
Análisis de modelo MA(1).

Caso de estudio: Recaudo en una entidad financiera



Presentado por

Laura Marcela Medellín Valbuena

LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Fundación Universitaria Los Libertadores

Facultad de Ingeniería y Ciencias Básicas

Especialización en Estadística Aplicada Bogotá D.C, Colombia

2020

Análisis de modelo MA(1).

Caso de estudio: Recaudo en una entidad financiera



Presentado por

Laura Marcela Medellín Valbuena

LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Asesor temático

José John Fredy González Veloza

Fundación Universitaria Los Libertadores

Facultad de Ingeniería y Ciencias Básicas

Especialización en Estadística Aplicada Bogotá D.C, Colombia

2020

Notas de aceptación

Firma del presidente del jurado



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Bogotá DC, Septiembre de 2020.

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

Agradecimientos

A todas las personas que han pasado a lo largo del camino, a todos los que han dejado huellas en mi crecimiento y formación diaria.

A Dios por permitirme prepararme y cumplir con esta meta

A mi mamá porque aunque no esté aquí, siempre me motivo a ser mejor, me apoyo, y me dio todas las enseñanzas más bonitas de la vida para ser quien soy.

A mi familia, porque son mi inspiración y mi motor para ser mejor cada día.

A mis amigos, por su ayuda y paciencia diaria.

A mis maestros por cumplir su labor con tan buena disposición, y por siempre ayudarme en el proceso.

Índice General

Introducción	10
2. Planteamiento del Problema	12
2.2.1 Objetivo General.....	14
2.2.2 Objetivos Específicos.....	14
3. Marco Teórico / Conceptual	16
3.1 Recaudos y desembolsos a nivel financiero	16
3.2 Series de tiempo	22
3.2.1 Clases de series de tiempo.....	25
3.2.2 Clasificación de series de tiempo.....	26
3.3 Modelos de series de tiempo	28
3.3.1 Modelos de Autorregresion.....	28
3.3.2 Modelo ARIMA (p,d,q).....	29
4. Marco Metodológico	33
5. Resultados	35
6. Discusión de resultados	49
7. Conclusiones y recomendaciones	50
8. Referencias Bibliográficas	51

Índice de Ilustraciones.

Ilustración 1. Ejemplo oficina 105 actual.....	18
Ilustración 2. Ejemplo oficina 105 ideal.....	19
Ilustración 3. Ejemplo de serie de tiempo. IPC años 1990 al 2011. Banrep (2012)	23
Ilustración 40. Residuales ARIMA (0.1.7)	39
Ilustración 51. Pronóstico modelo Arima (0.1.7).....	40
Ilustración 62. Residuales ARIMA (4.1.7)	40
Ilustración 7. Pronóstico modelo Arima (4.1.7).....	41
Ilustración 84. Residuales ARIMA (2.1.3)	42
Ilustración 96. Residuales ARIMA (3.1.3)	43
Ilustración 10. Pronóstico modelo Arima (3.1.3).....	44
Ilustración 11. Residuales ARIMA (3.1.7)	44
Ilustración 122. Resultados de todos los modelos	47

Resumen

Algunas entidades financieras generalmente deben cumplir con una serie de préstamos a sus clientes. Para ello, estas entidades deben tener una correcta planeación del efectivo que debería estar disponible sin excederse de sus cupos. Para esto las entidades financieras realizan una contratación de personas especializadas en la proyección de efectivo, estas personas se encargan de saber cuándo se debe inyectar más dinero en las bóvedas de las oficinas o cuando se debe recoger dinero y custodiar hasta sus fondos en las transportadoras de valores.

En el ejercicio diario de la planeación de cada oficina intervienen muchos factores internos y externos, varios de estos difíciles de predecir, como la fuerza comercial de cada oficina, el clima, la economía del país, entre otras

Actualmente el modelo está basado en la planeación de desembolsos que tiene cada oficina, pero para lograr que este modelo sea más eficaz y determine más variables que no se están teniendo en cuenta se decide analizar el comportamiento de los recaudos, los recaudos son un factor importante a la hora de necesitar efectivo para desembolsar un crédito, puesto que si la oficina tiene más dinero de recaudo (entrada) que pendientes de desembolso posiblemente no necesitará un servicio de provisión. De acuerdo a esto se usaron métodos de series de tiempo, modelo ARIMA y diferentes pruebas como Jarque Bera y Ljung Box para determinar cuál de los modelos planteados se desempeña mejor para el pronóstico del recaudo.

Al finalizar las pruebas de los modelos ARIMA realizados se observa que el modelo MA(1) obtenido no es un modelo de pronóstico, esto debido a que observamos que los resultados obtenidos muestran un modelo MA no invertible es decir que el efecto de las observaciones pasadas influye cada vez más con la distancia del tiempo; sin embargo con estos resultados se puede concluir que aunque el valor de los recaudos dependerán siempre del inicio de los mismos, y a medida que pase el tiempo seguirán variando sin poder obtener un pronóstico.

Introducción

En las entidades financieras se presenta mucha volatilidad en las entradas (recaudos) y salidas (desembolsos) de dineros, esto debido a la facilidad que brindan actualmente a la hora de otorgar créditos. Por esto una oficina bancaria debe contar con cierta disponibilidad de efectivo físico para poder cumplir a sus clientes con sus préstamos y que a la vez los clientes también cumplan con sus cuotas.

Debido a esto estas entidades financieras buscar tener una buena planeación de sus entradas y salidas de efectivo, por esto se deben buscar estrategias que permitan que las oficinas mantenga un valor óptimo en sus bóvedas asegurando el efectivo que colocan a disposición de sus clientes y evitando riesgos de sobrecupo o desabastecimiento de efectivo (agotamiento).

Cabe mencionar que es muy difícil poder acertar con lo que se debe tener en una oficina, e incluso es muy complicado poder desembolsar o recaudar todo lo esperado porque hay factores externos que pueden ocasionar que el cliente no realice el pago o recoja su préstamo.

Por esta razón el presente proyecto busca encontrar el mejor modelo de pronóstico a través de modelos de series de tiempo para los recaudos de una oficina, basándonos en la historia de los

recaudos de 154 semanas (2 años) con el fin de que esté pronóstico de una idea más cercana a lo que debe tener una oficina para cumplir sus desembolsos.

En el capítulo dos se hablará del planteamiento del problema que se trabajó, la justificación y los objetivos que impusieron para la solución, en el tercer capítulo se habla del marco teórico y conceptual, en el cual se define con detalle las series de tiempo y el manejo de créditos en las entidades bancarias, en el capítulo cuatro se hablara del marco metodológico, en el capítulo cinco se mostrarán los modelos y los resultados obtenidos con ayuda del software Rstudio, en los resultados veremos que el modelo obtenido será un MA1, el cual sugiere que la serie no se puede pronosticar y se realizara un respectivo análisis sobre ello y en el capítulo seis se hablarán de las conclusiones y recomendaciones.

2. Planteamiento del Problema

2.1 Descripción del problema

A nivel nacional las entidades financieras están dedicadas a otorgar créditos a sus clientes, detrás del otorgamiento de créditos se deben hacer diferentes estudios donde la entidad valida por ejemplo el perfil del cliente, la capacidad de deuda, etc. Para esto las entidades tienen un área comercial la cual debe cumplir con unas metas específicas de colocamiento de créditos mensualmente. Luego de estas validaciones cuando se decide aprobar el crédito, la entidad financiera debe preparar el dinero para ser desembolsado al cliente.

En algunos casos, el dinero de los desembolsos puede estar en la caja física de la entidad, pero en otros casos cuando son montos grandes puede no alcanzar el efectivo físico y es necesario solicitar a la transportadora de valores que lleve una provisión de efectivo para un día determinado o para el día que el cliente vaya a recoger su efectivo en calidad de préstamo.

El proceso suele verse sencillo pero el ejercicio de ofrecer créditos y disponer del efectivo en la brevedad posible tiene muchos factores y variables que hacen que esto sea un proceso poco

óptimo y algunas veces demorado. Hay varias variables que afectan de manera indirecta el proceso del tiempo y disposición de los créditos las cuales no podemos controlar ni pronosticar, como el tipo de cliente, la edad, el día de la semana, y hasta el clima.

Hay otro tipo de variables que por el contrario pueden ayudarnos a jugar con los datos y así poder evitar reprocesos en los desembolsos, una de estas variables es el valor de los recaudos. En el ejercicio actual conocemos que las oficinas hacen préstamos pero también reciben efectivo de dichos préstamos y por ende a la hora de desembolsar créditos se puede contar con el efectivo que ha entrado a la oficina.

Los recaudos no tienen un control como el que tiene los desembolsos, aunque estos afectan de manera directa la disponibilidad del efectivo con el que se puede contar, por lo tanto se debe analizar la variable de recaudos y así sacar un pronóstico de la misma para ayudar a predecir el efectivo que seguramente se requerirá.

2.1.1. Formulación del problema.

¿A través de las series de tiempo se puede determinar modelos precisos para pronosticar la entrada de efectivo (recaudos) de una oficina?

2.1.2. Sistematización del problema.

- ¿Cómo se puede analizar los resultados obtenidos?
- ¿Cuál de los modelos obtenidos será el más adecuado para el ejercicio de recaudos?

2.2 Objetivos

2.2.1 Objetivo General

Analizar los valores históricos de los recaudos de la oficina Principal de la ciudad de Bucaramanga de la entidad financiera

2.2.2 Objetivos Específicos

- Adecuar la información de la base de los movimientos de la oficina Principal de Bucaramanga.
- Aplicar diferentes modelos ARIMA estudiados durante la especialización
- Evaluar el desempeño de cada modelo ARIMA

2.3 Justificación

En el presente trabajo se analiza el manejo que se le da a la programación de entidades financieras a través de una transportadora de valores, en la cual no hay un parámetro relacionado con la estadística que me permita la programación (planeación diaria que se debe hacer para saber si se le debe llevar o no dinero a una oficina de un banco, para que así mismo

la oficina pueda desembolsar créditos a sus clientes) de efectivo sistemática para las oficinas de una entidad financiera, los encargados de dicha labor la realizan basándose en sus conocimientos tácitos, estos conocimientos no siempre son los más adecuados ni los más acertados debido a que cuando se hace una mala planeación de efectivo se puede incurrir en muchos servicios innecesarios, más costos para el cliente, y paradas no efectivas de la transportadora.

A partir del pronóstico de recaudos se podría realizar una planeación más eficiente de recorridos más efectivo y por ende menos paradas de poco valor, así se evita que una oficina quede sin dinero para realizar sus desembolsos a sus clientes y también se trata de lograr que en las pocas paradas o recorridos de dinero se envíe más efectivo, con esto se da más valor del dinero en el tiempo, logrando una planeación de efectivo mucho más efectiva y menos costosa.

Las series de tiempo son una representación de la información y sirve para pronosticar eventos futuros, para extraer información representativa sobre las relaciones entre las diferentes variables y para extrapolar e interpolar los datos. En este trabajo se usa Modelos ARIMA los cuales ayudan a encontrar patrones que se puedan usar para una predicción futura.

3. Marco Teórico / Conceptual

3.1 Recaudos y desembolsos a nivel financiero

Recaudo

El recaudo es el efectivo que entra a una entidad bancaria con la figura de recuperación y/o pago de intereses y cuotas.

Desembolsos

Se consideran desembolsos a las salidas de efectivo de una oficina bancaria para cumplir con los créditos otorgados a un cliente. En algunas oficinas se puede relacionar con el sinónimo pagos. (Préstamos o créditos). Según (Eco-finanzas. Sin fecha en las referencias, Desembolso) un desembolso es La entrega de Dinero en efectivo como consecuencia de una Compra o una obligación de pago. Es sinónimo de pago. Fase de una transacción financiera que mide los pagos efectuados. Representa una salida de fondos. En pocas palabras podemos referirnos a los desembolsos como un crédito.

Manejo de desembolsos

Para la entidad financiera de la investigación, la asignación de créditos supone inicialmente la solicitud de un cliente por una cantidad de efectivo determinada, cabe mencionar que esta entidad financiera otorga créditos para microempresas únicamente; la oficina debe realizar estudios de seguridad, de capacidad de pago, entre otros, para luego dar una respuesta afirmativa o negativa; en caso de que la respuesta sea afirmativa se sube el valor del crédito a un programa del banco en el cual se puede observar todos los movimientos que se han tenido en el día, entre algunos:

Bóveda Inicial: La bóveda inicial es el valor final del día de ayer, es decir el valor neto de como amaneció la oficina.

Recaudos: Son los pagos que han hecho los clientes en el día y/o abonos a sus créditos

Pagos efectivos: Son los créditos desembolsados en el día.

Efectivo Recibido: Es el valor de la provisión recibida por parte de la transportadora de valores.

Efectivo Enviado: Es el valor de la recolección enviada al cuarto de procesamiento que se tienen en la transportadora de valores

Pendiente de pago: Es el valor de los créditos aprobados que están pendientes de desembolsar

Fecha: Es el día, mes y año que se está consultando

Bóveda Final: Es el valor con el que cierra la oficina su operación al final del día, este mismo valor será el valor inicial para el siguiente día.

Historia de la programación de oficinas

La programación de oficinas consiste en enviar efectivo en un carro de valores a una oficina de una entidad financiera, de acuerdo a sus necesidades y su cupo, llamamos cupo al efectivo máximo permitido que puede tener una oficina en su bóveda física y las necesidades se definen como lo que una oficina podría llegar a necesitar diariamente para cumplir con créditos que se deben desembolsar.

El proceso se realiza actualmente basándonos en los pendientes de pago de las oficinas, pero, este método no está siendo el más útil, esto debido a que el comportamiento de las oficinas es volátil, y aunque se logra cumplir con lo que la oficina necesita (cumplir con desembolsos) lo ideal es evitar recorridos innecesarios de poco valor

Ejemplo 1: (Actualmente)

Ilustración 1. Ejemplo oficina 105 actual



Total transportes ejemplo 2 en 3 días: 3 RECORRIDOS

Ejemplo 2: (IDEAL)

Ilustración 2. Ejemplo oficina 105 ideal



Total transportes ejemplo 2 en 3 días: 2 RECORRIDOS

Como se puede observar en el ejemplo 1, se puede concluir que actualmente se cumple con una necesidad, pero se hacen recorridos o transportes no óptimos; en el ejemplo 2, se observa que en los 3 días de prueba se puede optimizar un transporte, entonces, en vez de realizar 3 recorridos en 3 días, se pueden hacer 2 o menos en la misma cantidad de tiempo.

Manejo de desembolsos en diferentes entidades financieras:

En el mercado financiero hay diferentes opciones de créditos para personas naturales y para personas jurídicas

Personas Jurídicas

En previas investigaciones se encontró que las personas jurídicas son aquellos grupos o entidades constituidas para la consecución de objetivos colectivos y duraderos a las que las leyes conceden capacidad de ser sujetos de derechos y deberes. (Acedo, 2007),

Personas Naturales

En Colombia, una persona natural se podría definir como, la responsable de las obligaciones y/o deudas de una empresa.

De acuerdo a la Cámara de Comercio, una persona natural es todo individuo de la especie humana, cualquiera que sea su raza, sexo, estirpe o condición. Ahora bien, en términos empresariales, la persona natural es aquel individuo que al actuar en su propio nombre se ocupa de alguna o algunas actividades que la ley considera mercantiles. En tal caso, la totalidad de su

patrimonio, que incluye el personal y el familiar, sirve como prenda de garantía por las obligaciones que adquiriera en desarrollo de su actividad económica. Cámara de Comercio, (2019) Colombia: <https://www.ccb.org.co/Preguntas-frecuentes/Tramites-registrales/Que-es-una-persona-natural>)

Luego de ver las diferencias entre una persona natural y una jurídica y teniendo en cuenta el tipo de clientes que maneja la entidad financiera en estudio, vamos a ver como otras entidades realizan sus operaciones crediticias o desembolsos.

De acuerdo con (Rojas, 2017) en América Latina se ha diversificado en los últimos años las diferentes medidas para levantar las barreras al financiamiento a la innovación y al crecimiento. Los países de ingresos medios como Colombia han adoptado un enfoque sistemático con el cual se da apoyo al financiamiento de innovaciones y a programas con entidades de educación públicas que apoyen los nuevos proyectos.

En ese orden de ideas se podría decir que en un país como Colombia los proyectos de pymes y pequeñas empresas son apoyadas por varias entidades financieras como la de nuestro estudio.

En algunas entidades financieras tienen diferentes requisitos para solicitudes de crédito, y adicionalmente, varios días de aprobación. Para la entidad de estudio se ven algunos casos donde se aprueba el crédito de un día al otro, o el mismo día incluso.

Por eso los movimientos en estudio suelen ser muy variables y difíciles de predecir.

En esta ocasión nos apoyaremos en las series de tiempo para poder analizar la información obtenida

3.2 Series de tiempo

Las series de tiempo, series temporales o series cronológicas son un modo estructurado, en el cual se recopilan, y se representa la evolución y secuencias de los datos en intervalos de tiempo, es decir, datos diarios, semanales, mensuales, semestrales, anuales, etc.

Los datos que comprende una serie de datos pueden ser iguales, usados por ejemplo en la temperatura en un observatorio meteorológico en días sucesivos al mediodía o desiguales como en la talla de una persona, el peso medido en diferentes etapas de su vida.

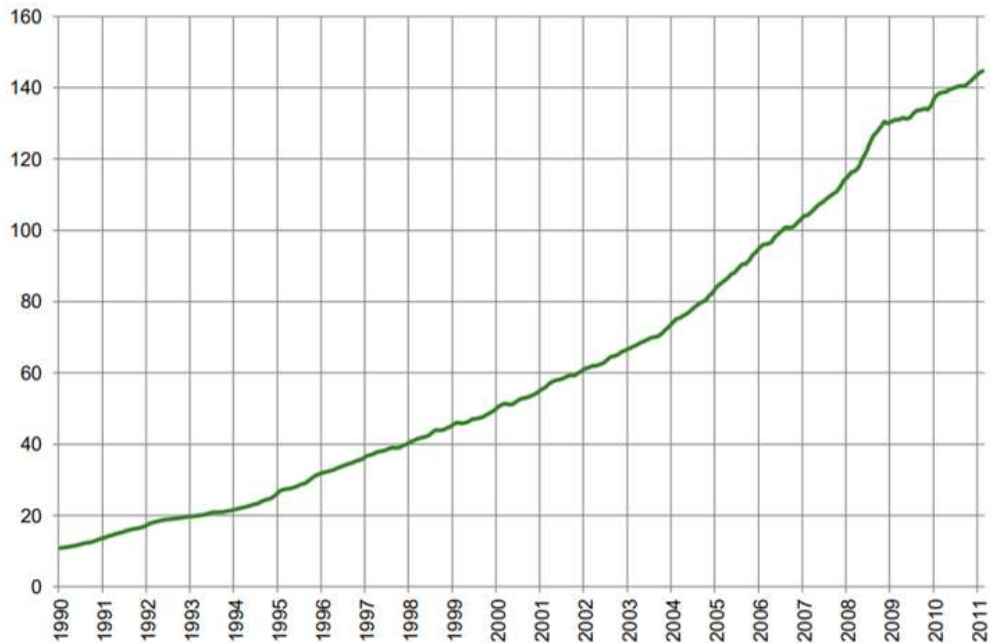
De acuerdo a (Levine, 1996, p. 859), una serie de tiempo “es un conjunto de datos numéricos que se obtienen en periodos regulares a través del tiempo” por ejemplo la tasa de inflación, la tasa de desempleo y los índices bursátiles.

(Antunes, 2011) nos explica que un ejemplo de series de tiempo son las observaciones anuales del PIB de un país, el IPC (Índice de Precios al consumidor mensual), las ventas mensuales de una empresa, entre otras.

Este es un ejemplo de series de tiempo

Ilustración 3. Ejemplo de serie de tiempo. IPC años 1990 al 2011. Banrep (2012)

IPC JULIO 1990 - 2011



(Hernández, 2015) dice que una serie de tiempo es una colección de observaciones sobre un determinado fenómeno efectuadas en sucesivos momentos del tiempo. Corresponde a una realización de un proceso generador de datos.

Las series se pueden denominar Serie Estocástica, una parte conocida (sistemática) susceptible de predecir y de una parte totalmente desconocida (aleatoria), y Serie determinística cuando el futuro se puede predecir sin error, es una variable que está determinada o fija y la cual no cambia de una muestra a otra

(Villavicencio, 2010) menciona varios ejemplos donde se puede utilizar las series temporales:

Economía y Marketing: Proyecciones del empleo y le desempleo, evolución del índice de precios de la leche, beneficios netos mensuales de las entidades bancarias, índices del precio del petróleo.

Demografía: Número de habitantes por año, tasa de mortalidad infantil por año.

Medioambiente: Temperatura media mensual, lluvia recogida diariamente en una localidad, medición diaria del contenido de residuos tóxicos en un río

Para concluir la definición de una serie de tiempo dada por (Carvajal.P, 2003), básicamente una serie de tiempo se le denomina a cualquier variable que conste de datos reunidos, registrados u observados sobre incrementos sucesivos de tiempo. Por lo tanto, se concluye que es una secuencia ordenada de observaciones sobre una variable en particular.

Las series de tiempo tienen componentes

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

- **Tendencia:** Una serie de tiempo con tendencia es aquella que contiene un componente de largo plazo que representa el crecimiento o declinación de la serie a través de un período amplio.
- **Estacional:** Se define como estacional una serie de tiempo con un patrón de cambio a si mismo año tras año. Por lo regular, el desarrollo de una técnica de pronóstico estacional

comprende la selección de un método multiplicativo o uno de adición y estimar después índices estacionales a partir de la historia de la serie.

- **Ciclo:** El efecto cíclico se define como la fluctuación en forma de onda alrededor de la tendencia. Los patrones cíclicos tienden a repetirse en los datos cada dos, tres o más años. Es difícil establecer un modelo para estos patrones cíclicos, ya que no son estables.
- **Irregular:** El componente irregular de la serie de tiempo es el factor residual, es decir, “todo lo que sobra” y toma en consideración las desviaciones de los valores reales de la serie de tiempo en comparación con los esperados; es el elemento aleatorio.

Basados en (Alfaro, 2009) cuando en una serie de tiempo hay datos ausentes se puede usar dos métodos:

- La descomposición de componentes principales de la matriz de correlación de datos en una misma variable entre estaciones cercanas y en periodos similares: Este método es multivariable y permite que se incorporen valores para rellenar los fenómenos de escala.
- Método cuando no hay estaciones cercanas y el relleno se debe realizar con la misma información de la estación: Se debe ajustar un modelo auto regresivo a la serie y así usar el modelo como estimador de los datos ausentes

3.2.1 Clases de series de tiempo

De acuerdo a (Hanke y Reitsch, 1996) los pronósticos se pueden clasificar por criterios:

1. Tiempo: Existen pronósticos a corto y a largo plazo. Los pronósticos a largo plazo ayudan al curso de la organización

2. Criterio: Posición (Entorno micro y macro): Con este criterio se definen el detalle del pronóstico, macro pronóstico y micro pronóstico, por ejemplo, el micro pronóstico podría ser el valor que se necesitaría para la producción anual de un producto, mientras que un ejemplo del macro pronóstico puede ser el incremento en la carga tributaria que realice el gobierno para el siguiente año.
3. Pronóstico Cualitativo y Cuantitativo: El primero hace alusión a cuando se basan en el juicio de una persona, y el segundo hace alusión al proceso mecánico que dé como resultado datos matemáticos.

Para (Sanchez, 2011), las series de tiempo se clasifican así:

3.2.2 Clasificación de series de tiempo

En función del número de respuestas:

- **Univariante:** Se centra en el análisis de una única característica.
- **Multivariante:** Función de transferencia. (Wikipedia, 2019), modelo matemático que a través de un cociente relaciona la respuesta de un sistema (modelada o señal de salida) con una señal de entrada o excitación (también modelada). En la teoría de control, a menudo se usan las funciones de transferencia para caracterizar las relaciones de entrada y salida de componentes o de sistemas que se describen mediante ecuaciones diferenciales lineales e invariantes en el tiempo.
- **Lineales:** La tendencia a largo plazo de muchas series de negocios (industriales y comerciales), como ventas, exportaciones y producción, con frecuencia se aproxima a una línea recta. Esta línea muestra que algo aumenta o disminuye a un ritmo constante.

Este método para obtener una línea recta de mejor ajuste es el método de mínimos cuadrado.

- **No Lineales:** Según (Leguizamón, 2014) la serie es lineal cuando se presenta un comportamiento curvilíneo. Dentro de las tendencias no lineales que pueden presentarse en una serie se encuentran, la polinomial, logarítmica, exponencial y potencial, entre otras.
- **Series Estacionarias:** (Jj.Marin, 2015) Menciona que una serie estacionaria es estable, la media y la variabilidad son constantes a lo largo del tiempo. Es una serie estable, sin que se presente aumentos o disminuciones sistemáticos de sus valores.
- **Series No Estacionarias:** Esta serie es contraria a las series estacionarias, con las series estacionarias la variabilidad y la media cambian con el tiempo.

(Chase, Aquilano y Jacobs , 2001), hacen una clasificación de pronósticos de acuerdo a lo que la importancia de lo que se quiere pronosticar:

1. Los pronósticos cualitativos se forman mediante:

Composición de fuerza de ventas, investigación de mercados, panel, analogía histórica, Método Delphi

2. Los pronósticos basados en el análisis de series de tiempo incluyen:

Promedios móviles, promedio móvil ponderado, suavización exponencial, análisis de regresión, box Jenkins, series de tiempo de Shiskin, proyecciones de tendencia

3. Los pronósticos basados en modelos causales se componen de: Análisis de regresión, modelos econométricos, modelos de entrada / salida, Indicadores guía
4. Pronósticos basados en modelos de simulación

3.3 Modelos de series de tiempo

3.3.1 Modelos de Autorregresion

Un modelo es autorregresivo si la variable endógena de un período t es explicada por las observaciones de ella misma correspondientes a períodos anteriores junto con un término de error.

En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error

En los análisis de series de tiempo se usa una metodología denominada “Box-Jenkins, su nombre se debe en origen a los estadísticos George E.P. Box y Gwilym Jenkin, este se usa para modelos autorregresivos de media móvil (ARMA) o para modelos autorregresivos integrados de media móvil (ARIMA), esto para encontrar el mejor ajuste de una serie temporal de valores con el fin de tener un mejor pronóstico.

(De la Fuente, 2015) resume en 4 fases el modelo Box Jenkins

Fase 1: Identificar el posible modelo ARIMA que la serie pueda requerir

Fase 2: Seleccionar provisionalmente un modelo para la serie estacionaria, se entra en el segundo paso de la estimación, donde los parámetros AR y MA del modelo se estiman por máxima verosimilitud y se obtienen sus errores estándar y los residuos del modelo.

Fase 3: En esta se realiza el diagnostico, con el diagnostico comprobamos que los residuos no tienen estructura de dependencia y siguen un proceso de ruido blanco. Si los residuos muestran estructura se modifica el modelo para incorporarla y se repiten las etapas anteriores hasta obtener el modelo más adecuado.

Fase 4: Es el paso final, es el pronóstico, una vez obtenido el modelo adecuado se puede empezar a realizar la predicción. El modelo seleccionado se usara como la formula inicial de la predicción.

3.3.2 Modelo ARIMA (p,d,q)

Un modelo ARIMA (p,d,q) es un Modelo Autorregresivo - Integrado de Medias Móviles de orden p, d, q. Este modelo es el instrumento que nos describe la evolución futura que seguiría la serie si en lo sucesivo no sufriese ninguna perturbación aleatoria.

De acuerdo a (Orjuela y Delgado, 2016) este modelo se representa por la siguiente ecuación:

$$Y_t = -(\Delta^d Y_t - Y_t) + \phi_0 + \sum_{i=1}^p \phi_i \Delta^d Y_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Los parámetros p , d , q son números enteros no negativos, estos indican como es el orden de las distintas componentes del modelo.

p = AR – Autorregresivo

d = I -Interacciones

q = MA – Media Móvil

De acuerdo con (Orejuela y Delgado, 2016), este modelo es muy sensible a la precisión con que se determinen sus coeficientes.

Para ajustar la serie se debe tener en cuenta:

1. Datos Atípicos (Departamento de Matemáticas Universidad de Coruña España, 2016): corresponde a aquel registro mayor o menor de lo esperado que se detecta por tener un residuo que es un valor “inusual”, muy grande o muy pequeño en relación con la distribución asociada a los residuos.
2. Transformación de Boxcox (Statgraphics, 2011): diseñado para determinar una transformación óptima para Y mientras se estima un modelo de regresión lineal, cuando la variabilidad de Y cambia como una función de X .

Los modelos de series de tiempo conocidos como modelos ARIMA pueden incluir términos autorregresivos y / o términos de promedio móvil. En la semana 1, aprendimos un término autorregresivo en un modelo de serie temporal para la variable es un valor rezagado de x_1 . Por

ejemplo, un término autorregresivo de retardo 1 es (multiplicado por un coeficiente). Esta lección define términos de promedio móvil.

Un término de media móvil en un modelo de serie temporal es un error pasado (multiplicado por un coeficiente)

El 1st orden de media móvil modelo, denotado por MA (1) es:

$$x_t = \mu + w_t + \theta_1 w_{t-1}$$

El 2nd orden de media móvil modelo, denotado por MA (2) es:

$$x_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}$$

El modelo de media móvil de q-ésimo orden, denotado por MA (q) es:

$$x_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}$$

Propiedades teóricas de una serie temporal con un modelo MA (1)

La media es $E(x_t) = \mu$

La varianza es $Var(x_t) = \sigma^2 w (1 + \theta_1^2)$

La función de autocorrelación (ACF) es: $\rho_1 = \frac{\theta_1}{1 + \theta_1^2}$, y $\rho_h = 0 * h \geq 2$

El único valor distinto de cero en el ACF teórico es para el retraso 1 . Todas las demás autocorrelaciones son 0. Por lo tanto, una muestra de ACF con una autocorrelación significativa solo en el retraso 1 es un indicador de un posible modelo MA (1).

4. Marco Metodológico

Este estudio busca encontrar un modelo estadístico para pronosticar las entradas de una oficina de una entidad bancaria en Colombia. La información contenida aquí es de carácter confidencial.

Con la investigación realizada y la información contenida aquí se pudo llegar a determinar que para poder realizar un pronóstico de recaudos (entradas) se tendría que usar series de tiempo y sobre estas mismas revisar cual de todas se acercaba más con la realidad de la oficina, elección de modelo óptimo para pronosticar el comportamiento futuro de la variable de estudio.

Se realizaron series semanales utilizando modelo arima, y suavizado exponencial.

Los datos se obtuvieron del sistema de la entidad financiera y se exportaron a Excel, se adecuaron de tal forma que estos tuvieran como variables, Fecha, Recaudo, Desembolso, Saldo Inicial y saldo final. De acuerdo a estas se decide trabajar con dos que son tiempo (Fecha-Semanas) y Recaudos.

Los datos se macearon desde Rstudio (R es un entorno de programación libre que se utiliza para el procesamiento y análisis estadístico de datos (matrixcpmsolutions.com, 2018). Se trata de uno de los lenguajes de programación más utilizados en investigación científica, siendo además muy popular en los campos de aprendizaje automático (machine learning), minería de datos,

investigación biomédica, bioinformática y matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y graficación), Excel y archivo txt.

Inicialmente se graficaron los datos de la base sin modelar con el fin de analizar el comportamiento inicial de la variable para identificar la tendencia de los datos y posteriormente se determinó la cantidad requerida de diferenciadores.

Se realizaron diagnósticos iniciales, definición de frecuencias, y posteriormente se empezó a analizar si la serie obtenida era o no estacionaria.

De acuerdo a esto se decide realizar la investigación y pruebas se realizan varios modelos arimas a los cuales se les aplicaron las pruebas Jarque Bera (Se realiza para comprobar si los residuos tiene la asimetría de una distribución normal, si esta da un valor $>5\%$ se puede definir que los residuos son normales, y la prueba Ljung Box (Esta prueba define si existe una correlacion serial, si esta prueba es $<5\%$ indica que no hay autocorrelacion serial).

5. Resultados

A continuación se presentan los modelos y pruebas realizadas a partir de la base “centro.abastos” bajo las series de tiempo, específicamente modelo Arima. La base de datos es semanal desde Enero del año 2018 al 31 de Diciembre de 2019.

También se decide particionar la base de la siguiente forma: semana 1 a la semana 96 para hacer el modelo y de la semana 97 a la 104 para diagnosticar los resultados.

Ilustración 4. Gráfico de dispersión dataprueba

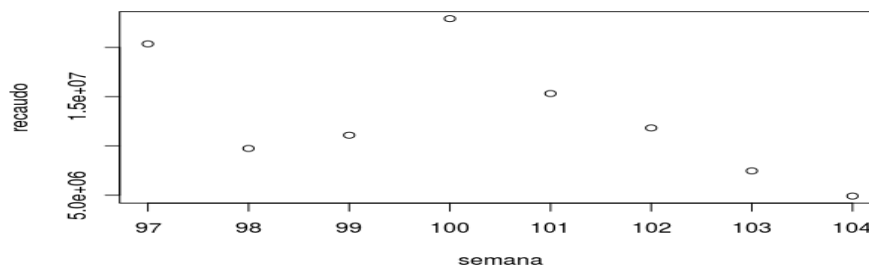
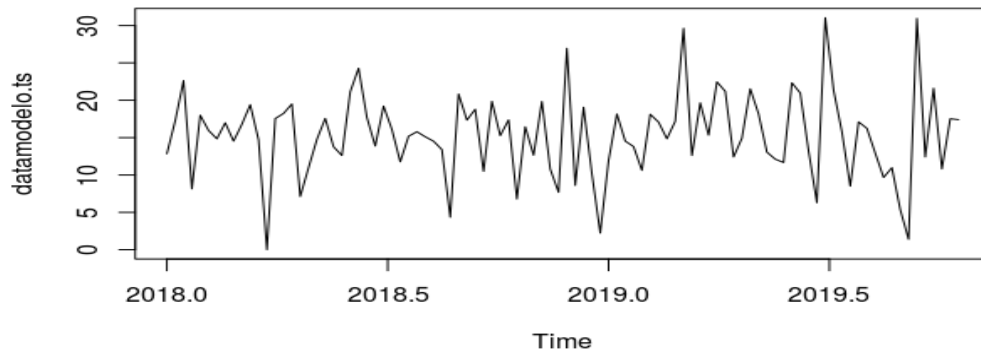


Ilustración 5. Gráfico de líneas datamodelo



La metodología usada fue por series de tiempo, y se usó el modelo ARIMA, este modelo se usa cuando la serie no es estacionaria, esto implica que el comportamiento de la serie no es homogénea y se debe ajustar las diferencias de tal forma para obtener una serie estacionaria.

Para esto se ejecutó el siguiente código en Rstudio y se obtuvo el ACF y el PACF

```
serie1=diff(datamodelo.ts,1)
```

```
adf.test(serie1)
```

```
acf(serie1, lag.max = 55)
```

```
pacf(serie1, lag.max = 55,)
```

Ilustración 6. ACF

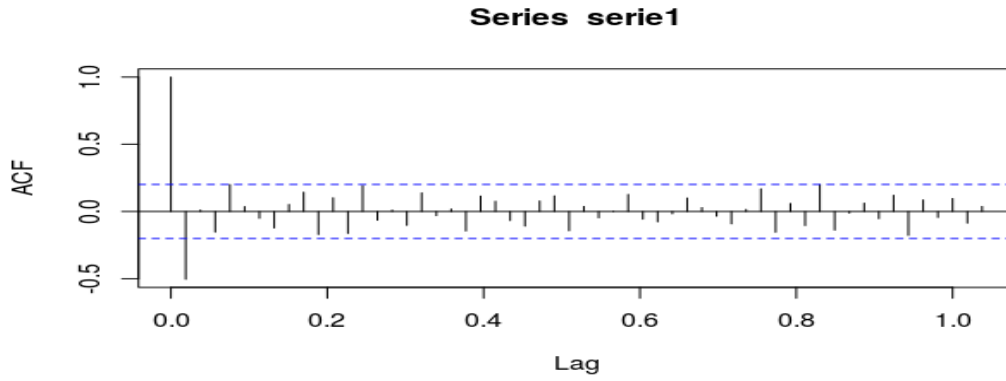
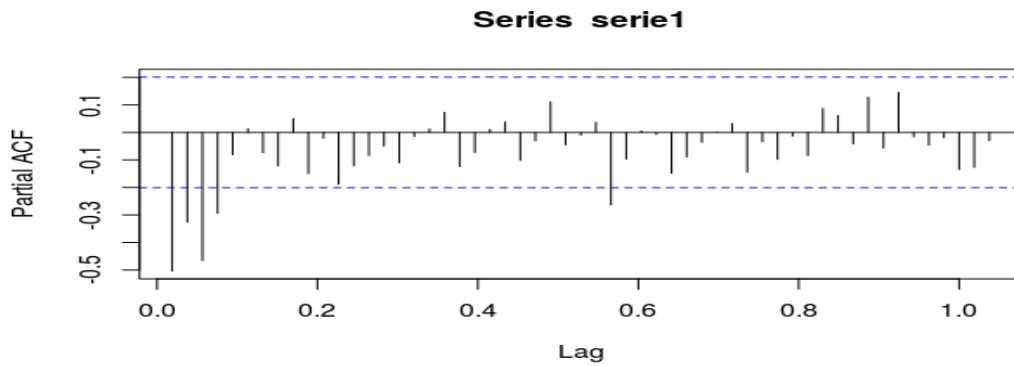


Ilustración 7. PACF

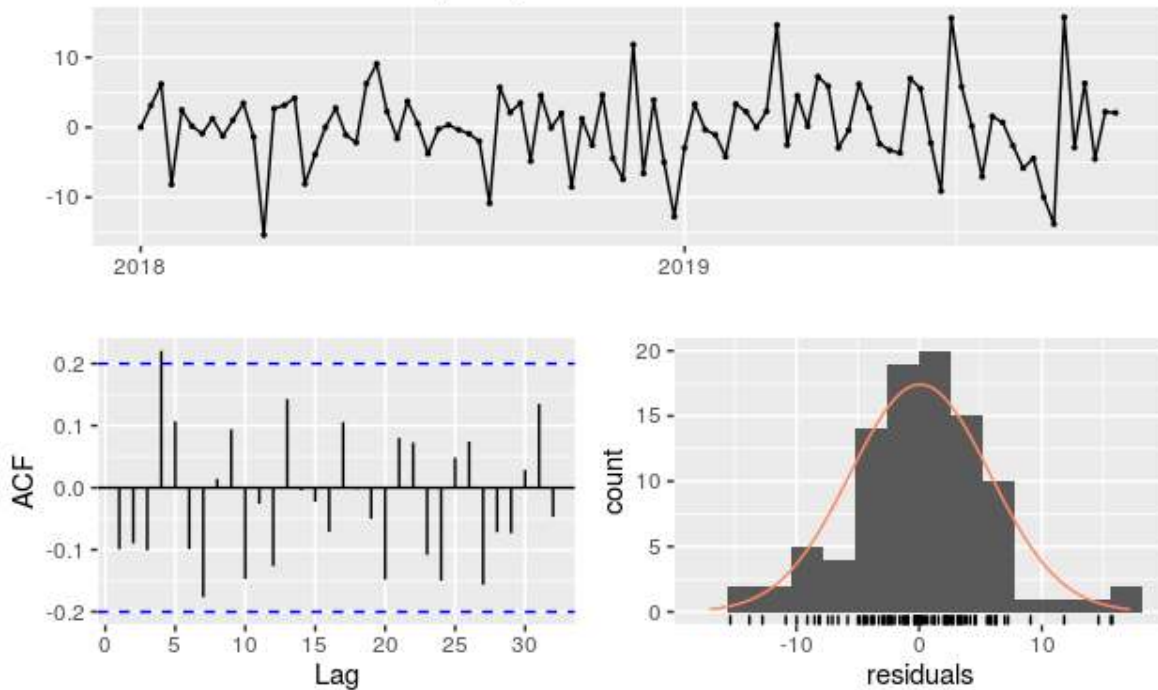


Según los resultados al realizar las diferenciaciones se seleccionaron los siguientes modelos

Modelo ARIMA (0.1.1)

Ilustración 8. Residuales ARIMA (0.1.1)

Residuals from ARIMA(0,1,1)



ARIMA(0,1,1)

```
Coefficients:
      ma1
      -1.0000
s.e.    0.0301
```

```
sigma^2 estimated as 32.85: log likelihood=-302.45
AIC=608.9  AICc=609.03  BIC=614.01
```

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.07248824	5.671808	4.244251	-Inf	Inf	0.7695668	-0.09920908

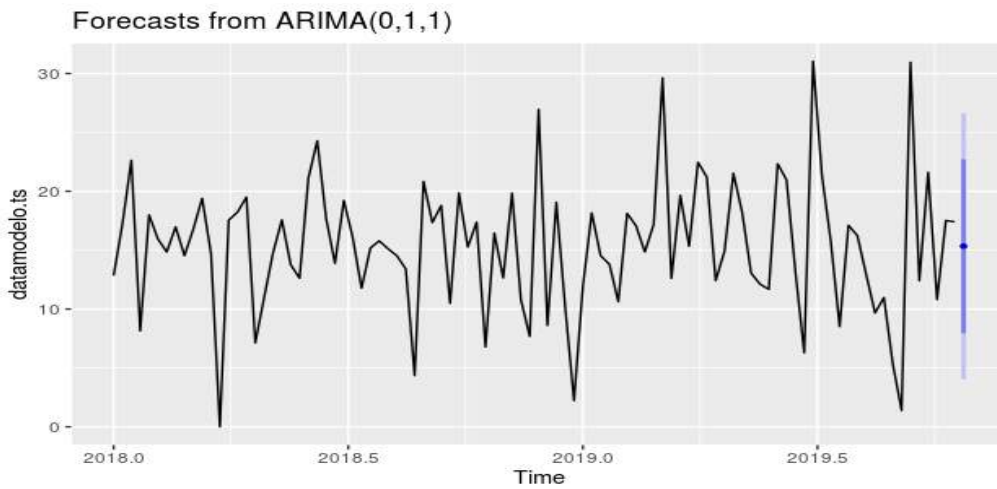
> [checkresiduals\(modelo1\)](#)

Ljung-Box test

```
data: Residuals from ARIMA(0,1,1)
Q* = 23.122, df = 18, p-value = 0.1859
```

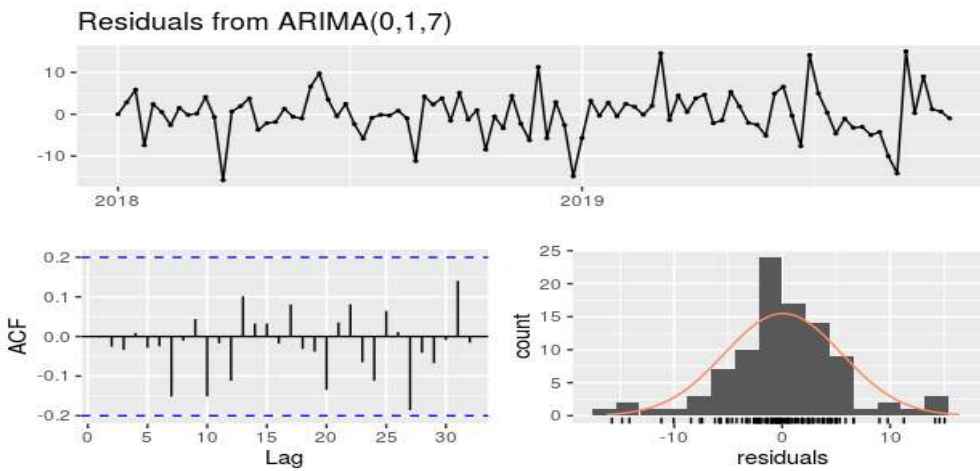
```
Model df: 1. Total lags used: 19
```

Ilustración 9. Pronóstico modelo Arima (0.1.1)



Modelo ARIMA (0.1.7)

Ilustración 40. Residuales ARIMA (0.1.7)



ARIMA(0,1,7)

Coefficients:

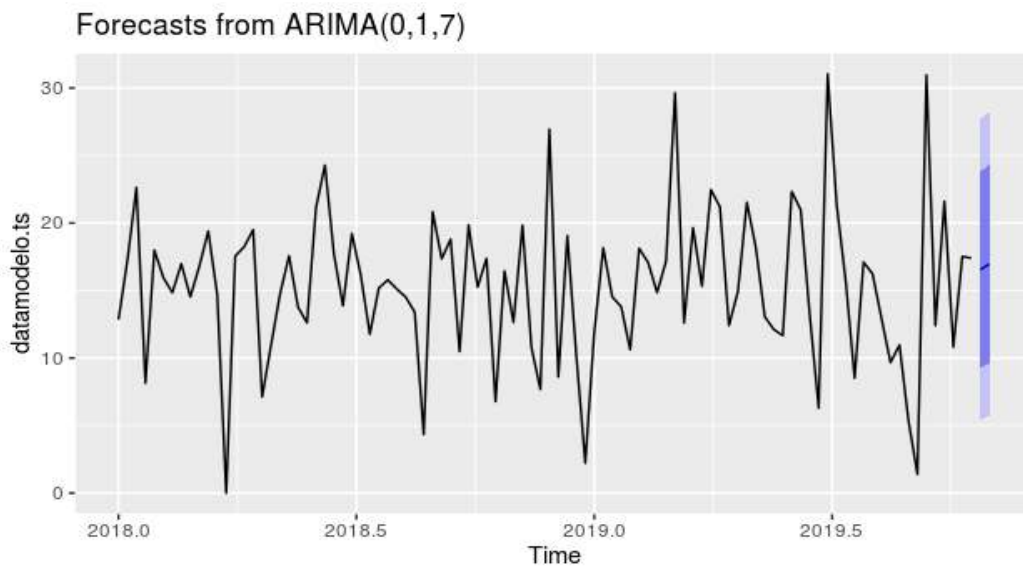
	ma1	ma2	ma3	ma4	ma5	ma6	ma7
	-1.1308	0.1151	0.0078	0.2308	-0.0730	-0.2138	0.0638
s.e.	0.1145	0.1549	0.1589	0.1439	0.1792	0.1951	0.1567

sigma^2 estimated as 32.02: log likelihood=-298.17
AIC=612.34 AICc=614.01 BIC=632.77

Training set error measures:

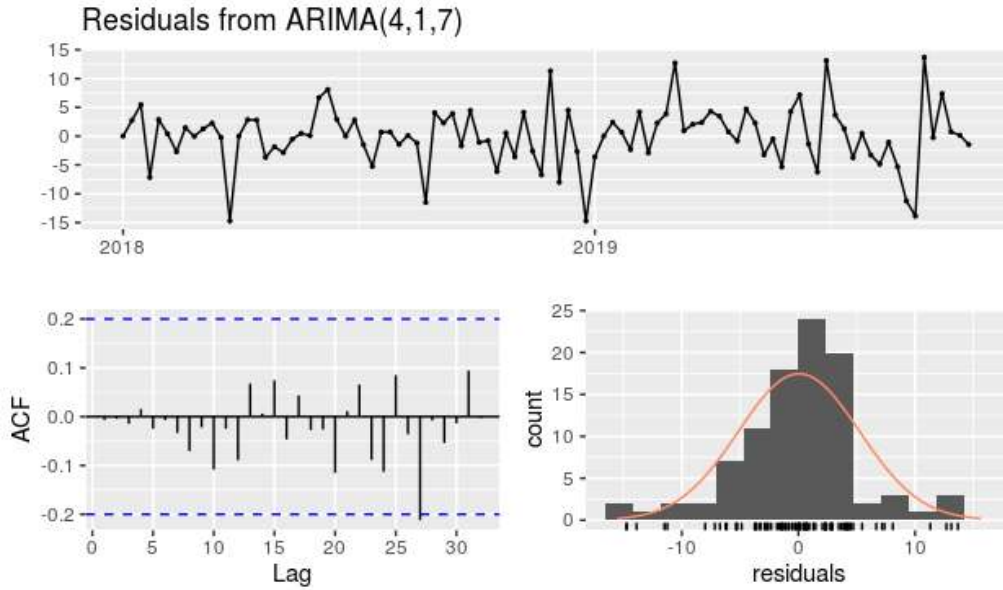
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.08292153	5.417841	3.840123	-Inf	Inf	0.6962904	-0.001840095

Ilustración 51. Pronóstico modelo Arima (0.1.7)



Modelo ARIMA (4.1.7)

Ilustración 62. Residuales ARIMA (4.1.7)



ARIMA(4,1,7)

Coefficients:

	ar1	ar2	ar3	ar4	ma1	ma2	ma3	ma4	m
a5	-1.2527	-1.2187	-0.6827	-0.5514	0.1869	-0.1030	-0.6129	0.1517	-0.39
21									
s.e.	0.2270	0.4286	0.4560	0.2568	0.2386	0.2871	0.2379	0.3221	0.29
51									
	ma6	ma7							
	0.0614	-0.2919							
s.e.	0.1177	0.1256							

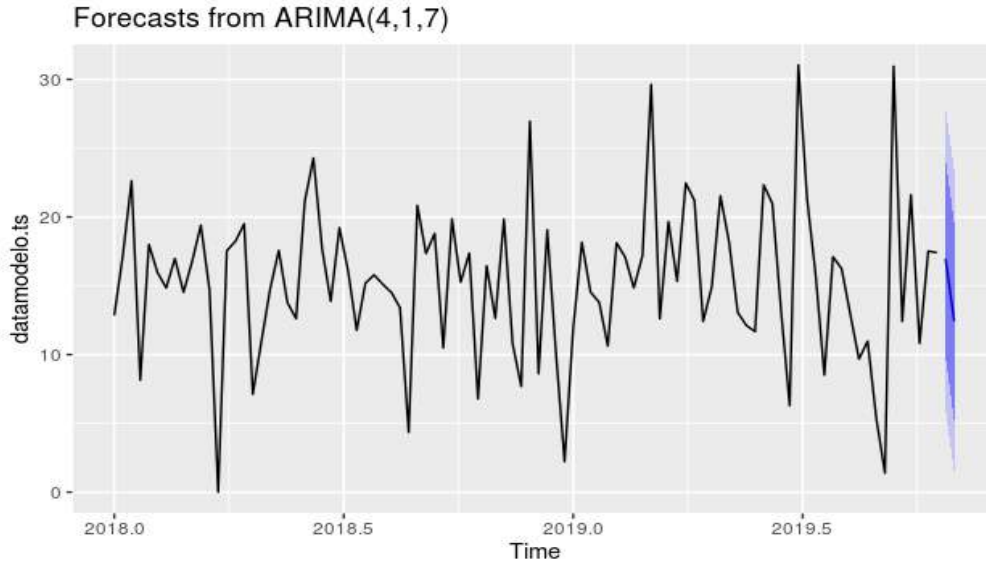
sigma² estimated as 30.52: log likelihood=-295.37

AIC=614.74 AICc=618.55 BIC=645.39

Training set error measures:

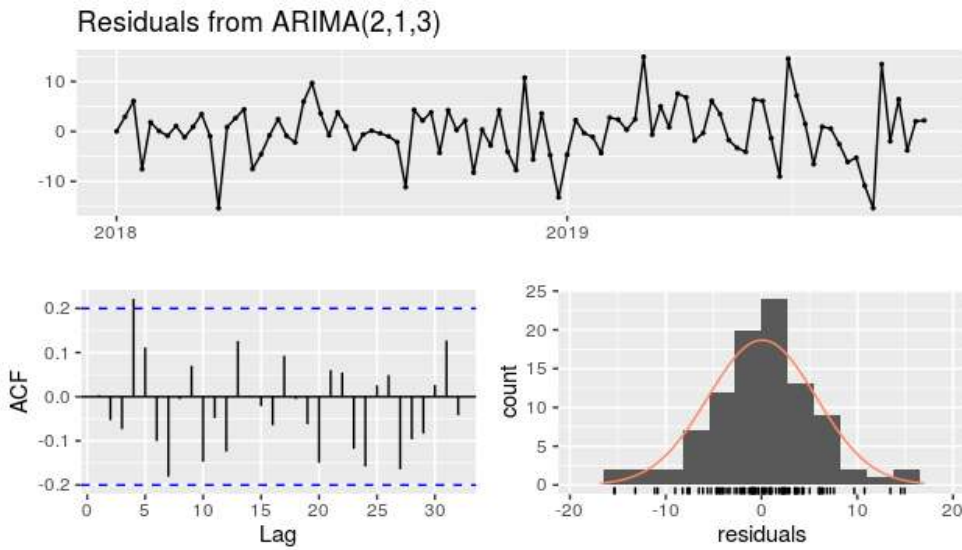
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.04987496	5.1678	3.678079	-Inf	Inf	0.6669086	-0.006902725

Ilustración 7. Pronóstico modelo Arima (4.1.7)



Modelo ARIMA (2.1.3)

Ilustración 84. Residuales ARIMA (2.1.3)



ARIMA(2,1,3)

Coefficients:

	ar1	ar2	ma1	ma2	ma3
	-0.2281	0.1245	-0.8897	-0.2997	0.1894
s.e.	NaN	NaN	NaN	NaN	NaN

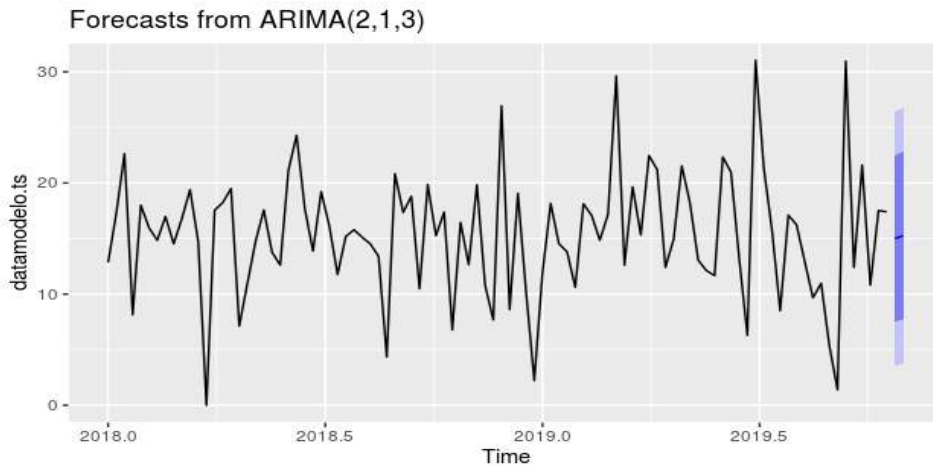
sigma² estimated as 33.68: log likelihood=-301.75

AIC=615.5 AICc=616.45 BIC=630.82

Training set error measures:

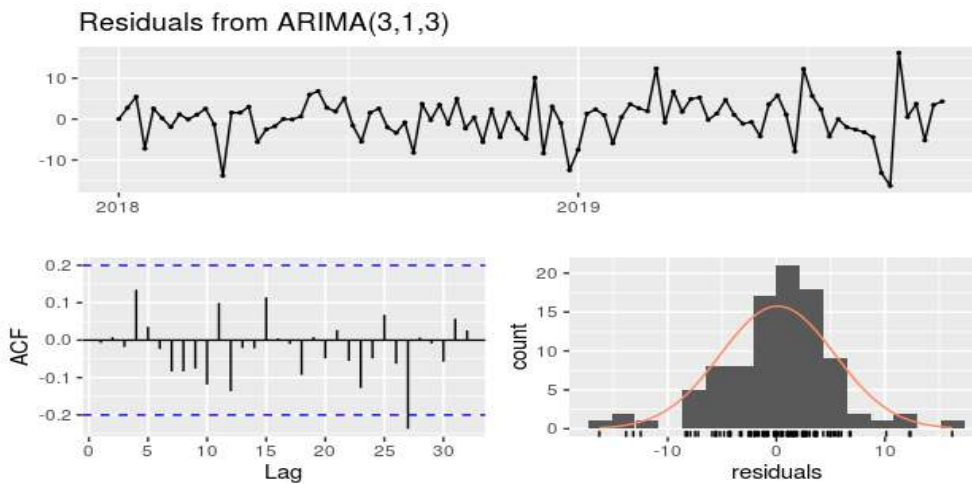
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.06378379	5.619038	4.171232	-Inf	Inf	0.756327	0.003991754

Ilustración 15. Pronóstico modelo Arima (2.1.3)



Modelo ARIMA (3.1.3)

Ilustración 96. Residuales ARIMA (3.1.3)



ARIMA(3,1,3)

Coefficients:

ar1	ar2	ar3	ma1	ma2	ma3
-----	-----	-----	-----	-----	-----

```

      0.0921  -0.9622  -0.1473  -1.2129  1.2129  -0.9999
s.e.  0.1107   0.0273   0.1058   0.0599   0.0946   0.0805

```

```

sigma^2 estimated as 29.63: log likelihood=-297.22
AIC=608.43  AICc=609.72  BIC=626.31

```

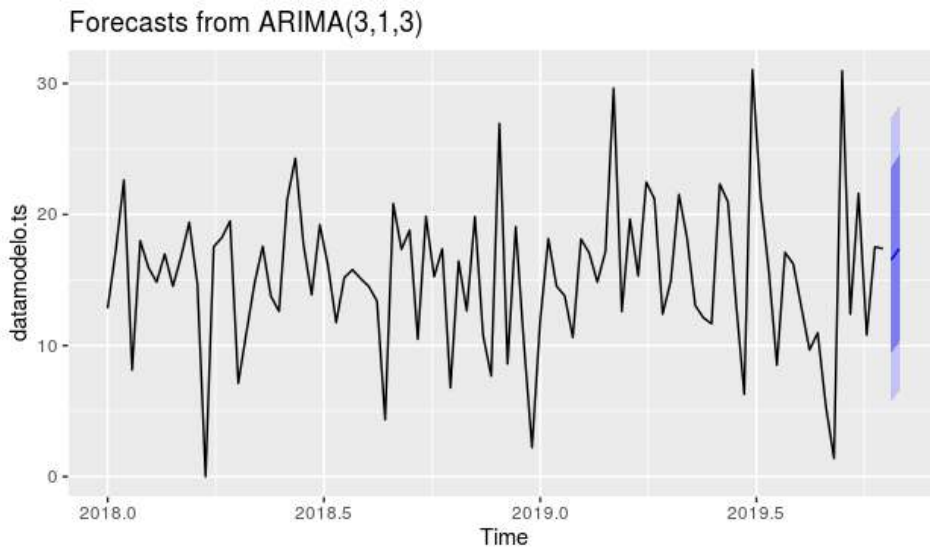
Training set error measures:

```

           ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set 0.1225873 5.241234 3.829045 -Inf  Inf 0.6942817 -0.007117792

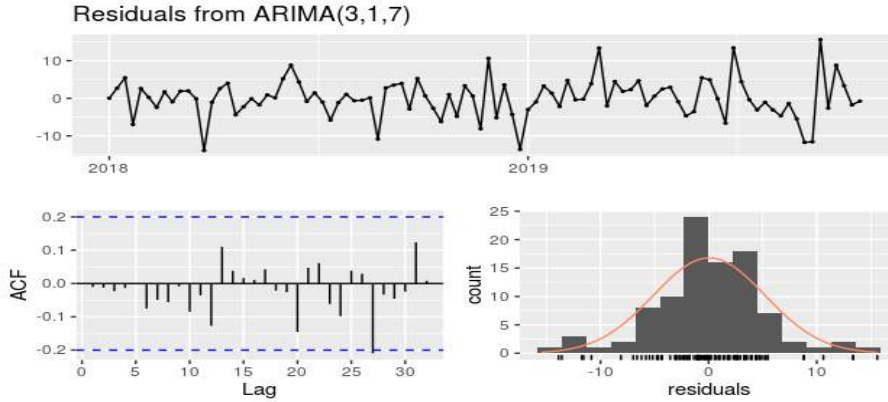
```

Ilustración 10. Pronóstico modelo Arima (3.1.3)



Modelo ARIMA (3.1.7)

Ilustración 11. Residuales ARIMA (3.1.7)



ARIMA(3,1,7)

Coefficients:

	ar1	ar2	ar3	ma1	ma2	ma3	ma4	ma5	ma
6	-0.9860	-0.9405	-0.2591	-0.0955	-0.0213	-0.7721	0.0441	0.1657	0.067
4									
s.e.	0.2947	0.2018	0.2497	0.2895	0.1892	0.1496	0.2929	0.1152	0.128
3									
		ma7							
		-0.3883							
s.e.		0.1490							

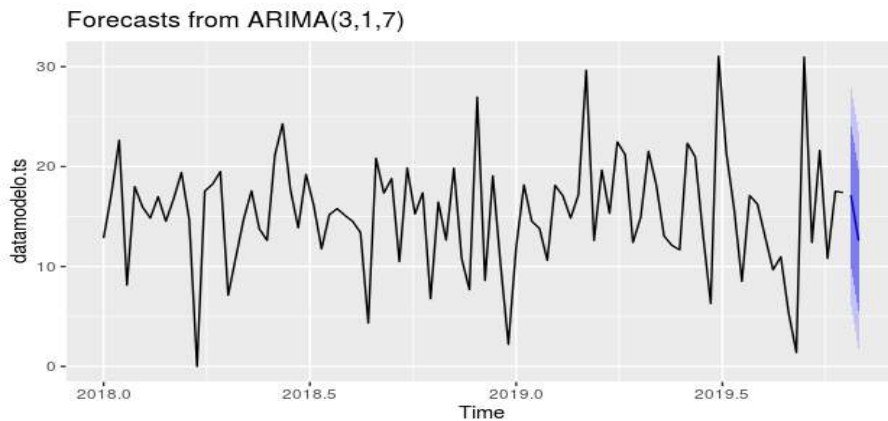
σ^2 estimated as 29.94: log likelihood=-295.38

AIC=612.75 AICc=615.93 BIC=640.84

Training set error measures:

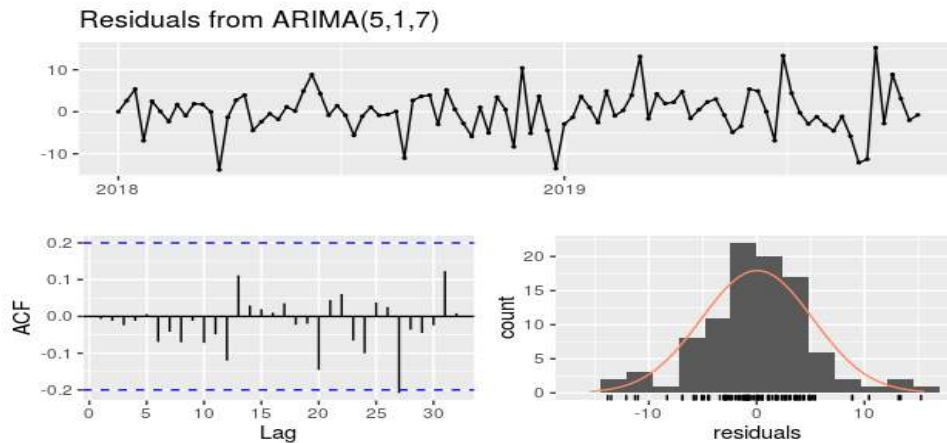
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.05185474	5.148746	3.735731	-Inf	Inf	0.6773621	-0.01024449

Ilustración 19. Pronóstico modelo Arima (3.1.7)



Modelo ARIMA (5.1.7)

Ilustración 20. Residuales ARIMA (5.1.7)



ARIMA(5,1,7)

Coefficients:

	ar1	ar2	ar3	ar4	ar5	ma1	ma2	ma3	ma4
ma4	-1.0800	-1.0504	-0.3381	-0.0310	0.0407	-0.0029	-0.0156	-0.8019	-0.0
137									
s.e.	0.4666	0.5260	0.5813	0.4522	0.4014	0.4590	0.3207	0.2566	0.3
636									
	ma5	ma6	ma7						
	0.1075	0.1269	-0.4005						
s.e.	0.2819	0.4473	0.1849						

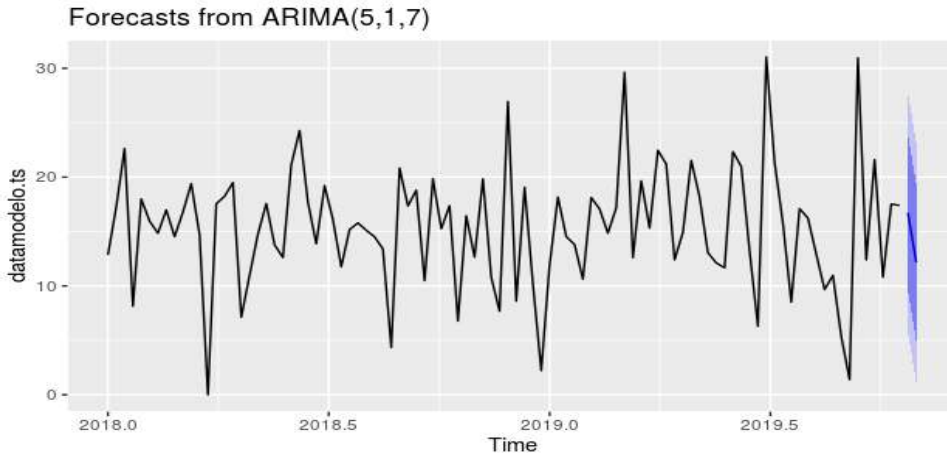
sigma² estimated as 30.57: log likelihood=-295.35

AIC=616.69 AICc=621.19 BIC=649.89

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.04307405	5.141003	3.745242	-Inf	Inf	0.6790866	-0.006550607

Ilustración 21. Pronóstico modelo Arima (5.1.7)



De acuerdo al análisis realizado a cada uno de los modelos anteriormente vistos obtuvimos los siguientes resultados:

Ilustración 122. Resultados de todos los modelos

MODELO		BIC	RMSE (Millones)	MAE (Millones)	>5% Jarque - Bera	>5% Ljung - Box
MODELO 1	ARIMA (0,1,1)	614.01	5.671.808	4.244.251	20,0%	18%
MODELO 2	ARIMA (0,1,7)	632.77	5.417.841	3.840.123	0,0%	66%
MODELO 3	ARIMA (4,1,7)	645.39	5.167.800	3.678.079	0,0%	76%
MODELO 4	ARIMA (2,1,3)	630.82	5.619.038	4.171.232	18,0%	11%
MODELO 5	ARIMA (3,1,3)	626.31	5.241.234	3.829.045	0,3%	55%
MODELO 6	ARIMA (3,1,7)	640.84	5.148.746	3.735.731	2,0%	74%
MODELO 7	ARIMA (5,1,7)	649.89	5.141.003	3.745.242	4,0%	58%

Como se puede observar para todos los modelos se les realizo sus respectivas pruebas de Ljung Box y la prueba Jarque vera, así como se evaluo el valor RMSE (Error cuadrático medio) y el MAE (diferencia absoluta promedio entre X y Y) se nota que los mejores resultados se obtuvieron en el Modelo 1 ARIMA (0.1.1), en el cual la prueba Jarque Bera da un resultado de p-value 20% y la prueba Ljung Box del p-value 18%.

El segundo modelo con mejor resultado fue el Modelo 4 ARIMA (2.1.3) este obtuvo el Jarque Vera del p-value 18% y una prueba Ljung Box del p-value 11%.

De acuerdo al diagnóstico de la ilustración 23, el mejor modelo es el ARIMA (0.1.1). Sin embargo el parámetro es -1 lo cual significa que: Los datos futuros no se pueden pronosticar, debido a que es una ecuación de pronóstico ARIMA (0.1.1) -sin constante con $\theta_1 = 1 - \alpha$. Esto significa que puede ajustar un suavizado exponencial simple especificándolo como un modelo ARIMA (0, 1,1) sin constante, y el coeficiente estimado MA (1) corresponde a 1-menos-alfa. Lo cual quiere decir que los pronósticos tenderán a quedarse atrás de las tendencias o puntos de inflexión en aproximadamente $1 / \alpha$ períodos. De ello se deduce que la edad promedio de los datos en los pronósticos de 1 período de un modelo ARIMA (0.1.1) sin constante es $\frac{1}{(1-\theta_1)}$ Robert Nau (2020), así como θ_1 se aproxima a 1, el ARIMA (0,1,1) modelo -sin-constante se convierte en un modelo de largo plazo de media móvil, y como θ_1 se acerca a θ se convierte en un modelo de caminata aleatoria sin deriva.

6. Discusión de resultados

Aunque se realizaron varios modelos ARIMA para poder lograr un pronóstico, se observó que el mejor modelo era el ARIMA (0.1.1), sin embargo este modelo no logra dar un pronóstico, esto es debido a que el parámetro es -1 lo cual quiere decir que los datos futuros no se pueden pronosticar ya que estos dependen del inicio de los recaudos y entre más tiempo pase más lejano será el valor del pronóstico a la realidad. Por esto aunque el modelo no funciono para el pronostico de recaudos, pudimos analizar las variables y se pudo identificar que hay variaciones al azar de los recaudos lo cual es difícil de predecir.

Cabe mencionar que los recaudos van muy de la mano a comportamientos no predecibles del ser humano y su entorno, es decir, al clima, a los horarios de las entidades financieras, el ambiente, e incluso hasta a temas emocionales del mismo.

7. Conclusiones y recomendaciones

- Se analizaron los datos inicialmente de la base de datos, en este análisis se puede observar que los datos muestran una serie de movimientos irregulares o al azar a corto plazo.
- Haciendo uso de la información obtenida de los recaudos semanales de la oficina Principal desde el año 2018 al 2019 se lograron 7 modelos estadísticos, sin embargo estos modelos no fueron del todo óptimos dado que obtuvimos modelos MA-1. De acuerdo al diagnóstico de cada modelo, el mejor es el ARIMA(0,1,1). Sin embargo, este modelo resultó no invertible e insuficiente para pronosticar.
- Para un próximo trabajo se recomienda realizar nuevas investigaciones sobre los modelos ARIMA con parámetro -1.

8. Referencias Bibliográficas

- Banco de la Republica. (1 de Septiembre de 2008). “Publicaciones Banrep”.
Obtenido de <http://www.banrep.gov.co>
- Giraldo, N. (2006). Notas de Clase. Series de tiempo con R. Medellín: Universidad Nacional de Colombia
- Botero Botero (2008) “Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia, Colombia
- Solórzano Tovar (2016) “Identificación y caracterización de patrones climáticos en la ciudad de Manizales, usando técnicas de series de tiempo y de conglomerados”, Colombia, IDEA
- Mogni (2013) “Módelos de Series de Tiempo con aplicaciones en la industria aerocomercial”, Colombia
- Méndez Bustamante (2013) Series de tiempo: “Una aplicación a registros hidrométricos en una cuenca del Estado de Oaxaca2”, México.
- Wei William (2006) “Time Series Analysis”, Oxford, Reino Unido, The Oxford Handbook of Quantitative Methods in Psychology

- Eberly Collage of Science (2020) “Análisis de series de tiempo aplicado”, Philadelphia, PA 19104, Estados Unidos, La Universidad Estatal de Pensilvania
- Equipo docente de Econometría II (2010), “Estimación de modelos ARIMA: Paro y empleo registrado” Madrid, España, Universidad Nacional de Educación a Distancia
- Forero, S. L. (2017), “Series de tiempo I”, Bogotá, Colombia, Fundación Universitaria Los Libertadores.
- Forero, S. L. (2017), “Series de tiempo II”, Bogotá, Colombia, Fundación Universitaria Los Libertadores.
- Alonso y Garcia Martos , (2012). “¿Los modelos MA no invertibles implican que el efecto de las observaciones pasadas aumenta con la distancia? , obtenido de <https://stats.stackexchange.com/questions/413997/do-non-invertible-ma-models-imply-that-the-effect-of-past-observations-increases>