Trabajo Práctico N° 1: **Análisis de series temporales**

Integrantes:

Ana Carla Fiori

Karina Roitman

Marcos Buccellato

Waldo Fattore

Facultad de Ingeniería, Universidad Austral

Curso: MEDGCV 2°

Profesores:

Braian Drago

Rodrigo Del Rosso

Sebastián Calcagno

14 de noviembre, 2023

Resumen Ejecutivo

El presente trabajo es un análisis exploratorio de tres series temporales mensuales que agrupan las ventas de 130 locales de un centro comercial del norte del conurbano bonaerense divididos en tres rubros. El objetivo planteado fue poder entender las características de estas series, analizar posibles modelos predictivos y ver la posibilidad de encontrar relaciones causales entre ellas.

En este análisis se planteó el uso de modelos de tipo ARIMA y VAR para realizar predicciones sobre las ventas futuras con cada una de las series. En el primer caso, en las tres series, se logró encontrar modelos con coeficientes significativos. Sin embargo, en todos los casos al realizar predicciones con los mismos, los resultados no fueron satisfactorios devolviendo bandas de confianza excesivamente amplias. Por otro lado, al analizar relaciones causales entre las series a partir del test de causalidad de Granger, no se encontraron relaciones significativas. A partir de este resultado fue inapropiado continuar con la aplicación de modelos VAR.

Consideramos que las series de tiempo elegidas no se prestan para un análisis satisfactorio con el tipo de modelos seleccionados y que es necesario analizar alternativas para encontrar modelos predictivos más performativos. Sin embargo, creemos que sería de interés poder realizar un análisis similar al planteado en este trabajo, pero utilizando series temporales con frecuencia diaria. Quizás en este caso los modelos puedan encontrar dependencias temporales más significativas y permitan también encontrar relaciones causales entre las series.

Índice de Contenido

Resumen Ejecutivo 1
Índice de Contenido 2
Introducción 3
Marco Teórico 7
Análisis de Resultados 13
Descripción de las series Originales 13
Análisis Serie Retail 14
FAC, FACP y FAS (función de autocovarianzas) 15
Tests de raíces unitarias 16
Ajuste de modelos 17
Análisis sobre los Residuos del Modelo 20
Predicciones 22
Análisis Serie Super 24
FAC, FACP y FAS 24
Tests de raíces unitarias 25
Ajuste de modelos 26
Análisis de los residuos del modelo 29
Predicciones 30
Análisis Serie Cowork 31
FAC, FACP y FAS 31
Tests de raíces unitarias 32
Ajuste de modelos 33
Análisis sobre los Residuos del Modelo 36
Predicciones 37
VAR 38
Conclusiones 39
Referencias bibliográficas 40
Apéndices 41

Introducción

Para este trabajo hemos elegido tres series de tiempo de ventas correspondientes a diferentes rubros de un centro comercial ubicado en el límite noroeste del complejo de barrios Nordelta. Es un centro comercial a cielo abierto que tiene 130 locales comerciales, 3 salas de cine, un supermercado Jumbo, un espacio de Cowork de 2100 metros cuadrados y un Fablab.

Los rubros principales de los locales del centro comercial son la indumentaria femenina y los locales gastronómicos, sin embargo, la oferta general es muy variada. El trade area del centro comercial abarca todo el complejo de barrio Nordelta, gran parte de Rincón de Milberg y Benavidez, su público principalmente consta de familias, siendo las principales consumidoras las mujeres. El centro comercial tiene plena ocupación de locales de primeras marcas y es un punto de referencia para los habitantes de la zona.

El centro comercial tiene ingresos a partir de los alquileres y las expensas que se cobra a los locales, al mismo tiempo, es el encargado de administrar los fondos comunes de promoción que son utilizados con fines de marketing y en beneficio de cada local. Si bien las negociaciones de los contratos con los locales son individuales y dependen de la marca, el rubro y otras variables, por lo general, los mismos están siempre en relación con el volumen de venta mensual de cada local. De esta forma, cada mes los locales deben reportar las ventas realizadas día por día o el total mensual dependiendo el rubro y el caso particular. Para este trabajo estaremos tomando las series temporales con las ventas de todos los locales del centro comercial como base para construir tres series temporales que agrupen tres rubros muy diferentes. El supermercado, la venta de los locales y el Cowork.

La elección de estos tres rubros responde a razones propias del negocio por un lado y al deseo de tener tres series de tiempo que tengan características diferentes para su análisis. El supermercado es uno de los locatarios principales del centro comercial, representando el mayor

volumen de facturación individual, lo que lo convierte, como cliente, en un activo importantísimo para el centro comercial. Entender y predecir las características de las ventas del mismo es vital para poder predecir cómo va a ser el rendimiento del negocio en términos generales. Por este motivo es importante estudiarlo por separado y ver qué efectos, influencias, o correlaciones pueden o no existir con el resto de las ventas. Hay que considerar que este supermercado, hasta hace poco, era la sucursal con mayor volumen de ventas de la argentina para esta marca pese a su modesta superficie y al mismo tiempo era el jugador dominante en el rubro en toda el área de influencia. Este supermercado representaba la principal fuente de abastecimiento de alimentos y productos de uso doméstico para todo el conglomerado de barrios de Nordelta y Villanueva y gran parte de rincón de Milberg. Sin embargo, en el último año han aparecido competidores importantes en la zona que han cambiado la ecuación y que implican riesgos a futuro. Esto presenta dos problemas, por un lado, es necesario entender como la competencia impacta en las ventas del supermercado diferenciando este efecto de los vaivenes propios de la estacionalidad y la tendencia general. Por eso resulta de suma importancia comprender las características de esta serie de tiempo.

El resto de los locales del centro comercial tienen características muy heterogéneas y problemáticas diferentes según el rubro. Si bien cada local no se acerca en importancia al volumen facturado por el supermercado, la combinación particular de locales (el “tenant mix”) es lo que le da al centro comercial su identidad y lo que genera el flujo de visitantes que sostiene al negocio y posiblemente potencia (algo a estudiar) las ventas del supermercado.

Pese a las diferencias que se dan en cada rubro, entender cómo es la evolución de ventas de los locales en general del centro comercial es de vital importancia. Por un lado, ayuda a entender la evolución general del negocio ya que las ventas están directamente en relación con lo que se factura. Comprender estacionalidades, ciclos y tendencias generales ayudan a entender cómo evoluciona el negocio y a tomar posibles medidas de ajuste. Por ejemplo, si el

nivel general de ventas decae, puede ser necesario reforzar las acciones de marketing o analizar a la competencia. Más allá de la facturación, el centro comercial renueva su tenant mix continuamente, en las nuevas negociaciones con marcas y locales, el volumen de ventas es muy importante a la hora de obtener nuevas operaciones, las marcas analizan mucho esto antes de firmar un contrato. Si bien hay especificidades por rubro, es muy importante entender cómo funcionan las ventas del centro comercial para poder tener información precisa a la hora de negociar estos nuevos contratos.

Una de las preguntas que desvela a muchas marcas es entender cuál puede ser su hipotético volumen de facturación y en qué momentos, esto es vital para poder dimensionar el tamaño de la inversión a realizar. Si bien es cierto que, desde la perspectiva del centro comercial, exagerar las cifras puede ser un buen argumento de ventas, lo cierto es que una vez firmado el contrato, si las mismas no se materializan, la relación entre el centro comercial y la marca se ven deterioradas lo que muchas veces lleva a incumplimientos y litigios. Por tal motivo, poder presentar de forma consistente y respaldada por datos y análisis las cifras del centro comercial es un factor importante para negociar la entrada y resolver litigios. Por otro lado, como ya mencionamos, las ventas de los locales son declaradas por cada local de forma individual, este proceso tiene varios mecanismos de control para evitar fraude, pero no es infalible. Poder entender cuando hay una sospecha de que los datos se están cargando de forma fraudulenta es importante y un paso para hacerlo es entender cuáles son las características generales de las series de ventas del centro comercial. Entendemos que este análisis, para poder ser de mayor utilidad, debería hacerse por rubro e incluso a nivel de local individual, creemos que un primer análisis general es importante.

Por último, presentamos la serie de ventas del Cowork. Este negocio en particular si bien, en volumen, no es tan significativo, es parte de un concepto importante dentro del negocio de los centros comerciales que son las “anclas”. Las “anclas” en un centro comercial son aquellos

negocios que no producen una facturación importante o un buen rendimiento por metro cuadrado, pero que generan una afluencia de gente que derrama ventas en el resto de los locales.

El Cowork en cuestión es un negocio que se desarrolló hace cinco años atrás con la idea de generar un ancla nueva para el centro comercial aprovechando una importante superficie que se liberó en la planta alta producto del cierre de una gran tienda de artículos generales. La idea de poner un Cowork dentro del centro comercial fue una apuesta que en ese momento se consideró riesgosa y generó mucha polémica dentro del grupo accionista.

Siendo el volumen de facturación por metro cuadrado muy inferior al resto de los locales, aún ahora, y pese a que el Cowork opera a máxima capacidad, se cuestiona la pertinencia del mismo. Por tal motivo es de interés entender si la existencia del Cowork genera algún efecto sobre la performance del centro comercial en general.

Una primera aproximación para estudiar esto es analizar su serie de ventas en relación con los otros grandes rubros del centro comercial. Seguramente el desagregado en rubros más específicos sea pertinente en este caso, pero nuevamente una primera aproximación general es necesaria al menos para ver posibles efectos generales.

Somos conscientes también que quizás la variable ventas del Cowork no sea la más apropiada para medir este efecto (tal vez la afluencia de personas es más importante), pero creemos importante hacer el análisis. Entendemos también que dadas las características tan diferentes de este negocio del resto de los locales y el supermercado, las características de la serie de tiempo seguramente sean muy diferentes y sea interesante el contraste en su análisis.

Marco Teórico

Una serie temporal es una secuencia de observaciones o medidas recopiladas a lo largo del tiempo, generalmente en intervalos regulares. Estas observaciones están ordenadas cronológicamente y se utilizan para analizar patrones, tendencias y variaciones temporales en datos.

Las series temporales pueden exhibir una variedad de patrones, y a menudo es útil dividir una serie temporal en varios componentes, cada uno de los cuales representa un patrón subyacente. Las series se pueden descomponer a partir de cuatro componentes que son inobservables; tendencia, estacionalidad, ciclos, y movimiento irregular (representa la variabilidad aleatoria o no sistemática en una serie temporal que no se puede atribuir a tendencias, estacionalidad o ciclos).

Una serie temporal estacionaria es aquella cuyas propiedades, como la media y la varianza, no dependen del momento en que se observa la serie. Por lo tanto, las series temporales con tendencia o estacionalidad no son estacionarias, ya que dichas características afectan al valor de la serie temporal en distintos momentos. (Hyndman & Athanasopoulos, 2018).

Dentro de la categoría de modelos de procesos estacionarios, los autorregresivos son considerados como los más simples. Estos modelos, denominados AR (Autoregressive Models), generalizan la noción de regresión para expresar la dependencia lineal entre dos variables aleatorias. (Peña, 2005).

El término autorregresión implica que se trata de una regresión de la variable contra sí misma. En contraste con un modelo de regresión múltiple, donde se pronostica la variable de interés mediante una combinación lineal de predictores, en un modelo de autorregresión, la

predicción se realiza a través de una combinación lineal de valores anteriores de la misma variable. (Hyndman & Athanasopoulos, 2018)

Un modelo autorregresivo de orden p tiene la siguiente ecuación:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

Donde Et es un ruido blanco, c es una constante o intercepto que representa el valor medio esperado de la serie temporal y ϕ son los coeficientes autorregresivos asociados con los valores pasados de la serie temporal. Cada coeficiente indica la contribución del valor de la serie en el tiempo t-*i*, al valor de la serie en el tiempo t. Al ajustar los parámetros $\phi_1, \ldots, \phi_p$ se pueden obtener diferentes patrones de series temporales. Es importante destacar que la varianza del término de error sólo afectará a la escala de la serie, no a los patrones inherentes a ésta. (Hyndman & Athanasopoulos, 2018)

Un enfoque alternativo es el uso de un modelo de media móvil MA (Moving Average Model), el cual, en lugar de incorporar valores anteriores de la variable objetivo en la predicción, utiliza los errores de predicción pasados. Este enfoque guarda similitudes con un modelo de regresión, pero en lugar de depender de variables anteriores, se enfoca en los errores de predicción previos. (Hyndman & Athanasopoulos, 2018)

Un modelo de medias móviles de orden q tiene la siguiente ecuación:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

Donde Et es un ruido blanco, c es una constante que representa el valor medio esperado de la serie temporal y θ son los coeficientes asociados con los términos de error pasados.

Cambiando los parámetros $\theta_1, \ldots, \theta_q$ se alcanzan distintos patrones de series temporales. Cómo en los modelos autorregresivos, la varianza del término de error no afecta a los patrones de la serie. (Hyndman & Athanasopoulos, 2018)

Un proceso MA (q) es siempre estacionario, por ser la suma de procesos estacionarios. Permanecen estacionarios independientemente de los valores específicos de los coeficientes, ya que se construyen como combinaciones lineales de errores, que se consideran ruido blanco.

Otro modelo es el modelo “ARMA", que es un acrónimo para referirse a "Autoregressive Moving Average". El modelo ARMA combina los dos componentes mencionados anteriormente: AR y MA.

Si el modelo ARMA tiene q=0, el proceso es AR(p) es puro. Asimismo, si p=0, el proceso es MA(q) puro. En el modelo ARMA, se limita la consideración a aquellos modelos donde todas las raíces características de ϕ están dentro del círculo unitario.

Una forma de hacer estacionaria una serie temporal no estacionaria es calcular las diferencias entre observaciones consecutivas, proceso que se conoce como diferenciación. Las transformaciones, como los logaritmos, pueden ayudar a estabilizar la varianza de una serie. La diferenciación puede ayudar a estabilizar la media, eliminando los cambios en el nivel de una serie temporal y, por lo tanto, descartando o reduciendo la tendencia y la estacionalidad.

Además de observar el gráfico de la serie original, el gráfico de función de autocorrelación, FAC es útil para identificar series temporales no estacionarias.

Un gráfico FAC muestras las autocorrelaciones que miden la relación entre yt y yt-k para diferentes valores de k. Si yt y yt-1 están correlacionadas, entonces yt-1 e yt-2 también deberían estarlo. De esta manera, yt y yt-2 pueden que estén correlacionadas, simplemente porque ambas

están conectadas a yt-1 y no porque exista una información contenida en yt-2 que pueda ser usada para predecir yt. Este problema se soluciona con la función de autocorrelación parcial.

La Función de Autocorrelación Parcial (FACP) evalúa la relación entre yt e yt−k después de eliminar los rezagos 1, 2,..., k-1. La primera autocorrelación parcial es idéntica a la primera autocorrelación, ya que no hay lags intermedios que eliminar. Cada autocorrelación parcial puede estimarse como el último coeficiente en un modelo autorregresivo.

Para una serie temporal estacionaria, la FAC caerá a cero con relativa rapidez, mientras que el gráfico de FAC de los datos no estacionarios disminuye lentamente. (Hyndman & Athanasopoulos, 2018).

Una forma de determinar más objetivamente si es necesaria la diferenciación, es utilizar una prueba de raíz unitaria. Una de las pruebas más comunes es la de Dickey-Fuller. La hipótesis nula indica que la serie temporal tiene una raíz unitaria, lo que implica que no es estacionaria. Si se rechaza dicha hipótesis, el test sugiere que la serie requiere diferenciación para transformarse en estacionaria. Además de las pruebas de Dickey-Fuller existen las pruebas de Phillips-Perron y KPSS (Kwiatlsowski, Phillips, Schmidt y Shin), en las cuales, por el contrario, se establece como hipótesis nula que la serie de tiempo es estacionaria y como hipótesis alternativa que tiene una raíz unitaria. (Peña, 2005)

Si se combina la diferenciación con la autorregresión y un modelo de medias móviles, se obtiene un modelo **ARIMA** no estacional. ARIMA es el acrónimo de AutoRegressive Integrated Moving Average (media móvil integrada autorregresiva), donde la integración es la inversa a la diferenciación. Si al analizar el test de raíces unitarias se concluye que una o más de estas raíces son mayores o iguales a la unidad, se clasifica la secuencia {yt} como un proceso integrado, y se le denomina modelo de media móvil autorregresiva integrada. (Enders, 2014).

El modelo completo puede escribirse de la siguiente forma,

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Donde y'_t es la serie diferenciada, que puede haberse diferenciado más de una vez. Los predictores de la derecha incluyen tanto los valores retardados (lags) como los errores de los lags. A este modelo se lo conoce como ARIMA (*p, d, q*), donde:

p= es el orden de la parte autorregresiva,

d= es el grado u orden de diferenciación,

q= es el orden de la parte de medias móviles

Los intervalos de predicción de los modelos ARIMA se basan en la hipótesis de que los residuos están incorrelacionados y se distribuyen normalmente. Si alguna de estas hipótesis no se cumple, los intervalos de predicción pueden ser incorrectos. Por esta razón, antes de elaborar intervalos de predicción, se debe trazar siempre la función de autocorrelación y el histograma de los residuos para comprobar los supuestos. (Hyndman & Athanasopoulos, 2018)

Cuando existe dependencia estacional, se puede generalizar el modelo ARIMA incorporando además de la dependencia regular, que es la asociada a los intervalos de medida de la serie, la dependencia estacional, que es la asociada a observaciones separadas por m períodos. (Peña, 2005)

Un modelo ARIMA estacional o **SARIMA** se forma incluyendo términos estacionales adicionales en los modelos ARIMA:

$$\text{ARIMA} \quad \underbrace{(p,d,q)}_{\uparrow} \quad \underbrace{(P,D,Q)_m}_{\uparrow}$$

Parte no estacional Parte estacional

P es el orden de autorregresión estacional,
D es el número de diferencias estacionales,
Q es el orden de la media móvil estacional
m es la longitud del período estacional.

La parte estacional del modelo consiste en términos que son similares a los componentes no estacionales del modelo, pero involucra desplazamientos hacia atrás del período estacional.

Cuando se quiere caracterizar las interacciones simultáneas entre un grupo de series, se utiliza un modelo del tipo vector autorregresivo (**VAR**). Si las series son estacionarias, se pronostican ajustando un VAR a los datos directamente, conocido como "VAR en niveles". Si las series no son estacionarias, se toman las diferencias de los datos para hacerlos estacionarios y ajustamos un modelo VAR, conocido como "VAR en diferencias". En ambos casos, los modelos se estiman ecuación por ecuación utilizando el principio de mínimos cuadrados.

La otra posibilidad es que las series sean no estacionarias pero estén cointegradas, lo que significa que existe una combinación lineal de ellas que es estacionaria. En este caso, debe incluirse una especificación VAR que incluya un mecanismo de corrección del error (normalmente denominado modelo vectorial de corrección de errores) y deben utilizarse métodos de estimación alternativos a la estimación por mínimos cuadrados. (Hyndman & Athanasopoulos, 2018)

Análisis de Resultados

Descripción de las series Originales

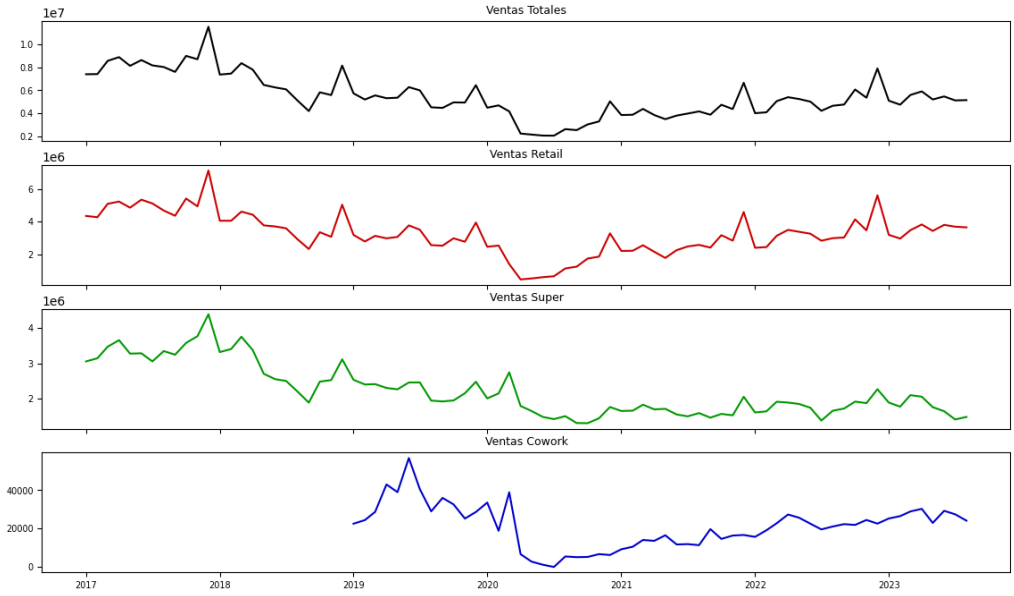


Figura 1

Gráficos de las Series Originales

Las series temporales elegidas corresponden a datos obtenidos de un centro comercial. Se eligieron los datos pertenecientes a las ventas para Retail, Supermercado, y Cowork para el análisis.

El período de tiempo estudiado para Retail y Supermercado comprende desde enero 2015 hasta agosto 2023. Para la serie temporal Cowork se cuenta con menos datos, abarcando enero 2019 hasta agosto 2023.

Al analizar los gráficos obtenidos, se intuye que para la serie Retail y Super se podría aplicar una primera diferenciación. Esto se estima porque si bien la serie no aparenta ser estacionaria, no presenta saltos bruscos en ningún momento.

Por otra parte, en las series Super y Retail se detecta una estacionalidad, con las ventas alcanzando niveles notablemente mayores sobre el fin de año. En base a esto se puede esperar una estacionalidad con período 12, ya que la información está mensualizada.

Una serie temporal estacionaria es aquella cuyas propiedades no dependen del momento en que se observa la serie. Por lo tanto, las series temporales con tendencias o con estacionalidad, no son estacionarias: la tendencia y la estacionalidad afectarán al valor de la serie temporal en distintos momentos. Por el contrario, una serie de ruido blanco es estacionaria: no importa cuándo se observe, debería tener el mismo aspecto en cualquier momento. (Hyndman & Athanasopoulos, 2018)

Análisis Serie Retail

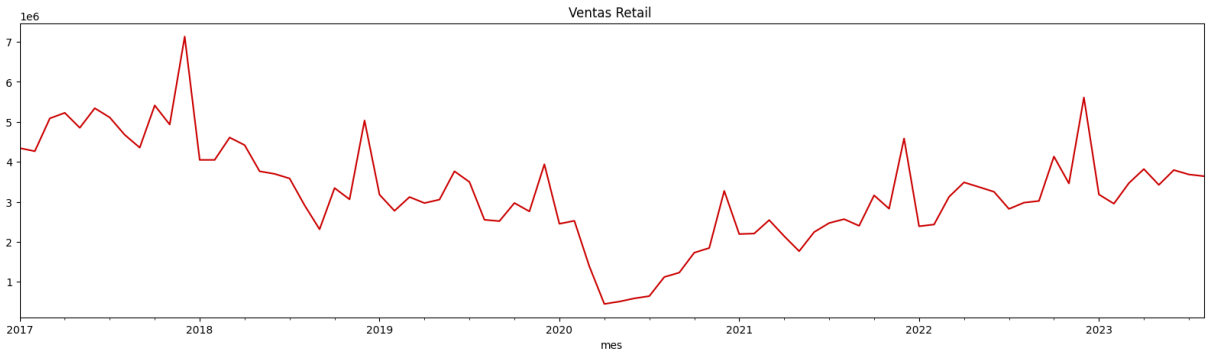


Figura 2 Gráfico Ventas Retail

FAC, FACP y FAS (función de autocovarianzas)

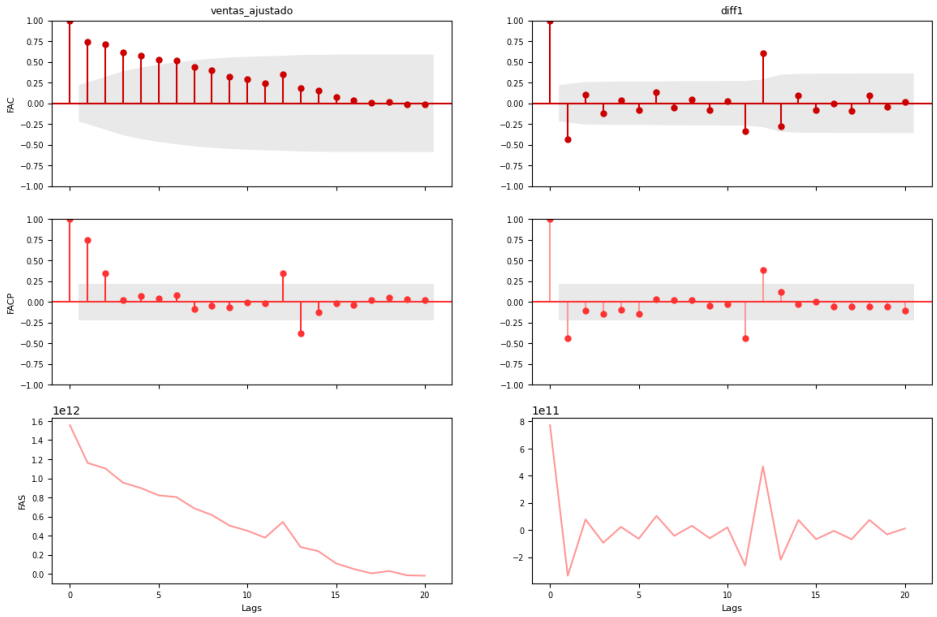


Figura 3

Análisis de la serie Retail

En la función de autocorrelación se observa un decrecimiento lineal lo que indicaría la no-estacionariedad de la serie. Para una serie temporal estacionaria, la FAC decaerá a cero con relativa rapidez, mientras que el FAC de los datos no estacionarios disminuye lentamente. (Hyndman & Athanasopoulos, 2018)

En contrapartida, al observar la serie de su primera diferencia, puede observarse un decaimiento inmediato, con correlaciones posibles en el primer y doceavo lag, lo cual sustenta la teoría de una posible estacionalidad anual. Se puede suponer que un modelo MA 1 sería un buen punto de partida para modelar un ajuste.

Observando la FACP, se nota una caída exponencial en la serie original y su primera diferencia, particularmente en los primeros dos lags más el doceavo y treceavo, lo cual sugiere quizás un modelo AR (2) y la misma estacionalidad antedicha.

La FACP para un proceso AR(p) puro debe cortar a cero para todos los rezagos mayores que p. Esta es una característica útil de la FACP que puede ayudar en la identificación de un modelo AR(p). El PACF de un proceso estacionario ARMA(p, q) debe decaer hacia cero, de manera directa u oscilatoria, a partir del lag p (Enders, 2014)

Tests de raíces unitarias

Se aplicaron las pruebas de hipótesis de Dickey Fuller, Phillips Perron y KPSS con el objetivo de poner a prueba la conjetura de que las raíces del Polinomio Característico son iguales a la unidad. Se trata de pruebas estadísticas de hipótesis de estacionariedad diseñadas para determinar si es necesaria la diferenciación.

Augmented Dickey-Fuller

Estadistico ADF	p-Valor	Estacionaridad	Modo
-2.1134	0.2391	No	Constante sola
-1.5532	0.8102	No	Constante y Tendencia Lineal
-2.7274	0.4387	No	Constante y Tendencia Lineal y Cuadratica
-0.7234	0.4027	No	Sin Contante ni Tendencia

Phillips-Perron

Estadistico PP	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
-0.87	0.3400	12	No	1	n-No incluye término independiente ni lineal
-3.62	0.0055	12	Si	1	c-Con término independiente, Sin término lineal
-3.96	0.0100	12	Si	1	ct-Incluye ambos términos

KPSS

Estad. KPSS	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
0.4973	0.0423	5	No	1	c - estacionarios alrededor de una constante.
0.2970	0.0100	5	No	1	ct - estacionarios alrededor de una tendencia.

Tabla 1 Salida Raíces Unitarias Ventas Retail

En el test de Dickey Fuller, la hipótesis nula es que la serie no es estacionaria, y al obtener p-valores mayores a 0.05, no se rechaza la hipótesis nula, por lo que se concluye que la serie no es estacionaria.

Por el contrario, para los tests de Phillips Perron y KPPS, la hipótesis nula indica que la serie es estacionaria. Se obtuvieron p-valores menores a 0.05, por lo que rechazamos la hipótesis nula y se deduce la no estacionariedad.

Para continuar con el análisis, aplicamos la primera diferencia a la serie.

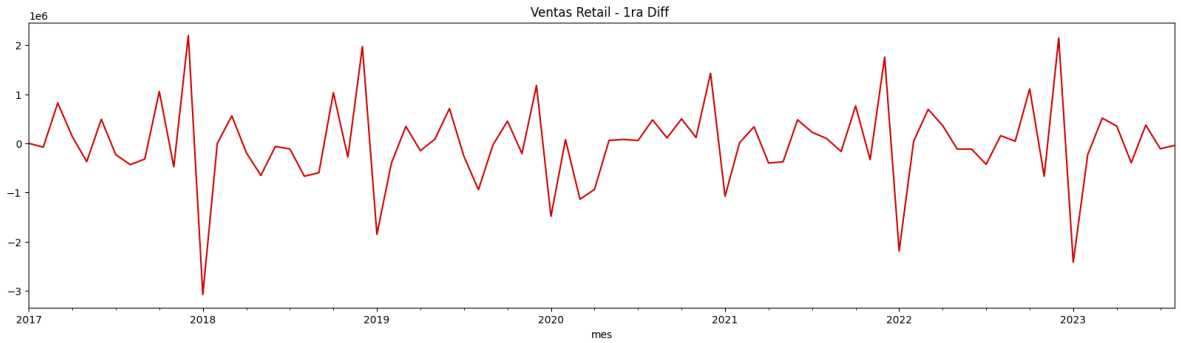


Figura 4 Gráficos Ventas Retail 1era Diff

Augmented Dickey-Fuller

Estadistico ADF	p-Valor	Estacionaridad	Modo
-2.7607	0.0641	No	Constante sola
-4.3799	0.0024	Si	Constante y Tendencia Lineal
-4.0866	0.0245	Si	Constante y Tendencia Lineal y Cuadratica
-2.6607	0.0076	Si	Sin Contante ni Tendencia

Phillips-Perron

Estadistico PP	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
-18.04	0.0000	12	Si	0	n-No incluye término independiente ni lineal
-17.99	0.0000	12	Si	0	c-Con término independiente, Sin término lineal
-19.16	0.0000	12	Si	0	ct-Incluye ambos términos

KPSS

Estad. KPSS	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
0.1135	0.1000	6	Si	0	c - estacionarios alrededor de una constante.
0.0596	0.1000	6	Si	0	ct - estacionarios alrededor de una tendencia.

Tabla 2 Salida Raíces Unitarias Ventas Retail 1era Diff

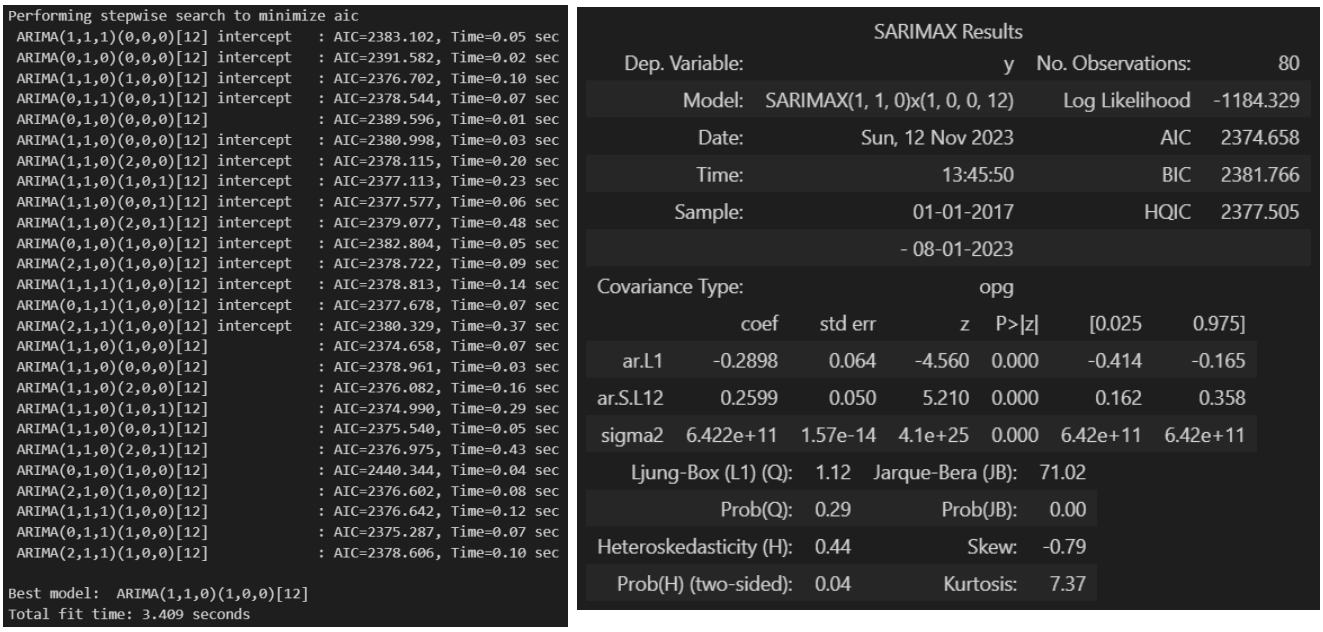
En relación con la serie diferenciada, se observa que la mayoría de los test indican que la serie es estacionaria por lo que no se requiere una nueva diferenciación.

Ajuste de modelos

El proceso auto-ARIMA trata de identificar los parámetros óptimos para un modelo ARIMA, estableciendo un único modelo ARIMA ajustado. Dicho proceso funciona realizando pruebas de diferenciación (Kwiatkowski-Phillips-Schmidt-Shin, Augmented Dickey-Fuller o Phillips-Perron)

para determinar el orden de diferenciación, d. Con el objetivo de encontrar el mejor modelo, auto-ARIMA optimiza un criterio de información dado, Criterio de Información de Akaike, Criterio de Información de Akaike Corregido, Criterio de Información Bayesiano, Criterio de Información de Hannan-Quinn, y devuelve el modelo que minimiza el valor. (pmdarima.arima.auto_arima, s.f.)

El mejor modelo obtenido implementando auto-ARIMA es ARIMA (1,1,0)(1,0,0)[12] con un Akaike de 2374.66. Se observa que todos los componentes de dicho modelo son significativos de manera individual.



```
Performing stepwise search to minimize aic
 ARIMA(1,1,1)(0,0,0)[12] intercept   : AIC=2383.102, Time=0.05 sec
 ARIMA(0,1,0)(0,0,0)[12] intercept   : AIC=2391.582, Time=0.02 sec
 ARIMA(1,1,0)(1,0,0)[12] intercept   : AIC=2376.702, Time=0.10 sec
 ARIMA(0,1,1)(0,0,1)[12] intercept   : AIC=2378.544, Time=0.07 sec
 ARIMA(0,1,0)(0,0,0)[12]             : AIC=2389.596, Time=0.01 sec
 ARIMA(1,1,0)(0,0,0)[12] intercept   : AIC=2380.998, Time=0.03 sec
 ARIMA(1,1,0)(2,0,0)[12] intercept   : AIC=2378.115, Time=0.20 sec
 ARIMA(1,1,0)(1,0,1)[12] intercept   : AIC=2377.113, Time=0.23 sec
 ARIMA(1,1,0)(0,0,1)[12] intercept   : AIC=2377.577, Time=0.06 sec
 ARIMA(1,1,0)(2,0,1)[12] intercept   : AIC=2379.077, Time=0.48 sec
 ARIMA(0,1,0)(1,0,0)[12] intercept   : AIC=2382.804, Time=0.05 sec
 ARIMA(2,1,0)(1,0,0)[12] intercept   : AIC=2378.722, Time=0.09 sec
 ARIMA(1,1,1)(1,0,0)[12] intercept   : AIC=2378.813, Time=0.14 sec
 ARIMA(0,1,1)(1,0,0)[12] intercept   : AIC=2377.678, Time=0.07 sec
 ARIMA(2,1,1)(1,0,0)[12] intercept   : AIC=2380.329, Time=0.37 sec
 ARIMA(1,1,0)(1,0,0)[12]             : AIC=2374.658, Time=0.07 sec
 ARIMA(1,1,0)(0,0,0)[12]             : AIC=2378.961, Time=0.03 sec
 ARIMA(1,1,0)(2,0,0)[12]             : AIC=2376.082, Time=0.16 sec
 ARIMA(1,1,0)(1,0,1)[12]             : AIC=2374.990, Time=0.29 sec
 ARIMA(1,1,0)(0,0,1)[12]             : AIC=2375.540, Time=0.05 sec
 ARIMA(1,1,0)(2,0,1)[12]             : AIC=2376.975, Time=0.43 sec
 ARIMA(0,1,0)(1,0,0)[12]             : AIC=2440.344, Time=0.04 sec
 ARIMA(2,1,0)(1,0,0)[12]             : AIC=2376.602, Time=0.08 sec
 ARIMA(1,1,1)(1,0,0)[12]             : AIC=2376.642, Time=0.12 sec
 ARIMA(0,1,1)(1,0,0)[12]             : AIC=2375.287, Time=0.07 sec
 ARIMA(2,1,1)(1,0,0)[12]             : AIC=2378.606, Time=0.10 sec

Best model:  ARIMA(1,1,0)(1,0,0)[12]
Total fit time: 3.409 seconds
```

SARIMAX Results

Dep. Variable:	y	No. Observations:	80
Model:	SARIMAX(1, 1, 0)x(1, 0, 0, 12)	Log Likelihood	-1184.329
Date:	Sun, 12 Nov 2023	AIC	2374.658
Time:	13:45:50	BIC	2381.766
Sample:	01-01-2017	HQIC	2377.505
	- 08-01-2023		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.2898	0.064	-4.560	0.000	-0.414	-0.165
ar.S.L12	0.2599	0.050	5.210	0.000	0.162	0.358
sigma2	6.422e+11	1.57e-14	4.1e+25	0.000	6.42e+11	6.42e+11

Ljung-Box (L1) (Q):	1.12	Jarque-Bera (JB):	71.02
Prob(Q):	0.29	Prob(JB):	0.00
Heteroskedasticity (H):	0.44	Skew:	-0.79
Prob(H) (two-sided):	0.04	Kurtosis:	7.37

Salida 1

Se dividió la serie considerando un 80% para train y 20% para test. La salida obtenida proporciona un Akaike de 1899.02 y todos los componentes del modelo son significativos de manera individual.

```
                                SARIMAX Results
==========================================================================================
Dep. Variable:                    ventas_ajustado   No. Observations:                   64
Model:            SARIMAX(1, 1, 0)x(1, 0, 0, 12)   Log Likelihood                -946.511
Date:                           Sat, 11 Nov 2023   AIC                           1899.022
Time:                                   21:36:30   BIC                           1905.452
Sample:                             01-01-2017     HQIC                          1901.551
                                  - 04-01-2022
Covariance Type:                            opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.2737      0.075     -3.627      0.000      -0.422      -0.126
ar.S.L12       0.1545      0.056      2.738      0.006       0.044       0.265
sigma2      6.541e+11   1.64e-14   3.99e+25      0.000    6.54e+11    6.54e+11
===================================================================================
Ljung-Box (L1) (Q):                   0.99   Jarque-Bera (JB):                 20.52
Prob(Q):                              0.32   Prob(JB):                          0.00
Heteroskedasticity (H):               0.45   Skew:                             -0.51
Prob(H) (two-sided):                  0.07   Kurtosis:                          5.60
===================================================================================
```

```
MSE:    309256970025
MAE:    357021
RMSE:   556109
MAPE:   0.095
```

Salida 2

Adicionalmente se construyeron modelos de manera manual, en base a lo observado a las funciones de autocorrelación y autocorrelación parcial.

```
Modelo 0-Akaike: 1939.3972-     AR 1
Modelo 1-Akaike: 2072.397-      MA 1
Modelo 2-Akaike: 1929.4622-     ARMA 1-1
Modelo 3-Akaike: 1930.729-      AR 1,2
Modelo 4-Akaike: 2057.2077-     MA 1,2
Modelo 5-Akaike: 8.0-   AR 1,2,12
Modelo 6-Akaike: 2028.2274-     MA 1,2,12
Modelo 7-Akaike: 1898.9437-     ARIMA 1-1-0
Modelo 8-Akaike: 1899.373-      ARIMA 0-1-1
Modelo 9-Akaike: 1900.9254-     ARIMA 1-1-1
Modelo 10-Akaike: 1900.967-     ARIMA 1,2-1-0
Modelo 11-Akaike: 1901.1304-    ARIMA 0-1-1,2
Modelo 12-Akaike: 51.0803-      ARIMA 1,2,12-1-0
Modelo 13-Akaike: 1902.9348-    ARIMA 0-1-1,2,12
Modelo 14-Akaike: 1899.0221-    SARIMA 1-1-0 Season AR 1
Modelo 15-Akaike: 1899.1972-    SARIMA 1-1-0 Season MA 1
Modelo 16-Akaike: 1900.5622-    SARIMA 1-1-0 Season ARMA 1-1
Modelo 17-Akaike: 6.0-  AR 1,12
Modelo 18-Akaike: 1938.6846-    ARIMA 1,12-1-0
Modelo 19-Akaike: 2045.4012-    MA 1,12
Modelo 20-Akaike: 1900.9043-    ARIMA 0-1-1,12
Modelo 21-Akaike: 2025.9619-    ARIMA 0-0-0 Season AR 1
Modelo 22-Akaike: 2025.9619-    AR 12
Modelo 23-Akaike: 2078.8583-    MA 12
Modelo 24-Akaike: 1952.9135-    ARIMA 12-1-0
Modelo 25-Akaike: 1904.7435-    ARIMA 0-1-12
```

```
                                SARIMAX Results
==========================================================================================
Dep. Variable:                    ventas_ajustado   No. Observations:                   64
Model:                 SARIMAX(0, 1, [12])   Log Likelihood                -950.372
Date:                           Sat, 11 Nov 2023   AIC                           1904.744
Time:                                   21:36:39   BIC                           1909.030
Sample:                             01-01-2017     HQIC                          1906.429
                                  - 04-01-2022
Covariance Type:                            opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L12         0.2075      0.038      5.518      0.000       0.134       0.281
sigma2      6.077e+11    2.4e-14   2.54e+25      0.000    6.08e+11    6.08e+11
===================================================================================
Ljung-Box (L1) (Q):                   9.31   Jarque-Bera (JB):               115.17
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.45   Skew:                            -1.31
Prob(H) (two-sided):                  0.07   Kurtosis:                         9.08
===================================================================================
```

Salida 3

	Modelo	Akaike	MSE	MAE	RMSE	MAPE
0	AR 1	1939	1.32E+12	8.73E+05	1.15E+06	23%
1	MA 1	2072	1.23E+13	3.42E+06	3.51E+06	96%
2	ARMA 1-1	1929	6.91E+11	5.75E+05	8.31E+05	14%
3	AR 1,2	1931	7.37E+11	6.09E+05	8.58E+05	15%
4	MA 1,2	2057	1.24E+13	3.44E+06	3.52E+06	97%
5	AR 1,2,12	8	1.29E+13	3.54E+06	3.60E+06	100%
6	MA 1,2,12	2028	1.12E+13	3.24E+06	3.34E+06	91%
7	ARIMA 1-1-0	1899	4.25E+11	4.14E+05	6.52E+05	11%
8	ARIMA 0-1-1	1899	4.52E+11	4.31E+05	6.72E+05	11%
9	ARIMA 1-1-1	1901	4.24E+11	4.14E+05	6.51E+05	11%
10	ARIMA 1,2-1-0	1901	4.24E+11	4.14E+05	6.51E+05	11%
11	ARIMA 0-1-1,2	1901	4.32E+11	4.19E+05	6.57E+05	11%
12	ARIMA 1,2,12-1-0	51	1.17E+35	1.87E+17	3.42E+17	5184765550687%
13	ARIMA 0-1-1,2,12	1903	3.24E+11	3.67E+05	5.69E+05	10%
14	SARIMA 1-1-0 Season AR 1	1899	3.09E+11	3.57E+05	5.56E+05	10%
15	SARIMA 1-1-0 Season MA 1	1899	3.28E+11	3.63E+05	5.72E+05	10%
16	SARIMA 1-1-0 Season ARMA 1-1	1901	2.77E+11	3.55E+05	5.26E+05	10%
17	AR 1,12	6	1.29E+13	3.54E+06	3.60E+06	100%
18	ARIMA 1,12-1-0	1939	2.98E+11	3.51E+05	5.46E+05	9%
19	MA 1,12	2045	1.02E+13	3.10E+06	3.19E+06	87%
20	ARIMA 0-1-1,12	1901	3.42E+11	3.68E+05	5.85E+05	10%
21	ARIMA 0-0-0 Season AR 1	2026	1.23E+12	1.01E+06	1.11E+06	29%
22	AR 12	2026	1.23E+12	1.01E+06	1.11E+06	29%
23	MA 12	2079	5.53E+12	2.02E+06	2.35E+06	56%
24	ARIMA 12-1-0	1953	2.82E+11	3.95E+05	5.31E+05	11%
25	ARIMA 0-1-12	1905	3.03E+11	3.67E+05	5.50E+05	10%

Tabla 3 Métricas de error de modelos manuales

Analizando los modelos que construimos de manera manual, se obtuvieron modelos con valores de Akaike muy similares. Al evaluar las métricas de performance y teniendo en cuenta el criterio de Parsimonia, nos quedamos con el modelo ARIMA (0,1,12) que tiene un Akaike 1904.74.

Finalmente, comparando el mejor modelo obtenido de manera manual versus el modelo elegido por el auto-ARIMA, se eligió este último ya que tiene mejores métricas de performance y teniendo en cuenta la estructura de los datos, es razonable considerar la estacionalidad de 12 meses.

Análisis sobre los Residuos del Modelo

Los residuos son útiles para comprobar si un modelo ha captado adecuadamente la información de los datos. Un buen método de previsión producirá residuos con las siguientes propiedades:

1. Los residuos no están correlacionados. Si existen correlaciones entre los residuos entonces queda información en los residuos que debe utilizarse para calcular las previsiones.
2. Los residuos tienen media cero. Si los residuos tienen una media distinta de cero, las previsiones están sesgadas.

Además de estas propiedades esenciales, es útil pero no necesario, que los residuos tengan también las dos propiedades siguientes.

1. Los residuos tienen una varianza constante.
2. Los residuos tienen una distribución normal.

Para evaluar la autocorrelación en los residuos se realizó la prueba de Ljung-Box. Su objetivo es determinar si hay evidencia significativa de autocorrelación en los residuos hasta cierto rezago. Esta prueba es especialmente útil después de ajustar un modelo a una serie temporal para asegurarse que los residuos no exhiban patrones de autocorrelación.

La hipótesis nula ($H0$) de la prueba de Ljung-Box es que los rezagos hasta un cierto punto son independientes, lo que implica que no hay autocorrelación. La hipótesis alternativa ($H1$) es que hay autocorrelación en al menos uno de los rezagos. (Hyndman & Athanasopoulos, 2018).

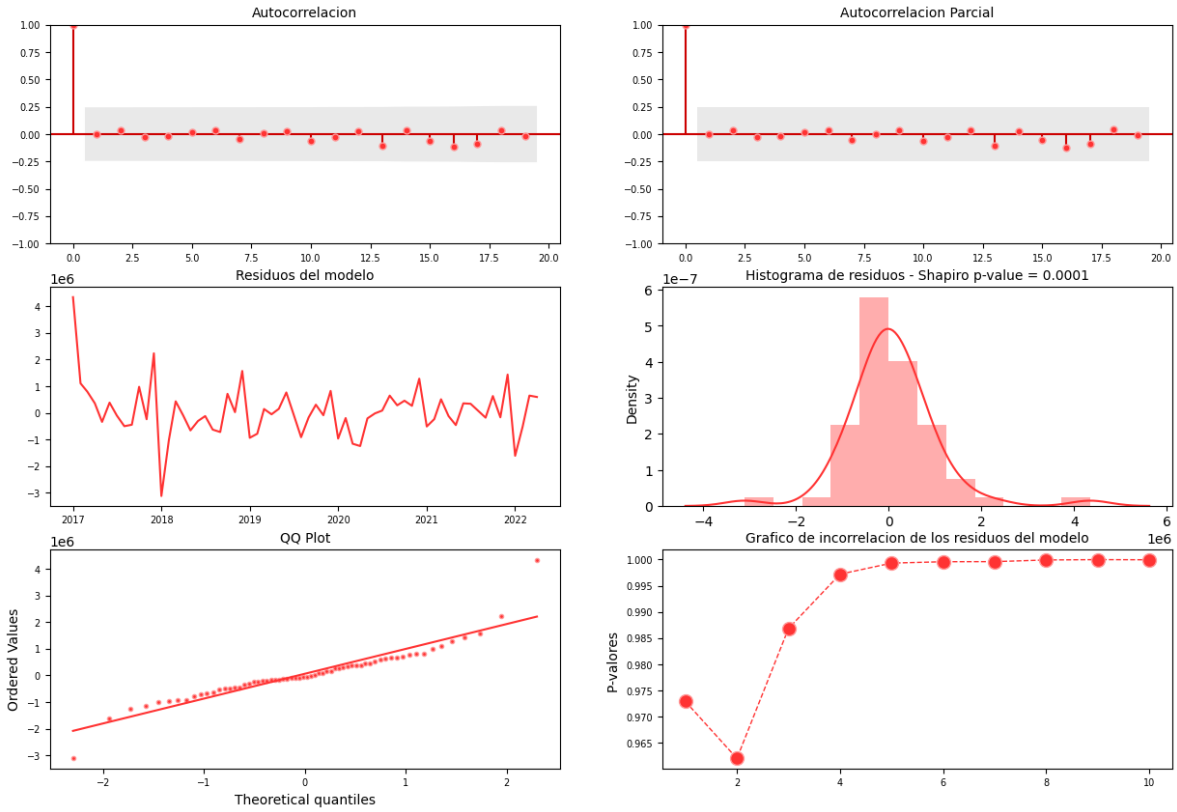


Figura 5 Analisis de Residuos

En los gráficos FAC y FACP no se observan picos, todos los lags se encuentran dentro del intervalo de confianza. Además, en el test de Ljung-Box se obtuvo el p-valor es cercano a uno en todos los lags, por lo que corroboramos la incorrelación de los mismos.

El gráfico de residuos del modelo estos oscilan en una media cero, en consecuencia, las predicciones no estarían sesgadas.

Respecto a la normalidad de los residuos, el test de Shapiro rechaza la normalidad, por ende las predicciones pueden no ser correctas.

Predicciones

A continuación, se encuentran los gráficos de las predicciones obtenidas, y podemos notar que el modelo no fue tan preciso dado que no alcanzó un buen ajuste.

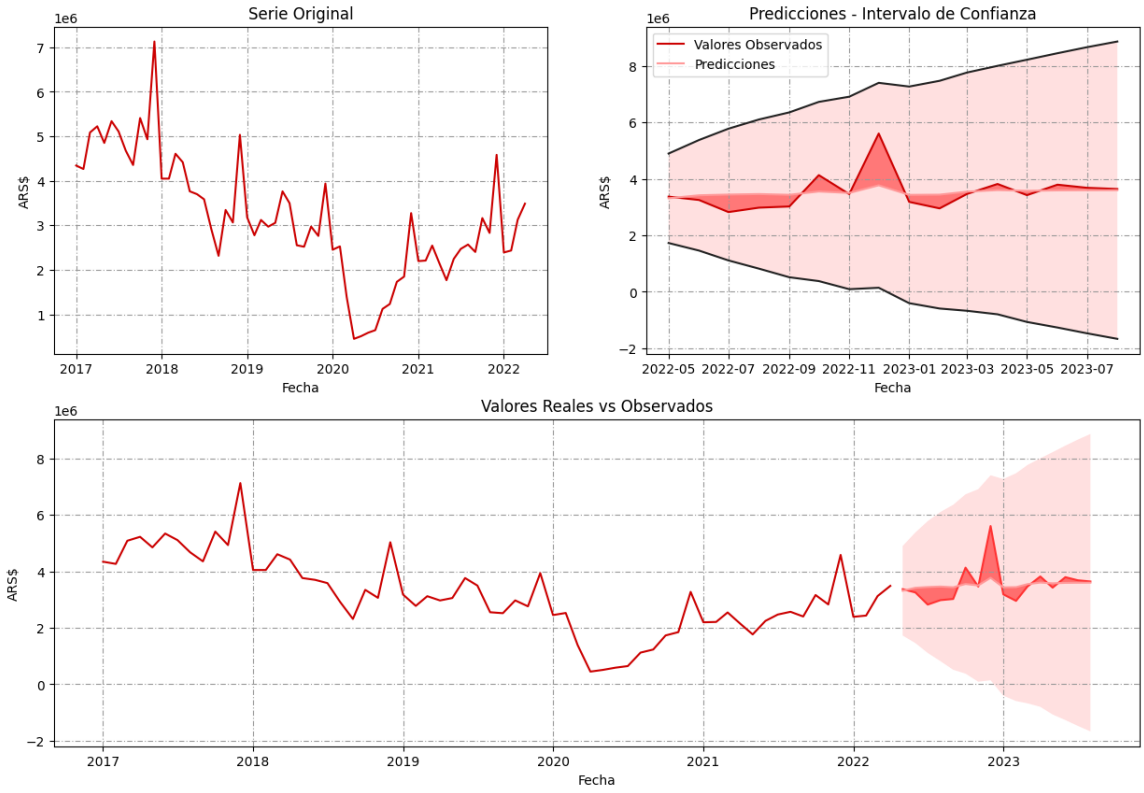


Figura 6 Modelo predictivo

Análisis Serie Super

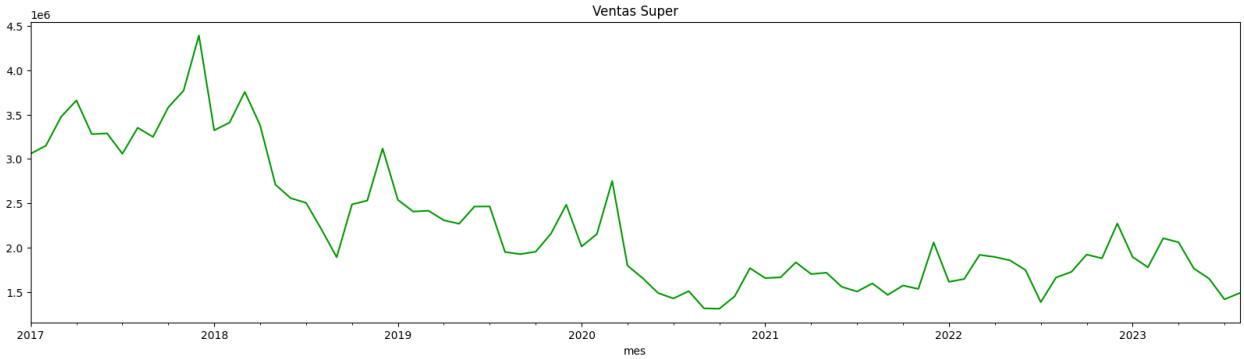


Figura 7 Grafico Ventas Super

FAC, FACP y FAS

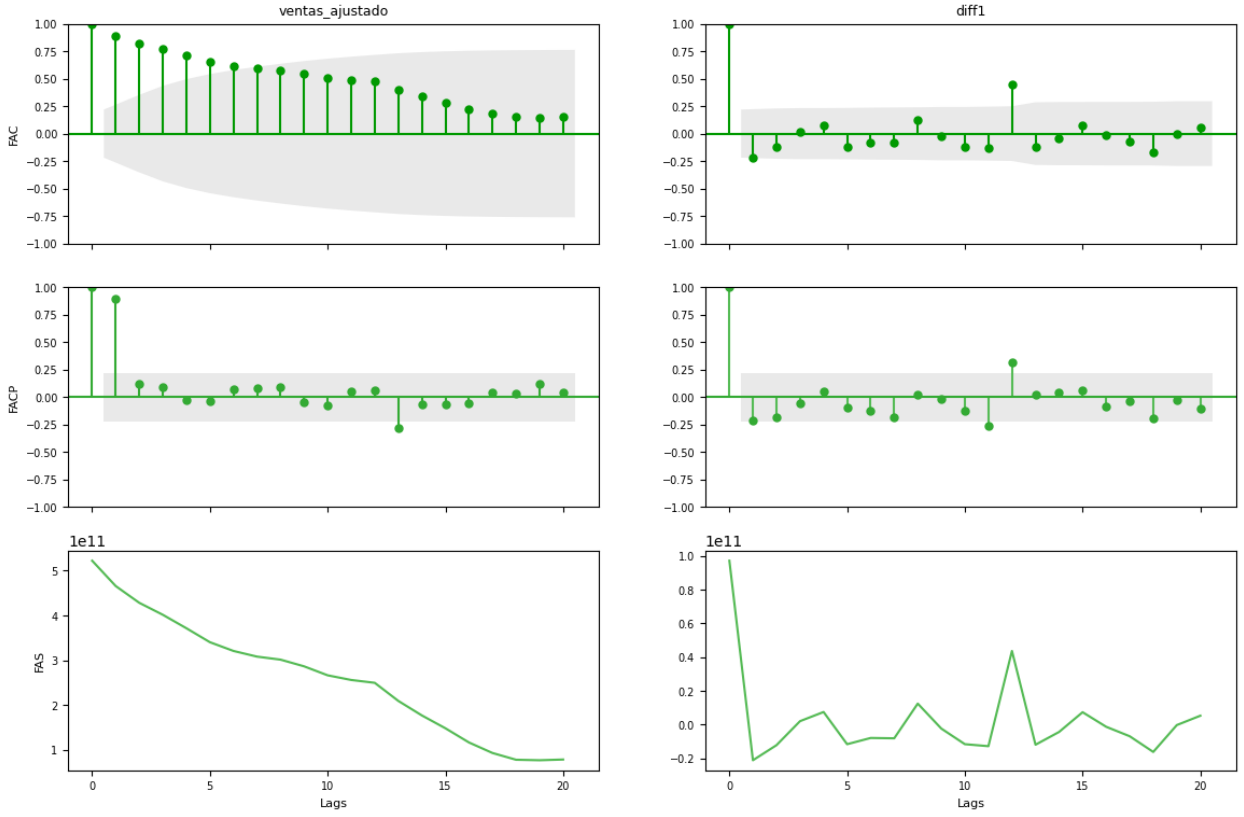


Figura 8 Analisis Serie Super

Es importante destacar que la serie Super tiene un comportamiento muy similar a la serie Retail. La función de autocorrelación se exhibe un patrón de decrecimiento lineal, sugiriendo la no-estacionariedad de la serie temporal.

Analizando la serie en su primera diferencia, se evidencia una disminución abrupta, con correlaciones en los lags primero y duodécimo, indicando una estacionalidad anual. Se podría considerar que un modelo de media móvil (MA 1) sería un punto de inicio adecuado para realizar un ajuste en la modelización.

Al analizar la FACP, se aprecia una disminución exponencial en la serie diferenciada en los lags onceavo y doceavo. Este patrón sugiere la posibilidad de un modelo autorregresivo de orden 12 respaldando la presencia de la estacionalidad mencionada anteriormente.

Tests de raíces unitarias

Augmented Dickey-Fuller

Estadistico ADF	p-Valor	Estacionaridad	Modo
-2.7403	0.0673	No	Constante sola
-1.4806	0.8357	No	Constante y Tendencia Lineal
-2.9235	0.3329	No	Constante y Tendencia Lineal y Cuadratica
-2.2243	0.0251	Si	Sin Contante ni Tendencia

Phillips-Perron

Estadistico PP	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
-1.33	0.1699	12	No	1	n-No incluye término independiente ni lineal
-1.62	0.4717	12	No	1	c-Con término independiente, Sin término lineal
-3.46	0.0439	12	Si	1	ct-Incluye ambos términos

KPSS

Estad. KPSS	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
1.0901	0.0100	5	No	1	c - estacionarios alrededor de una constante.
0.2543	0.0100	5	No	1	ct - estacionarios alrededor de una tendencia.

Tabla 4 Salida Raíces Unitarias Ventas Super

Mediante el test de Dickey Fuller se concluye que no se rechaza la hipótesis nula, por lo que la serie no es estacionaria.

Los resultados de los tests de Phillips Perron y KPPS indican que no puede rechazarse la hipótesis nula y consecuentemente se deduce la no estacionariedad.

Para continuar con el análisis, se realiza la primera diferencia a la serie.

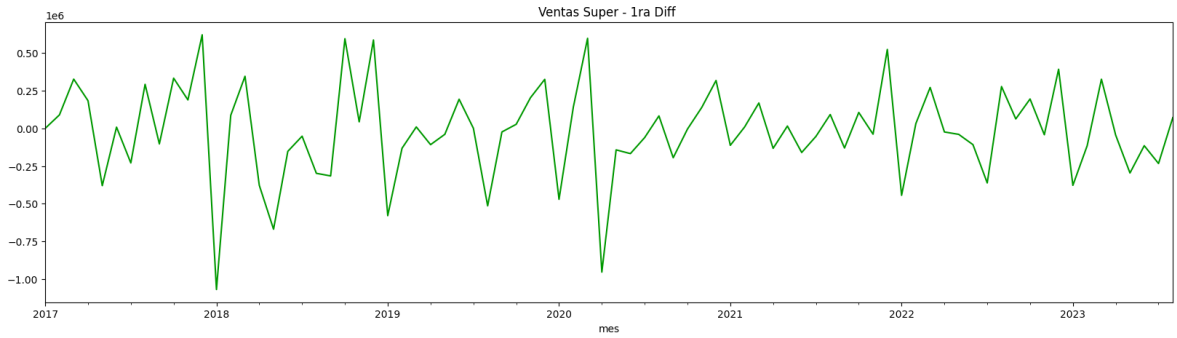


Figura 9 Grafico Serie Super Primera Diferencia

En relación con la serie diferenciada, se observa que la mayoría de los test de raíces unitarias indican que la serie es estacionaria por lo que no se requiere una nueva diferenciación.

Augmented Dickey-Fuller

Estadistico ADF	p-Valor	Estacionaridad	Modo
-3.7467	0.0035	Si	Constante sola
-5.1607	0.0001	Si	Constante y Tendencia Lineal
-4.5201	0.0061	Si	Constante y Tendencia Lineal y Cuadratica
-2.9103	0.0035	Si	Sin Contante ni Tendencia

Phillips-Perron

Estadistico PP	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
-12.75	0.0000	12	Si	0	n-No incluye término independiente ni lineal
-13.45	0.0000	12	Si	0	c-Con término independiente, Sin término lineal
-13.38	0.0000	12	Si	0	ct-Incluye ambos términos

KPSS

Estad. KPSS	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
0.0907	0.1000	10	Si	0	c - estacionarios alrededor de una constante.
0.0878	0.1000	10	Si	0	ct - estacionarios alrededor de una tendencia.

Tabla 5 Salida Raíces Unitarias Serie Super Primera Diferencia

Ajuste de modelos

El resultado del auto-ARIMA como mejor modelo es ARIMA (0,1,0)(0,0,0)[12] sin estacionalidad, con un Akaike de 2226.23, y el componente de ese modelo es significativo.

```
Performing stepwise search to minimize aic
 ARIMA(1,1,1)(0,0,0)[12] intercept   : AIC=2231.296, Time=0.05 sec
 ARIMA(0,1,0)(0,0,0)[12] intercept   : AIC=2227.901, Time=0.02 sec
 ARIMA(1,1,0)(1,0,0)[12] intercept   : AIC=2230.850, Time=0.06 sec
 ARIMA(0,1,1)(0,0,1)[12] intercept   : AIC=2231.331, Time=0.07 sec
 ARIMA(0,1,0)(0,0,0)[12]             : AIC=2226.223, Time=0.02 sec
 ARIMA(0,1,0)(1,0,0)[12] intercept   : AIC=2229.851, Time=0.05 sec
 ARIMA(0,1,0)(0,0,1)[12] intercept   : AIC=2229.855, Time=0.04 sec
 ARIMA(0,1,0)(1,0,1)[12] intercept   : AIC=2231.811, Time=0.09 sec
 ARIMA(1,1,0)(0,0,0)[12] intercept   : AIC=2228.867, Time=0.02 sec
 ARIMA(0,1,1)(0,0,0)[12] intercept   : AIC=2229.365, Time=0.03 sec

Best model:  ARIMA(0,1,0)(0,0,0)[12]
Total fit time: 0.457 seconds
```

SARIMAX Results

Dep. Variable:	y	No. Observations:	80
Model:	SARIMAX(0, 1, 0)	Log Likelihood	-1112.112
Date:	Sun, 12 Nov 2023	AIC	2226.223
Time:	13:46:03	BIC	2228.593
Sample:	01-01-2017	HQIC	2227.173
	- 08-01-2023		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
sigma2	9.76e+10	1.17e+10	8.366	0.000	7.47e+10	1.2e+11

Ljung-Box (L1) (Q):	3.90	Jarque-Bera (JB):	10.08
Prob(Q):	0.05	Prob(JB):	0.01
Heteroskedasticity (H):	0.35	Skew:	-0.57
Prob(H) (two-sided):	0.01	Kurtosis:	4.32

Salida 4

Se particiona la serie, 80% train y 20% para el conjunto de prueba. La salida generada arroja un valor Akaike de 1782.8, y el elemento del modelo tiene significatividad individual.

SARIMAX Results

Dep. Variable:	ventas_ajustado	No. Observations:	64
Model:	SARIMAX(0, 1, 0)	Log Likelihood	-890.400
Date:	Sun, 12 Nov 2023	AIC	1782.800
Time:	13:46:03	BIC	1784.943
Sample:	01-01-2017	HQIC	1783.643
	- 04-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
sigma2	1.088e+11	1.48e+10	7.377	0.000	7.99e+10	1.38e+11

Ljung-Box (L1) (Q):	3.22	Jarque-Bera (JB):	8.09
Prob(Q):	0.07	Prob(JB):	0.02
Heteroskedasticity (H):	0.25	Skew:	-0.63
Prob(H) (two-sided):	0.00	Kurtosis:	4.22

MSE:	66443660197
MAE:	204158
RMSE:	257767
MAPE:	0.123

Salida 5

Se observa un MAPE de un 12% para el modelo elegido por auto-ARIMA.

```
Modelo 0-Akaike: 1816.8454-     AR 1
Modelo 1-Akaike: 2024.3879-     MA 1
Modelo 2-Akaike: 1814.813-      ARMA 1-1
Modelo 3-Akaike: 1815.9623-     AR 1,2
Modelo 4-Akaike: 2011.0293-     MA 1,2
Modelo 5-Akaike: 1817.2586-     AR 1,2,12
Modelo 6-Akaike: 1978.0866-     MA 1,2,12
Modelo 7-Akaike: 1783.9531-     ARIMA 1-1-0
Modelo 8-Akaike: 1783.9858-     ARIMA 0-1-1
Modelo 9-Akaike: 1785.9939-     ARIMA 1-1-1
Modelo 10-Akaike: 1786.0385-    ARIMA 1,2-1-0
Modelo 11-Akaike: 1785.9368-    ARIMA 0-1-1,2
Modelo 12-Akaike: 1796.7789-    ARIMA 1,2,12-1-0
Modelo 13-Akaike: 1833.9891-    ARIMA 0-1-1,2,12
Modelo 14-Akaike: 1785.8297-    SARIMA 1-1-0 Season AR 1
Modelo 15-Akaike: 1785.8338-    SARIMA 1-1-0 Season MA 1
Modelo 16-Akaike: 1787.7616-    SARIMA 1-1-0 Season ARMA 1-1
Modelo 17-Akaike: 6.0-  AR 1,12
Modelo 18-Akaike: 1794.0012-    ARIMA 1,12-1-0
Modelo 19-Akaike: 1998.9021-    MA 1,12
Modelo 20-Akaike: 1796.9266-    ARIMA 0-1-1,12
Modelo 21-Akaike: 1955.1976-    ARIMA 0-0-0 Season AR 1
Modelo 22-Akaike: 1955.1976-    AR 12
Modelo 23-Akaike: 2024.6908-    MA 12
Modelo 24-Akaike: 1792.7593-    ARIMA 12-1-0
Modelo 25-Akaike: 1787.816-     ARIMA 0-1-12
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:        ventas_ajustado   No. Observations:                   64
Model:               SARIMAX(1, 1, 0)   Log Likelihood                -889.977
Date:                Sun, 12 Nov 2023   AIC                          1783.953
Time:                        16:00:39   BIC                          1788.239
Sample:                  01-01-2017     HQIC                         1785.639
                       - 04-01-2022
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0757      0.074     -1.023      0.306      -0.221       0.069
sigma2      1.069e+11   1.63e-13   6.54e+23      0.000    1.07e+11    1.07e+11
===================================================================================
Ljung-Box (L1) (Q):                   1.41   Jarque-Bera (JB):                 4.86
Prob(Q):                              0.24   Prob(JB):                         0.09
Heteroskedasticity (H):               0.23   Skew:                            -0.51
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.89
===================================================================================
```

Salida 6

En base a lo observado en las FAC y FACP se diseñaron métodos manuales. El modelo con menor Akaike fue ARIMA (1,1,0)

	Modelo	Akaike	MSE	MAE	RMSE	MAPE
0	AR 1	1817	5.80E+10	1.88E+05	2.41E+05	11%
1	MA 1	2024	3.07E+12	1.72E+06	1.75E+06	96%
2	ARMA 1-1	1815	5.54E+10	1.84E+05	2.35E+05	11%
3	AR 1,2	1816	5.57E+10	1.84E+05	2.36E+05	11%
4	MA 1,2	2011	3.05E+12	1.72E+06	1.75E+06	96%
5	AR 1,2,12	1817	6.59E+10	2.05E+05	2.57E+05	12%
6	MA 1,2,12	1978	2.92E+12	1.68E+06	1.71E+06	94%
7	ARIMA 1-1-0	1784	6.68E+10	2.05E+05	2.58E+05	12%
8	ARIMA 0-1-1	1784	6.65E+10	2.04E+05	2.58E+05	12%
9	ARIMA 1-1-1	1786	6.68E+10	2.05E+05	2.58E+05	12%
10	ARIMA 1,2-1-0	1786	6.67E+10	2.05E+05	2.58E+05	12%
11	ARIMA 0-1-1,2	1786	6.72E+10	2.06E+05	2.59E+05	12%
12	ARIMA 1,2,12-1-0	1797	6.72E+10	2.06E+05	2.59E+05	12%
13	ARIMA 0-1-1,2,12	1834	6.81E+10	2.08E+05	2.61E+05	13%
14	SARIMA 1-1-0 Season AR 1	1786	6.74E+10	2.07E+05	2.60E+05	12%
15	SARIMA 1-1-0 Season MA 1	1786	6.75E+10	2.07E+05	2.60E+05	12%
16	SARIMA 1-1-0 Season ARMA 1-1	1788	7.63E+10	2.24E+05	2.76E+05	13%
17	AR 1,12	6	3.25E+12	1.79E+06	1.80E+06	100%
18	ARIMA 1,12-1-0	1794	6.74E+10	2.07E+05	2.60E+05	12%
19	MA 1,12	1999	2.62E+12	1.60E+06	1.62E+06	90%
20	ARIMA 0-1-1,12	1797	6.73E+10	2.06E+05	2.59E+05	12%
21	ARIMA 0-0-0 Season AR 1	1955	4.49E+10	1.88E+05	2.12E+05	10%
22	AR 12	1955	4.49E+10	1.88E+05	2.12E+05	10%
23	MA 12	2025	1.12E+12	9.74E+05	1.06E+06	56%
24	ARIMA 12-1-0	1793	6.69E+10	2.05E+05	2.59E+05	12%
25	ARIMA 0-1-12	1788	6.70E+10	2.05E+05	2.59E+05	12%

Tabla 6 Metricas de error de modelos manuales

Si bien hay modelos manuales con Akaike y métricas de performance muy similares a los obtenidos por auto-ARIMA, se elige este último por el criterio de parsimonia.

Análisis de los residuos del modelo

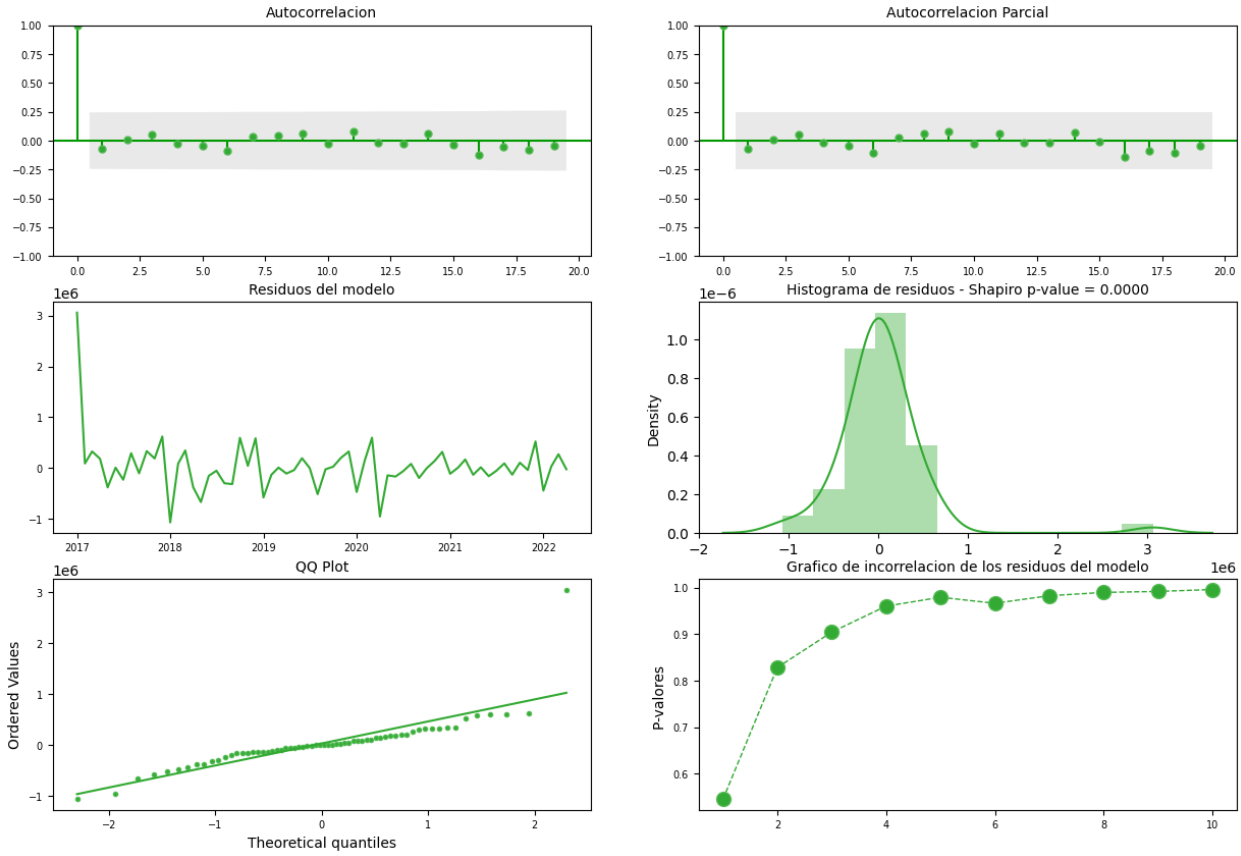


Figura 10 Analisis de Residuos

Al observar los gráficos FAC y FACP se nota que los lags se encuentran dentro del intervalo de confianza. Asimismo, el gráfico de incorrelación refleja los resultados obtenidos por el test de Ljung-Box, evidenciando que el p-valor se acerca a uno en todos los lags, respaldando la incorrelación de los mismos.

En el gráfico de residuos del modelo, se observa que oscilan entorno a cero, por lo tanto, podría sugerirse que las predicciones no presentan sesgo aparente.

En relación con la normalidad de los residuos, el resultado del test de Shapiro indica un rechazo de la hipótesis de normalidad, por ende, las predicciones podrían no ser precisas.

Predicciones

A continuación, se presentan los gráficos de las predicciones obtenidas. A partir de estos, se puede concluir que el modelo no demostró precisión ya que no logró un ajuste adecuado.

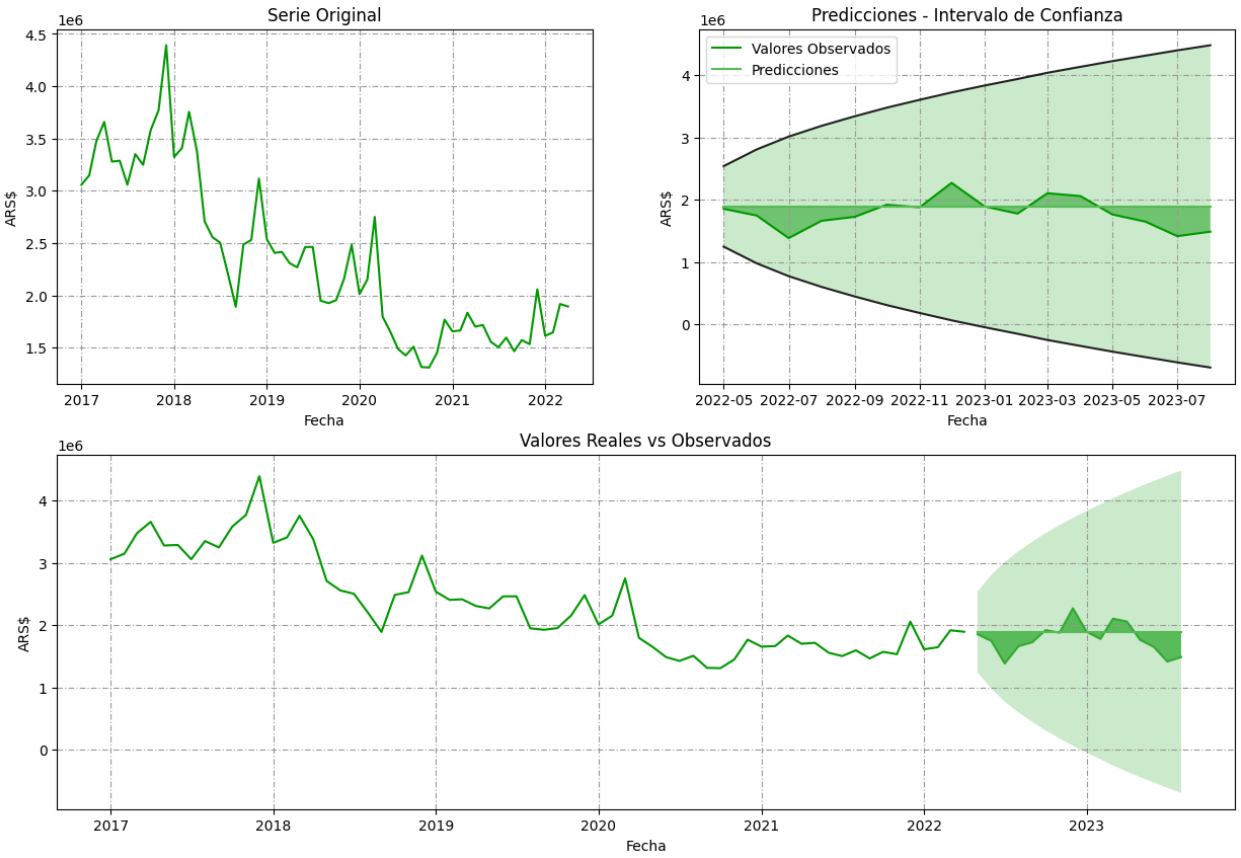


Figura 11 Modelo Predictivo

Análisis Serie Cowork

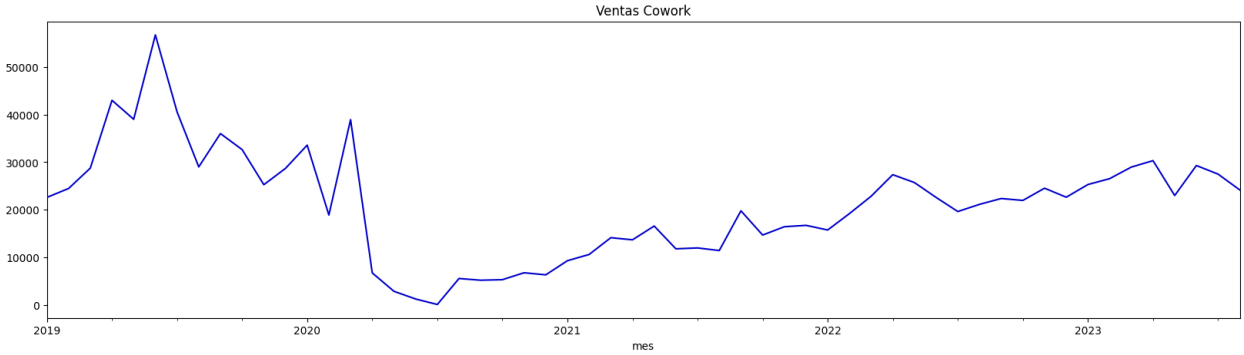


Figura 12 Grafico Serie Cowork

FAC, FACP y FAS

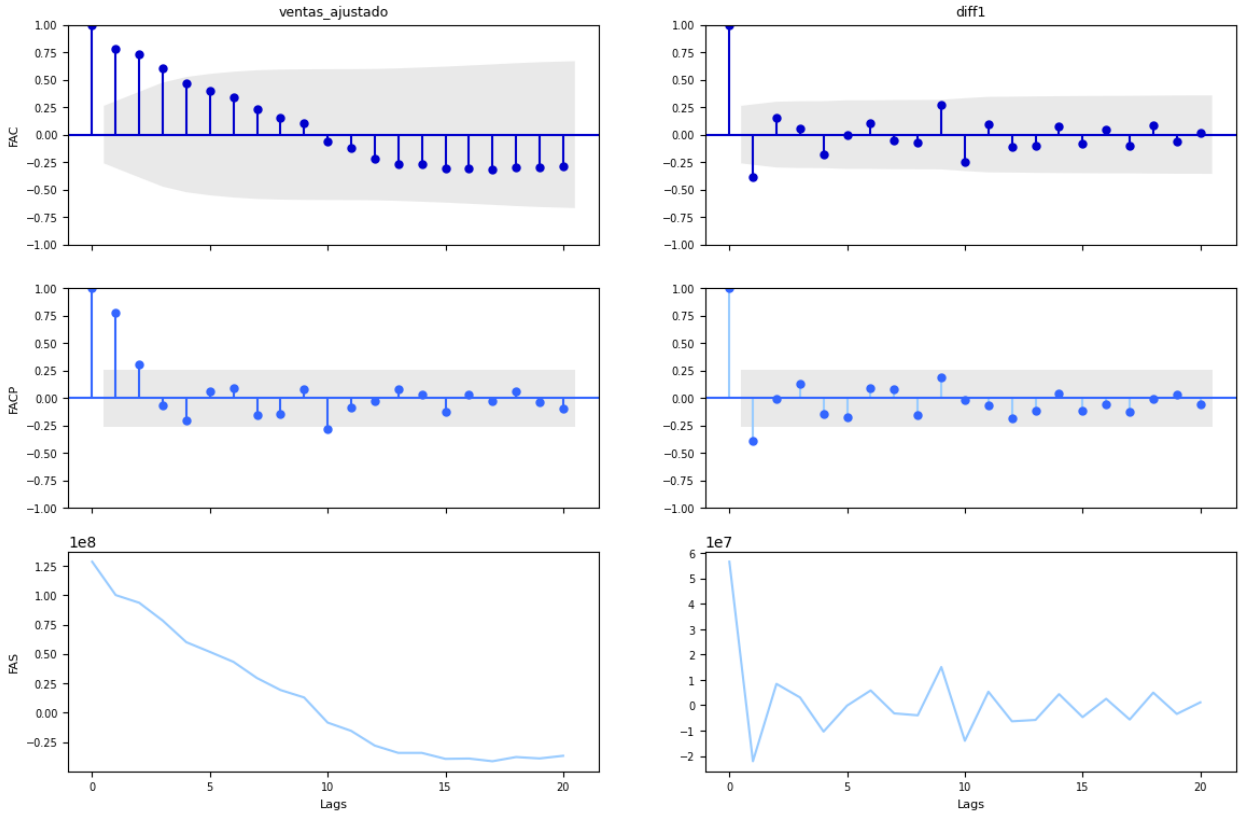


Figura 13 Analisis Serie Cowork

La función de autocorrelación se observa un patrón de decrecimiento lineal, lo que implica la no-estacionariedad de la serie.

Analizando FAC en la serie en su primera diferencia, se evidencia un decrecimiento inmediato, con correlaciones en el primer lag. Se podría suponer que un modelo MA de orden 1 sería una elección apropiada.

Observando la FACP, se aprecia una disminución exponencial en la serie diferenciada. Es te patrón sugiere la posibilidad de un modelo autorregresivo de orden 1.

Tests de raíces unitarias

Augmented Dickey-Fuller

Estadistico ADF	p-Valor	Estacionaridad	Modo
-1.7020	0.4301	No	Constante sola
-1.6563	0.7696	No	Constante y Tendencia Lineal
-2.9501	0.3195	No	Constante y Tendencia Lineal y Cuadratica
-0.7441	0.3940	No	Sin Contante ni Tendencia

Phillips-Perron

Estadistico PP	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
-0.91	0.3234	11	No	1	n-No incluye término independiente ni lineal
-2.72	0.0707	11	No	1	c-Con término independiente, Sin término lineal
-2.74	0.2196	11	No	1	ct-Incluye ambos términos

KPSS

Estad. KPSS	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
0.2322	0.1000	4	Si	0	c - estacionarios alrededor de una constante.
0.2218	0.0100	4	No	0	ct - estacionarios alrededor de una tendencia.

Tabla 7 Salida Raíces Unitarias Serie Cowork

Empleando el test de Dickey Fuller no se rechaza la hipótesis nula, por lo que la serie no es estacionaria.

Los resultados de los tests de Phillips Perron y KPPS indican que no puede rechazarse la hipótesis nula, lo que implicaría no estacionariedad de la serie.

Para proseguir con el análisis, se realiza la primera diferenciación a la serie.

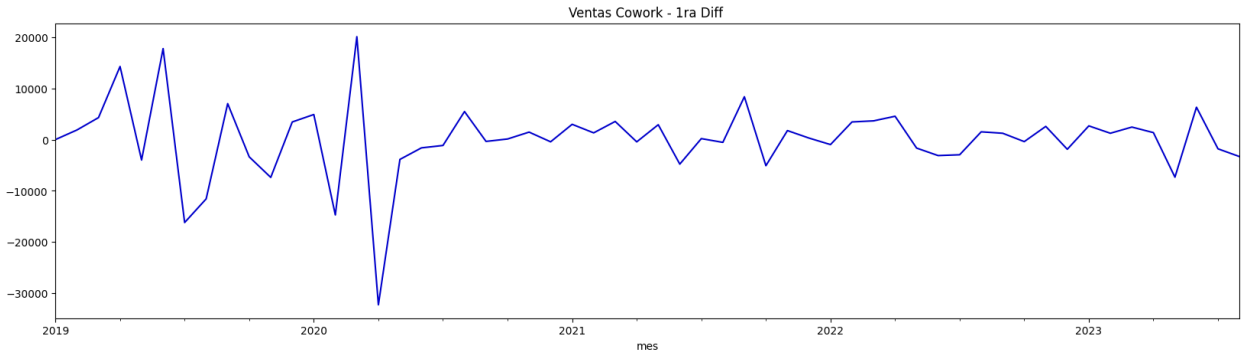


Figura 14 Grafico Serie Cowork Primera Diferencia

En cuanto a la serie diferenciada, se observa que todos los tests señalan que la serie es estacionaria, por lo que no es necesario realizar más diferenciaciones.

Augmented Dickey-Fuller

Estadistico ADF	p-Valor	Estacionaridad	Modo
-10.9820	0.0000	Si	Constante sola
-10.8865	0.0000	Si	Constante y Tendencia Lineal
-10.8108	0.0000	Si	Constante y Tendencia Lineal y Cuadratica
-11.0846	0.0000	Si	Sin Contante ni Tendencia

Phillips-Perron

Estadistico PP	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
-11.25	0.0000	11	Si	0	n-No incluye término independiente ni lineal
-11.14	0.0000	11	Si	0	c-Con término independiente, Sin término lineal
-11.17	0.0000	11	Si	0	ct-Incluye ambos términos

KPSS

Estad. KPSS	p-Valor	NumLags	Estacionaridad	nDiffs	Tipo_Regresion
0.0734	0.1000	2	Si	0	c - estacionarios alrededor de una constante.
0.0636	0.1000	2	Si	0	ct - estacionarios alrededor de una tendencia.

Tabla 8 Salida Raíces Unitarias Serie Cowork Primera Diferencia

Ajuste de modelos

El modelo óptimo obtenido mediante auto-ARIMA es ARIMA (1,0,2) (0,0,1) [12], con un Akaike de 1156.9. La mayoría de los componentes del modelo son significativos individualmente.

```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,0,0)[12] intercept   : AIC=1157.112, Time=0.03 sec
 ARIMA(0,0,0)(0,0,0)[12] intercept   : AIC=1208.573, Time=0.01 sec
 ARIMA(1,0,0)(1,0,0)[12] intercept   : AIC=1160.276, Time=0.08 sec
 ARIMA(0,0,1)(0,0,1)[12] intercept   : AIC=1189.069, Time=0.06 sec
 ARIMA(0,0,0)(0,0,0)[12]             : AIC=1290.621, Time=0.02 sec
 ARIMA(1,0,1)(1,0,0)[12] intercept   : AIC=1157.054, Time=0.09 sec
 ARIMA(1,0,1)(2,0,0)[12] intercept   : AIC=1158.530, Time=0.13 sec
 ARIMA(1,0,1)(1,0,1)[12] intercept   : AIC=1158.943, Time=0.11 sec
 ARIMA(1,0,1)(0,0,1)[12] intercept   : AIC=1156.972, Time=0.06 sec
 ARIMA(1,0,1)(0,0,2)[12] intercept   : AIC=1158.861, Time=0.18 sec
 ARIMA(1,0,1)(1,0,2)[12] intercept   : AIC=1160.665, Time=0.20 sec
 ARIMA(1,0,0)(0,0,1)[12] intercept   : AIC=1159.857, Time=0.05 sec
 ARIMA(2,0,1)(0,0,1)[12] intercept   : AIC=1157.452, Time=0.09 sec
 ARIMA(1,0,2)(0,0,1)[12] intercept   : AIC=1156.897, Time=0.09 sec
 ARIMA(1,0,2)(0,0,0)[12] intercept   : AIC=1157.326, Time=0.04 sec
 ARIMA(1,0,2)(1,0,1)[12] intercept   : AIC=1158.893, Time=0.16 sec
 ARIMA(1,0,2)(0,0,2)[12] intercept   : AIC=1158.864, Time=0.15 sec
 ARIMA(1,0,2)(1,0,0)[12] intercept   : AIC=1157.230, Time=0.11 sec
 ARIMA(1,0,2)(1,0,2)[12] intercept   : AIC=1160.859, Time=0.28 sec
 ARIMA(0,0,2)(0,0,1)[12] intercept   : AIC=1175.631, Time=0.13 sec
 ARIMA(2,0,2)(0,0,1)[12] intercept   : AIC=1158.908, Time=0.13 sec
 ARIMA(1,0,3)(0,0,1)[12] intercept   : AIC=1158.939, Time=0.10 sec
 ARIMA(0,0,3)(0,0,1)[12] intercept   : AIC=1163.382, Time=0.10 sec
 ARIMA(2,0,3)(0,0,1)[12] intercept   : AIC=1161.023, Time=0.24 sec
 ARIMA(1,0,2)(0,0,1)[12]             : AIC=1162.008, Time=0.13 sec

Best model:  ARIMA(1,0,2)(0,0,1)[12] intercept
Total fit time: 2.797 seconds
```

SARIMAX Results

Dep. Variable:	y	No. Observations:	56
Model:	SARIMAX(1, 0, 2)x(0, 0, [1], 12)	Log Likelihood	-572.449
Date:	Sun, 12 Nov 2023	AIC	1156.897
Time:	16:09:05	BIC	1169.049
Sample:	01-01-2019	HQIC	1161.608
	- 08-01-2023		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	4584.2328	2527.447	1.814	0.070	-369.472	9537.938
ar.L1	0.7788	0.100	7.759	0.000	0.582	0.976
ma.L1	-0.2806	0.135	-2.074	0.038	-0.546	-0.015
ma.L2	0.3507	0.113	3.095	0.002	0.129	0.573
ma.S.L12	-0.2261	0.121	-1.864	0.062	-0.464	0.012
sigma2	3.353e+07	0.298	1.12e+08	0.000	3.35e+07	3.35e+07

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	5.11
Prob(Q):	0.96	Prob(JB):	0.08
Heteroskedasticity (H):	0.12	Skew:	-0.04
Prob(H) (two-sided):	0.00	Kurtosis:	4.48

Salida 7

Se separa train y test (80%-20%), y observamos que el Akaike es de 943.85, la mayoría de los componentes tienen significatividad individual y el MAPE es de un 16.3%.

SARIMAX Results

Dep. Variable:	ventas_ajustado	No. Observations:	45
Model:	SARIMAX(1, 0, 2)x(0, 0, [1], 12)	Log Likelihood	-466.924
Date:	Sun, 12 Nov 2023	AIC	943.849
Time:	16:09:05	BIC	952.882
Sample:	01-01-2019	HQIC	947.216
	- 09-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.9737	0.043	22.532	0.000	0.889	1.058
ma.L1	-0.3696	0.131	-2.826	0.005	-0.626	-0.113
ma.L2	0.2418	0.160	1.508	0.131	-0.072	0.556
ma.S.L12	-0.2061	0.177	-1.162	0.245	-0.554	0.142
sigma2	4.314e+07	1.56e-09	2.77e+16	0.000	4.31e+07	4.31e+07

Ljung-Box (L1) (Q):	0.05	Jarque-Bera (JB):	5.21
Prob(Q):	0.82	Prob(JB):	0.07
Heteroskedasticity (H):	0.18	Skew:	-0.56
Prob(H) (two-sided):	0.00	Kurtosis:	4.24

MSE:	30923348
MAE:	4540
RMSE:	5561
MAPE:	0.163

Salida 8

Considerando lo que se ha notado en los gráficos FAC y FACP, se diseñaron modelos manuales. Encontramos en el listado que el modelo con menor Akaike es el modelo 7 de ARIMA (1,1,0), con un valor de Akaike de 917.5.

```
Modelo 0-Akaike: 944.573-       AR 1
Modelo 1-Akaike: 1010.5959-     MA 1
Modelo 2-Akaike: 941.6765-      ARMA 1-1
Modelo 3-Akaike: 940.3918-      AR 1,2
Modelo 4-Akaike: 997.8068-      MA 1,2
Modelo 5-Akaike: 8.0-   AR 1,2,12
Modelo 6-Akaike: 1010.3023-     MA 1,2,12
Modelo 7-Akaike: 917.4981-      ARIMA 1-1-0
Modelo 8-Akaike: 918.7738-      ARIMA 0-1-1
Modelo 9-Akaike: 919.5203-      ARIMA 1-1-1
Modelo 10-Akaike: 919.3858-     ARIMA 1,2-1-0
Modelo 11-Akaike: 919.5265-     ARIMA 0-1-1,2
Modelo 12-Akaike: 923.0307-     ARIMA 1,2,12-1-0
Modelo 13-Akaike: 923.8917-     ARIMA 0-1-1,2,12
Modelo 14-Akaike: 918.3739-     SARIMA 1-1-0 Season AR 1
Modelo 15-Akaike: 918.3923-     SARIMA 1-1-0 Season MA 1
Modelo 16-Akaike: 920.3715-     SARIMA 1-1-0 Season ARMA 1-1
Modelo 17-Akaike: 6.0-  AR 1,12
Modelo 18-Akaike: 921.098-      ARIMA 1,12-1-0
Modelo 19-Akaike: 1005.6936-    MA 1,12
Modelo 20-Akaike: 919.7596-     ARIMA 0-1-1,12
Modelo 21-Akaike: 1033.9313-    ARIMA 0-0-0 Season AR 1
Modelo 22-Akaike: 1033.9313-    AR 12
Modelo 23-Akaike: 1032.9516-    MA 12
Modelo 24-Akaike: 923.2701-     ARIMA 12-1-0
Modelo 25-Akaike: 940.6649-     ARIMA 0-1-12
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:        ventas_ajustado   No. Observations:                   45
Model:               SARIMAX(1, 1, 0)   Log Likelihood                -456.749
Date:                Sun, 12 Nov 2023   AIC                           917.498
Time:                        23:45:39   BIC                           921.066
Sample:                  01-01-2019     HQIC                          918.821
                       - 09-01-2022
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.3203      0.072     -4.448      0.000      -0.461      -0.179
sigma2      6.004e+07    2.1e-10   2.85e+17      0.000       6e+07       6e+07
==============================================================================
Ljung-Box (L1) (Q):                   0.12   Jarque-Bera (JB):                12.22
Prob(Q):                              0.73   Prob(JB):                         0.00
Heteroskedasticity (H):               0.08   Skew:                            -0.77
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.07
==============================================================================
```

Salida 9

	Modelo	Akaike	MSE	MAE	RMSE	MAPE
0	AR 1	945	1.31E+08	1.02E+04	1.15E+04	37%
1	MA 1	1011	6.52E+08	2.51E+04	2.55E+04	96%
2	ARMA 1-1	942	6.64E+07	7.16E+03	8.15E+03	26%
3	AR 1,2	940	7.21E+07	7.47E+03	8.49E+03	27%
4	MA 1,2	998	6.14E+08	2.40E+04	2.48E+04	91%
5	AR 1,2,12	8	6.84E+08	2.60E+04	2.62E+04	100%
6	MA 1,2,12	1010	5.18E+08	2.18E+04	2.28E+04	83%
7	ARIMA 1-1-0	917	2.35E+07	3.95E+03	4.85E+03	14%
8	ARIMA 0-1-1	919	2.40E+07	4.00E+03	4.90E+03	14%
9	ARIMA 1-1-1	920	2.32E+07	3.92E+03	4.82E+03	14%
10	ARIMA 1,2-1-0	919	2.27E+07	3.86E+03	4.77E+03	14%
11	ARIMA 0-1-1,2	920	1.90E+07	3.41E+03	4.36E+03	12%
12	ARIMA 1,2,12-1-0	923	2.37E+07	3.94E+03	4.86E+03	14%
13	ARIMA 0-1-1,2,12	924	1.18E+08	9.64E+03	1.09E+04	36%
14	SARIMA 1-1-0 Season AR 1	918	3.00E+07	4.46E+03	5.48E+03	16%
15	SARIMA 1-1-0 Season MA 1	918	3.09E+07	4.54E+03	5.56E+03	16%
16	SARIMA 1-1-0 Season ARMA 1-1	920	3.04E+07	4.50E+03	5.52E+03	16%
17	AR 1,12	6	6.84E+08	2.60E+04	2.62E+04	100%
18	ARIMA 1,12-1-0	921	2.44E+07	4.02E+03	4.94E+03	14%
19	MA 1,12	1006	4.07E+08	1.95E+04	2.02E+04	74%
20	ARIMA 0-1-1,12	920	8.50E+07	7.21E+03	9.22E+03	26%
21	ARIMA 0-0-0 Season AR 1	1034	1.25E+08	1.08E+04	1.12E+04	42%
22	AR 12	1034	1.25E+08	1.08E+04	1.12E+04	42%
23	MA 12	1033	3.00E+08	1.52E+04	1.73E+04	57%
24	ARIMA 12-1-0	923	2.29E+07	3.84E+03	4.78E+03	14%
25	ARIMA 0-1-12	941	1.13E+08	8.63E+03	1.06E+04	32%

Tabla 9 Métricas de error de modelos manuales

Si bien varios modelos manuales tienen un Akaike similar, notamos que entre el modelo del auto-ARIMA y el menor de los manuales hay cierta diferencia (943.85 vs 917.5). Por este motivo, preferimos el modelo diseñado de manera manual.

Análisis sobre los Residuos del Modelo

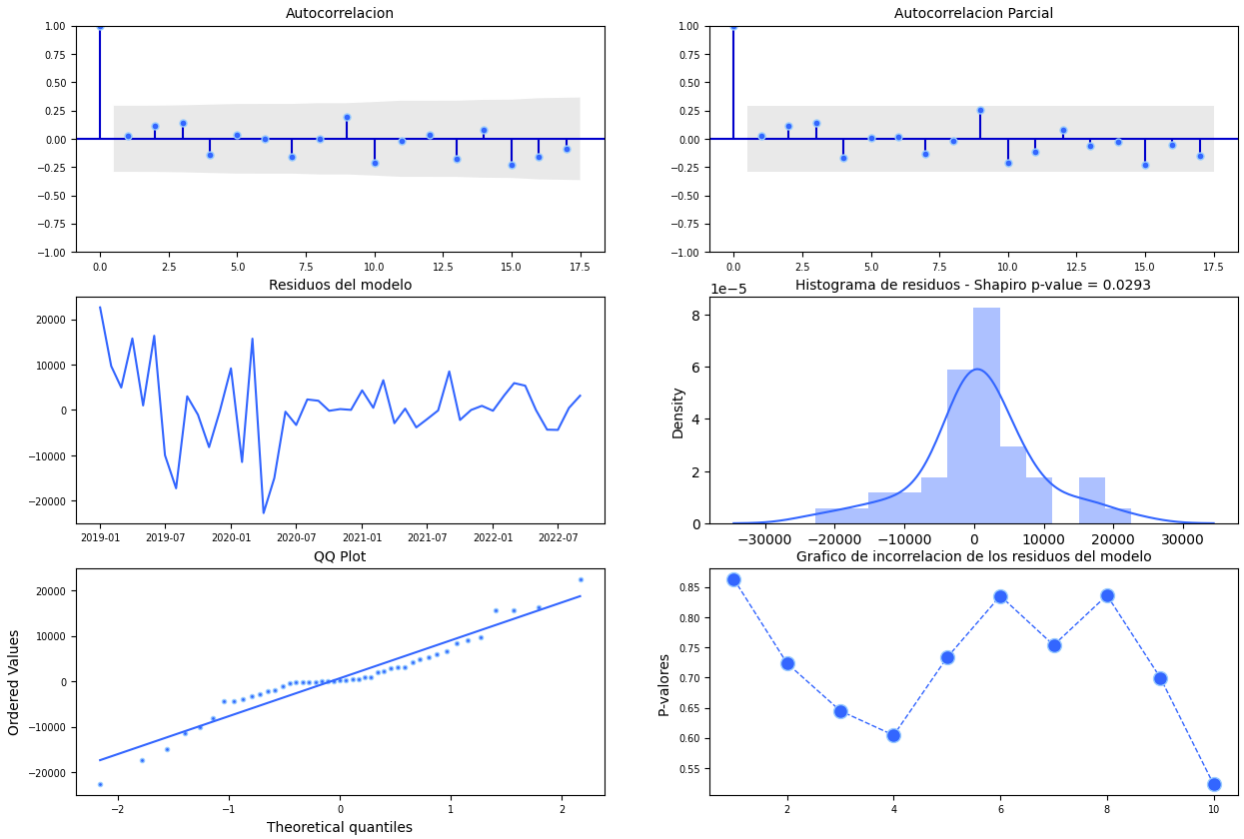


Figura 15 Analisis de Residuos

Al examinar los gráficos FAC y FACP se nota que los lags se encuentran dentro del intervalo de confianza. Además, en el gráfico de incorrelación de los residuos del modelo, notamos que el p-valor se acerca a uno en todos los lags, confirmando la incorrelación de los mismos.

En el gráfico de residuos del modelo, se podría ver que oscilan entorno a cero, por lo que, las predicciones no presentan sesgo aparente.

El resultado del test de Shapiro indica un rechazo de la hipótesis de normalidad (p-valor: 0.0293), entonces las predicciones podrían no ser precisas.

Predicciones

A continuación, se presentan los gráficos de las predicciones y podemos concluir que el modelo no demostró precisión ya que no logró un ajuste adecuado.

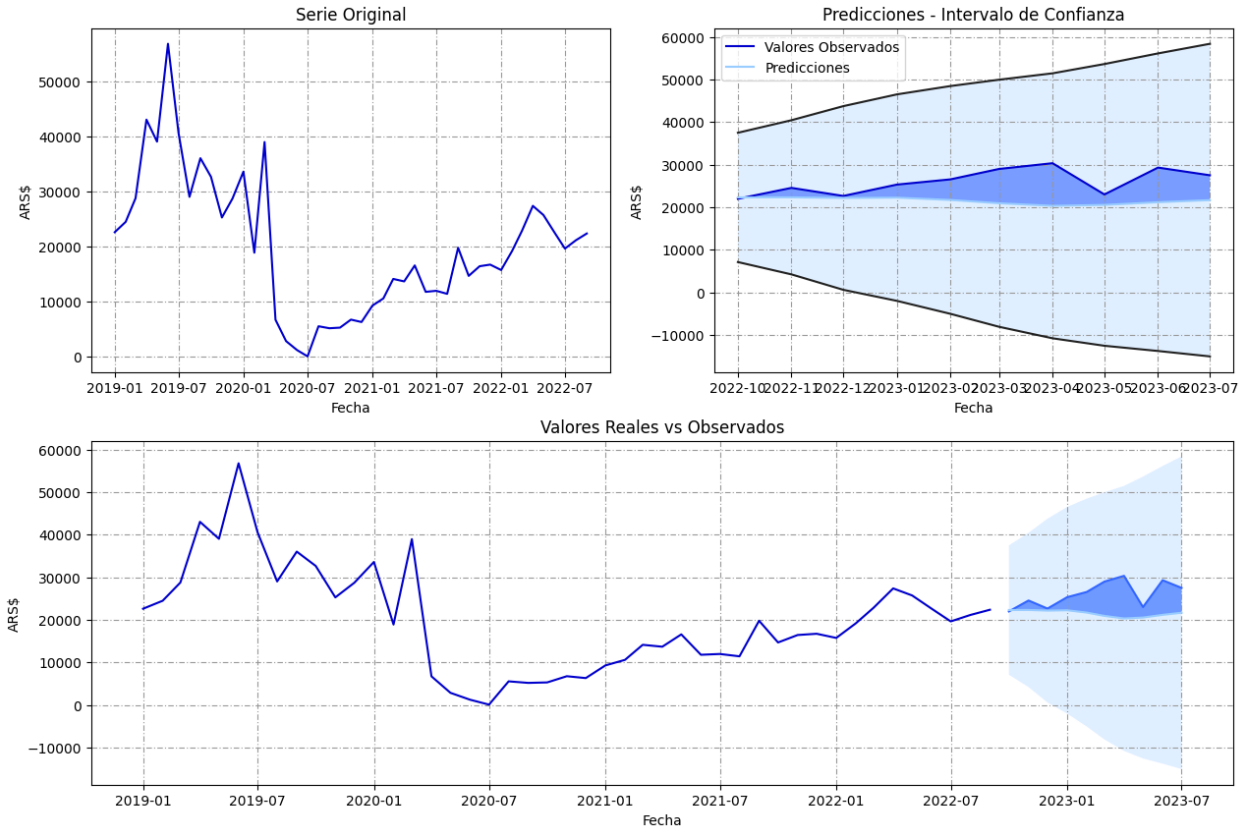


Figura 16 Modelo Predictivo

VAR

Como siguiente paso en el análisis, se considera la utilización de un modelo de Vectores Autorregresivos (VAR). Al no conocerse si existe causalidad entre nuestras variables, se procede a efectuar el test de causalidad de Granger entre las variables pertinentes a fin de descubrirlo.

Se consideraron para el análisis las series en su primera diferencia, ya que las originales no son estacionarias, basado en los cálculos previos. Cabe destacar que para la serie Cowork se cuenta con datos de un período más corto de tiempo.

La hipótesis nula (*H*0) es que una variable no causa (según Granger) a la otra, es decir, la inclusión los rezagos de la primera variable no mejora la predicción de la segunda. La hipótesis alternativa (*H*1) es que sí existe causalidad según Granger.

Estos son los resultados del test, se puede apreciar que los p-valores son mayores a 0.05 en los dos casos, con lo cual no se puede rechazar la hipótesis nula. Es decir, que no hay causalidad de Granger entre las variables nuestro modelo.

	retail_x	super_x	cowork_x
retail_y	1.000000	0.282890	0.593671
super_y	0.336060	1.000000	0.621837
cowork_y	0.083099	0.486424	1.000000

Tabla 10 Causalidad de Granger

Con esto se concluye que no podemos continuar con la elaboración de un modelo VAR, ya que las variables no contribuyen a mejorar la capacidad predictiva.

Conclusiones

En la introducción planteamos realizar un análisis preliminar sobre las series de ventas de los tres rubros para comprender las características de estas series en general y la posibilidad de construir modelos predictivos a partir de las mismas. A partir de los resultados del análisis para las tres series, podemos concluir que a pesar de que surgen algunos modelos significativos la capacidad predictiva de los mismos es limitada debido a la excesiva banda de confianza. Incluso en el caso del supermercado el modelo elegido no muestra relaciones temporales en la variable. De esta manera, un modelo sobre la serie de tiempo no ofrece capacidad predictiva.

Por otro lado, cuando realizamos el análisis de causalidad se encontró la problemática de que no había una relación causal significativa según Granger por lo cual no tuvo sentido realizar un modelo VAR.

Se concluyo que probablemente este tipo de modelos no son adecuados para modelar estas series, por lo que se deberían explorar otras alternativas que permitan un mejor ajuste. Como un siguiente análisis se podrían evaluar las ventas diarias para examinar si los modelos resultan relevantes en dicho caso.

Referencias bibliográficas

Enders, W. (2014). *Applied econometric time series.* Wiley

Hyndman, R., Athanasopoulos, G. (2018) *Forecasting: Principles and Practice.* OTexts

Peña, D. (2005). *Análisis de series temporales*. Alianza Editorial

Documentacion: pmdarima.arima.auto_arima (s.f.)

https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html

Apéndices

Link a Repositorio: https://github.com/waldof86/AST/tree/main/TP1

Link a Datos: https://github.com/waldof86/AST/tree/main/TP1/Data