio): El error medio es la diferencia entre el valor estimado y el valor verdadero, se calcula de la siguiente manera:

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_t - \hat{Y}_t)$$

¹⁵ Definiciones obtenidas de *StackExchange*, la URL del sitio es: <https://stats.stackexchange.com/questions/194453/interpreting-accuracy-results-for-an-arima-model-fit>

MAE (error absoluto medio): El error absoluto medio, mide el promedio de la magnitud de los errores en un conjunto de pronósticos, indica qué tan grande es el error que se puede esperar del pronóstico en promedio, se calcula de la siguiente manera:

$$MAE = \frac{\sum_{i=1}^n |Y_t - \hat{Y}_t|}{n}$$

RMSE (error cuadrático medio): Siempre será más grande o igual a *MAE*, cuanto mayor es la diferencia entre ellos, mayor es la varianza en los errores individuales de la muestra.

Si $RMSE = MAE$, entonces todos los errores son de la misma magnitud; tanto el *MAE* como el *RMSE* pueden variar desde 0 a ∞ . Los valores más bajos son mejores. Esta medida se calcula de la siguiente manera:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y} - Y_t)^2}{n}}$$

EPM (error de porcentaje medio): Es el promedio calculado de los errores porcentuales por el cual los predictores de un modelo difieren de los valores reales de la cantidad que se pronostica, se calcula de la siguiente manera:

$$EPM = \sum_{i=1}^n \left[\frac{(Y_t - \hat{Y}_t)}{Y_t} \right] \times 100$$

MAPE (error porcentual absoluto medio): Expresa la precisión como un porcentaje del error. Debido a que el *MAPE* es un porcentaje, puede ser más fácil de entender que otros estadísticos de medición de precisión, por ejemplo, si el *MAPE* es 5, entonces en promedio, el pronóstico está errado en un 5 % (Minitab, 2019). Es posible que algunas veces se observe un valor del *MAPE* muy elevado, aunque el modelo parezca ajustarse a los datos adecuadamente hay que examinar la gráfica para ver si los valores de los datos se aproximan a cero, debido a que el *MAPE* divide el valor absoluto entre los datos observados, los valores que se aproximan a cero pueden aumentar significativamente el *MAPE* (Minitab, 2019). Esta medida se calcula de la siguiente manera:

$$MAPE = \frac{\sum_{i=1}^n \frac{|Y_t - \hat{Y}_t|}{|Y_t|}}{n}$$

MASE (error medio absoluto a escala): Es independiente de la escala de los datos, por lo que puede usarse para comparar pronósticos entre conjuntos de datos con diferentes escalas. Al comparar métodos de pronóstico, el método con el *MASE* más bajo es el que se prefiere. Esta medida se calcula de la siguiente manera:

$$MASE = \frac{MAE}{* MAE}$$

donde;

MAE = Error absoluto medio

* *MAE* = Error absoluto medio del modelo de pronóstico de punto de referencia calculado como muestra.

4.3.2 Análisis de los residuos

Una de las condiciones necesarias para realizar una predicción mediante una serie de tiempo es que la variable de error siga una distribución aproximadamente normal.

4.3.2.1 Distribución normal

La distribución normal es una distribución de probabilidad teórica, tiene forma de campana. Se dice que una variable aleatoria (continua) X se distribuye de manera normal si su función de distribución de probabilidad tiene la siguiente forma (Gujarati, 2010, pág. 817):

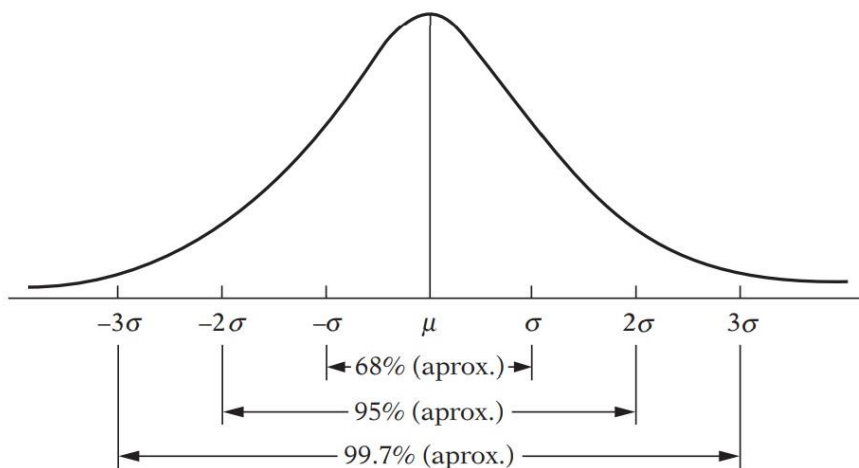
$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right); \quad -\infty < x < \infty$$

donde μ y σ^2 se denominan parámetros de la distribución, media y varianza de la distribución, respectivamente. Gujarati identifica que la distribución normal tiene algunas propiedades importantes:

1. Es simétrica alrededor de su valor promedio.
2. Aproximadamente 68 % del área bajo la curva normal se encuentra entre los valores de $\mu \pm \sigma$, alrededor de 95 % del área se encuentra entre $\mu \pm 2\sigma$ y alrededor de 99.7 % del área se encuentra entre $\mu \pm 3\sigma$.
3. La distribución normal depende de los parámetros, μ y σ^2 . Por lo tanto, una vez especificados se puede determinar la probabilidad de que X esté dentro de cierto intervalo mediante la función de distribución de probabilidad de la distribución normal.

En la siguiente imagen se observan estas propiedades:

Imagen 1. Áreas bajo la curva normal (Gujarati, 2010, pág. 817).



4.3.2.2 Sesgo

Medida de asimetría¹⁶ de una distribución, considerando que la distribución normal es simétrica,¹⁷ el sesgo puede ser positivo¹⁸ o negativo.¹⁹

¹⁶ Es una medida que provee información acerca de la distribución uniforme de los registros que se están analizando.

¹⁷ Cuando la mayoría de los registros recolectados se distribuyen de la misma manera en ambos lados.

¹⁸ Cuando la mayoría de los registros recolectados se encuentran por encima del valor del promedio.

¹⁹ Cuando la mayoría de los datos recolectados se encuentran en los valores más bajos que su valor promedio.

4.3.2.3 Curtosis

Es una medida que determina el grado de concentración que presentan los datos en la región central de la distribución (SPSS FREE, 2015). Si los datos están muy concentrados en su valor medio se denomina leptocurtosis, si los datos se encuentran concentrados normalmente se denomina mesocurtosis y si los datos tienen una concentración baja se denomina platicurtosis.

4.3.2.4 Estadístico Cuantil-Cuantil (Q-Q *plot*)

El gráfico Cuantil-Cuantil (Q-Q *plot*) permite comparar la distribución empírica de un conjunto de datos, con la distribución normal, por lo que, este gráfico se considera como una técnica gráfica para la prueba de normalidad de un conjunto de datos (Catillo Gutierrez & Lozano Aguilera, 2007).

4.4 Pronóstico

Teniendo el modelo de ajuste adecuado, es decir, que el mismo haya pasado la prueba de diagnóstico de ruido blanco, se procede a realizar la proyección a futuro de los valores de la serie de tiempo, esta es la última fase que señalan Gujarati y Porter.

Capítulo 5. Aplicación de la metodología Box-Jenkins al INPC con el lenguaje «R»²⁰

Tabla 1. INPC general de México²¹

Período quincenal	Índice general
01/01/1988	4.636980106
02/01/1988	4.799511568
01/02/1988	5.041704344
02/02/1988	5.181864637
01/03/1988	5.332375151
02/03/1988	5.414721544
01/04/1988	5.526642661
02/04/1988	5.551240429
01/05/1988	5.615513439
02/05/1988	5.676704021
01/06/1988	5.739709868
02/06/1988	5.782874209
01/07/1988	5.844765201
02/07/1988	5.870150931
01/08/1988	5.908970863
02/08/1988	5.913717283
01/09/1988	5.944982419
02/09/1988	5.945298433
01/10/1988	5.973526731
02/10/1988	6.007448971
01/11/1988	6.057258403
02/11/1988	6.084051689
01/12/1988	6.176145805
02/12/1988	6.218489547

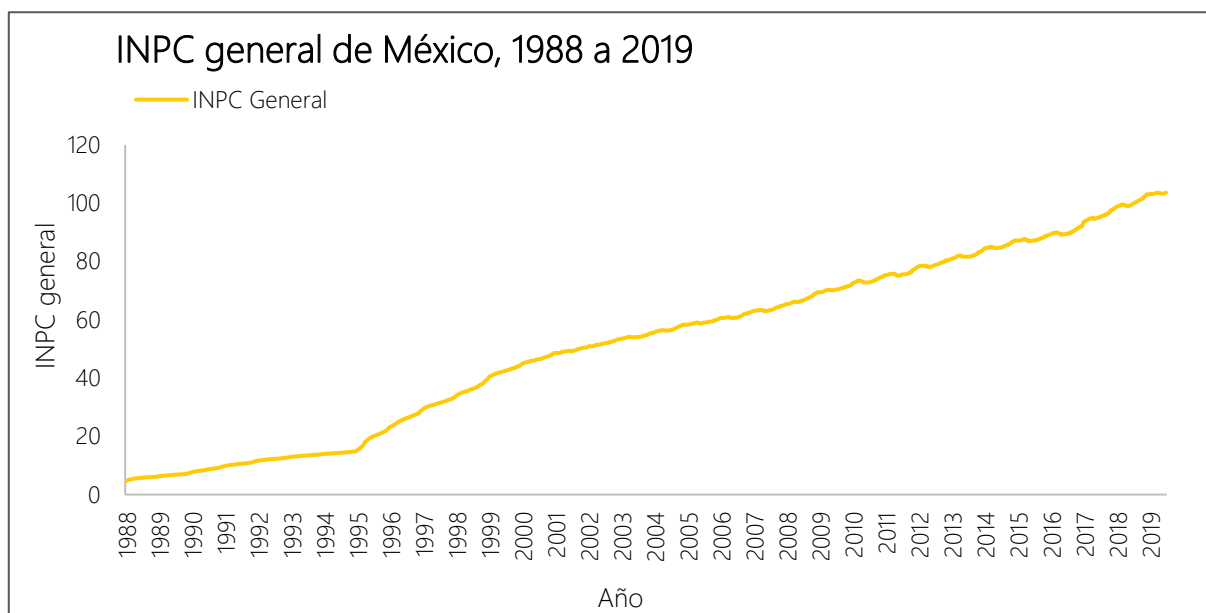
Fuente: Elaboración propia en Excel con datos del Inegi.

La Tabla 1 muestra un subconjunto de los datos analizados, la serie temporal tiene 732 observaciones del INPC general en periodo quincenal, a partir de la primera quincena de 1988 hasta la primera quincena de julio de 2019. Esta serie de datos se graficó para observar sus componentes.

²⁰ Ver Anexo 1.

²¹ Esta serie de datos se encuentra disponible en el Banco de Información Económica (BIE) del Inegi en la siguiente ruta:
<https://www.inegi.org.mx/sistemas/bie/?idserPadre=10000500001500600040>

Gráfica 6. INPC general quincenal de México, 1988 a 2019



Fuente: Elaboración propia en Excel con datos del Inegi.

La Gráfica 6 muestra la evolución del INPC, se distingue la tendencia creciente determinista y falta de estacionariedad en media; debido a que las observaciones no tienen una trayectoria alrededor de su valor promedio, es decir, la media de las observaciones de la serie de tiempo cambia a lo largo del periodo, se descartó falta de estacionariedad en varianza debido a que, mientras aumenta el periodo, la varianza de las observaciones no cambia a lo largo del tiempo, para corroborar que se necesita diferenciar la serie de tiempo y estabilizar su varianza se aplicó a las observaciones la prueba *Augmented Dickey-Fuller* (*adf.test*)²² en el programa estadístico «R», esta prueba tiene como hipótesis nula (H_0) la no estacionariedad de la serie y como hipótesis alternativa (H_a) la estacionariedad de la serie; formalmente:

- H_0 : serie de tiempo no estacionaria (raíz unitaria).
- H_a : serie de tiempo estacionaria.

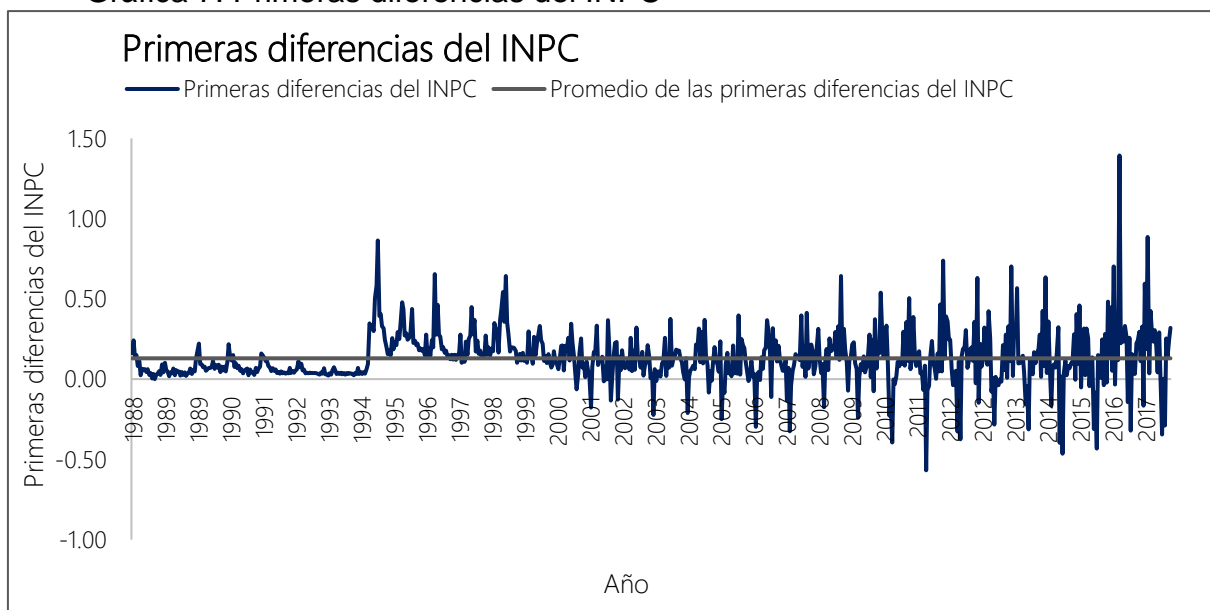
²² La prueba *Augmented Dickey Fuller* (Dickey-Fuller Aumentada) puede usarse con serie de datos autocorrelacionados, la hipótesis nula de esta señala que hay una raíz unitaria y la hipótesis alternativa señala que la serie de tiempo es estacionaria (Stephanie, 2016).

Para rechazar la H_0 frente a la H_a , el valor p tiene que ser menor que 0.05; de lo contrario, se acepta la H_0 frente a la H_a y se aplican las primeras diferencias a la serie. El resultado de la prueba ADF fue el siguiente:

```
## Raíz unitaria (serie de tiempo no estacionaria) el valor p es:
0.68
## Debido a que la serie no es estacionaria entonces se estabiliza la
serie de tiempo
## antes de estimar el modelo.
```

El resultado de la prueba generó un valor p mayor que 0.05, por ello, se acepta la H_0 frente a la H_a y se diferencia una vez la serie de tiempo, se vuelve a aplicar la prueba para comprobar que ya no hay raíz unitaria, antes de ver los resultados de la prueba, se graficó la serie del INPC diferenciada para observar si tiene un comportamiento de ruido blanco:

Gráfica 7. Primeras diferencias del INPC



Fuente: Elaboración propia en Excel con datos del INPC.

La Gráfica 7 muestra que las observaciones del INPC diferenciadas tienen una trayectoria alrededor del cero (valor medio de la serie del INPC) y con varianza constante, es decir, es ruido blanco, lo que significa que la serie es estacionaria en varianza, el resultado de la prueba ADF fue:

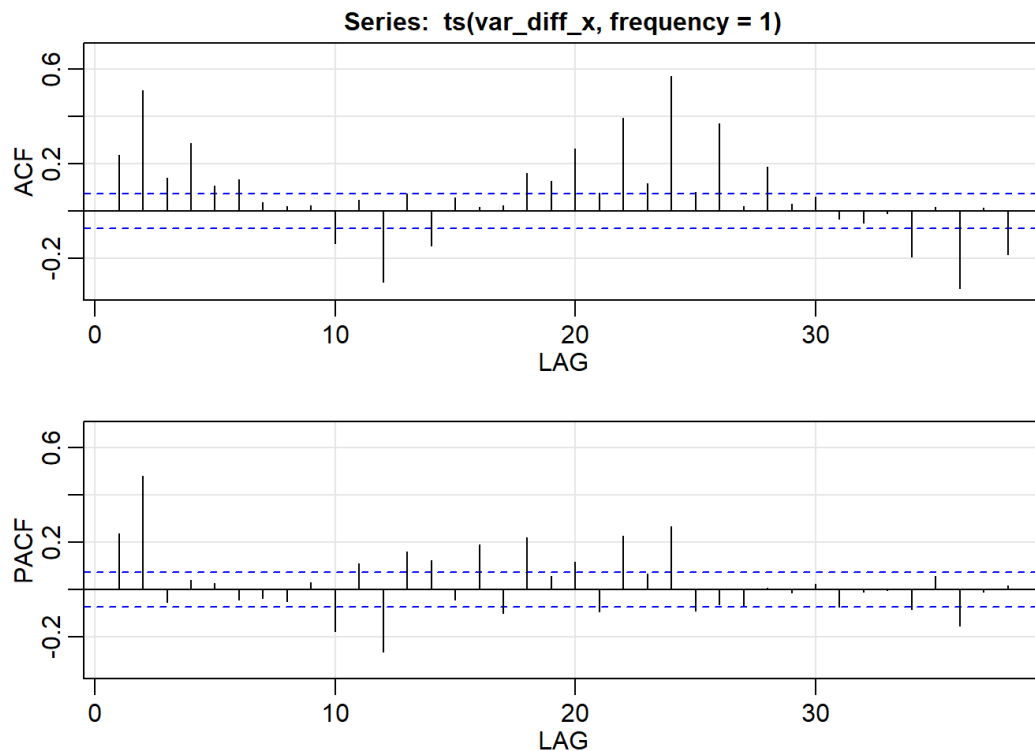
```
## Serie de tiempo estacionaria el valor p es: 0.01
## por lo tanto, se rechaza la H0 de no estacionariedad.
## Se puede entonces calcular el modelo de ajuste
```

En efecto, el valor p es menor que 0.05, por lo que se rechazó la H_0 de no estacionariedad frente a la H_a de estacionariedad; como la serie de tiempo es estacionaria, se prosiguió a la fase de identificación de los órdenes *ARIMA*.

5.1 Identificación de los órdenes no estacionales y estacionales

En esta fase las funciones de autocorrelación (*ACF*) y autocorrelación parcial (*PACF*) contribuyen a identificar los órdenes *ARMA*(p, q) de la serie de tiempo.

Gráfica 8. Funciones de autocorrelación y autocorrelación parcial



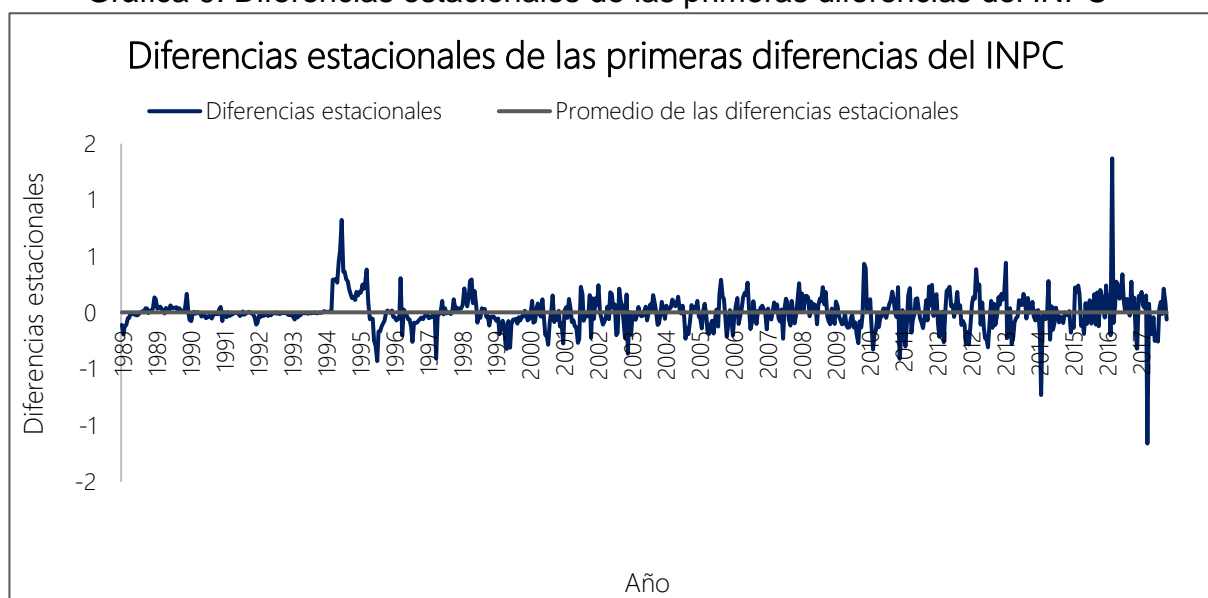
Fuente: Elaboración propia en RStudio.

La Gráfica 8 está compuesta por dos gráficas; la gráfica de la función *ACF* (parte superior) muestra un comportamiento estacional, crece y decrece a lo largo de los rezagos, lo que significa que la serie de tiempo tiene un componente estacional, por lo tanto, se debe de utilizar un modelo *SARIMA*. La gráfica de la función *PACF* (parte inferior) también muestra ese patrón de comportamiento creciente y decreciente a lo largo de los rezagos. Los rezagos de la función de autocorrelación *ACF* son estadísticamente significativos hasta el segundo, por lo que se identificó un proceso de promedios móviles de segundo orden, es decir, un *MA*(2); por otro lado, la función de autocorrelación parcial (*PACF*) muestra que

el segundo rezago es estadísticamente significativo, por ello se identificó un proceso autorregresivo de segundo orden, es decir, un $AR(2)$.

La serie de tiempo se diferenció una vez, entonces se tiene un proceso $ARIMA(2,1,2)$. Además, se detectó un componente estacional, lo que implicó calcular las diferencias estacionales sobre las primeras diferencias del INPC siendo estas de orden 24 porque la frecuencia de la serie es quincenal. Posteriormente se graficaron a fin de observar si se comportan como ruido blanco, después sobre las diferencias estacionales, se aplicó la prueba ADF con el objetivo de corroborar que no exista raíz unitaria:

Gráfica 9. Diferencias estacionales de las primeras diferencias del INPC

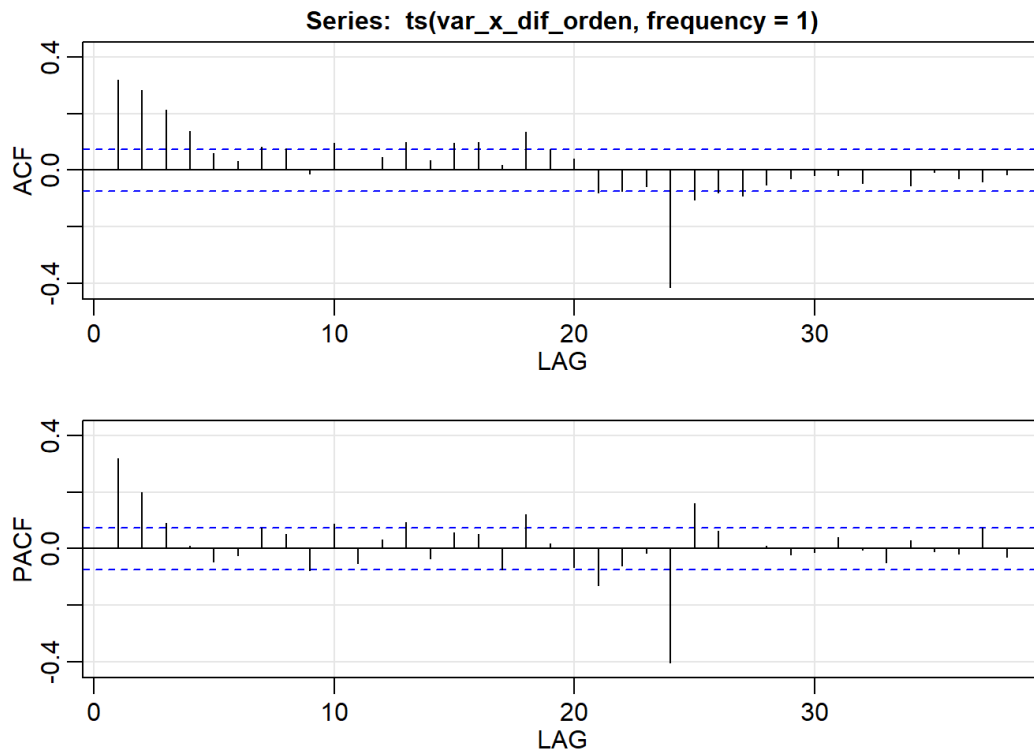


Fuente: Elaboración propia en Excel con datos del INPC.

La Gráfica 9 muestra que las diferencias estacionales de las primeras diferencias del INPC se comportan como ruido blanco; su media es cero y su varianza constante, al aplicar la prueba ADF se corroboró que no hay raíz unitaria.

Como no hay raíz unitaria se graficaron las funciones de autocorrelación (ACF) y autocorrelación parcial ($PACF$) para identificar los procesos AR y MA estacionales:

Gráfica 10. Funciones de autocorrelación y autocorrelación del componente estacional de la serie de tiempo



Fuente: Elaboración propia en RStudio.

La función de autocorrelación (*ACF*) de las diferencias estacionales muestra un patrón de comportamiento en el que los rezagos van decreciendo a lo largo de la gráfica y se acercan a cero, por ello, se identificó un proceso de promedios móviles estacionales de segundo orden, es decir, un *SMA*(2). En la función de autocorrelación parcial (*PACF*) se observa un patrón de comportamiento donde los rezagos van decreciendo y creciendo a lo largo de la gráfica, por ello, se identificó un proceso autorregresivo estacional de primer orden, es decir, un *SAR*(1), debido a que se aplicaron a la serie de tiempo las diferencias de orden 24, es decir, se diferenció un periodo quincenal, formando los órdenes estacionales *SARIMA*(*p, d, q*) (*P, D, Q*) [*S*], esto es, *SARIMA*(2,1,2) (1,1,2)₂₄

Identificados los órdenes no estacionales como los estacionales, se pasó a la fase de estimación de los parámetros autorregresivos y de promedios móviles.

5.2 Estimación de los modelos de ajuste

La estimación de los parámetros se realizó mediante máxima verosimilitud,²³ generé diversos modelos de ajuste y elegí el que tuvo los menores valores de Akaike, Schwarz y de medidas de precisión. La estimación de los modelos de ajuste se hizo en «R», los resultados los muestro enseguida:

Criterios AIC y BIC de los modelos de ajuste

```
##      Órdenes_pdq_y_PDQ AIC      BIC
## [1,] "(0,1,0) (1,1,2)" "-920.987620138287" "-902.604498082442"
## [2,] "(1,1,0) (1,1,2)" "-1058.10642288303" "-1035.12752031322"
## [3,] "(0,1,1) (1,1,2)" "-1001.72847709152" "-978.749574521712"
## [4,] "(1,1,1) (1,1,2)" "-1115.85677673985" "-1088.28209365609"
## [5,] "(0,1,2) (1,1,2)" "-1063.43005438567" "-1035.8553713019"
## [6,] "(2,1,0) (1,1,2)" "-1110.79289752918" "-1083.21821444541"
## [7,] "(1,1,2) (1,1,2)" "-1114.12942203252" "-1081.9589584348"
## [8,] "(2,1,1) (1,1,2)" "-1114.39764767385" "-1082.22718407612"
## [9,] "(2,1,2) (1,1,2)" "-1116.1732056978" "-1079.40696158611"
```

```
## Modelo de ajuste con valor Akaike más bajo: 9
```

```
## Valor Akaike: -1116.2
```

```
## Modelo de ajuste con valor BIC más bajo: 4
```

```
## Valor BIC: -1088.3
```

Medidas de precisión de los modelos de ajuste

```
##      MAPE      ME      RMSE      MAE      MPE      MASE
## [1,] 0.20796 0.0094089 0.12484 0.080528 0.043476 0.51007
## [2,] 0.16669 0.0052839 0.11331 0.071353 0.025043 0.45196
## [3,] 0.18435 0.0074426 0.11795 0.074910 0.034549 0.47449
## [4,] 0.15311 0.0029299 0.10855 0.068127 0.014362 0.43152
## [5,] 0.17173 0.0063236 0.11271 0.071891 0.029207 0.45536
## [6,] 0.15581 0.0040521 0.10894 0.069158 0.018878 0.43806
## [7,] 0.15322 0.0030848 0.10852 0.068220 0.014939 0.43211
## [8,] 0.15339 0.0032080 0.10850 0.068306 0.015408 0.43266
## [9,] 0.15225 0.0030495 0.10821 0.068110 0.014771 0.43142
```

```
## Modelo de ajuste con MAPE más bajo: 9
```

```
## Valor MAPE: 0.15225
```

```
## Modelo de ajuste con ME más bajo: 4
```

```
## Valor ME: 0.0029299
```

```
## Modelo de ajuste con RMSE más bajo: 9
```

```
## Valor RMSE: 0.10821
```

```
## Modelo de ajuste con MAE más bajo: 9
```

²³ Es una técnica para estimar los valores de θ dada una muestra finita de datos (Gómez Cadenas, 2005).


```
## Valor MAE: 0.06811
## Modelo de ajuste con MPE más bajo: 4
## Valor MPE: 0.014362
## Modelo de ajuste con MASE más bajo: 9
## Valor MASE: 0.43142
```

Tanto el criterio de Akaike como las medidas de precisión: *MAPE*, *RMSE*, *MAE* y *MASE* sugirieron que el modelo de ajuste con valor más bajo es el número 9; por otra parte, el criterio de Schwarz (*BIC*) y las medidas de precisión: *ME* y *MPE*, sugirieron que el modelo de ajuste con valor más bajo es el número 4. Debido a los resultados obtenidos por los criterios de selección, se eligieron los modelos de ajuste número 4 y número 9: antes de pasar a la fase de examen de diagnóstico se generaron cinco modelos de ajuste adicionales con el objetivo de identificar que, mientras aumenten los órdenes autorregresivos y de promedios móviles, disminuyen los valores de los criterios de selección, el resultado en «R» fue el siguiente:

```
## Modelo_Ajustado Órdenes_pdq_y_PDQ      AIC      BIC
## 1  Ajuste_INPC_1  (0,1,0) (1,0,0)      -920.99 -902.6
## 2  Ajuste_INPC_2  (1,1,1) (1,0,0)     -1115.86 -1088.3
## 3  Ajuste_INPC_3  (2,1,2) (1,0,0)     -1116.17 -1079.4
## 4  Ajuste_INPC_4  (3,1,3) (1,0,0)     -1119.22 -1073.3
## 5  Ajuste_INPC_5  (4,1,4) (1,0,0)     -1120.20 -1065.0
```

```
## Modelo de ajuste con valor Akaike más bajo: 5
## Valor Akaike: -1120.2
## Modelo de ajuste con valor BIC más bajo: 2
## Valor BIC: -1088.3
```

Medidas de precisión de los modelos de ajuste

```
## Modelo_Ajustado      MAPE      ME      RMSE      MAE      MPE      MASE
## 1  Ajuste_INPC_1  0.20796  0.0094089  0.12484  0.080528  0.043476  0.51007
## 2  Ajuste_INPC_2  0.15311  0.0029299  0.10855  0.068127  0.014362  0.43152
## 3  Ajuste_INPC_3  0.15225  0.0030495  0.10821  0.068110  0.014771  0.43142
## 4  Ajuste_INPC_4  0.15287  0.0030495  0.10768  0.068239  0.014636  0.43224
## 5  Ajuste_INPC_5  0.15120  0.0020802  0.10732  0.067396  0.012158  0.42689
```

```
## Modelo de ajuste con MAPE más bajo: 5
## Valor MAPE: 0.1512
## Modelo de ajuste con ME más bajo: 5
## Valor ME: 0.0020802
```

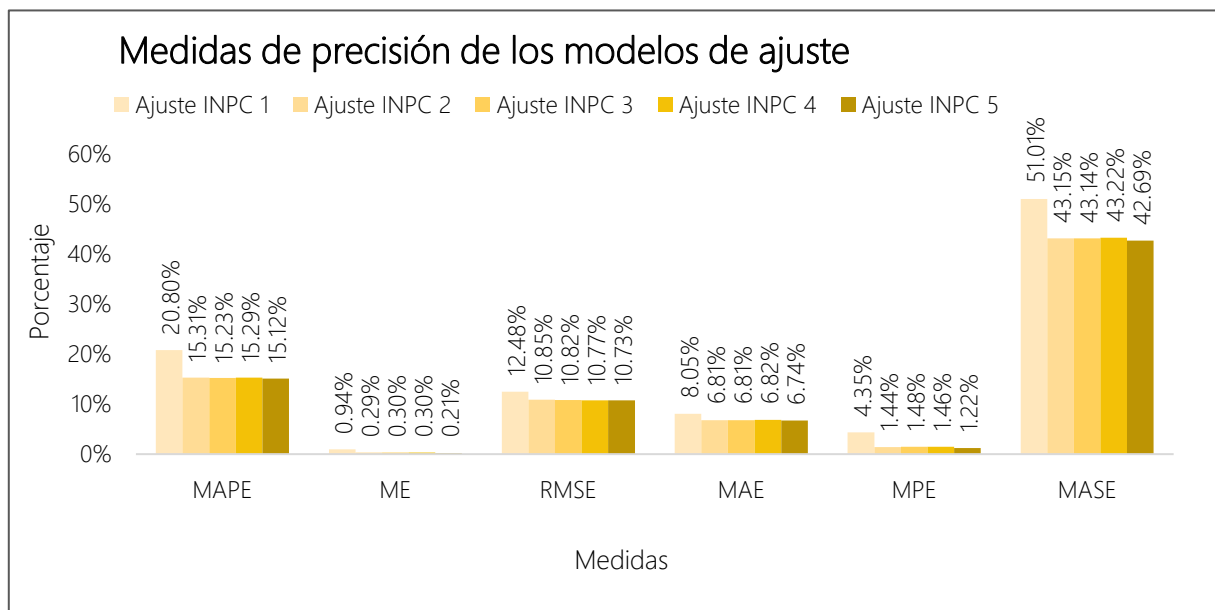
```

## Modelo de ajuste con RMSE más bajo: 5
## Valor RMSE: 0.10732
## Modelo de ajuste con MAE más bajo: 5
## Valor MAE: 0.067396
## Modelo de ajuste con MPE más bajo: 5
## Valor MPE: 0.012158
## Modelo de ajuste con MASE más bajo: 5
## Valor MASE: 0.42689

```

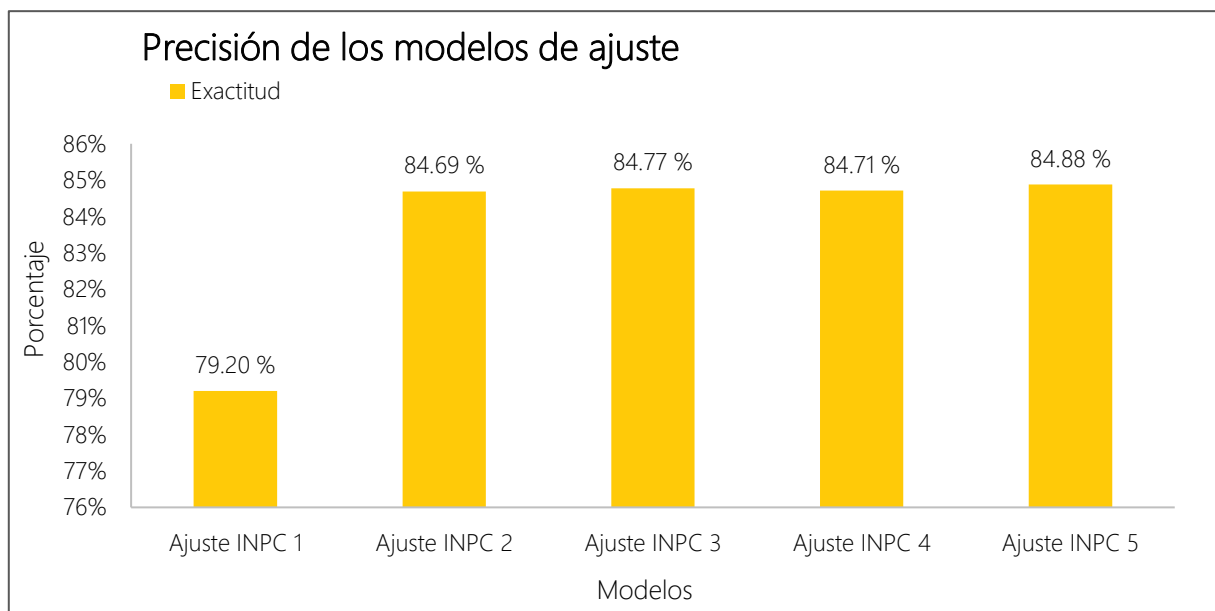
Se graficaron las medidas de precisión para identificar visualmente que sus valores disminuyen conforme aumentan los órdenes no estacionales, los resultados se muestran en las Gráficas 11 y 12.

Gráfica 11. Medidas de precisión de los modelos de ajuste



Fuente: Elaboración propia en Excel.

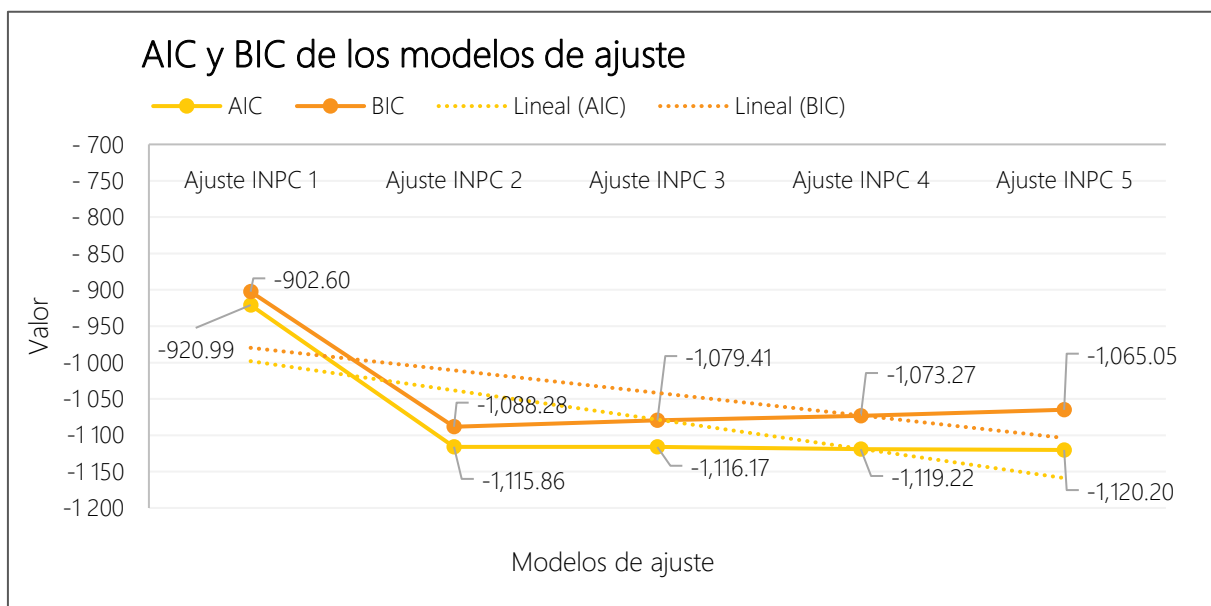
Gráfica 12. Precisión de los modelos de ajuste²⁴



Fuente: Elaboración propia en Excel.

Para identificar que los valores de los criterios de Akaike y Schwarz disminuyen conforme aumentan los órdenes no estacionales se graficaron sus valores.

Gráfica 13. Criterios Akaike y Schwarz



Fuente: Elaboración propia en Excel.

²⁴ Para formar esta medida se restó 1 a la medida de precisión *MAPE* y se multiplicó por 100.

La Gráfica 11 muestra que el valor de las medidas de precisión de los modelos de ajuste disminuye mientras aumentan los órdenes no estacionales, esto sugiere que se puede estimar otro modelo de ajuste mejor; sin embargo, los valores disminuyen pocas décimas porcentuales. La Gráfica 12 muestra que, mientras aumentan los procesos *AR* y *MA*, la precisión del pronóstico no sube a más de 85 %. la Gráfica 13 muestra el comportamiento de los criterios de Akaike y de Schwarz, el valor del primero disminuye conforme aumentan los órdenes no estacionales, en tanto que no sucede lo mismo con el segundo; la línea de tendencia lineal de ambos es decreciente, sugiriendo que se puede estimar un mejor modelo de ajuste.

Con base en los resultados obtenidos se generaron tres modelos de ajuste adicionales con los siguientes órdenes no estacionales y estacionales:

Ajuste INPC 1, *SARIMA*(3,1,1) (1,1,2)₂₄

Ajuste INPC 2, *SARIMA*(1,1,2) (1,1,2)₂₄

Ajuste INPC 3, *SARIMA*(2,1,2) (1,1,2)₂₄

Sobre los tres modelos de ajuste se aplicaron los criterios de selección con el objetivo de elegir uno y compararlo con los dos modelos de ajuste escogidos anteriormente; el resultado de los criterios fue el siguiente:

```
##      Modelo_Ajustado      AIC      BIC
## 1  SARIMA(3,1,1) (1,1,2) [24] -1127.9 -1091.2
## 2  SARIMA(3,1,2) (1,1,2) [24] -1126.0 -1084.6
## 3  SARIMA(3,1,3) (1,1,2) [24] -1119.2 -1073.3
```

```
## Modelo de ajuste con valor Akaike más bajo: 1
```

```
## Valor Akaike: -1127.9
```

```
## Modelo de ajuste con valor BIC más bajo: 1
```

```
## Valor BIC: -1091.2
```

Medidas de precisión de los modelos de ajuste

```
##      MAPE      ME      RMSE      MAE      MPE      MASE
## 1  0.15174 0.0017592 0.10729 0.067463 0.011386 0.42732
## 2  0.15178 0.0017635 0.10729 0.067482 0.011400 0.42744
## 3  0.15287 0.0030495 0.10768 0.068239 0.014636 0.43224
```

```

## Modelo de ajuste con MAPE más bajo: 1

## Valor MAPE: 0.15174

## Modelo de ajuste con ME más bajo: 1

## Valor ME: 0.0017592

## Modelo de ajuste con RMSE más bajo: 2

## Valor RMSE: 0.10729

## Modelo de ajuste con MAE más bajo: 1

## Valor MAE: 0.067463

## Modelo de ajuste con MPE más bajo: 1

## Valor MPE: 0.011386

## Modelo de ajuste con MASE más bajo: 1

## Valor MASE: 0.42732

```

La mayoría de los criterios de selección sugirieron elegir el primer modelo, por ello se compararon con los modelos escogidos anteriormente, siendo el resultado el siguiente:

```

##      Modelo_Ajustado  Órdenes_pdq_y_PDQ    AIC      BIC
## 1  Ajuste_INPC_1      (3,1,1) (1,1,2) -1127.9 -1091.2
## 2  Ajuste_INPC_2      (1,1,2) (1,1,2) -1114.1 -1082.0
## 3  Ajuste_INPC_3      (2,1,2) (1,1,2) -1116.2 -1079.4

```

```

## Modelo de ajuste con valor Akaike más bajo: 1

## Valor Akaike: -1127.9

## Modelo de ajuste con valor BIC más bajo: 1

## Valor BIC: -1091.2

```

Medidas de precisión de los modelos de ajuste

```

##      Modelo_Ajustado    MAPE      ME    RMSE    MAE      MPE    MASE
## 1  Ajuste_INPC_1 0.15174 0.0017592 0.10729 0.067463 0.011386 0.42732
## 2  Ajuste_INPC_2 0.15322 0.0030848 0.10852 0.068220 0.014939 0.43211
## 3  Ajuste_INPC_3 0.15225 0.0030495 0.10821 0.068110 0.014771 0.43142

```

```

## Modelo de ajuste con MAPE más bajo: 1

## Valor MAPE: 0.15174

## Modelo de ajuste con ME más bajo: 1

## Valor ME: 0.0017592

## Modelo de ajuste con RMSE más bajo: 1

## Valor RMSE: 0.10729

```

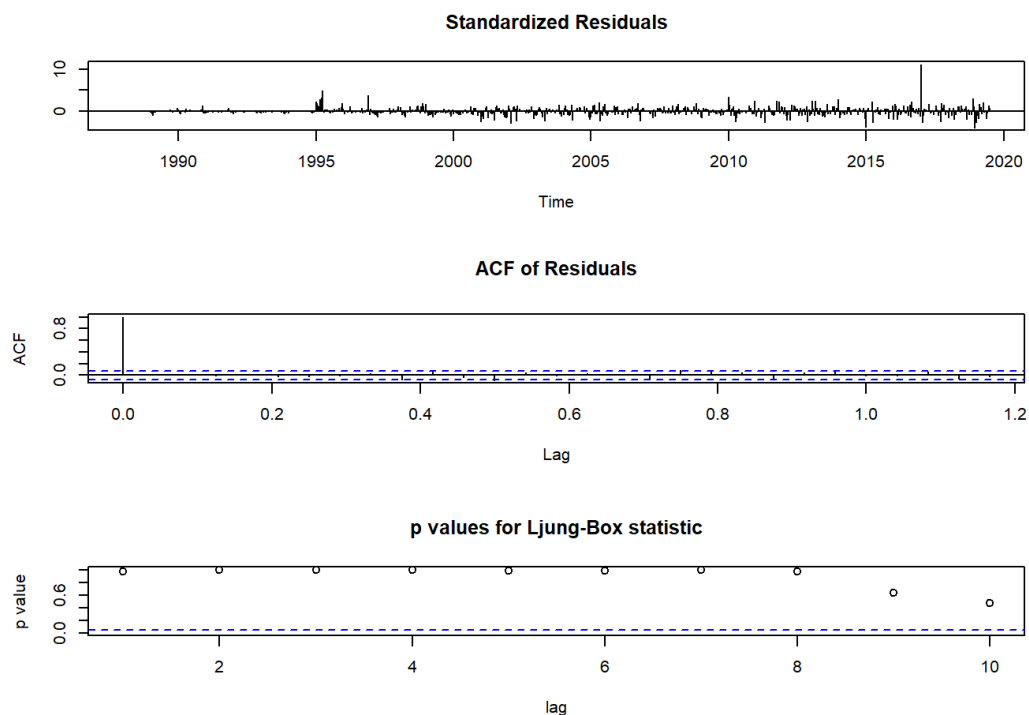
```
## Modelo de ajuste con MAE más bajo: 1
## Valor MAE: 0.067463
## Modelo de ajuste con MPE más bajo: 1
## Valor MPE: 0.011386
## Modelo de ajuste con MASE más bajo: 1
## Valor MASE: 0.42732
```

Los resultados de los criterios de selección sugirieron que el mejor modelo de ajuste es el primero, y antes de elegirlo, se realizó el examen de diagnóstico para observar si los residuos estimados eran ruido blanco.

5.3 Examen diagnóstico de los residuos de los modelos de ajuste

Se compararon los residuos de los tres modelos de ajuste elegidos como los mejores para corroborar que no estuvieran autocorrelacionados, se distribuyeran de manera normal y se comportaran como ruido blanco:

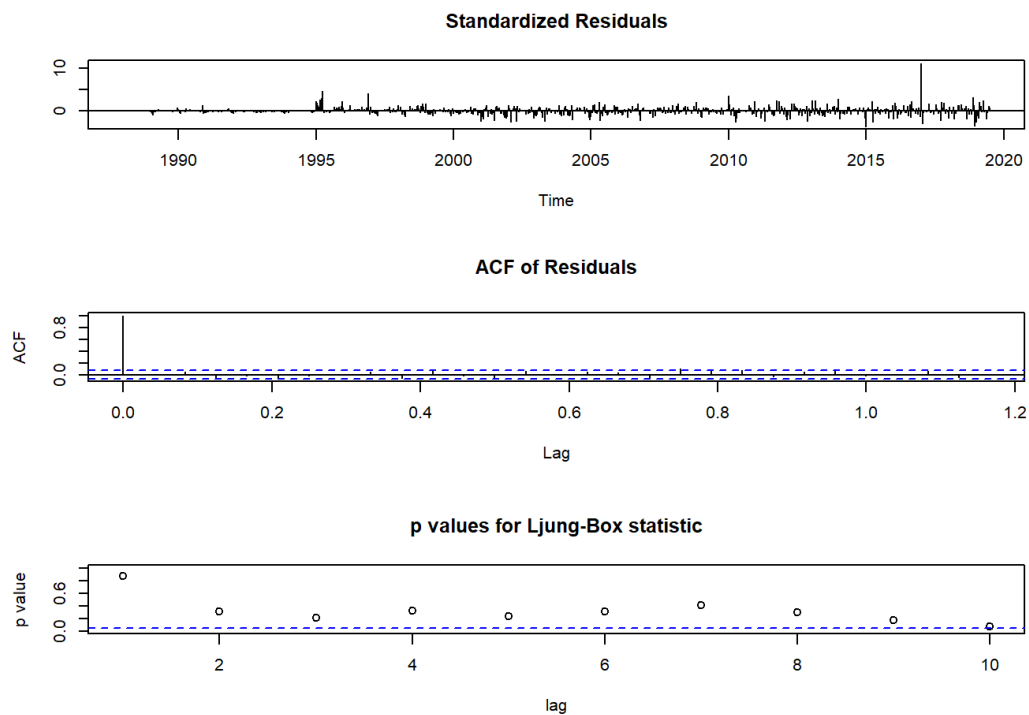
Gráfica 14. Ljung-Box del modelo de ajuste $SARIMA(3,1,1) (1,1,2)_{24}$



Fuente: Elaboración propia en RStudio.

La Gráfica 14 muestra que los rezagos están por arriba del valor p de 0.05 (línea punteada azul) del estadístico Ljung-Box, sugiriendo que se tiene suficiente evidencia estadística para rechazar la hipótesis nula de dependencia de los errores; es decir, se puede afirmar con un porcentaje alto de confianza que los errores no están autocorrelacionados.

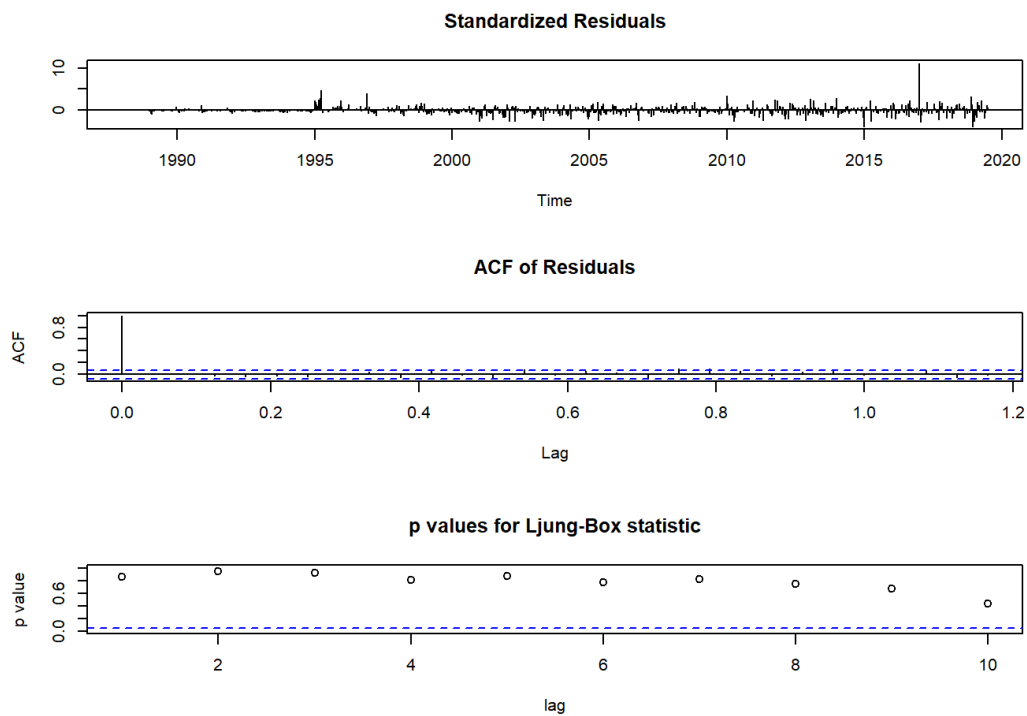
Gráfica 15. Ljung-Box del modelo de ajuste $SARIMA(1,1,2) (1,1,2)_{24}$



Fuente: Elaboración propia en RStudio.

La Gráfica 15 muestra que los rezagos están levemente por arriba del valor p de 0.05 (línea punteada azul) del estadístico Ljung-Box, por lo que se tiene poca evidencia estadística para rechazar la hipótesis nula de dependencia de los errores; es decir, se puede afirmar con un porcentaje pequeño de confianza que los errores no están autocorrelacionados.

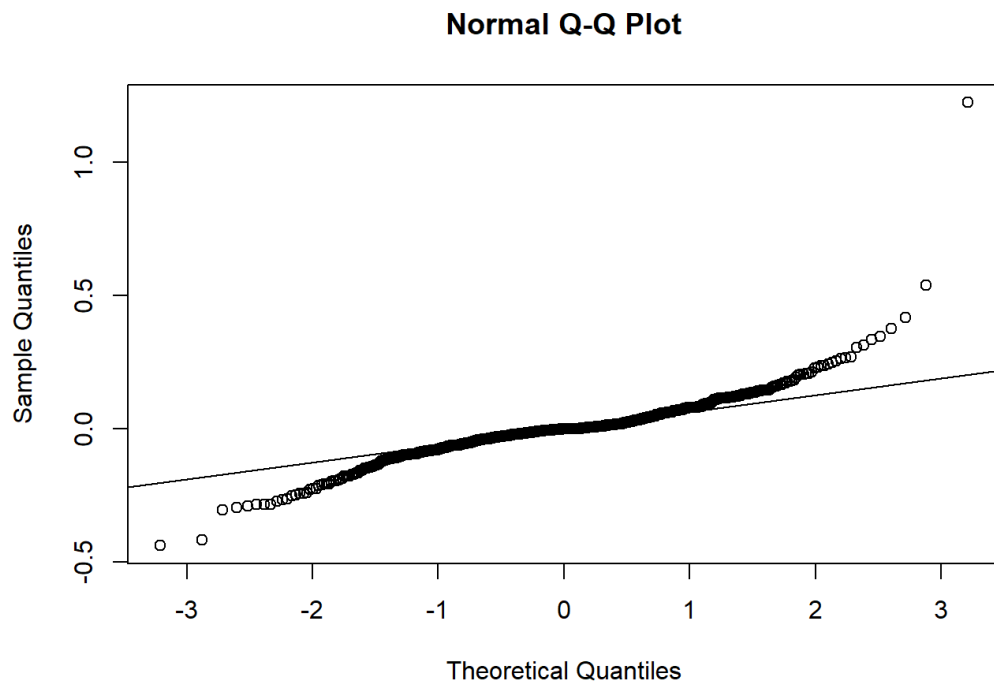
Gráfica 16. Ljung-Box del modelo de ajuste $SARIMA(2,1,2) (1,1,2)_{24}$



Fuente: Elaboración propia en RStudio.

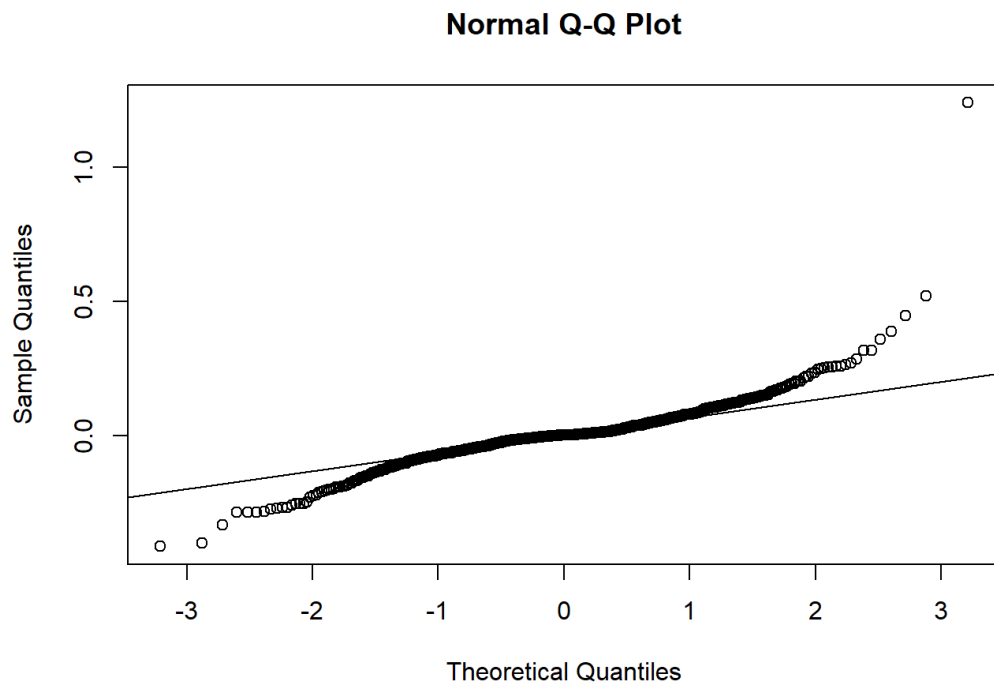
La Gráfica 16 muestra que los rezagos están por arriba del valor p de 0.05 (línea punteada azul) del estadístico Ljung-Box, de ahí que se tenga suficiente evidencia estadística para rechazar la hipótesis nula de dependencia de los errores; en otras palabras, se puede sustentar con un porcentaje alto de confianza que los errores no están autocorrelacionados.

Gráfica 17. Q-Q *Plot* del modelo de ajuste $SARIMA(3,1,1) (1,1,2)_{24}$



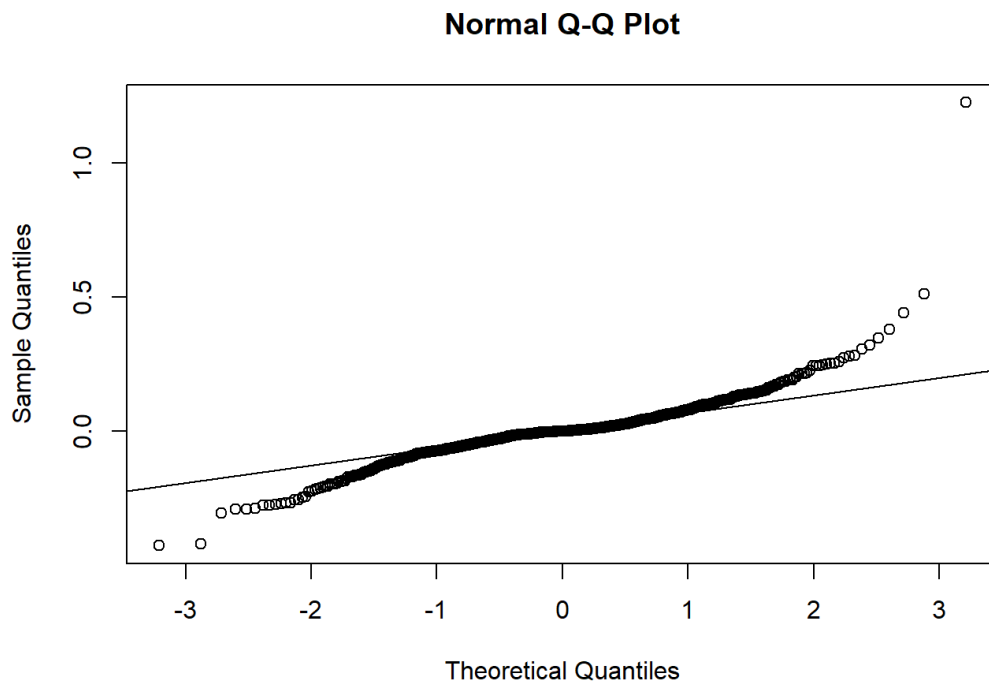
Fuente: Elaboración propia en RStudio.

Gráfica 18. Q-Q *Plot* del modelo de ajuste $SARIMA(1,1,2) (1,1,2)_{24}$



Fuente: Elaboración propia en RStudio.

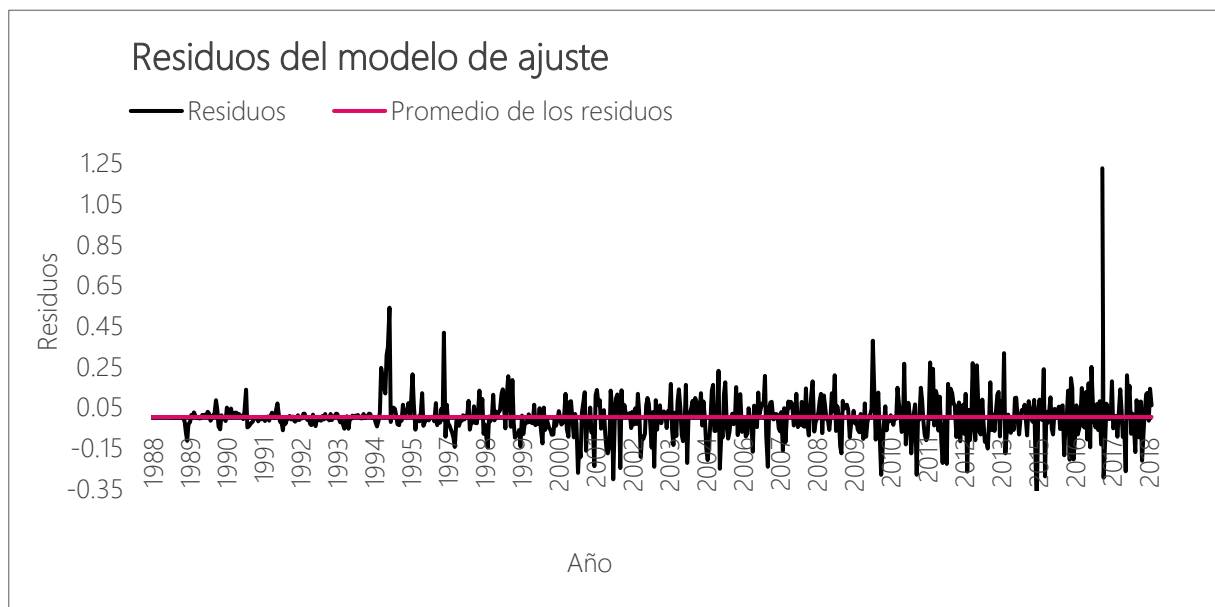
Gráfica 19. Q-Q Plot del modelo de ajuste $SARIMA(2,1,2) (1,1,2)_{24}$



Fuente: Elaboración propia en RStudio.

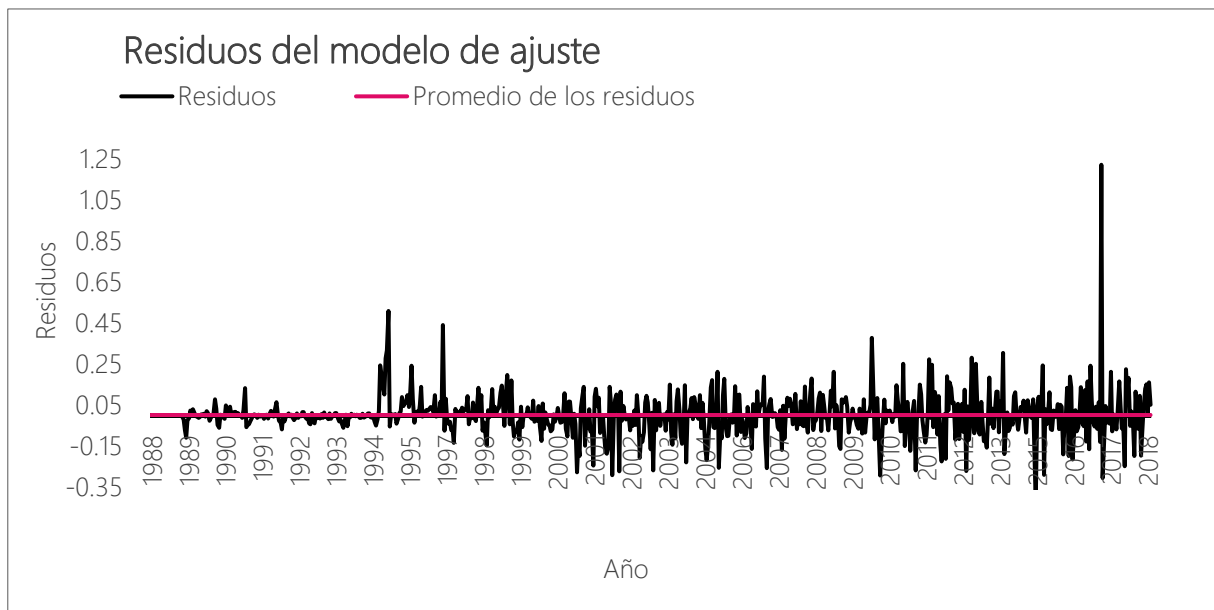
Las Gráficas 17, 18 y 19 muestran que los cuantiles muestrales están posicionados sobre la línea de los cuantiles teóricos.

Gráfica 20. Residuos del modelo de ajuste $SARIMA(3,1,1) (1,1,2)_{24}$



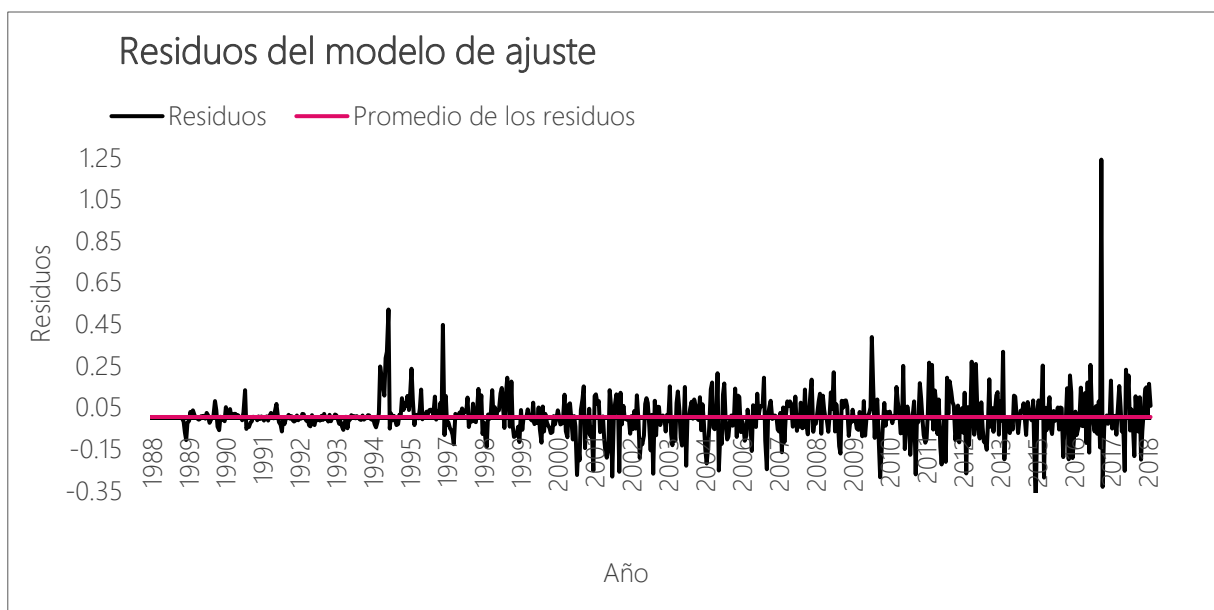
Fuente: Elaboración propia en Excel con datos del INPC.

Gráfica 21. Residuos del modelo de ajuste $SARIMA(1,1,2) (1,1,2)_{24}$



Fuente: Elaboración propia en Excel con datos del INPC.

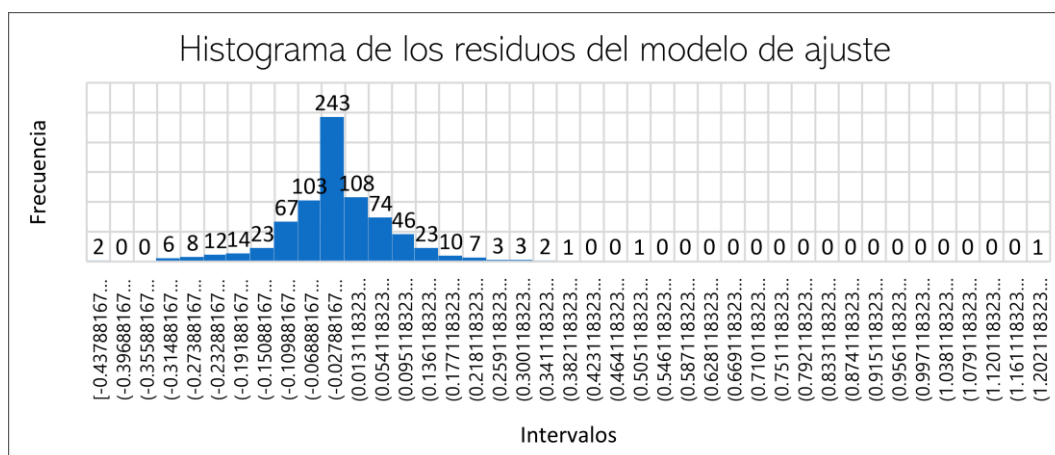
Gráfica 22. Residuos del modelo de ajuste $SARIMA(2,1,2) (1,1,2)_{24}$



Fuente: Elaboración propia en Excel con datos del INPC.

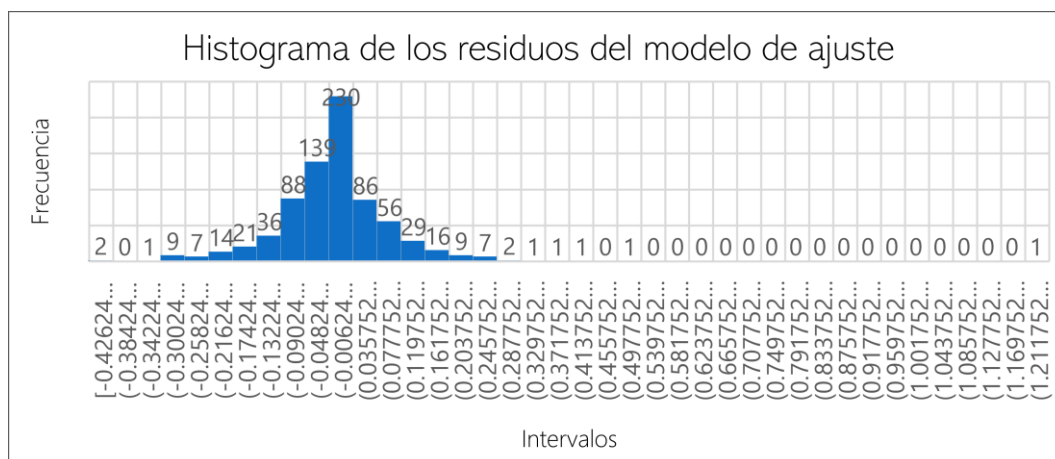
Las Gráficas 20, 21 y 22 muestran que los residuos de los tres modelos tienen una trayectoria alrededor del cero y su varianza es casi constante.

Gráfica 23. Histograma de los residuos del modelo de ajuste $SARIMA(3,1,1) (1,1,2)_{24}$



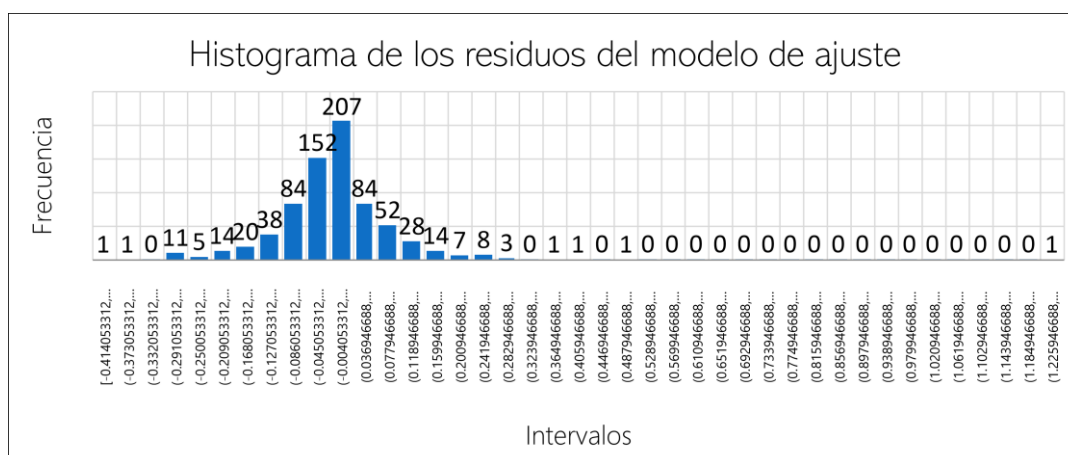
Fuente: Elaboración propia en Excel con datos del INPC.

Gráfica 24. Histograma de los residuos del modelo de ajuste $SARIMA(1,1,2) (1,1,2)_{24}$



Fuente: Elaboración propia en Excel con datos del INPC.

Gráfica 25. Histograma de los residuos del modelo de ajuste $SARIMA(2,1,2) (1,1,2)_{24}$



Fuente: Elaboración propia en Excel con datos del INPC.

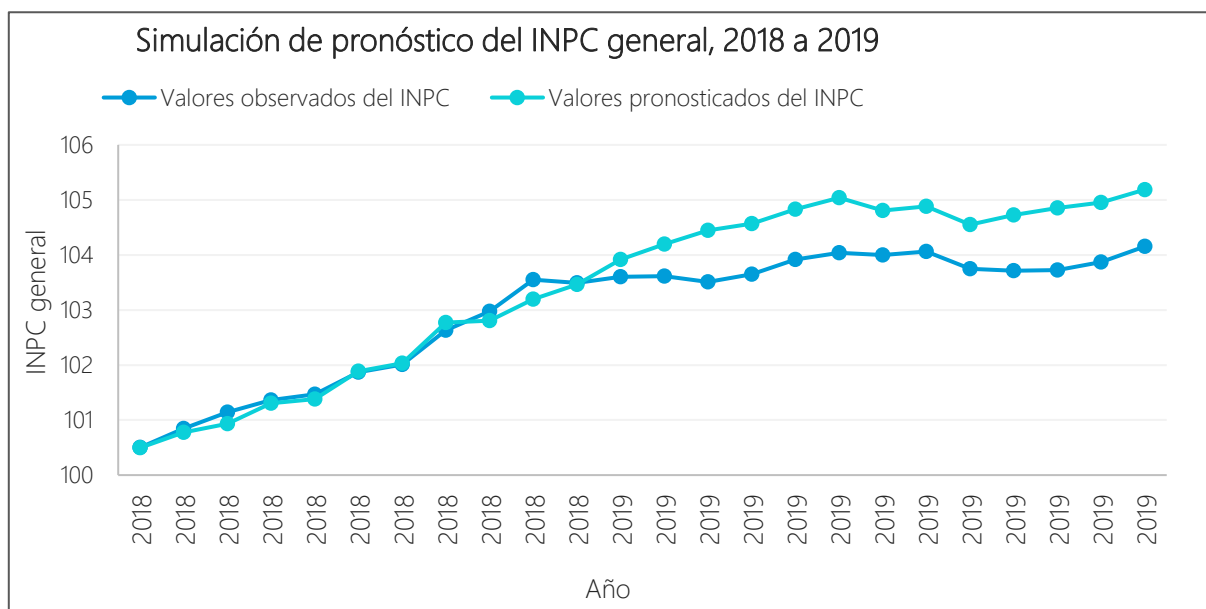
Las Gráficas 23, 24 y 25 muestran que los histogramas de los residuos presentan forma de distribución normal.

Del análisis de los residuos aplicado a los tres modelos de ajuste, se concluye que estos se distribuyen de manera normal.

Los residuos del primer modelo de ajuste presentan, además de normalidad, independencia, lo que lo hace ideal para esta metodología de pronóstico.

Se realizó el pronóstico de 24 quincenas del INPC, a partir de la quincena 14 de 2018 hasta la quincena 13 de 2019. Enseguida se muestra una simulación del pronóstico con el modelo de ajuste seleccionado: $SARIMA(3,1,1)(1,1,2)_{24}$.

Gráfica 26. Simulación de pronóstico del INPC general, 2018-2019

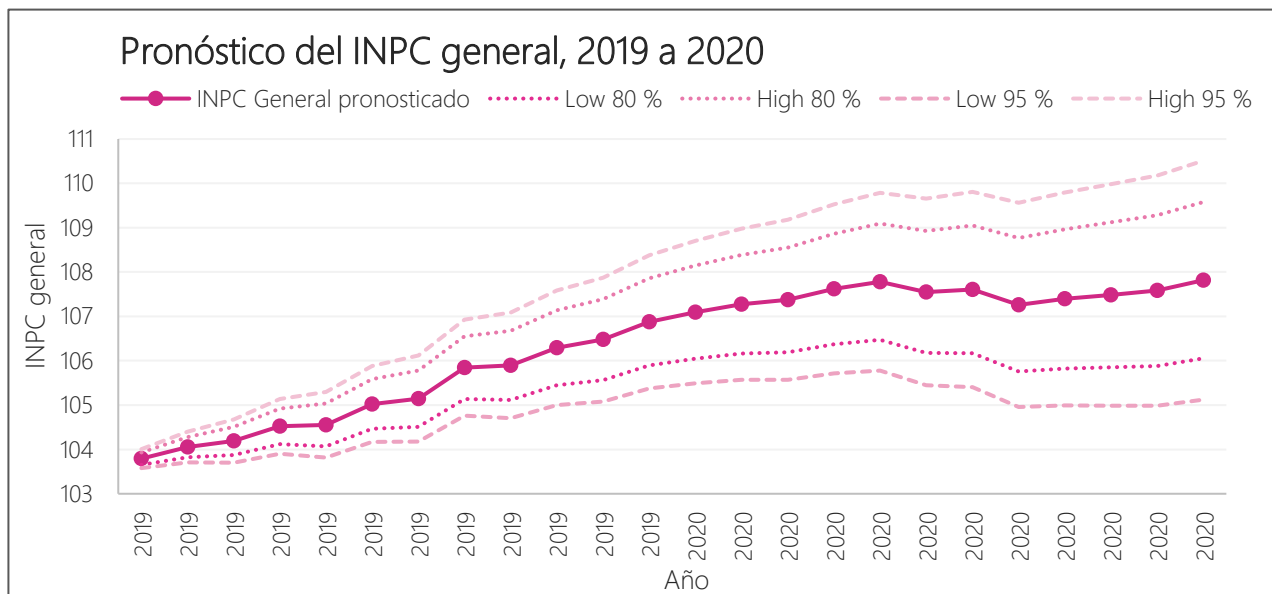


Fuente: Elaboración propia en Excel con datos del INPC observado y pronosticado.

La Gráfica 26 muestra que los valores observados del INPC y los valores pronosticados del mismo se ajustan bien hasta el inicio de la primera quincena de 2019; de ahí en adelante, el modelo sobreestima los valores pronosticados, sin embargo, tienen el mismo comportamiento que los valores observados.

5.4 Pronóstico del INPC general, 2019-2020

Gráfica 27. Pronóstico del INPC general, 2019-2020



Fuente: Elaboración propia en Excel con datos del INPC pronosticado.

Valores del pronóstico del INPC general

```
## Time Series:
## Start = c(2019, 14)
## End = c(2020, 13)
## Frequency = 24
## [1] 103.79 104.05 104.19 104.52 104.55 105.02 105.14 105.84 105.90
106.29
## [11] 106.48 106.88 107.10 107.27 107.37 107.62 107.78 107.55 107.61
107.26
## [21] 107.39 107.48 107.58 107.82
```

Valores de los coeficientes pertenecientes a los procesos autorregresivos y de promedios móviles, estacionales y no estacionales

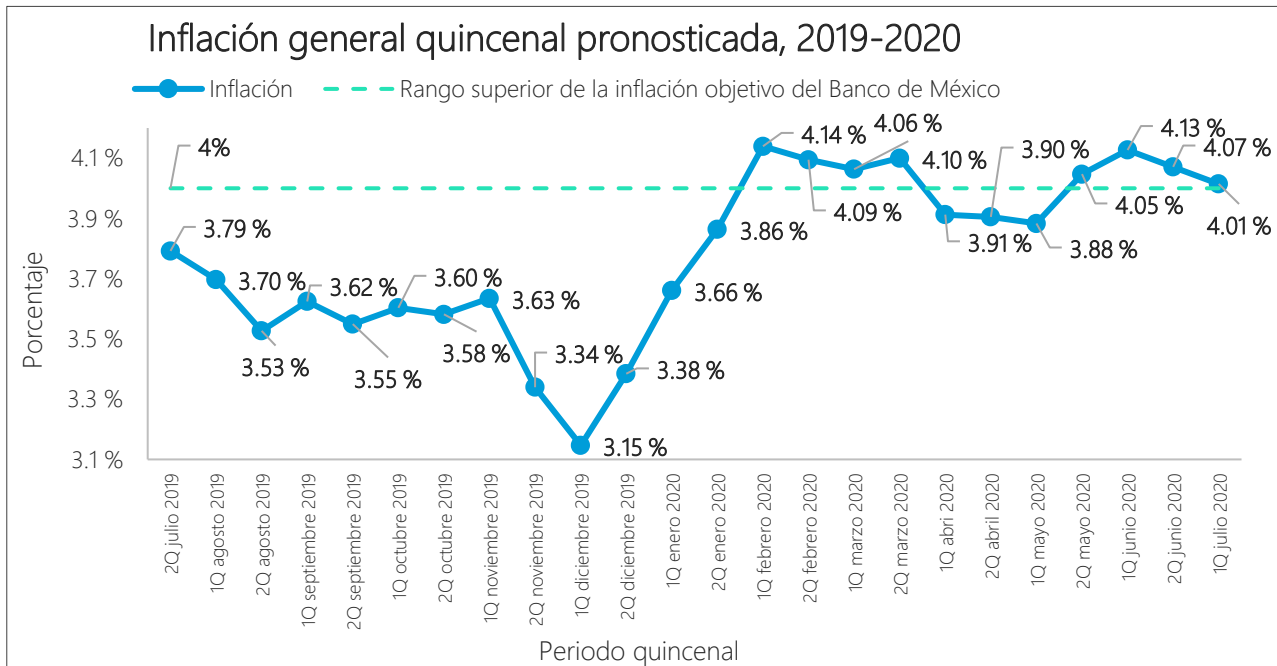
```
##          ar1          ar2          ar3          ma1          sar1          sma1          sma2
## [1,] 1.2114 -0.045444 -0.17817 -0.92436 0.55785 -1.4097 0.55199
```

5.5 Pronóstico de la inflación anualizada, 2019-2020

La inflación se calculó a partir de los valores pronosticados del INPC general, con base en la siguiente fórmula:

$$\text{Inflación} = \left(\frac{INPC_{1Q\ m\ t}}{INPC_{1Q\ m\ t-1}} - 1 \right) \times 100$$

Gráfica 28. Inflación general anualizada en México, 2019-2020



Fuente: Elaboración propia en Excel con datos del INPC pronosticado.

Notación del modelo $SARIMA(3,1,1) (1,1,2)_{24}$

$$\widehat{INPC}_{Ddt} = 1.2114INPC_{t-1} - 0.045444INPC_{t-2} - 0.17817INPC_{t-3} - 0.92436e_{t-1} + 0.55785INPC_{t-24} - 1.4097e_{t-24} + 0.55199e_{t-48} + u_t$$

donde

D = Diferencias estacionales

d = Diferencias simples

u_t = Error aleatorio

Conclusiones

La serie de tiempo presentó no estacionariedad en media y problema de raíz unitaria.

Los autocorrelogramas mostraron un comportamiento estacional, debido a ello, se utilizó un modelo *SARIMA* y, a través de él, se obtuvieron los órdenes *SARIMA*(2,1,2) (1,1,2)₂₄, generándose diversos modelos de ajuste para seleccionar aquel cuyos valores de criterios de información fueran los más bajos. Así, a partir de los resultados de los criterios de selección, se eligió un modelo de ajuste adicional con órdenes *SARIMA*(1,1,2) (1,1,2)₂₄, detectándose un comportamiento ascendente de los criterios de selección conforme aumentan los procesos *AR* y *MA*; sin embargo, solo se aumentaron dos órdenes al proceso autorregresivo y se disminuyó uno al proceso de promedios móviles, por lo que el modelo quedó de la siguiente forma: *SARIMA*(3,1,1) (1,1,2)₂₄.

Se compararon los tres modelos de ajuste *SARIMA* mencionados anteriormente, y solo uno de ellos presentó independencia, ausencia de autocorrelación en sus errores, varianza constante y media igual a cero, el modelo es: *SARIMA*(3,1,1) (1,1,2)₂₄. Según la medida de precisión *MAPE*, el pronóstico que realiza este modelo de ajuste está errado en 15.17 %, es decir, el pronóstico del *INPC* mediante este modelo se ajusta a los valores reales del *INPC* general con un 84.83 % de precisión.

La simulación muestra un adecuado ajuste de los valores pronosticados del *INPC* respecto a los valores reales de este; sin embargo, se observa una sobreestimación del pronóstico al iniciar el año 2019.

La inflación en México en 2019 se acercará gradualmente a 3 % (objetivo del Banco de México), pero se prevé una trayectoria de alrededor de 4 % (rango límite superior de la inflación objetivo), a partir de la primera quincena de febrero de 2020, hasta el término del periodo pronosticado.

En este análisis predictivo no hay variables que afectan el comportamiento de la inflación, es decir, este análisis es univariante, no se consideran factores políticos, sociales y económicos del país, ni externos, tampoco se toman en cuenta las devaluaciones del peso mexicano ni cambios estructurales en la serie de

tiempo derivados de cambios en la metodología de medición del *INPC*, es decir, los valores pronosticados del *INPC* general y, por ende, el comportamiento de la inflación general pronosticada puede no ser similar a la real a pesar de que el modelo de ajuste seleccionado tiene 84.83 % de precisión.

Con base en los planteamientos anteriores, es plausible que la metodología de Box y Jenkins pudiera considerarse de utilidad para pronosticar la inflación, por lo que podría tomarse en cuenta lo siguiente:

- Ampliar el modelo predictivo de univariante a multivariante.
- Analizar los efectos que provocan las devaluaciones del peso mexicano en la inflación.
- Analizar los cambios estructurales en la serie de tiempo derivados de cambios en la metodología de medición del *INPC*.
- Analizar los cambios estructurales en la serie de tiempo derivados de los objetivos de inflación seguidos por el Banco de México.

Para solucionar el problema de los cambios estructurales en la serie de tiempo del *INPC*, se recomienda: disminuir el tamaño de la muestra y comenzar a analizarla, ya sea a partir del momento en el que el Banco de México comenzó con el esquema de inflación objetivo o, a partir de un periodo posterior a la devaluación del peso mexicano más reciente, teniendo en cuenta que si la serie de tiempo contiene componentes estacionales, la muestra debe ser, por lo menos, de cinco a seis años, según lo señalado por el subgobernador del Banco de México Jonathan Heath, en su libro *Lo que indican los indicadores*: «Las estimaciones de los ajustes estacionales se basan en la experiencia económica de años anteriores. Por esto, es necesario que un indicador tenga, por lo menos, cinco o seis años de observaciones antes de poder realizar los ajustes necesarios. La ventaja principal es que se puede analizar mucho mejor el cambio en el margen (en el corto plazo) y así tomar decisiones más oportunas.» (Heath, 2012, pág. 19).

Anexos

Anexo 1. Código «R»²⁵ que se ejecuta en la consola de Rstudio para el pronóstico del INPC general. Modelo *Seasonal Autoregressive Integrated Moving Average SARIMA(p, d, q)(P, D, Q)_[S]*

■ El siguiente código que comienza a partir de la página 57 y termina en la página 76 de este documento, se ejecuta en la consola de RStudio®

```
# /
# /-----/
# / Script para pronosticar el INPC general /
# /-----/
# /

if (!"astsa" %in% rownames(installed.packages())) {
  install.packages("astsa")
}
if (!"tseries" %in% rownames(installed.packages())) {
  install.packages("tseries")
}
if (!"timeSeries" %in% rownames(installed.packages())) {
  install.packages("timeSeries")
}
if (!"fGarch" %in% rownames(installed.packages())) {
  install.packages("fGarch")
}
if (!"forecast" %in% rownames(installed.packages())) {
  install.packages("forecast")
}
if (!"dplyr" %in% rownames(installed.packages())) {
  install.packages("dplyr")
}
if (!"stats" %in% rownames(installed.packages())) {
  install.packages("stats")
}
if (!"zoo" %in% rownames(installed.packages())) {
  install.packages("zoo")
}
if (!"moments" %in% rownames(installed.packages())) {
  install.packages("moments")
}

ruta <- "F:/"
ruta_exp <- "F:../DATA"

setwd(ruta)
INPC_General <- read.csv('INPC General quincenal.csv', header = TRUE, dec = '.',
stringsAsFactors = F)
```

²⁵ Parte de este código en lenguaje «R» se programó basándose principalmente en cuatro fuentes: la Unidad de Enseñanza-Aprendizaje (UEA) *Modelos de Políticas Macroeconómicas I*, impartida por el profesor Jorge Antonio Ludlow Wiechers; un video en *YouTube* llamado *Series Temporales* subido en 2017 por el usuario Loren Araujo; un curso en línea de la plataforma *edX* llamado *Data Science: R Basics, Build a foundation in R and learn how to wrangle, analyze, and visualize data*, impartido en 2019 por el profesor de Bioestadística de la Universidad de Harvard Rafael Irizarry; y, finalmente, en la entrada de un *blog* sobre *Ciencia de datos* de la empresa *Oracle* que se titula *Introduction to Forecasting with ARIMA in R* escrito en 2017 por la científica de datos Ruslana Dalinina.

```

library(astsa)
library(tseries)
library(timeSeries)
library(fGarch)
library(stats)
library(zoo)
library(ggplot2)
library(moments)

### Particionando la serie de tiempo en dos conjuntos
## Conjunto 1 -- Base de entrenamiento (train)
## Conjunto 2 -- Base de prueba (test)
var_x <- INPC_General

longitud <- length(var_x[,2])
diferencias_1 <- 1
diferencias_1 <- diferencias_1 + 1
diferencias_2 <- 24
diferencias_2 <- diferencias_2 + 1
tail_diferencias_2 <- longitud - 24

periodo_1diff <- var_x[diferencias_1:tail_diferencias_2,1]
periodo_diffOrd <- var_x[diferencias_2:tail_diferencias_2,1]

longitud_x <- length(var_x[,2])
numTrain_x <- longitud_x - 24

train_x <- var_x[1:numTrain_x,]
test_x <- tail(var_x, 24) # Valores reales de la serie
# que van a ser contrastados con los pronosticados.

var_x_contraste <- train_x # Asignando valores del entrenamiento a la variable
# x que será contrastada.
var_x_contraste <- ts(var_x_contraste, start = c(1988, 1), frequency = 24)
var_x_contraste <- var_x_contraste[, -1]
var_x_contraste <- zoo::na.approx(var_x_contraste)

#-----
# CALCULANDO LA PERIODICIDAD DE LA SERIE DE TIEMPO
#-----
#-----
# CALCULANDO PERIODICIDAD QUINCENAL
#-----
num_char_x <- (var_x[1,1])
num_char_x <- nchar(num_char_x)

Q1 <- substr(var_x[1, 1], 1, 2)
Q3 <- substr(var_x[3, 1], 1, 2)

num_char_x_prob_Q <- if (num_char_x == 10) {
  print("Q")
} else {
  print("NQ")
}

quin_x <- if (Q1 == Q3) {
  print("Q")
}

quincenal_x <- if (num_char_x_prob_Q == quin_x) {
  print("Serie quincenal")
  mes_x <- 1
  frec_x <- 24
  frec_x_chunk <- 24
  anhos_x <- substr(var_x[, 1], 7, 10)
  anhos_x_chunk <- substr(var_x[, 1], 7, 10)
}

```

```

anho.min_x <- min(as.numeric(anhos_x))
anho.min_x_chunk <- min(as.numeric(anhos_x_chunk))

var_x <- ts(var_x, start = c(anho.min_x), frequency = frec_x)
var_x <- var_x[, -1]

} else {
  print("Serie no quincenal")
  #-----
  # CALCULANDO PERIODICIDAD DIFERENTE A UNA QUINCENA
  #-----
  anho.min_x_chunk <- min(anhos_x_chunk)
  frec_x_chunk <- as.data.frame(anhos_x_chunk )$anhos_x_chunk == anho.min_x_chunk + 1
  frec_x_chunk <- sum(frec_x_chunk == TRUE)

  longitud_x <- length(var_x[,2])
  numTrain_x <- longitud_x - frec_x_chunk

  train_x <- var_x[1:numTrain_x,]
  test_x <- tail(var_x, frec_x_chunk) # Valores reales de la serie
  # que van a ser contrastados con los pronosticados.

  var_x <- train_x # Asignando valores del entrenamiento a la variable x

  anhos <- as.numeric(substr(var_x[, 1], 1, 4))
  anho.min_x <- min(anhos)
  anho <- anho.min_x
  anho.max_x <- max(anhos)
  mes_x <- cbind(as.numeric(gsub("(.*)/", "", var_x[, 1])))
  mes_x <- mes_x[1,]
  frec_x <- as.data.frame(anhos)$anhos == anho.min_x + 1
  frec_x <- sum(frec_x == TRUE)
  #-----
  # CALCULANDO PERIODICIDAD ANUAL
  #-----
  list <- as.list(var_x)
  list <- as.data.frame(list[1])
  list_long_x <- lengths(strsplit(as.character(list[1,1]), ''))

  var_x <- if (list_long_x == 4) {
    var_x <- ts(var_x, start = c(anho.min_x), end = c(anho.max_x))
    var_x <- var_x[, -1]
  } else {
    var_x <- ts(var_x, start = c(anho.min_x, mes_x), frequency = frec_x)
    var_x <- var_x[, -1]
  }
}

# Creando objeto de serie de tiempo
var_x <- ts(var_x, start = c(frec_x, mes_x), frequency = frec_x)
var_x <- var_x[, -1]
var_x <- zoo::na.approx(var_x)

# Exportar datos
setwd(ruta_exp)

write.csv(periodo_1diff, "primeras_diff_xyz.csv")
write.csv(periodo_diffOrd, "diff_orden_xyz.csv")

test <- test_x[,1]
write.csv(test, "Periodo_simulac_xyz.csv")

setwd(ruta)

```

```

### Graficando INPC General de México de 1988 a 2019.

plot(var_x, type = 'l', main = 'INPC General de México de 1988 a 2019', sub =
'Fuente: INEGI', ylab= 'INPC General', xlab='Año')

test.raiz <- adf.test(var_x)
options(digits=2)
if (test.raiz$p.value < 0.05)
{
  valor <- test.raiz$p.value
  cat("No hay Raíz unitaria (serie de tiempo estacionaria) el valor p es: ", valor,
      "\npor lo tanto, se rechaza la H0 de no estacionariedad.",
      "\nSe puede entonces calcular el modelo de ajuste.")
} else {
  valor <- test.raiz$p.value
  cat("Raíz unitaria (serie de tiempo no estacionaria) el valor p es: " , valor,

      "\nDebido a que la serie no es estacionaria entonces se estabiliza la serie
de tiempo",
      "\nantes de estimar el modelo.");
  n_diff_x <- forecast::ndiffs(var_x, test = c("adf"))
  var_diff_x <- diff(var_x, n_diff_x)
}
}
#### Eliminando la no estacionariedad en media de la serie de tiempo mediante el
cálculo de las primeras diferencias

# Exportar datos
setwd(ruta_exp)

write.csv(var_diff_x, "var_diff_log_x.csv")
write.csv(train_x , "train.csv")

setwd(ruta)

plot(var_diff_x, type = 'l', main = 'INPC General de México de 1988 a 2019', sub =
'Fuente: INEGI', ylab= 'Primeras diferencias del INPC', xlab='Año')
abline(h = mean(var_diff_x), col = 'blue')

### Se vuelve a aplicar la prueba; Augmented Dickey-Fuller (adf.test) para comprobar
que la serie es estacionaria

test.raiz <- adf.test(var_diff_x)
options(digits=2)
if (test.raiz$p.value < 0.05)
{
  valor <- test.raiz$p.value
  cat("Serie de tiempo estacionaria el valor p es: ", valor,
      "
por lo tanto, se rechaza la H0 de no estacionariedad.
Se puede entonces calcular el modelo de ajuste")
} else {
  valor <- test.raiz$p.value
  cat("Serie de tiempo no estacionaria.
",
      "Debido a que la serie sigue presentando no estacionariedad",
      "se calculan las primeras diferencias de la serie",
      "junto con el rezago de orden 24, el valor p es el siguiente: ")
  diff_orden <- diff(var_x, lag = 24)
  test.raiz <- adf.test(diff_orden)
  options(digits=2)
  if (test.raiz$p.value < 0.05)
  {
    valor <- test.raiz$p.value
    cat(valor,
        "

```

```

por lo tanto, la serie es estacionaria.
Se puede entonces calcular el modelo de ajuste")
    } else {
        cat("Serie de tiempo no estacionaria.")
    }
}
options(digits=5)

## 1. Identificación

### Gráficas de la función de autocorrelación (ACF) y de la función de autocorrelación parcial (PACF) para obtener los patrones del proceso AR() y MA().

acf2(ts(var_diff_x, frequency = 1))

var_x_dif_orden <- diff(diff(var_x), lag = 24) ### Diferencias estacionales

# Exportar datos
setwd(ruta_exp)

write.csv(var_x_dif_orden, "var_x_dif_orden_x.csv")

plot(var_x_dif_orden, type = 'l', main = 'INPC General de México de 1988 a 2019',
sub = 'Fuente: INEGI', ylab= 'Diferencias estacionales del INPC', xlab='Año')
abline(h = mean(diff(var_x,lag = 24)),col = 'blue')

setwd(ruta)

### Gráficas de la función de autocorrelación (ACF) y de la función de autocorrelación parcial (PACF) para obtener los patrones del proceso AR() y MA() del componente estacional.

acf2(ts(var_x_dif_orden, frequency = 1))

#### Criterios AIC y BIC de los modelos de ajuste

INPC <- var_x

ajuste_INPC_1 <- arima(INPC, order=c(0,1,0), seasonal = list(order = c(1,1,2), period = 24),method="ML", include.mean = TRUE)
ajuste_INPC_2 <- arima(INPC, order=c(1,1,0), seasonal = list(order = c(1,1,2), period = 24),method="ML", include.mean = TRUE)
ajuste_INPC_3 <- arima(INPC, order=c(0,1,1), seasonal = list(order = c(1,1,2), period = 24),method="ML", include.mean = TRUE)
ajuste_INPC_4 <- arima(INPC, order=c(1,1,1), seasonal = list(order = c(1,1,2), period = 24),method="ML", include.mean = TRUE)
ajuste_INPC_5 <- arima(INPC, order=c(0,1,2), seasonal = list(order = c(1,1,2), period = 24),method="ML", include.mean = TRUE)
ajuste_INPC_6 <- arima(INPC, order=c(2,1,0), seasonal = list(order = c(1,1,2), period = 24),method="ML", include.mean = TRUE)
ajuste_INPC_7 <- arima(INPC, order=c(1,1,2), seasonal = list(order = c(1,1,2), period = 24),method="ML", include.mean = TRUE)
ajuste_INPC_8 <- arima(INPC, order=c(2,1,1), seasonal = list(order = c(1,1,2), period = 24),method="ML", include.mean = TRUE)
ajuste_INPC_9 <- arima(INPC, order=c(2,1,2), seasonal = list(order = c(1,1,2), period = 24),method="ML", include.mean = TRUE)

Modelo_Ajustado <-
c("ajuste_INPC_1","ajuste_INPC_2","ajuste_INPC_3","ajuste_INPC_4","ajuste_INPC_5","ajuste_INPC_6","ajuste_INPC_7","ajuste_INPC_8", "ajuste_INPC_9")
Patrones_pdq_y_PDQ <- c("(0,1,0) (1,1,2)", "(1,1,0) (1,1,2)", "(0,1,1) (1,1,2)", "(1,1,1) (1,1,2)", "(0,1,2) (1,1,2)", "(2,1,0) (1,1,2)", "(1,1,2) (1,1,2)", "(2,1,1) (1,1,2)", "(2,1,2) (1,1,2)")
AIC <- c(ajuste_INPC_1$aic, ajuste_INPC_2$aic, ajuste_INPC_3$aic,
ajuste_INPC_4$aic,ajuste_INPC_5$aic, ajuste_INPC_6$aic,
ajuste_INPC_7$aic,ajuste_INPC_8$aic, ajuste_INPC_9$aic)

```

```

BIC <- c(BIC(ajuste_INPC_1), BIC(ajuste_INPC_2), BIC(ajuste_INPC_3),
BIC(ajuste_INPC_4), BIC(ajuste_INPC_5), BIC(ajuste_INPC_6),
      BIC(ajuste_INPC_7), BIC(ajuste_INPC_8), BIC(ajuste_INPC_9))

Summary_stat <- cbind(Patrones_pdq_y_PDQ, AIC, BIC)
Summary_stat

# Exportar datos
setwd(ruta_exp)
write.csv(Summary_stat, "AIC_BIC.csv")
setwd(ruta)

Modelo_Ajustado <- as.data.frame(Modelo_Ajustado)

AIC <- as.data.frame(AIC)
BIC <- as.data.frame(BIC)

Summary_stat <- cbind(Modelo_Ajustado, AIC, BIC)

# AIC
index_Summary_stat_Akaike_bajo <- which(Summary_stat[,2] == min(Summary_stat[,2]))

Akaike_mas_bajo <- cbind(Summary_stat[index_Summary_stat_Akaike_bajo,1])
Ajuste_Akaike_mas_bajo <- cbind(Summary_stat[index_Summary_stat_Akaike_bajo,2])

# BIC
index_Summary_stat_BIC_bajo <- which(Summary_stat[,3] == min(Summary_stat[,3]))

BIC_mas_bajo <- cbind(Summary_stat[index_Summary_stat_BIC_bajo,1])
Ajuste_BIC_mas_bajo <- cbind(Summary_stat[index_Summary_stat_BIC_bajo,3])

cat("Modelo de ajuste con valor Akaike más bajo:", Akaike_mas_bajo[1,1])
cat("Valor Akaike:", Ajuste_Akaike_mas_bajo[1,1])

cat("Modelo de ajuste con valor BIC más bajo:", BIC_mas_bajo[1,1])
cat("Valor BIC:", Ajuste_BIC_mas_bajo[1,1])

#### Medidas de precisión de los modelos de ajuste

accuracy_fit_INPC_1 <- forecast::accuracy(ajuste_INPC_1)
accuracy_fit_INPC_2 <- forecast::accuracy(ajuste_INPC_2)
accuracy_fit_INPC_3 <- forecast::accuracy(ajuste_INPC_3)
accuracy_fit_INPC_4 <- forecast::accuracy(ajuste_INPC_4)
accuracy_fit_INPC_5 <- forecast::accuracy(ajuste_INPC_5)
accuracy_fit_INPC_6 <- forecast::accuracy(ajuste_INPC_6)
accuracy_fit_INPC_7 <- forecast::accuracy(ajuste_INPC_7)
accuracy_fit_INPC_8 <- forecast::accuracy(ajuste_INPC_8)
accuracy_fit_INPC_9 <- forecast::accuracy(ajuste_INPC_9)

MAPE <- c(accuracy_fit_INPC_1[5], accuracy_fit_INPC_2[5], accuracy_fit_INPC_3[5],
          accuracy_fit_INPC_4[5], accuracy_fit_INPC_5[5], accuracy_fit_INPC_6[5],
          accuracy_fit_INPC_7[5], accuracy_fit_INPC_8[5], accuracy_fit_INPC_9[5])
ME <- c(accuracy_fit_INPC_1[1], accuracy_fit_INPC_2[1], accuracy_fit_INPC_3[1],
        accuracy_fit_INPC_4[1], accuracy_fit_INPC_5[1], accuracy_fit_INPC_6[1],
        accuracy_fit_INPC_7[1], accuracy_fit_INPC_8[1], accuracy_fit_INPC_9[1])
RMSE <- c(accuracy_fit_INPC_1[2], accuracy_fit_INPC_2[2], accuracy_fit_INPC_3[2],
          accuracy_fit_INPC_4[2], accuracy_fit_INPC_5[2], accuracy_fit_INPC_6[2],
          accuracy_fit_INPC_7[2], accuracy_fit_INPC_8[2], accuracy_fit_INPC_9[2])
MAE <- c(accuracy_fit_INPC_1[3], accuracy_fit_INPC_2[3], accuracy_fit_INPC_3[3],
          accuracy_fit_INPC_4[3], accuracy_fit_INPC_5[3], accuracy_fit_INPC_6[3],
          accuracy_fit_INPC_7[3], accuracy_fit_INPC_8[3], accuracy_fit_INPC_9[3])
MPE <- c(accuracy_fit_INPC_1[4], accuracy_fit_INPC_2[4], accuracy_fit_INPC_3[4],
          accuracy_fit_INPC_4[4], accuracy_fit_INPC_5[4], accuracy_fit_INPC_6[4],

```

```

accuracy_fit_INPC_7[4],accuracy_fit_INPC_8[4],accuracy_fit_INPC_9[4])
MASE <- c(accuracy_fit_INPC_1[6],accuracy_fit_INPC_2[6],accuracy_fit_INPC_3[6],
accuracy_fit_INPC_4[6],accuracy_fit_INPC_5[6],accuracy_fit_INPC_6[6],
accuracy_fit_INPC_7[6],accuracy_fit_INPC_8[6],accuracy_fit_INPC_9[6])

Summary_accuracy <- cbind(MAPE,ME,RMSE,MAE,MPE,MASE)
Summary_accuracy

# Exportar datos
setwd(ruta_exp)
write.csv(Summary_accuracy, "Medidas_de_precisión.csv")
setwd(ruta)

Summary_accuracy <- cbind(Modelo_Ajustado,MAPE,ME,RMSE,MAE,MPE,MASE)

# Calculando índices para obtener los valores más bajos del resumen estadístico de
precisión de los modelos de ajuste

index_MAPE <- which(Summary_accuracy[,2] == min(Summary_accuracy[,2]))

MAPE_mas_bajo <- cbind(Summary_accuracy[index_MAPE ,1])
Valor_MAPE_mas_bajo <- cbind(Summary_accuracy[index_MAPE,2])

cat("Modelo de ajuste con MAPE más bajo:", MAPE_mas_bajo[1,1])
cat("Valor MAPE:", Valor_MAPE_mas_bajo[1,1])

index_ME <- which(Summary_accuracy[,3] == min(Summary_accuracy[,3]))

ME_mas_bajo <- cbind(Summary_accuracy[index_ME ,1])
Valor_ME_mas_bajo <- cbind(Summary_accuracy[index_ME,3])

cat("Modelo de ajuste con ME más bajo:", ME_mas_bajo[1,1])
cat("Valor ME:", Valor_ME_mas_bajo[1,1])

index_RMSE <- which(Summary_accuracy[,4] == min(Summary_accuracy[,4]))

RMSE_mas_bajo <- cbind(Summary_accuracy[index_RMSE ,1])
Valor_RMSE_mas_bajo <- cbind(Summary_accuracy[index_RMSE,4])

cat("Modelo de ajuste con RMSE más bajo:", RMSE_mas_bajo[1,1])
cat("Valor RMSE:", Valor_RMSE_mas_bajo[1,1])

index_MAE <- which(Summary_accuracy[,5] == min(Summary_accuracy[,5]))

MAE_mas_bajo <- cbind(Summary_accuracy[index_MAE ,1])
Valor_MAE_mas_bajo <- cbind(Summary_accuracy[index_MAE,5])

cat("Modelo de ajuste con MAE más bajo:", MAE_mas_bajo[1,1])
cat("Valor MAE:", Valor_MAE_mas_bajo[1,1])

index_MPE <- which(Summary_accuracy[,6] == min(Summary_accuracy[,6]))

MPE_mas_bajo <- cbind(Summary_accuracy[index_MPE ,1])
Valor_MPE_mas_bajo <- cbind(Summary_accuracy[index_MPE,6])

cat("Modelo de ajuste con MPE más bajo:", MPE_mas_bajo[1,1])
cat("Valor MPE:", Valor_MPE_mas_bajo[1,1])

index_MASE <- which(Summary_accuracy[,7] == min(Summary_accuracy[,7]))

MASE_mas_bajo <- cbind(Summary_accuracy[index_MASE,1])

```



```

Valor_MASE_mas_bajo <- cbind(Summary_accuracy[index_MASE,7])

cat("Modelo de ajuste con MASE más bajo:", MASE_mas_bajo[1,1])
cat("Valor MASE:", Valor_MASE_mas_bajo[1,1])

INPC <- var_x
ajuste_x <- arima(INPC, order=c(1,1,1), seasonal = list(order = c(1,1,2), period =
24),method="ML", include.mean = TRUE)
tsdiag(ajuste_x)
INPC_General <- var_x

# Exportar datos

setwd(ruta_exp)
h_x <- forecast::forecast(ajuste_x)
h_x <- as.data.frame(h_x)
write.csv(h_x, "h_arima_x.csv")
setwd(ruta)

qqnorm(ajuste_x$resid, main = "Normal Q-Q Plot",
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles")
qqline(ajuste_x$resid)

#### Los valores de los cuantiles están posicionados sobre la línea.

resid_ajuste_x <- ajuste_x$residuals
plot(resid_ajuste_x, main = "Graficando los residuos del ajuste del modelo.", ylab =
"Residuos", type = "o")
abline(h = mean(resid_ajuste_x), col = "blue")

hist(resid_ajuste_x, main = "Histograma de los residuos del ajuste del modelo.", col=
"black", xlab = "Residuos")

# Exportar datos
setwd(ruta_exp)

write.csv(resid_ajuste_x, "residuos_x.csv")

setwd(ruta)

#### Resumen estadístico del ajuste del modelo

cat("Sigma Cuadrado:", ajuste_x$sigma2)
cat("Media de los residuos:", mean(resid_ajuste_x))

cat("Akaike:", ajuste_x$aic)

### Simulación de pronóstico: INPC General.

INPC_General <- var_x_contraste
pronostico <- sarima.for(INPC_General, n.ahead = 24, p = 1, d =1, q = 1, P = 1, D =
1, Q = 2, S = 24, no.constant = FALSE)
values.predict_x <- pronostico$pred

##### Línea **roja** valores observados de la serie de tiempo
##### Línea **azul** valores pronosticados de la serie de tiempo

Periodo_x <- gsub("/(*)", " ", test_x[,1])
Observed_x <- test_x[,2]
Forecasted_x <- values.predict_x

simulation_x <- data.frame(Periodo_x, Observed_x, Forecasted_x)

simulation_plot_x <- ggplot(data= simulation_x, aes(x=Periodo_x, y=Observed_x,
group=1)) +
  geom_line(col="red2", size = 1.1) +

```

```

    geom_point(col="red2", size = 3) +
    geom_line(col="deeppink2", size = 3.5, alpha = 0.18) +
    labs(title = "Valores observados y valores pronosticados.", x = "Fecha", y = "INPC
General")+
    geom_line(col = "Cyan4", size = 1.1, aes(y=Forecasted_x)) +
    geom_point(col = "Cyan4", size = 3, aes(y=Forecasted_x)) +
    geom_line(col="Cyan", size = 3.5, alpha = 0.18, aes(y=Forecasted_x))
simulation_plot_x

# Exportar datos
setwd(ruta_exp)

write.csv(Observed_x, "Observed_x.csv")
write.csv(Forecasted_x, "Forecasted_x.csv")

setwd(ruta)

### **Pronóstico de 24 quincenas para el INPC General de México.**

setwd(ruta)
INPC_General <- read.csv('INPC General quincenal.csv', header = T, dec = '.', string-
sAsFactors = F)
var_x <- INPC_General

var_x <- ts(var_x, start = c(1988, 1), frequency = 24)
var_x <- var_x[, -1]
var_x <- zoo::na.approx(var_x)

INPC_General <- var_x

pronostico_x <- sarima.for(INPC_General, n.ahead = 24, p = 1, d = 1, q = 1, P = 1, D =
1, Q = 2, S = 24, no.constant = FALSE)

### Valor de los predictores

(rbind(ajuste_x$coef))

### Valores de la predicción

(pronostico_x$pred)

# Exportar datos
setwd(ruta_exp)

write.csv(pronostico_x$pred, "Hforecasted_x.csv")

setwd(ruta)

### Modelo II

setwd(ruta)
INPC_General <- read.csv('INPC General quincenal.csv', header = T, dec = '.', string-
sAsFactors = F)

### Particionando la serie de tiempo en dos conjuntos
## Conjunto 1 -- Base de entrenamiento (train)
## Conjunto 2 -- Base de prueba (test)
var_y <- INPC_General

longitud_y <- length(var_y[,2])
numTrain_y <- longitud_y - 24

train_y <- var_y[1:numTrain_y,]

```

```

test_y <- tail(var_y, 24) # Valores reales de la serie que van a ser contrastados
con los pronosticados.

var_y_contraste <- train_y # Asignando valores del entrenamiento a la variable y
que será contrastada
var_y_contraste <- ts(var_y_contraste, start = c(1988, 1), frequency = 24)
var_y_contraste <- var_y_contraste[, -1]
var_y_contraste <- zoo::na.approx(var_y_contraste)

var_y <- ts(var_y, start = c(1988, 1), frequency = 24)
var_y <- var_y[, - 1]
var_y <- zoo::na.approx(var_y)

#### El segundo modelo contiene el siguiente patrón **AR(2)I(1)MA(2)** junto con el
patrón de residuos (P,D,Q): **AR(1)I(1)MA(2)** . Se deduce el siguiente modelo de
ajuste:

##### **ajusteARIMA <- arima(INPC_General, order=c(2,1,2), seasonal = list(order =
c(1,1,2), period = 24), method="ML", include.mean = TRUE)**

INPC <- var_y
ajuste_y <- arima(INPC, order=c(2,1,2), seasonal = list(order = c(1,1,2), period =
24),method="ML", include.mean = TRUE)
tsdiag(ajuste_y)
INPC_General <- var_y

# Exportar datos

setwd(ruta_exp)
h_y <- forecast::forecast(ajuste_y)
h_y <- as.data.frame(h_y)
write.csv(h_y, "h_arima_y.csv")
setwd(ruta)

#### Graficando los cuantiles de la distribución normal de los residuos del modelo
ajustado.

qqnorm(ajuste_y$resid, main = "Normal Q-Q Plot",
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles")
qqline(ajuste_y$resid)

#### Los valores de los cuantiles están posicionados sobre la línea.

resid_ajuste_y <- ajuste_y$residuals
plot(resid_ajuste_y, main = "Graficando los residuos del ajuste del modelo.", ylab =
"Residuos", type = "o")
abline(h = mean(resid_ajuste_y), col = "blue")

hist(resid_ajuste_y, main = "Histograma de los residuos del ajuste del modelo.", col=
"black", xlab = "Residuos")

# Exportar datos
setwd(ruta_exp)

write.csv(resid_ajuste_y, "residuos_y.csv")

setwd(ruta)

#### Resumen estadístico del ajuste del modelo

cat("Sigma Cuadrado:", ajuste_y$sigma2)
cat("Media de los residuos:", mean(resid_ajuste_y))

cat("Akaike:", ajuste_y$aic)

#### Simulación de pronóstico: INPC General.

```

```

INPC_General <- var_y_contraste
pronostico <- sarima.for(INPC_General, n.ahead = 24, p = 2, d =1, q = 2, P = 1, D =
1, Q = 2, S = 24, no.constant = FALSE)

values.predict_y <- pronostico$pred

##### Línea **roja** valores observados de la serie de tiempo
##### Línea **azul** valores pronosticados de la serie de tiempo

Periodo_y <- gsub("/(*)", " ", test_y[,1])
Observed_y <- test_y[,2]
Forecasted_y <- values.predict_y

simulation_y <- data.frame(Periodo_y, Observed_y, Forecasted_y)

simulation_plot_y <- ggplot(data= simulation_y, aes(x=Periodo_y, y=Observed_y,
group=1)) +
  geom_line(col="red2", size = 1.1) +
  geom_point(col="red2", size = 3)+
  geom_line(col="deeppink2", size = 3.5, alpha = 0.18) +
  labs(title = "Valores observados y valores pronosticados.",x = "Fecha", y = "INPC
General") +
  geom_line(col = "Cyan4", size = 1.1, aes(y=Forecasted_y)) +
  geom_point(col = "Cyan4", size = 3, aes(y=Forecasted_y)) +
  geom_line(col="Cyan", size = 3.5, alpha = 0.18,aes(y=Forecasted_y))
simulation_plot_y

# Exportar datos
setwd(ruta_exp)

write.csv(Observed_y, "Observed_y.csv")
write.csv(Forecasted_y, "Forecasted_y.csv")

setwd(ruta)

### **Pronóstico de 24 quincenas para el INPC General de México.**

setwd(ruta)
INPC_General <- read.csv('INPC General quincenal.csv', header = T, dec = '.', string-
sAsFactors = F)
var_y <- INPC_General

  var_y <- ts(var_y, start = c(1988, 1), frequency = 24)
  var_y <- var_y[, - 1]
var_y <- zoo::na.approx(var_y)

INPC_General <- var_y

pronostico_y <- sarima.for(INPC_General, n.ahead = 24, p = 2, d =1, q = 2, P = 1, D =
1, Q = 2, S = 24, no.constant = FALSE)

### Valor de los predictores

(rbind(ajuste_y$coef))

### Valores de la predicción

(pronostico_y$pred)

# Exportar datos
setwd(ruta_exp)

write.csv(pronostico_y$pred, "Hforecasted_y.csv")

setwd(ruta)

```

```

### Modelo III

setwd(ruta)
INPC_General <- read.csv('INPC General quincenal.csv', header = T, dec = '.', string-
sAsFactors = F)

### Particionando la serie de tiempo en dos conjuntos
## Conjunto 1 -- Base de entrenamiento (train)
## Conjunto 2 -- Base de prueba (test)
var_z <- INPC_General

longitud_z <- length(var_z[,2])
numTrain_z <- longitud_z - 24

train_z <- var_z[1:numTrain_z,]
test_z <- tail(var_z, 24) # Valores reales de la serie
                           # que van a ser contrastados con los pronosticados.

var_z_contraste <- train_z # Asignando valores del entrenamiento a la variable z
que será contrastada
var_z_contraste <- ts(var_z_contraste, start = c(1988, 1), frequency = 24)
var_z_contraste <- var_z_contraste[, -1]
var_z_contraste <- zoo::na.approx(var_z_contraste)

var_z <- ts(var_z, start = c(1988, 1), frequency = 24)
var_z <- var_z[, - 1]
var_z <- zoo::na.approx(var_z)

### El tercer modelo contiene el siguiente patrón **AR(3)I(1)MA(1)** junto con el pa-
trón de residuos (P,D,Q): **AR(1)I(1)MA(2)** . Se deduce el siguiente modelo de
ajuste:

##### **ajusteARIMA <- arima(INPC_General, order=c(3,1,1), seasonal = list(order =
c(1,1,2), period = 24), method="ML", include.mean = TRUE)**

INPC <- var_z
ajuste_z <- arima(INPC, order=c(3,1,1), seasonal = list(order = c(1,1,2), period =
24),method="ML", include.mean = TRUE)
tsdiag(ajuste_z)
INPC_General <- var_z

# Exportar datos

setwd(ruta_exp)
h_z <- forecast::forecast(ajuste_z)
h_z <- as.data.frame(h_z)
write.csv(h_z, "h_arima_z.csv")
setwd(ruta)

### Graficando los cuantiles de la distribución normal de los residuos del modelo
ajustado.

qqnorm(ajuste_z$resid, main = "Normal Q-Q Plot",
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles")
qqline(ajuste_z$resid)

#### Los valores de los cuantiles están posicionados sobre la línea.

resid_ajuste_z <- ajuste_z$residuals
plot(resid_ajuste_z, main = "Graficando los residuos del ajuste del modelo.", ylab =
"Residuos", type = "o")
abline(h = mean(resid_ajuste_z), col = "blue")

hist(resid_ajuste_z, main = "Histograma de los residuos del ajuste del modelo.", col=
"black", xlab = "Residuos")

```

```

# Exportar datos
setwd(ruta_exp)

write.csv(resid_ajuste_z, "residuos_z.csv")

setwd(ruta)

#### Resumen estadístico del ajuste del modelo

cat("Sigma Cuadrado:", ajuste_z$sigma2)
cat("Media de los residuos:", mean(resid_ajuste_z))

cat("Akaike:", ajuste_z$aic)

### Simulación de pronóstico: INPC General.

INPC_General <- var_z_contraste
pronostico <- sarima.for(INPC_General, n.ahead = 24, p = 3, d = 1, q = 1, P = 1, D =
1, Q = 2, S = 24, no.constant = FALSE)

values.predict_z <- pronostico$pred

##### Línea **roja** valores observados de la serie de tiempo
##### Línea **azul** valores pronosticados de la serie de tiempo

Periodo_z <- gsub("/(*)", " ", test_z[,1])
Observed_z <- test_z[,2]
Forecasted_z <- values.predict_z

simulation_z <- data.frame(Periodo_z, Observed_z, Forecasted_z)

simulation_plot_z <- ggplot(data= simulation_z, aes(x=Periodo_z, y=Observed_z,
group=1)) +
  geom_line(col="red2", size = 1.1) +
  geom_point(col="red2", size = 3)+
  geom_line(col="deeppink2", size = 3.5, alpha = 0.18) +
  labs(title = "Valores observados y valores pronosticados.", x = "Fecha", y = "INPC
General") +
  geom_line(col = "Cyan4", size = 1.1, aes(y=Forecasted_z)) +
  geom_point(col = "Cyan4", size = 3, aes(y=Forecasted_z)) +
  geom_line(col="Cyan", size = 3.5, alpha = 0.18, aes(y=Forecasted_z))
simulation_plot_z

# Exportar datos
setwd(ruta_exp)

write.csv(Observed_z, "Observed_z.csv")
write.csv(Forecasted_z, "Forecasted_z.csv")

setwd(ruta)

### **Pronóstico de 24 quincenas para el INPC General de México.**

setwd(ruta)
INPC_General <- read.csv('INPC General quincenal.csv', header = T, dec = '.', string-
sAsFactors = F)
var_z <- INPC_General

var_z <- ts(var_z, start = c(1988, 1), frequency = 24)
var_z <- var_z[, - 1]
var_z <- zoo::na.approx(var_z)

INPC_General <- var_z

```

```

pronostico_z <- sarima.for(INPC_General, n.ahead = 24, p = 3, d = 1, q = 1, P = 1, D =
1, Q = 2, S = 24, no.constant = FALSE)

### Valor de los predictores

(rbind(ajuste_z$coef))

### Valores de la predicción

(pronostico_z$pred)

# Exportar datos
setwd(ruta_exp)

write.csv(pronostico_z$pred, "Hforecasted_z.csv")

setwd(ruta)

#-----
# -Proyecto Terminal      Modelos de ajuste I      AR() MA()
#                          Auto Regressive Integrate Moving Average
#-----

setwd(ruta)
INPC_General <- read.csv('INPC General quincenal.csv', header = T, dec = '.', string-
sAsFactors = F)

var_x <- INPC_General
var_x <- ts(var_x, start = c(1988, 1), frequency = 24)
var_x <- var_x[, -1]
var_x <- zoo::na.approx(var_x)

INPC <- var_x

Ajuste_INPC_1 <- arima(INPC, order=c(0,1,0), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)
Ajuste_INPC_2 <- arima(INPC, order=c(1,1,1), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)
Ajuste_INPC_3 <- arima(INPC, order=c(2,1,2), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)
Ajuste_INPC_4 <- arima(INPC, order=c(3,1,3), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)
Ajuste_INPC_5 <- arima(INPC, order=c(4,1,4), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)

Modelo_Ajustado <-
c("Ajuste_INPC_1", "Ajuste_INPC_2", "Ajuste_INPC_3", "Ajuste_INPC_4", "Ajuste_INPC_5")
#Patrones_pdq_y_PDQ <- c("(0,1,0) (1,0,0)", "(1,1,1) (1,0,0)", "(2,1,2)
(1,0,0)", "(3,1,3) (1,0,0)", "(4,1,4) (1,0,0)")
AIC <- c(Ajuste_INPC_1$aic, Ajuste_INPC_2$aic, Ajuste_INPC_3$aic,
Ajuste_INPC_4$aic, Ajuste_INPC_5$aic)
BIC <- c(BIC(Ajuste_INPC_1), BIC(Ajuste_INPC_2), BIC(Ajuste_INPC_3),
BIC(Ajuste_INPC_4), BIC(Ajuste_INPC_5))

Modelo_Ajustado <- as.data.frame(Modelo_Ajustado)
#Patrones_pdq_y_PDQ <- cbind(Patrones_pdq_y_PDQ)

AIC <- as.data.frame(AIC)
BIC <- as.data.frame(BIC)

```

```

Summary_stat <- cbind(Modelo_Ajustado, AIC, BIC)
Summary_stat

# AIC

index_Summary_stat_Akaike_bajo <- which(Summary_stat[,2] == min(Summary_stat[,2]))

Akaike_mas_bajo <- cbind(Summary_stat[index_Summary_stat_Akaike_bajo,1])
Ajuste_Akaike_mas_bajo <- cbind(Summary_stat[index_Summary_stat_Akaike_bajo,2])

# BIC

index_Summary_stat_BIC_bajo <- which(Summary_stat[,3] == min(Summary_stat[,3]))

BIC_mas_bajo <- cbind(Summary_stat[index_Summary_stat_BIC_bajo,1])
Ajuste_BIC_mas_bajo <- cbind(Summary_stat[index_Summary_stat_BIC_bajo,3])

cat("Modelo de ajuste con valor Akaike más bajo:", Akaike_mas_bajo[1,1])
cat("Valor Akaike:", Ajuste_Akaike_mas_bajo[1,1])

cat("Modelo de ajuste con valor BIC más bajo:", BIC_mas_bajo[1,1])
cat("Valor BIC:", Ajuste_BIC_mas_bajo[1,1])

#### Medidas de precisión de los modelos de ajuste

accuracy_fit_INPC_1 <- forecast::accuracy(Ajuste_INPC_1)
accuracy_fit_INPC_2 <- forecast::accuracy(Ajuste_INPC_2)
accuracy_fit_INPC_3 <- forecast::accuracy(Ajuste_INPC_3)
accuracy_fit_INPC_4 <- forecast::accuracy(Ajuste_INPC_4)
accuracy_fit_INPC_5 <- forecast::accuracy(Ajuste_INPC_5)

MAPE <- c(accuracy_fit_INPC_1[5],accuracy_fit_INPC_2[5],accuracy_fit_INPC_3[5],
          accuracy_fit_INPC_4[5],accuracy_fit_INPC_5[5])
ME <- c(accuracy_fit_INPC_1[1],accuracy_fit_INPC_2[1],accuracy_fit_INPC_3[1],
        accuracy_fit_INPC_4[1],accuracy_fit_INPC_5[1])
RMSE <- c(accuracy_fit_INPC_1[2],accuracy_fit_INPC_2[2],accuracy_fit_INPC_3[2],
          accuracy_fit_INPC_4[2],accuracy_fit_INPC_5[2])
MAE <- c(accuracy_fit_INPC_1[3],accuracy_fit_INPC_2[3],accuracy_fit_INPC_3[3],
          accuracy_fit_INPC_4[3],accuracy_fit_INPC_5[3])
MPE <- c(accuracy_fit_INPC_1[4],accuracy_fit_INPC_2[4],accuracy_fit_INPC_3[4],
          accuracy_fit_INPC_4[4],accuracy_fit_INPC_5[4])
MASE <- c(accuracy_fit_INPC_1[6],accuracy_fit_INPC_2[6],accuracy_fit_INPC_3[6],
          accuracy_fit_INPC_4[6],accuracy_fit_INPC_5[6])

Summary_accuracy <- cbind(Modelo_Ajustado, MAPE,ME,RMSE,MAE,MPE,MASE)
Summary_accuracy

#Summary_accuracy <- cbind(Modelo_Ajustado,MAPE,ME,RMSE,MAE,MPE,MASE)

# Calculando índices para obtener los valores más bajos del resumen estadístico de
precisión de los modelos de ajuste

index_MAPE <- which(Summary_accuracy[,2] == min(Summary_accuracy[,2]))

MAPE_mas_bajo <- cbind(Summary_accuracy[index_MAPE ,1])
Valor_MAPE_mas_bajo <- cbind(Summary_accuracy[index_MAPE,2])

cat("Modelo de ajuste con MAPE más bajo:", MAPE_mas_bajo[1,1])
cat("Valor MAPE:", Valor_MAPE_mas_bajo[1,1])

index_ME <- which(Summary_accuracy[,3] == min(Summary_accuracy[,3]))

```



```

ME_mas_bajo <- cbind(Summary_accuracy[index_ME ,1])
Valor_ME_mas_bajo <- cbind(Summary_accuracy[index_ME,3])

cat("Modelo de ajuste con ME más bajo:", ME_mas_bajo[1,1])
cat("Valor ME:", Valor_ME_mas_bajo[1,1])

index_RMSE <- which(Summary_accuracy[,4] == min(Summary_accuracy[,4]))

RMSE_mas_bajo <- cbind(Summary_accuracy[index_RMSE ,1])
Valor_RMSE_mas_bajo <- cbind(Summary_accuracy[index_RMSE,4])

cat("Modelo de ajuste con RMSE más bajo:", RMSE_mas_bajo[1,1])
cat("Valor RMSE:", Valor_RMSE_mas_bajo[1,1])

index_MAE <- which(Summary_accuracy[,5] == min(Summary_accuracy[,5]))

MAE_mas_bajo <- cbind(Summary_accuracy[index_MAE ,1])
Valor_MAE_mas_bajo <- cbind(Summary_accuracy[index_MAE,5])

cat("Modelo de ajuste con MAE más bajo:", MAE_mas_bajo[1,1])
cat("Valor MAE:", Valor_MAE_mas_bajo[1,1])

index_MPE <- which(Summary_accuracy[,6] == min(Summary_accuracy[,6]))

MPE_mas_bajo <- cbind(Summary_accuracy[index_MPE ,1])
Valor_MPE_mas_bajo <- cbind(Summary_accuracy[index_MPE,6])

cat("Modelo de ajuste con MPE más bajo:", MPE_mas_bajo[1,1])
cat("Valor MPE:", Valor_MPE_mas_bajo[1,1])

index_MASE <- which(Summary_accuracy[,7] == min(Summary_accuracy[,7]))

MASE_mas_bajo <- cbind(Summary_accuracy[index_MASE,1])
Valor_MASE_mas_bajo <- cbind(Summary_accuracy[index_MASE,7])

cat("Modelo de ajuste con MASE más bajo:", MASE_mas_bajo[1,1])
cat("Valor MASE:", Valor_MASE_mas_bajo[1,1])

# Exportar datos

setwd(ruta_exp)
write.csv(Summary_accuracy, "Medidas de precisión.csv")
write.csv(Summary_stat, "AIC BIC.csv")
setwd(ruta)

#-----
# -Proyecto Terminal      Modelos de ajuste II      AR() MA()
#                          Auto Regressive Integrate Moving Average
#-----

setwd(ruta)
INPC_General <- read.csv('INPC General quincenal.csv', header = T, dec = '.', string-
sAsFactors = F)

var_x <- INPC_General
var_x <- ts(var_x, start = c(1988, 1), frequency = 24)
var_x <- var_x[, -1]
var_x <- zoo::na.approx(var_x)

INPC <- var_x

```

```

Ajuste_INPC_1 <- arima(INPC, order=c(3,1,1), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)
Ajuste_INPC_2 <- arima(INPC, order=c(3,1,2), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)
Ajuste_INPC_3 <- arima(INPC, order=c(3,1,3), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)

Modelo_Ajustado <- c("Ajuste_INPC_1","Ajuste_INPC_2","Ajuste_INPC_3")
#Patrones_pdq_y_PDQ <- c("(3,1,1) (1,1,2)","(3,1,2) (1,1,2)","(3,1,3) (1,1,2)")
AIC <- c(Ajuste_INPC_1$aic, Ajuste_INPC_2$aic, Ajuste_INPC_3$aic)
BIC <- c(BIC(Ajuste_INPC_1), BIC(Ajuste_INPC_2), BIC(Ajuste_INPC_3))

Modelo_Ajustado <- as.data.frame(Modelo_Ajustado)
#Patrones_pdq_y_PDQ <- cbind(Patrones_pdq_y_PDQ)

AIC <- as.data.frame(AIC)
BIC <- as.data.frame(BIC)

Summary_stat <- cbind(Modelo_Ajustado, AIC, BIC)
Summary_stat

# AIC

index_Summary_stat_Akaike_bajo <- which(Summary_stat[,2] == min(Summary_stat[,2]))

Akaike_mas_bajo <- cbind(Summary_stat[index_Summary_stat_Akaike_bajo,1])
Ajuste_Akaike_mas_bajo <- cbind(Summary_stat[index_Summary_stat_Akaike_bajo,2])

# BIC

index_Summary_stat_BIC_bajo <- which(Summary_stat[,3] == min(Summary_stat[,3]))

BIC_mas_bajo <- cbind(Summary_stat[index_Summary_stat_BIC_bajo,1])
Ajuste_BIC_mas_bajo <- cbind(Summary_stat[index_Summary_stat_BIC_bajo,3])

cat("Modelo de ajuste con valor Akaike más bajo:", Akaike_mas_bajo[1,1])
cat("Valor Akaike:", Ajuste_Akaike_mas_bajo[1,1])

cat("Modelo de ajuste con valor BIC más bajo:", BIC_mas_bajo[1,1])
cat("Valor BIC:", Ajuste_BIC_mas_bajo[1,1])

#### Medidas de precisión de los modelos de ajuste

accuracy_fit_INPC_1 <- forecast::accuracy(Ajuste_INPC_1)
accuracy_fit_INPC_2 <- forecast::accuracy(Ajuste_INPC_2)
accuracy_fit_INPC_3 <- forecast::accuracy(Ajuste_INPC_3)

MAPE <- c(accuracy_fit_INPC_1[5],accuracy_fit_INPC_2[5],accuracy_fit_INPC_3[5])
ME <- c(accuracy_fit_INPC_1[1],accuracy_fit_INPC_2[1],accuracy_fit_INPC_3[1])
RMSE <- c(accuracy_fit_INPC_1[2],accuracy_fit_INPC_2[2],accuracy_fit_INPC_3[2])
MAE <- c(accuracy_fit_INPC_1[3],accuracy_fit_INPC_2[3],accuracy_fit_INPC_3[3])
MPE <- c(accuracy_fit_INPC_1[4],accuracy_fit_INPC_2[4],accuracy_fit_INPC_3[4])
MASE <- c(accuracy_fit_INPC_1[6],accuracy_fit_INPC_2[6],accuracy_fit_INPC_3[6])

Summary_accuracy <- cbind(Modelo_Ajustado, MAPE,ME,RMSE,MAE,MPE,MASE)
Summary_accuracy

#Summary_accuracy <- cbind(Modelo_Ajustado,MAPE,ME,RMSE,MAE,MPE,MASE)

# Calculando índices para obtener los valores más bajos del resumen estadístico de
precisión de los modelos de ajuste

index_MAPE <- which(Summary_accuracy[,2] == min(Summary_accuracy[,2]))

MAPE_mas_bajo <- cbind(Summary_accuracy[index_MAPE ,1])

```

```

Valor_MAPE_mas_bajo <- cbind(Summary_accuracy[index_MAPE,2])

cat("Modelo de ajuste con MAPE más bajo:", MAPE_mas_bajo[1,1])
cat("Valor MAPE:", Valor_MAPE_mas_bajo[1,1])

index_ME <- which(Summary_accuracy[,3] == min(Summary_accuracy[,3]))

ME_mas_bajo <- cbind(Summary_accuracy[index_ME ,1])
Valor_ME_mas_bajo <- cbind(Summary_accuracy[index_ME,3])

cat("Modelo de ajuste con ME más bajo:", ME_mas_bajo[1,1])
cat("Valor ME:", Valor_ME_mas_bajo[1,1])

index_RMSE <- which(Summary_accuracy[,4] == min(Summary_accuracy[,4]))

RMSE_mas_bajo <- cbind(Summary_accuracy[index_RMSE ,1])
Valor_RMSE_mas_bajo <- cbind(Summary_accuracy[index_RMSE,4])

cat("Modelo de ajuste con RMSE más bajo:", RMSE_mas_bajo[1,1])
cat("Valor RMSE:", Valor_RMSE_mas_bajo[1,1])

index_MAE <- which(Summary_accuracy[,5] == min(Summary_accuracy[,5]))

MAE_mas_bajo <- cbind(Summary_accuracy[index_MAE ,1])
Valor_MAE_mas_bajo <- cbind(Summary_accuracy[index_MAE,5])

cat("Modelo de ajuste con MAE más bajo:", MAE_mas_bajo[1,1])
cat("Valor MAE:", Valor_MAE_mas_bajo[1,1])

index_MPE <- which(Summary_accuracy[,6] == min(Summary_accuracy[,6]))

MPE_mas_bajo <- cbind(Summary_accuracy[index_MPE ,1])
Valor_MPE_mas_bajo <- cbind(Summary_accuracy[index_MPE,6])

cat("Modelo de ajuste con MPE más bajo:", MPE_mas_bajo[1,1])
cat("Valor MPE:", Valor_MPE_mas_bajo[1,1])

index_MASE <- which(Summary_accuracy[,7] == min(Summary_accuracy[,7]))

MASE_mas_bajo <- cbind(Summary_accuracy[index_MASE,1])
Valor_MASE_mas_bajo <- cbind(Summary_accuracy[index_MASE,7])

cat("Modelo de ajuste con MASE más bajo:", MASE_mas_bajo[1,1])
cat("Valor MASE:", Valor_MASE_mas_bajo[1,1])

# Exportar datos
setwd(ruta_exp)
write.csv(Summary_accuracy, "Medidas de precisión_3.csv")
write.csv(Summary_stat, "AIC BIC_3.csv")
setwd(ruta)

#-----
# -Proyecto Terminal      Modelos de ajuste III      AR() MA()
#                          Auto Regressive Integrate Moving Average
#-----

setwd(ruta)
INPC_General <- read.csv('INPC General quincenal.csv', header = T, dec = '.', string-
sAsFactors = F)

var_x <- INPC_General
var_x <- ts(var_x, start = c(1988, 1), frequency = 24)
var_x <- var_x[, -1]

```

```

var_x <- zoo::na.approx(var_x)

INPC <- var_x

Ajuste_INPC_1 <- arima(INPC, order=c(3,1,1), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)
Ajuste_INPC_2 <- arima(INPC, order=c(1,1,2), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)
Ajuste_INPC_3 <- arima(INPC, order=c(2,1,2), seasonal = list(order = c(1,1,2), pe-
riod = 24),method="ML", include.mean = TRUE)

Modelo_Ajustado <- c("Ajuste_INPC_1","Ajuste_INPC_2","Ajuste_INPC_3")
#Patrones_pdq_y_PDQ <- c("(3,1,1) (1,1,2)","(1,1,2) (1,1,2)","(2,1,2) (1,1,2)")
AIC <- c(Ajuste_INPC_1$aic, Ajuste_INPC_2$aic, Ajuste_INPC_3$aic)
BIC <- c(BIC(Ajuste_INPC_1), BIC(Ajuste_INPC_2), BIC(Ajuste_INPC_3))

Modelo_Ajustado <- as.data.frame(Modelo_Ajustado)
#Patrones_pdq_y_PDQ <- cbind(Patrones_pdq_y_PDQ)

AIC <- as.data.frame(AIC)
BIC <- as.data.frame(BIC)

Summary_stat <- cbind(Modelo_Ajustado, AIC, BIC)
Summary_stat

# AIC

index_Summary_stat_Akaike_bajo <- which(Summary_stat[,2] == min(Summary_stat[,2]))

Akaike_mas_bajo <- cbind(Summary_stat[index_Summary_stat_Akaike_bajo,1])
Ajuste_Akaike_mas_bajo <- cbind(Summary_stat[index_Summary_stat_Akaike_bajo,2])

# BIC

index_Summary_stat_BIC_bajo <- which(Summary_stat[,3] == min(Summary_stat[,3]))

BIC_mas_bajo <- cbind(Summary_stat[index_Summary_stat_BIC_bajo,1])
Ajuste_BIC_mas_bajo <- cbind(Summary_stat[index_Summary_stat_BIC_bajo,3])

cat("Modelo de ajuste con valor Akaike más bajo:", Akaike_mas_bajo[1,1])
cat("Valor Akaike:", Ajuste_Akaike_mas_bajo[1,1])

cat("Modelo de ajuste con valor BIC más bajo:", BIC_mas_bajo[1,1])
cat("Valor BIC:", Ajuste_BIC_mas_bajo[1,1])

#### Medidas de precisión de los modelos de ajuste

accuracy_fit_INPC_1 <- forecast::accuracy(Ajuste_INPC_1)
accuracy_fit_INPC_2 <- forecast::accuracy(Ajuste_INPC_2)
accuracy_fit_INPC_3 <- forecast::accuracy(Ajuste_INPC_3)

MAPE <- c(accuracy_fit_INPC_1[5],accuracy_fit_INPC_2[5],accuracy_fit_INPC_3[5])
ME <- c(accuracy_fit_INPC_1[1],accuracy_fit_INPC_2[1],accuracy_fit_INPC_3[1])
RMSE <- c(accuracy_fit_INPC_1[2],accuracy_fit_INPC_2[2],accuracy_fit_INPC_3[2])
MAE <- c(accuracy_fit_INPC_1[3],accuracy_fit_INPC_2[3],accuracy_fit_INPC_3[3])
MPE <- c(accuracy_fit_INPC_1[4],accuracy_fit_INPC_2[4],accuracy_fit_INPC_3[4])
MASE <- c(accuracy_fit_INPC_1[6],accuracy_fit_INPC_2[6],accuracy_fit_INPC_3[6])

Summary_accuracy <- cbind(Modelo_Ajustado, MAPE,ME,RMSE,MAE,MPE,MASE)
Summary_accuracy

#Summary_accuracy <- cbind(Modelo_Ajustado,MAPE,ME,RMSE,MAE,MPE,MASE)

# Calculando índices para obtener los valores más bajos del resumen estadístico de
precisión de los modelos de ajuste

```

```

index_MAPE <- which(Summary_accuracy[,2] == min(Summary_accuracy[,2]))
MAPE_mas_bajo <- cbind(Summary_accuracy[index_MAPE,1])
Valor_MAPE_mas_bajo <- cbind(Summary_accuracy[index_MAPE,2])

cat("Modelo de ajuste con MAPE más bajo:", MAPE_mas_bajo[1,1])
cat("Valor MAPE:", Valor_MAPE_mas_bajo[1,1])

index_ME <- which(Summary_accuracy[,3] == min(Summary_accuracy[,3]))
ME_mas_bajo <- cbind(Summary_accuracy[index_ME,1])
Valor_ME_mas_bajo <- cbind(Summary_accuracy[index_ME,3])

cat("Modelo de ajuste con ME más bajo:", ME_mas_bajo[1,1])
cat("Valor ME:", Valor_ME_mas_bajo[1,1])

index_RMSE <- which(Summary_accuracy[,4] == min(Summary_accuracy[,4]))
RMSE_mas_bajo <- cbind(Summary_accuracy[index_RMSE,1])
Valor_RMSE_mas_bajo <- cbind(Summary_accuracy[index_RMSE,4])

cat("Modelo de ajuste con RMSE más bajo:", RMSE_mas_bajo[1,1])
cat("Valor RMSE:", Valor_RMSE_mas_bajo[1,1])

index_MAE <- which(Summary_accuracy[,5] == min(Summary_accuracy[,5]))
MAE_mas_bajo <- cbind(Summary_accuracy[index_MAE,1])
Valor_MAE_mas_bajo <- cbind(Summary_accuracy[index_MAE,5])

cat("Modelo de ajuste con MAE más bajo:", MAE_mas_bajo[1,1])
cat("Valor MAE:", Valor_MAE_mas_bajo[1,1])

index_MPE <- which(Summary_accuracy[,6] == min(Summary_accuracy[,6]))
MPE_mas_bajo <- cbind(Summary_accuracy[index_MPE,1])
Valor_MPE_mas_bajo <- cbind(Summary_accuracy[index_MPE,6])

cat("Modelo de ajuste con MPE más bajo:", MPE_mas_bajo[1,1])
cat("Valor MPE:", Valor_MPE_mas_bajo[1,1])

index_MASE <- which(Summary_accuracy[,7] == min(Summary_accuracy[,7]))
MASE_mas_bajo <- cbind(Summary_accuracy[index_MASE,1])
Valor_MASE_mas_bajo <- cbind(Summary_accuracy[index_MASE,7])

cat("Modelo de ajuste con MASE más bajo:", MASE_mas_bajo[1,1])
cat("Valor MASE:", Valor_MASE_mas_bajo[1,1])

# Exportar datos

setwd(ruta_exp)
write.csv(Summary_accuracy, "Medidas de precisión_III.csv")
write.csv(Summary_stat, "AIC BIC_III.csv")
setwd(ruta)

```

Bibliografía

- Alonso, J. C., & Rivera, A. F. (8 de junio de 2017). Pronosticando la inflación mensual en Colombia un paso hacia adelante: una aproximación "de abajo hacia arriba". *Revista de métodos cuantitativos para la economía y la empresa*, 98-118.
- Amazon. (2019). *¿Qué son los datos de streaming?* Obtenido de Amazon: <https://aws.amazon.com/es/streaming-data/>
- Angulo, P. (2011). *Regresión y ajuste de modelos*. Obtenido de Departamento de matemáticas de la Universidad Autónoma de Madrid: <http://verso.mat.uam.es/~pablo.angulo/doc/laboratorio/b6regresion.html>
- Araujo, L. (2017). *Series temporales*. Obtenido de YouTube: <https://www.youtube.com/watch?v=O0DLO0oWd5c&list=LL8WlIM1cdFXn2mMiuQ5KtMg&index=435&t=0s>
- Banco de México. (s.f.). *¿Qué es inflación? ¿Cómo se mide?* Obtenido de Banxico Educa: http://educa.banxico.org.mx/infografias_y_fichas/inflacion_infografias_/que-es-inflacion-como-se-mide.html
- Banco de México. (s.f.). *Inflación*. Obtenido de Banxico Educa: <http://educa.banxico.org.mx/economia/inflacion-economia.html>
- Banco de México. (s.f.). *Política Monetaria*. Obtenido de Banxico Educa: http://educa.banxico.org.mx/banco_mexico_banca_central/politica-monetaria-banco-mexi.html
- Banco de México. (s.f.). *Preguntas frecuentes sobre la inflación*. Obtenido de Banxico Educa: <http://educa.banxico.org.mx/economia/preguntas-inflacion.html>
- Brownlee, J. (17 de agosto de 2018). *A Gentle Introduction to SARIMA for Time Series Forecasting in Python*. Obtenido de Machine Learning Mastery: <https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/?fbclid=IwAR0lp2pqdZ-n9LrmbjeISd2OXxYONsNjqgdDWfIPoadN3YA4HoUjGEu8Pg>
- Cabrer Borrás, B. (marzo de 2014). *Análisis económico. Econometría empresarial II*. Recuperado el 28 de 11 de 2019, de Universidad de Valencia: <https://www.uv.es/~cabrer/Espanyol/material/Tema10/Tema10.pdf>
- Capistrán, C., Constandse, C., & Ramos-Francia, M. (s.f.). *Uso de Modelos Estacionales para Pronosticar la Inflación de Corto Plazo en México*. Ciudad de México, México.
- Características de series temporales*. (s.f.). Obtenido de IBM Knowledge Center: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/dt/timeseries_characteristics.html
- Catillo Gutierrez, S., & Lozano Aguilera, E. D. (2007). Q-Q Plot Normal. Los puntos de posición gráfica. . *Iniciación a la investigación*, 1.
- Dalinina, R. (10 de enero de 2017). *Introduction to Forecasting with ARIMA in R*. Obtenido de Oracle Data Science Blog: <https://blogs.oracle.com/datascience/introduction-to-forecasting-with-arima-in-r>
- Durán, R., Garrido, E., Godoy, C., & Tena, J. d. (2012). Predicción de la inflación en México con modelos desagregados por componentes. *Estudios Económicos*, 133-167.

- Espinal Rodríguez, T., Giraldo Botero, P. A., Peña Córdoba, N., & Lotero, L. (2016). Análisis, modelizado y pronóstico de la serie de índices de precios al consumidor en Medellín 2010-2015. *Encuentro de Investigación Formativa - Memorias VII versión*, 127-142.
- Flores Castillo, L. A. (2007). Pronóstico del Índice Nacional de Precios al Consumidor. *Revista iberoamericana de contaduría, economía y administración*, 1-25.
- Funciones de autocorrelación y autocorrelación parcial*. (s.f.). Obtenido de IBM Knowledge Center: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/dt/timeseries_acf_pacf.html
- García Serrano, A. (2015). *Inteligencia Artificial. Fundamentos, práctica y aplicaciones*. México: Alfaomega.
- Gómez Cadenas, J. (2005). *Curso de Estadística*. Obtenido de Centro de Ciencias de Benasque Pedro Pascual: <http://benasque.org/benasque/2005tae/2005tae-talks/233s6.pdf>
- Gujarati, D. N. (2010). Econometría de series de tiempo: pronósticos. En D. N. Gujarati, *Econometría* (págs. 773-800). McGraw-Hill.
- Heath, J. (2012). Ajustes estacionales. En J. Heath, *Lo que indican los indicadores. Cómo utilizar la información estadística para entender la realidad económica de México* (pág. 19). Ciudad de México: INEGI.
- IBM Knowledge Center. (s.f.). *Transformaciones de series*. Obtenido de IBM Knowledge Center: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/dt/timeseries_transformations.html
- Inflación*. (s.f.). Obtenido de Banxico Educa: <http://educa.banxico.org.mx/economia/inflacion-economia.html>
- Instituto Nacional de Estadística y Geografía. (16 de junio de 2019). *Índice nacional de precios al consumidor; índice general quincenal*. Obtenido de Banco de Información Económica: <https://www.inegi.org.mx/sistemas/bie/?idserPadre=10000500001500600040>
- Instituto Nacional de Estadística y Geografía. (s.f.). *Preguntas frecuentes (FAQs) del INPC*. Obtenido de INEGI: https://www.inegi.org.mx/contenidos/proyectos/precios/inpc/doc/Preguntas_frecuentes_INPC.pdf
- Instituto Nacional de Estadística y Geografía. (s.f.). *Preguntas frecuentes (FAQs) del INPC*. Obtenido de INEGI: https://www.inegi.org.mx/contenidos/proyectos/precios/inpc/doc/Preguntas_frecuentes_INPC.pdf
- International Business Machines. (s.f.). *Autorregresión (estacional)*. Obtenido de IBM Knowledge Center: https://www.ibm.com/support/knowledgecenter/es/SSEP7J_11.0.0/com.ibm.swg.ba.cognos.pp_usr_gd.doc/c_ppcug_autoregressionseasonal.html#AutoregressionSeasonal
- International Business Machines. (s.f.). *Criterios ARIMA de series temporales*. Obtenido de IBM Knowledge Center: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/clementine/timeseries_arima_criteria.html

- International Business Machines. (s.f.). *Descomposición estacional*. Obtenido de IBM Knowledge Center:
https://www.ibm.com/support/knowledgecenter/es/SSLVMB_sub/statistics_mainhelp_ddita/spss/trends/idh_seas.html
- International Business Machines. (s.f.). *Funciones de autocorrelación y autocorrelación parcial*. Obtenido de IBM Knowledge Center:
https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/dt/timeseries_acf_pacf.html
- International Business Machines. (s.f.). *Transformaciones de series*. Obtenido de IBM Knowledge Center:
https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/dt/timeseries_transformations.html
- Interpreting accuracy results for an ARIMA model fit*. (s.f.). Obtenido de Stackexchange:
<https://stats.stackexchange.com/questions/194453/interpreting-accuracy-results-for-an-arma-model-fit>
- Irizarry, R. (marzo de 2019). *Data Science: R Basics. Build a foundation in R and learn how to wrangle, analyze, and visualize data*. Obtenido de edX:
<https://www.edx.org/es/course/data-science-r-basics-2>
- Jason, B. (18 de septiembre de 2019). *How to Create an ARIMA Model for Time Series Forecasting in Python*. Obtenido de Machine Learning Mastery:
https://machinelearningmastery.com/arma-for-time-series-forecasting-with-python/?fbclid=IwAR0Vmwj6_diHf7dC7rAU92acChJKGnEww22zoPrQQgWlaeipV2Uqj12N9u4
- Lexico. (2019). *Inflación*. Obtenido de Lexico: <https://www.lexico.com/es/definicion/inflacion>
- Ljung-box statistics for arima residuals in R*. (s.f.). Obtenido de Stackexchange:
<https://stats.stackexchange.com/questions/64711/ljung-box-statistics-for-arma-residuals-in-r-confusing-test-results>
- Ludlow Wiechers, J. A. (2018). Notas de clase. *Modelos de Políticas Macroeconómicas I*. Ciudad de México, México: Universidad Autónoma Metropolitana Unidad Azcapotzalco.
- Microsoft. (11 de febrero de 2018). *Qué son las bases de datos no relacionales y No SQL*. Obtenido de Microsoft Azure: <https://docs.microsoft.com/es-es/azure/architecture/data-guide/big-data/non-relational-data>
- Minitab. (2019). *Interpretar todos los estadísticos y gráficas para Promedio móvil*. Obtenido de Minitab: <https://support.minitab.com/es-mx/minitab/19/help-and-how-to/modeling-statistics/time-series/how-to/moving-average/interpret-the-results/all-statistics-and-graphs/>
- Nate Wadson. (23 de Enero de 2012). *Using Mean Absolute Error for Forecast Accuracy*. Obtenido de canworksmart: <https://canworksmart.com/using-mean-absolute-error-forecast-accuracy/>
- Política Monetaria*. (s.f.). Obtenido de Banxico Educa:
http://educa.banxico.org.mx/banco_mexico_banca_central/politica-monetaria-banco-mexi.html
- PRICING revenue management . (s.f.). *Precisión del Pronóstico*. Obtenido de PRICING - revenue management : <https://www.pricing.cl/conocimiento/precision-del-pronostico/>

- Ramírez Buitrago, D. J., & Huérfano Ochoa, H. I. (2014). Protocolo para realizar seguimiento a variables macroeconómicas periodo a periodo. *Econógrafos*.
- Regression Model Accuracy (MAE, MSE, RMSE, R-squared) Check in R*. (14 de Febrero de 2019). Obtenido de Datatechnotes: <https://www.datatechnotes.com/2019/02/regression-model-accuracy-mae-mse-rmse.html>
- Rodríguez Báez, G. (2006). Análisis de la inflación en México con series de tiempo utilizando como indicador el Índice de Precios al Consumidor (INPC). Ciudad de México, México.
- Rodríguez N., N., & Siado C., P. (junio de 2003). Un pronóstico no paramétrico de la inflación colombiana. Bogotá, Colombia.
- Series Temporales: Procesos ARIMA estacionales*. (s.f.). Obtenido de uc3m: <http://halweb.uc3m.es/esp/Personal/personas/imolina/MiDocencia/SeriesTemporales/TeMa6SeriesEstud.pdf>
- SPSS FREE. (2015). *Medidas de Distribución - Asimetría y Curtosis*. Obtenido de SPSS FREE: <http://www.spssfree.com/curso-de-spss/analisis-descriptivo/medidas-de-distribucion-curtosis-asimetria.html>
- Stephanie. (7 de junio de 2016). *ADF — Augmented Dickey Fuller Test*. Obtenido de Statistics How To: <https://www.statisticshowto.datasciencecentral.com/adf-augmented-dickey-fuller-test/>
- Torres Remon, M. Á. (2012). Qué es una base de datos. En M. A. Torres Remon, *Programación Transact con SQL Server 2012* (pág. 45). Perú: MACRO.
- Unidad Académica de Matemáticas. (s.f.). *UNA SERIE DE TIEMPO TIENE LAS SIGUIENTES COMPONENTES*. Obtenido de Unidad Académica de Matemáticas: <http://matematicas.reduaz.mx/home/Docentes/ltrueba/Series/admon4.htm>
- Universidad Carlos III de Madrid. (s.f.). *Capítulo 3. Procesos autorregresivos*. Obtenido de Universidad Carlos III de Madrid: <https://www.uc3m.es/ss/Satellite/UC3MInstitucional/es/PortadaMiniSiteA/1371229065435/>
- Villavicencio, J. (s.f.). *Introducción a series de tiempo*. Obtenido de Instituto de Estadísticas de PR: http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D
- Woody, B., Dean, D., GuhaThakurta, D., Bansal, G., Conners, M., & Tok, W.-H. (2016). Forecasting. En B. Woody, D. Dean, D. GuhaThakurta, G. Bansal, M. Conners, & W.-H. Tok, *Data Science with Microsoft SQL Server 2016* (págs. 66-74). Washington: Octal Publishing, Inc.