# Estadística III 3009137, semestre 01 de 2022

Equipo de Trabajo No. 2 Serie No. 3 Curso: Ma – Ju, Horario: 10.00 am – 12.00 m

## **ANÁLISIS DE SERIES DE TIEMPO. AJUSTE DE TENDENCIA Y ESTACIONALIDAD**

Sofía Cuartas García[[1]](#footnote-1), Simón Cuartas Rendón[[2]](#footnote-2) y Deivid Zhang Figueroa[[3]](#footnote-3)

*Fecha de entrega: 25 – 06 – 2022*

1. **ANÁLISIS DESCRIPTIVO Y TEST HEGY DE RAÍCES UNITARIAS**

En este trabajo se van a plantear modelos SARIMA para ajustar la serie temporal del ***índice de ventas del sector manuacturero colombiano en pesos nominales*** que publica el Departamento Administrativo Nacional de Estadística (DANE) cada mes, con el objetivo de poder realizar pronósticos para este índice de cara al futuro, y a su vez, se van a contrastar estos modelos con modelos ARMA y modelos globales y locales que no realizan ningún ajuste a los errores y que han sido estudiados anteriormente. Para ello, es importante tener presente que la serie temporal de índice de ventas en pesos nominales del sector manufacturero colombiano con la que se trabajará va desde enero de 2001 hasta noviembre de 2021, lo cual implica que se tienen observaciones para esta serie de tiempo (teniendo presente que aquellos datos asociados a los meses de marzo de 2020 en adelante fueron imputados para ignorar los efectos que tuvo la pandemia por la *COVID-19* en este índice, entre otros indicadores económicos), si bien para el ajuste de los todos los modelos fueron empleadas los valores del índice de ventas en pesos nominales para los primeros , dejando los últimas doce observaciones como parte del periodo *ex post* para poder realizar validación cruzada.

Mencionado lo anterior, se comienza revisando las características de la serie de tiempo a trabajar en la ***figura 1*** considerando las observaciones de esta.

|  |  |
| --- | --- |
| Chart  Description automatically generated**(a)** | Chart, line chart  Description automatically generated**(b)** |

***Figura 1.*** Serie temporal de ***índice de ventas del sector manufacturero colombiano en pesos nominales*** desde enero de 2001 a noviembre de 2021. **(a)** Serie con el índice de ventas en pesos nominales en escala original. **(b).** Serie temporal con el índice de ventas en pesos nominales en escala logarítmica.

En la ***figura 1*** ***(a)*** el índice de ventas del sector manufacturero colombiano en pesos nominales se presenta en escala original. En esta se puede observar que la serie presenta tendencia creciente, ya que el índice bajo estudio tiende a aumentar con el tiempo; asimismo, se puede evidenciar como este índice va presentando una mayor varianza conforme para el tiempo, lo cual señala que esta serie es multiplicativa; de igual manera se aprecia que la tendencia es global en tanto es posible emplear una función suave del tiempo para modelarla, y además, se puede decir que la tendencia es determinística, puesto que la evolución sucede de forma perfectamente predecible. Por otro lado, es clara la presencia de la componente estacional en esta serie, debido a que existe un comportamiento repetitivo en cada año de la serie, y se puede decir que esta es exacta con meses, pues en general se observa que el índice de ventas en pesos nominales inicia con un mínimo en los meses de enero, y luego muestra un incremento rápido hasta alcanzar sus máximos anuales en los meses de diciembre, para luego mostrar una reducción al comenzar enero del siguiente año y comenzar así un nuevo periodo anual, y por lo anterior es también razonable decir que la estacionalidad es determinística. No obstante, vale la pena señalar que en algunos años este comportamiento estacional no tiene la misma apariencia que la mayoría de los años, según como se acabó de describir, por la presencia de ciclos en la serie, como ocurre por ejemplo entre los años 2008 a 2011. Para finalizar esta descripción inicial, se debe destacar que en la ***figura 1 (b)*** se aprecia como la variabilidad alrededor de la tendencia media de largo plazo se estabiliza. **XXXX ¿estacionalidad global?**

A continuación, es necesario observar la ***figura 2,*** para la cual se considera la serie recortada, esto es, excluyendo el periodo *ex post,* en la cual se presentan lo gráficos la serie recortada, su primera diferencia regular, su primera diferencia estacional y de la serie diferenciada por tendencia y estacionalidad (en todos los casos empleando la transformación logarítmica por tener la varianza estabilizada), así como de los gráficos de las funciones de autocorrelación asociadas.

|  |  |  |  |
| --- | --- | --- | --- |
| **Chart, line chart  Description automatically generated(a)** | **Chart  Description automatically generated(c)** | **Chart  Description automatically generated with medium confidence(e)** | **Chart  Description automatically generated(g)** |
| **Chart, histogram  Description automatically generated(b)** | **Chart, box and whisker chart  Description automatically generated(d)** | **Chart, histogram  Description automatically generated(f)** | **Chart, box and whisker chart  Description automatically generated(h)** |

***Figura 2.* (a).** Serie recortada, **(b)** ACF muestral de la serie recortada; **(c)** primera diferencia regular de la serie recortada, **(d)** ACF de la primera diferencia regular de la serie recortada; **(e) p**rimera diferencia estacional serie recortada, **(f)** ACF muestral de la primera diferencia estacional de la serie recortada; **(g)** primera diferencia regular y estacional de la serie recortada, **(h)** ACF muestral de la serie recortada diferenciada por tendencia y estacionalidad.

Comenzando con la serie temporal recortada y transformada a escala logarítmica de la ***figura 2 (a)*** se evidencia que la media no es constante, sino que tiende a crecer con el tiempo en tanto la tendencia es positiva, y dentro de cada año también se evidencia un incremento de enero a diciembre, como se analizó previamente; empero, esta transformación permitió estabilizar la varianza del índice de ventas en pesos nominales, y al pasar a su ACF muestral en la ***figura 2 (b)*** se concluye rápidamente que el proceso no es ergódico, pues la ACF disminuye muy lentamente tanto para *k* consecutivos en el caso de la parte regular, como para *k* múltiplos de doce (), que es el número de periodos en cada año. Además, debe destacarse que , lo que corrobora que esta serie se ve afectada por ciclos. Por lo anterior, se tiene que el logaritmo del proceso asociado al índice de ventas del sector manufacturero colombiano en pesos nominales **no es estacionario**.

Respecto a la primera diferencia de la serie temporal recortada en escala logarítmica, visible en la ***figura 2 (c)***, es claro que en este caso la media es constante y es aproximadamente nula, pero en cuanto a la varianza, no es muy claro que esta sea constante, ya que entre el 2015 y el 2020 parece que esta incrementa ligeramente, pero no de forma significativa respecto al resto de periodos **XXXX.** Luego, analizando la ***figura 2 (d),*** donde se observa su ACF muestral, donde refleja para la parte regular que esta es ergódica en tanto la autocorrelación muestral toma valores muy cercanos a cero rápidamente para los primeros valores de rezago; en cambio, la parte estacional se observa un patrón de cola positiva, si bien no puede ser considerada ergódica porque para se puede ver que la función de autocorrelación muestra toma un valor superior a 0.6, para la ACF es aproximadamente 0.5 y para la ACF es nuevamente cercana a 0.5. De esta forma, se puede afirmar que la primera diferencia regular del logaritmo del índice de ventas del sector manufacturero en pesos nominales **no es estacionario.**

Después, considerando la primera diferencia estacional de la serie recortada, cuya gráfica se halla en la ***figura 2 (e)***, se puede ver que la media no es constante, dado que existen ciclos en esta primera diferencia regular, pero se observa que los valores fluctúan dentro de un rango de tamaño muy semejante en todos los periodos, por lo que la varianza es constante. En cuanto a su función de autocorrelación muestral, visible en la ***figura 2 (f)***, se tiene para su parte regular se tiene que esta no es ergódica, ya que la ACF en el segundo rezago es mayor que para el primero, pero a continuación comienza a disminuir hasta , para presentar luego un crecimiento en y luego volver a decrecer en , y en cualquier caso, cuando decrece no lo hace lo suficientemente rápido como para sugerir que la ACF de la primera diferencia estacional en su parte estacional es ergódica. Respecto a la componente estacional, se tiene que tampoco se cumple el supuesto de ergodicidad, ya que si bien para la autocorrelación muestral es muy próxima a cero, esta se incrementa para . Así, se concluye que la primera diferencia estacional del índice de ventas del sector manufacturero colombiano en pesos nominales **no es estacionario.**

Finalmente, se estudiará la serie temporal diferenciada por tendencia y estacionalidad, esto es, a , cuyo gráfico se presenta en la ***figura 2 (g)*** y en la que se muestra que la media es nula y que la varianza es aproximadamente constante para toda la serie temporal**.** También resulta relevante analizar su función de autocorrelación muestral, exhibida en la ***figura 2 (h)***, en la cual se muestra que parta la parte regular se tiene un patrón tipo de cola exponencial sinusoidal que se acerca a cero rápidamente en . Respecto a la parte estacional, se puede ver que la ACF muestral toma inicialmente valores grandes pero para este es próximo a cero. Además, resulta también útil analizar la función de autocorrelación parcial, PACF, de este proceso, observado en la ***figura 3,*** en la cual se observa, para la parte regular, que esta parece ser una cola exponencial sinusoidal que a partir de se conserva dentro de los límites de Bartlett, por lo que estas autocorrelaciones parciales se toman como significativas y concluye que se cumple el supuesto de ergodicidad. De igual forma, para la parte estacional, se tiene que esta toma valores cercanos a -0.15 para , pero se debe tener presente que para rezagos grandes las estimaciones de la PACF son menos confiables, por lo que se puede asumir que estas no son significativas y por tanto se Chart, box and whisker chart

Description automatically generatedconcluye, tanto por la parte estacional como para la regular, que hay ergodicidad, coincidiendo con la conclusión derivada de la ACF en la ***figura 2 (h).***

***Figura 3.*** Función de autocorrelación parcial (PACF) muestral del logaritmo de la serie del índice de ventas del sector manufacturero colombiano en pesos nominales recortada y diferenciada por estacionalidad y tendencia.

De esta manera, cumpliéndose los criterios de media y varianza constantes, así como la ergodicidad, se concluye que el proceso asociado a la serie temporal del índice de ventas del sector manufacturero colombiano en pesos nominales **es un *proceso estacionario en covarianza.***

Ahora, se va a realizar el ***test HEGY*** para esta serie, el cual es un test tipo *Dickey Fuller aumentado, ADF,* con la salvedad de que la serie temporal será representando como un proceso autorregresivo infinito que supone que este es invertible. Así, se tiene que un proceso ARMA tendrá raíz unitaria estacional de periodo doce si su polinomio AR tiene como raíces una o más raíces periódicas del polinomio , las cuales se expresan como , y de este modo, si se asume que el proceso dado por logaritmo del índice de ventas del sector manufacturero colombiano en pesos nominales es invertible, entonces tendría representación AR(∞), donde el polinomio autorregresivo será , de manera que la representación AR de la serie será dada por:

Y así, la representación del modelo de regresión lineal múltiple apropiado para el test ADF se muestra en la ***ecuación 1:***

**[1]**

.

Donde las variables resultan de aplicar filtros tales que los coeficientes de regresión estén asociadas a las raíces , y tales variables son las siguientes:

Así pues, el juego de hipótesis para las pruebas del test HEGY son las presentadas en la ***tabla 1.*** Para cada una de las siete pruebas presentadas en tal prueba se tiene que sus estadísticos de prueba bajo poseen la distribución de un proceso estocástico complejo. Con esto presente, se acude a ***R*** para poder realizar estas pruebas y los resultados son presentados en la ***tabla 2***, y como se puede observar, no se rechazó la hipótesis nula para el primer, tercer, quinto, sexto y séptimo test, lo cual indica que el proceso asociado a la serie temporal de interés **presenta raíces unitarias regulares y estacionales,** con una significancia de , lo cual es un indicador de que es necesario aplicarle filtro de diferencia mixto , como el resultado en la ***figura 2 (g)***, que como se pudo observar en su momento, resultó en una serie estacionaria en sentido débil.

1. **IDENTIFICACIÓN DE MODELOS *SARIMA(p, d, q)(P, D, Q)[12]***

Realizados los análisis anteriores, ahora se van a identificar modelos **SARIMA(p, d, q)(P, D, Q)[s],** esto partiendo del proceso derivado de la serie temporal del índice de ventas del sector manufacturero colombiano con precios en pesos nominales que ha sido diferenciada por tendencia y estacionalidad, de manera que ya se sabe que para los modelos a presentar se cumple que . Mencionado esto, se partirá de las evaluaciones que se pueden realizar con la ACF y la PACF de las figuras ***2 (g)*** y ***3*** respectivamente.

***Tabla 1.*** Juego de hipótesis para las pruebas del test HEGY y sus resultados calculados con el paquete estadístico ***R.***  En la última columna se usa “✘” para indicar que se rechaza la hipótesis nula.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Test** | **Hipótesis** |  | **Frecuencia angular (*θ*)** | **Ciclos por año (*j*)** | **Estadístico de prueba calculado** | **Valor p ()** | **¿Rechazo?** |
| 1 |  | Hay raíz unitaria regular o no estacional |  | 0 | 2.9135488 | 0.10000000 |  |
| No hay raíz unitaria regular o no estacional |
| 2 |  | Hay raíz unitaria estacional bimensual |  | 6 | -2.8854438 | 0.01000000 | ✘ |
| No hay raíz unitaria estacional bimensual |
| 3 |  | Hay raíz unitaria estacional cuatrimestral |  | 3 | 0.8104567 | 0.10000000 |  |
| No hay raíz unitaria estacional cuatrimestral |
| 4 |  | Hay raíz unitaria estacional trimestral |  | 4 | 2.4693601 | 0.09062608 | ✘ |
| No hay raíz unitaria estacional trimestral |
| 5 |  | Hay raíz unitaria estacional semestral |  | 2 | 1.7683136 | 0.10000000 |  |
| No hay raíz unitaria estacional semestral |
| 6 |  | Hay raíz unitaria estacional de cinco ciclos al año |  | 5 | 2.2239858 | 0.10000000 |  |
| No hay raíz unitaria estacional de cinco ciclos al año |
| 7 |  | Hay raíz unitaria estacional anual |  | 1 | 2.4880179 | 0.10000000 |  |
| No hay raíz unitaria estacional anual |

Así pues, iniciando con la parte regular y chequeando únicamente para ambas funciones, se puede distinguir en la ACF que se da un patrón tipo cola, que podría ser de tipo exponencial sinusoidal bajo las justificaciones dadas en la primera sección, mientras que con la PACF se aprecia un patrón tipo corte, donde el primer rezago para el cual se anula estadísticamente la autocorrelación parcial es ; así, teniendo un patrón cola en la ACF y corte en la PACF, se obtiene que la parte regular de la serie temporal de interés diferenciada por estacionalidad y tendencia se puede modelar como un proceso de medias móviles con , por lo que de momento se tiene un **ARIMA(2, 1, 0).**

Por otro lado, para la parte estacional, evaluada para los rezagos , se ve un patrón contrario al caso regular, es decir, la ACF muestra un patrón corte, donde para se puede considerar que la ACF es estadísticamente significativa, mientras que para no lo es, toda vez que en la PACF se observa un patrón tipo cola negativa. Por lo anterior, se tiene que la parte estacional puede ser modelada como un proceso de medias móviles con , esto es, un **ARIMA(0, 1, 2)[12].** Entonces, al unir estos dos análisis, se obtiene que el modelo *completo*, por así decirlo, sería un **ARIMA(2, 1, 0)(0, 1, 2)[12].** La ecuación de este modelo es la ***ecuación 2:***

**[2]**

Hecho esto, se procederá a la identificación de modelos **SARIMA(p, d, q)(P, D, Q)[s]** empleando métodos automáticos, comenzando con la función auto.arima() de la librería **forecast** (versión 8.16) del paquete estadístico ***R***, y cuyas salidas se visualizan en la ***figura 4.***

|  |  |  |
| --- | --- | --- |
| **A screenshot of a computer  Description automatically generated with low confidence(a)** | **Text  Description automatically generated(b)** | **A screenshot of a computer  Description automatically generated with low confidence(c)** |
| **Text  Description automatically generated(d)** | **Text  Description automatically generated(e)** | **Text, letter  Description automatically generated(f)** |

***Figura 4.*** Identificación de métodos **SARIMA(p, d, q)(P, D, Q)[12]** usando: **(a)** AIC y *ocsb*, **(b)** AIC y *ch*, **(c)** AIC y *seas*, **(d)** BIC y *ocsb*, **(e)** BIC y *ch*, **(f)** BIC y *seas.* El objeto lny que usa en el primer argumento de cada función es la serie temporal de interés transformada usando logaritmo natural.

Así, se presenta en la ***tabla 2*** los modelos que han sido identificados, así como sus ecuaciones. Debe notarse que para los modelos que se identificarán como **uno** y ***dos,*** sus salidas en ***R*** tienen el indicador *with drift* (con deriva), por lo que se le debe agregar el sumando , que corresponde a tal deriva. Para el resto de modelos no se da esta situación.

***Tabla 2.*** Ecuaciones teóricas de los modelos identificados de forma automática con la función auto.arima().

|  |
| --- |
| **Modelo uno.** Usando el criterio de información AIC y método de selección de diferencias estacionales *ocsb.*  **ARIMA(0, 1, 2)(2, 0, 0)[12] con deriva.** |
| **Modelo dos.** Usando el criterio de información AIC y método de selección de diferencias estacionales *ch.*  **ARIMA(0, 1, 2)(2, 0, 0)[12] con deriva.** |
| **Modelo tres.** Usando el criterio de información AIC y método de selección de diferencias estacionales *seas.*  **ARIMA(4, 1, 0)(1, 1, 2)[12] sin deriva.** |
| **Modelo cuatro.** Usando el criterio de información BIC y método de selección de diferencias estacionales *ocsb.*  **ARIMA(1, 1, 1)(2, 0, 0)[12] sin deriva.** |
| **Modelo cinco.** Usando el criterio de información BIC y método de selección de diferencias estacionales *ch.*  **ARIMA(1, 1, 1)(2, 0, 0)[12] sin deriva.** |
| **Modelo seis.** Usando el criterio de información BIC y método de selección de diferencias estacionales *seas.*  **ARIMA(0, 1, 2)(0, 1, 1)[12] sin deriva.** |

Así, se debe notar que los modelos que se identifican como ***uno, dos, cuatro*** y ***cinco*** no resultan apropiados debido a que solo solo están considerando a la primera diferencia regular y no a la primera diferencia estacional, lo cual no resulta apropiado teniendo en cuenta lo arrojado por el test HEGY, que indica que hay evidencia muestral suficiente para sugerir que existen raíces unitarias regulares y estacionales en el proceso asociado a la serie temporal del índice de ventas del sector manufacturero colombiano en pesos nominales, de manera que es necesario incluir tanto la primera diferencia regular gracias a la tendencia, como la primera diferencia estacional, tal y como hacen los modelos ***tres*** y ***seis,*** que justamente se realizaron bajo el mismo método de selecciones de diferencias estacionales: ***“seas”***, pero diferenciándose en el criterio de información, que fue AIC para el modelo ***tres*** y BIC para el modelo ***seis.*** Lo anterior también se evidencia en las gráficas de la ***figura 2***: cuando solo se diferencia por tendencia (primera diferencia regular) o por estacionalidad (primera diferencia regular), los procesos resultantes ***no*** son estacionarios, mientras que el proceso derivado de diferenciar tanto por estacionalidad como por tendencia ***sí*** lo es.

Ahora, se va a utilizar otra función, que es armasubsets, de la librería TSA (versión 1.3), y la cual genera un tablero en el que se resaltan los parámetros estadísticamente significativos que, de acuerdo con esta función, deben ser incorporados en el modelo. Esto se hará de dos modos: con un tablero , visible en la ***figura 5*** (por lo que se ignora la parte estacional) y otro tablero (que sí incorpora términos asociados a la parte estacional de la serie), presentado en la ***figura 6.***

A picture containing text, crossword puzzle

Description automatically generatedA picture containing text, crossword puzzle

Description automatically generated***Figura 5.*** Gráfico arrojado por armasubsets sobre usando el método “*ols*”.

***Figura 6.*** Gráfico arrojado por armasubsets sobre usando el método “*ols*”.

Así pues, teniendo en cuenta que estos tableros han sido obtenidos con la serie diferenciada por estacionalidad y tendencia, por lo que se obtendrán modelos ARIMA con , se tiene que, a partir del tablero de armasubsets de la ***figura 5***, y al ceñirse únicamente el primer renglón, se obtiene un modelo ***ARIMA(9, 1, 12)*** que solo incluye a los términos y , por lo que su ecuación teórica es la ***ecuación 3:***

**[3]**

Sin embargo, si se incorpora el término , si bien continuaría siendo un ***ARIMA(9, 1, 12)***, la ecuación sería como se muestra en la ***ecuación 4***:

**[4]**

Por otro lado, con el tablero de armasubsets de la ***figura 6***, se identifica un modelo ***ARIMA(9, 1, 10)(0, 1, 1)[12]*** que solo emplea los términos y , por lo que ecuación sería la ***ecuación 5:***

**[5]**

Hecho lo anterior, se puede proceder ahora con el ajuste algunos modelos, como se muestra en la tercera sección.

1. **AJUSTE DE MODELOS CON VALIDACIÓN CRUZADA**

Se considerará cuatro modelos SARIMA: **ARIMA(2,1,0)(0,1,2)[12], ARIMA(4,1,0)(1,1,2)[12], ARIMA(6,1,10)(0,1,1)[12]** y **ARIMA(9,1,10)(0,1,1)[12].** En este sentido, primeramente, se presentan la ecuación de los modelos propuestos en la ***tabla 6***.

***Tabla 6.*** Ecuaciones teóricas de los modelos propuestos.

|  |
| --- |
| **Modelo 1. ARIMA(2, 1, 0)(0, 1, 2)[12]** |
| **Modelo 2. ARIMA(4, 1, 0)(1, 1, 2)[12]** |
| **Modelo 3. ARIMA(6, 1, 10)(0, 1, 1)[12]** |
| **Modelo 4. ARIMA(9, 1, 10)(0, 1, 1)[12]** |

Así, es importante realizar la evaluación de la significancia estadística de los parámetros ajustados para cada uno de los modelos, para lo cual es útil valerse de la ***tabla 7***.

|  |  |
| --- | --- |
| **Tabla 7:** Parámetros estimados modelos SARIMA | |
| **Tabla 4a.** Parámetros estimados en Modelo 1   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Parámetros | Estimación | Error Std |  |  | |  | -0.721755 | 0.063428 | -11.3791 | < 2.2×10-16 | |  | -0.332388 | 0.063474 | -5.2366 | 1.636×10-7 | |  | -0.784245 | 0.082040 | -9.5593 | < 2.2×10-16 | |  | -0.125373 | 0.086879 | -1.4431 | 0.149 | | AIC= 5.514306, BIC= 5.844665 | | | | | | **Tabla 4b.** Parámetros estimados en Modelo 2   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Parámetros | Estimación | Error Std |  |  | |  | -0.684906 | 0.067280 | -10.1799 | < 2.2×10-16 | |  | -0.292794 | 0.081324 | -3.6003 | 0.0003178 | |  | 0.023450 | 0.081146 | 0.2890 | 0.7725958 | |  | -0.069487 | 0.066520 | -1.0446 | 0.2962056 | |  | -0.668689 | 0.264019 | -2.5327 | 0.0113177 | |  | -0.091140 | 0.248068 | -0.3674 | 0.7133203 | |  | -0.704732 | 0.192679 | -3.6575 | 0.0002546 | | AIC= 5.644018, BIC= 6.248974 | | | | | |
| **Tabla 4c.** Parámetros estimados en Modelo 3   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Parámetros | Estimación | Error Std |  |  | |  | -0.742123 | 0.062208 | -11.9298 | < 2.2×10-16 | |  | -0.344376 | 0.064892 | -5.3069 | 1.115×10-7 | |  | 0.099804 | 0.054453 | 1.8329 | 0.06682 | |  | 0.079693 | 0.069183 | 1.1519 | 0.24936 | |  | -0.860131 | 0.056870 | -15.1245 | < 2.2×10-16 | | AIC= 5.622459, BIC= 6.046614 | | | | | | **Tabla 4d.** Parámetros estimados en Modelo 4   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Parámetros | Estimación | Error Std |  |  | |  | -0.757956 | 0.061594 | -12.3056 | < 2.2×10-16 | |  | -0.367751 | 0.063044 | -5.8333 | 5.436×10-9 | |  | 0.120641 | 0.066607 | 1.8112 | 0.07010 | |  | 0.154306 | 0.087548 | 1.7625 | 0.07798 | |  | -0.849464 | 0.054735 | -15.5196 | < 2.2×10-16 | | AIC= 5.588376, BIC= 6.009960 | | | | | |

Entonces, teniendo en cuenta los valores *p* mostrados en la quinta columna de las tablas 4a, 4b, 4c y 4d, se puede determinar la significancia estadística de cada uno de los parámetros de interés. Así pues, usando una significancia de 5%, en el modelo 1, todos los parámetros son significativos excepto ; en el modelo 2, todos los parámetros son significativos excepto , y ; en el modelo 3, solo y son no significativos y por último en el modelo 4, solo y son no significativos.

Por otro lado, vale la pena observar la ***figura 7***, en la cual se observa cómo es el ajuste de cada uno de los modelos comparado con la serie temporal. De las gráficas de ajuste se observa que al modelar log() se logra seguir la componente cíclica en todos los modelos SARIMA. Gráficamente no se logra identificar cuál modelo tiene mejor ajuste porque todas siguen la dinámica de la serie de forma adecuada. Además, se puede observar la presencia de datos atípicos que no logran ser explicados por ningún modelo, esto debido a que son observaciones que se comportan diferente, sin embargo, según las medidas de ajuste AIC y BIC, los modelos con mejor ajuste son el modelo 1 y el modelo 4, siendo mejor el modelo 1 que tiene menor AIC y BIC.

|  |  |
| --- | --- |
| **(a)** | **(b)**Gráfico, Histograma, Gráfico de dispersión  Descripción generada automáticamente |
| **(c)** | **(d)**Gráfico, Histograma, Gráfico de dispersión  Descripción generada automáticamente |

**Figura 7.** Graficas con ayuda de ***R*** de la serie original (en negro) y la serie ajustada (en rojo) para cada uno de los modelos, a saber: **(a)** modelo uno: ARIMA(2,1,0)(0,1,2)[12], **(b)** modelo dos: ARIMA(4, 1,0)(1,1,2)[12], **(c)** modelo tres: ARIMA(6, 1, 10)(0, 1, 1)[12] y **(d)** modelo cuatro: ARIMA(9, 1, 10)(0, 1, 1)[12]

1. **ANÁLISIS DE RESIDUALES Y VALIDACIÓN DE SUPUESTOS**

Ahora bien, se debe tener en cuenta que para el planteamiento de cada uno de los modelos se partieron de varios supuestos sobre implícitos al suponer que estos son un ruido blanco, los cuales son: media cero, varianza constante y ausencia de patrones contrarios a la independencia y la distribución normal. No obstante, teniendo en cuenta que estos no son conocidos, para poder validar los supuestos se va a realizar esta revisión mediante los residuales, que son los estimadores de , para lo cual se tiene la ***figura 8,*** en la que se pueden visualizar a los residuales contra el tiempo, y en la ***figura 9,*** los residuales contra los valores ajustados de cada modelo.

Así pues, vale la pena comenzar evaluando si se cumple el supuesto de media cero y varianza constante para cada modelo, y al revisar cada uno de los gráficos se valida que ninguna aporta evidencia en contra de los supuestos de homocedasticidad y media constante en cero, y tampoco se detecta que hayan patrones contrarios a la independencia mediante estos gráficos como ciclos o rachas de signos positivos y negativos.

|  |  |
| --- | --- |
| **(a)**Gráfico  Descripción generada automáticamente | **(b)**Gráfico  Descripción generada automáticamente con confianza baja |
| **(c)** | **(d)**Gráfico  Descripción generada automáticamente con confianza media |

**Figura 8.** Graficas hechas con ayuda de ***R*** de los residuales contra el tiempo para cada modelo, a saber: **(a)** modelo uno: ARIMA(2,1,0)(0,1,2)[12], **(b)** modelo dos: ARIMA(4,1,0)(1,1,2)[12], **(c)** modelo tres: ARIMA(6,1,10)(0,1,1)[12] y **(d)** modelo cuatro: ARIMA(9,1,10)(0,1,1)[12]. Nótese las líneas rojas horizontales presentes en cada modelo. En el medio se marca el residual nulo y las otras dos correspondes a las delimitaciones del positivo y el negativo del doble de la varianza de los residuales asociados con cada modelo.

|  |  |
| --- | --- |
| **(a)**Gráfico, Gráfico de dispersión  Descripción generada automáticamente | **(b)**Gráfico, Gráfico de dispersión  Descripción generada automáticamente |
| **(c)** | **(d)** |

**Figura 9.** Graficas hechas con ayuda de ***R*** de los residuales contra el índice nominal de ventas para cada modelo, a saber: **(a)** modelo uno: ARIMA(2,1,0)(0,1,2)[12], **(b)** modelo dos: ARIMA(4,1,0)(1,1,2)[12], **(c)** modelo tres: ARIMA(6,1,10)(0,1,1)[12] y **(d)** modelo cuatro: ARIMA(9,1,10)(0,1,1)[12]. Nótese las líneas rojas horizontales presentes en cada modelo. En el medio se marca el residual nulo y las otras dos correspondes a las delimitaciones del positivo y el negativo del doble de la varianza de los residuales asociados con cada modelo.

Asimismo, es necesario evaluar las gráficas del ACF y PACF para cada uno de los modelos en la ***figura 10,*** esto con el fin de validar el supuesto de ruido blanco para , donde para el **ACF** el estadístico de prueba es , y que con una significancia de aproximadamente se rechaza la hipótesis nula si , y las hipótesis a contrastar en esta son:

|  |  |
| --- | --- |
| **(a)** | **(b)** |
| **(c)**Gráfico  Descripción generada automáticamente | **(d)**Gráfico  Descripción generada automáticamente |

**Figura 10.** Graficas hechas con ayuda de ***R*** de las funciones de autocorrelación, **ACF,** estimadas para cada modelo, a saber: **(a)** modelo uno: ARIMA(2,1,0)(0,1,2)[12], **(b)** modelo dos: ARIMA(4,1,0)(1,1,2)[12], **(c)** modelo tres: ARIMA(6,1,10)(0,1,1)[12] y **(d)** modelo cuatro: ARIMA(9,1,10)(0,1,1)[12]. Nótese las líneas rojas horizontales presentes en cada modelo, las cuales corresponden a los límites de *Bartlett* para evaluar gráficamente la significancia de la ACF para cada rezago (graficados en el eje de las abscisas).

Y de igual forma, se presentan las gráficas de autocorrelación parcial, PACF, de cada uno de los modelos en la ***figura 11,*** donde el estadístico de prueba es y que con una significancia de aproximadamente se rechaza la hipótesis nula si , y las hipótesis a contrastar en la **PACF** es:

|  |  |
| --- | --- |
| **(a)**Gráfico, Gráfico de cajas y bigotes  Descripción generada automáticamente | **(b)**Gráfico  Descripción generada automáticamente |
| **(c)**Gráfico  Descripción generada automáticamente | **(d)**Gráfico  Descripción generada automáticamente |

**Figura 11.** Graficas hechas con ayuda de ***R*** de las funciones de autocorrelación parciales, P**ACF,** estimadas para cada modelo, a saber: **(a)** modelo uno: ARIMA(2,1,0)(0,1,2)[12], **(b)** modelo dos: ARIMA(4,1,0)(1,1,2)[12], **(c)** modelo tres: ARIMA(6,1,10)(0,1,1)[12] y **(d)** modelo cuatro: ARIMA(9,1,10)(0,1,1)[12]. Nótese las líneas rojas horizontales presentes en cada modelo, las cuales corresponden a los límites de *Bartlett* para evaluar gráficamente la significancia de la PACF para cada rezago (graficados en el eje de las abscisas).

De las anteriores gráficas de ACF y PACF, se puede observar que en ninguno de los modelos se rechaza el supuesto de incorrelación sobre el proceso asociado al , ya que ninguno de estos superan las líneas azules, que son los límites de *Bartlett* para rezagos pequeños, para valores pequeños de rezagos, si bien esto sí sucede para rezagos mayores, pero que pueden ser descartados teniendo presente que la probabilidad de cometer error tipo I aumenta con el valor del rezago.

De la misma forma, se recurre a las pruebas Ljung-Box, cuyos resultados son presentados en la ***tabla 5*** para evaluar el supuesto de incorrelación sobre , de manera que, considerando , se tiene que las hipótesis son:

Y se tiene que el estadístico de prueba es y que tiene como criterio de rechazo que el valor p sea pequeño. Con esto claro, se debe tener presente que se va a realizar seis veces este test conjunto para

Así pues, usando un nivel de significancia de 5%, se tiene a la luz de los resultados que en ningún escenario se rechaza la hipótesis nula, por lo que hay evidencia muestral suficiente para sugerir que son incorrelacionados.

**Tabla 5.** Test de *Ljung-Box* para para los modelos uno, dos, tres y cuatro con .

|  |  |
| --- | --- |
|  | |
| |  |  |  |  | | --- | --- | --- | --- | | ***Tabla 5a.*** Modelo uno. | | | | | **m** | **QLB** | **Grados de libertad** | **P(χ2>QLB)** | | 6 | 1.982275 | 6 | 0.9213216 | | 12 | 9.194579 | 12 | 0.6862271 | | 18 | 16.002607 | 18 | 0.5923654 | | 24 | 28.494640 | 24 | 0.2397309 | | 30 | 32.057837 | 30 | 0.3648423 | | 36 | 42.033452 | 36 | 0.2258933 | | |  |  |  |  | | --- | --- | --- | --- | | ***Tabla 5b.*** Modelo dos. | | | | | **m** | **QLB** | **Grados de libertad** | **P(χ2>QLB)** | | 6 | 0.7928482 | 6 | 0.9922641 | | 12 | 6.3790678 | 12 | 0.8957813 | | 18 | 13.5704190 | 18 | 0.7566350 | | 24 | 23.8393628 | 24 | 0.4708133 | | 30 | 27.9345550 | 30 | 0.5739049 | | 36 | 37.1007297 | 36 | 0.4180456 | |
| |  |  |  |  | | --- | --- | --- | --- | | ***Tabla 5c.*** Modelo tres. | | | | | **m** | **QLB** | **Grados de libertad** | **P(χ2>QLB)** | | 6 | 3.137233 | 6 | 0.7914361 | | 12 | 6.694835 | 12 | 0.8771039 | | 18 | 11.050634 | 18 | 0.8921956 | | 24 | 26.252068 | 24 | 0.3405000 | | 30 | 29.341449 | 30 | 0.4997173 | | 36 | 40.612890 | 36 | 0.2742943 | | |  |  |  |  | | --- | --- | --- | --- | | ***Tabla 5d.*** Modelo cuatro. | | | | | **m** | **QLB** | **Grados de libertad** | **P(χ2>QLB)** | | 6 | 1.788685 | 6 | 0.9380714 | | 12 | 6.832079 | 12 | 0.8685080 | | 18 | 12.671327 | 18 | 0.8107167 | | 24 | 26.435448 | 24 | 0.3314466 | | 30 | 28.376683 | 30 | 0.5504829 | | 36 | 39.258612 | 36 | 0.3259529 | |

Por último, sobre la normalidad de , donde se quiere contrastar las siguientes hipótesis: vs , se puede observar que en los tests de Shapiro-Wilk en la ***tabla 5***, todas las pruebas se rechazan la normalidad usando una significancia de 5%, además, en los gráficos de normalidad de la ***figura 12,*** se puede observar que los residuos de los modelos **uno, dos** y **cuatro** se da un alejamiento de las colas, lo que indica que la distribución real de los errores podría ser alguna que tenga mayor probabilidad de valores atípicos en comparación con una distribución normal. Por otro lado, en el modelo **tres,** se observa que sus residuales presentan un mejor ajuste para ambas colas, de manera que se concluye que los modelos **uno, dos** y **cuatro** tienen errores de ajuste que no se distribuyen normal, mientras que los errores de ajuste del **modelo tres** sí se distribuyen normal.

|  |  |
| --- | --- |
| Gráfico, Gráfico de líneas  Descripción generada automáticamente**(a)** | **(b)**Gráfico, Gráfico de líneas, Gráfico de dispersión  Descripción generada automáticamente |
| **Gráfico, Gráfico de líneas  Descripción generada automáticamente(c)** | **(d)**Gráfico, Gráfico de líneas  Descripción generada automáticamente |

***Figura 12.*** Tests de *Shapiro-Wilks* para verificar normalidad de los para los modelos uno, dos, tres y cuatro.

***Tabla 5.*** Tests de *Shapiro-Wilks* para verificar normalidad de los para los modelos uno, dos, tres y cuatro.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Modelo** | **Errores estructurales** | **Estadístico W** | **Valor p** | **Validez R.B.** | **Validez normalidad** |
| Uno | ARIMA(2,1,0)(0,1,2)[12] | 0.98552 | 0.01577 | Sí | No |
| Dos | ARIMA(4,1,0)(1,1,2)[12] | 0.98415 | 0.009155 | Sí | No |
| Tres | ARIMA(6,1,10)(0,1,1)[12] | 0.98811 | 0.04529 | Sí | Sí |
| Cuatro | ARIMA(9,1,10)(0,1,1)[12] | 0.98572 | 0.01711 | Sí | No |
| **NA.** No aplica la evaluación del test de normalidad por no cumplirse la incorrelación. | | | | | |

De esta manera, se concluye que el mejor modelo es el **tres,** dado que cumple con todos los supuestos sobre los errores de ajuste, ya que estos son independientes e idénticamente distribuidos como una normal con media nula y varianza constante.

1. Estudiante de Estadística, Universidad Nacional de Colombia – Sede Medellín. [↑](#footnote-ref-1)
2. Estudiante de Estadística, Universidad Nacional de Colombia – Sede Medellín [↑](#footnote-ref-2)
3. Estudiante de Estadística, Universidad Nacional de Colombia – Sede Medellín [↑](#footnote-ref-3)