# 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 manufacturero 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 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 las últimas observaciones como parte del periodo *ex post* para verificar la calidad del pronóstico mediante la estrategia de 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)*** se encuentra el índice de ventas del sector manufacturero colombiano en pesos nominales presentado en escala original. En esta se puede observar que la serie muestra tendencia creciente, ya que el índice bajo estudio tiende a aumentar con el tiempo; asimismo, se puede evidenciar cómo este índice va presentando una mayor varianza conforme pasa 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 su evolución se puede calificar como 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 aproximadamente 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 brusca 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 en comparación con la mayoría de los años, según como se acabó de describir, esto gracias a la presencia de ciclos en la serie, como ocurre por ejemplo entre los años 2008 a 2011, lo cual implica que una modelación global de la estacionalidad pueda tener inconvenientes en contraste con un modelo local que puede tener mejores resultados en este respecto. 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 al hacer una transformación logarítmica en la escala del índice de ventas del sector manufacturero colombiano en pesos nominales.

A continuación, es necesario observar la ***figura 2,*** para la que se considera la serie recortada con las primeras observaciones, 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  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. Nótese que los rezagos asociados a la estacionalidad han sido delineados con una línea punteada azul oscura vertical.

Se comienza entonces discutiendo la serie temporal transformada mediante logaritmo natural y sin diferenciar por tendencia ni estacionalidad, graficada en la ***figura 2 (a).*** En esta salta a la vista inicialmente que la media no es constante gracias a la componente estructural de la tendencia, que hace que la media sea creciente, pero también gracias a la estacionalidad, que provoca que disminuya al comenzar cada año respecto al último periodo del año anterior y aumente hasta finalizar el año. No obstante, parece que la varianza es constante, lo cual se debe a la transformación logarítmica. Luego, estudiando la función de autocorrelación muestral (en adelante ***ACF***) asociada a esta serie e ilustrada en la ***figura 2 (b),*** definida en este caso como , se observa para primeros seis rezagos que la función de autocorrelación toma valores muy próximos a uno, lo que hace factible la existencia de raíz unitaria regular en esta serie, lo cual hace necesario diferenciar por tendencia a la serie, y de igual manera, se observa que los valores de la función de autocorrelación disminuyen muy lentamente, lo cual implica que la parte regular no es un proceso ergódico; después, al analizar la parte estacional en , se evidencia una cola positiva que decae de forma lenta, lo cual indica que también sea muy probable la existencia de una raíz unitaria estacional, por lo que resultará también conveniente hallar la primera diferencia estacional para el logaritmo del índice de ventas en pesos nominales del sector manufacturero, y como también se observa que estos valores se acercan a cero de forma lenta, se concluye que la parte estacional no es ergódica. Con todo lo anterior, se concluye que este primer proceso **no es estacionario.**

Después, tomando la primera diferencia del logaritmo del índice de interés cuya gráfica se tiene en la ***figura 2 (c)***, se puede evidenciar que el nivel se ha estabilizado pero la media no es constante ya que persiste el patrón periódico en la serie, reflejado en los picos que se observan a lo largo de la serie y que ocurren justo en el comienzo y final de cada año (nótense los picos que se ciernen sobre las líneas punteadas verticales de los años 2005, 2010 y 2015); la varianza por su parte parece ser constante. Luego, pasando a la ACF de esta serie, definida como y retratada en la ***figura 2 (d)***, y comenzando con la parte regular, se puede evidenciar que la ACF, para los primeros seis rezagos, toma valores que son cercanos a cero, por lo que podría ser un proceso ergódico; ahora, respecto a la parte estacional, se puede notar que la ACF toma sus valores más altos en y con disminuciones progresivas muy pequeñas, lo que es señal de que el proceso, en su parte estacional, no es ergódico. Así, se concluye que el proceso diferenciado por tendencia **no es estacionario.** No obstante, es importante notar que basta con diferenciar una vez la serie por tendencia, ya que la tendencia es estabilizada y la parte regular del proceso es ergódico.

A continuación, se va a analizar el proceso diferenciado por estacionalidad, graficado en la ***figura 2 (e)***, en la cual se puede que ya no hay un patrón periódico exacto, si bien el nivel no es estable, siendo claro entonces que la media no es constante; de la misma manera, se evidencia que la variación del proceso diferenciado por estacionalidad no es igual para todos los periodos (evidenciado por ejemplo el primer quinquenio con una alta variabilidad en contraste con el último), pero es importante señalar que ya no se observa un patrón periódico en la serie. Así pues, avanzando con la ACF muestral, definida en este caso como , resalta para la parte regular con los primeros cinco rezagos que estos toman valores muy cercanos a uno y con un decaimiento lento, por lo que se concluye que la parte regular no es ergódica y, además, parece que su patrón es tipo cola sinusoidal exponencial. Luego, en cuanto a la parte regular, se observa que siempre toma valores que pueden ser considerados estadísticamente iguales a cero, por lo que se puede tomar como ergódico. Con todo esto se concluye que la serie diferenciada por primera vez por estacionalidad **no es estacionaria,** aunque se debe señalar que es suficiente con diferenciarla una vez por estacionalidad en tanto esto hace desaparecer las componentes periódicas exactas o aproximadamente exactas y la parte estacional es ergódica, eliminando entonces la raíz unitaria estacional.

**Tabla 1.** Resumen de los análisis realizados sobre cada una de las cuatro series

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Característica** | **Serie uno** | **Serie dos** | **Serie tres** | **Serie cuatro** |
|  |  |  |  |
| **Media constante** | ✘ | ✘ |  |  |
| **Varianza constante** | ✔ | ✔ |  |  |
| **Ergodicidad en la tendencia** | ✘ | ✔ |  |  |
| **Ergodicidad en la estacionalidad** | ✘ | ✘ |  |  |
| **Estacionariedad en covarianza** | **✘** | ✘ |  |  |

**XXXX. BORRAR >> TEST HEGY**

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 | 1.6332640 | 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(2, 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(6, 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.

|  |  |  |  |
| --- | --- | --- | --- |
| ***Parte AR*** | | ***Parte MA*** | |
| **Celda sombreada** | **Parámetros identificados** | **Celda sombreada** | **Parámetros identificados** |
| 1 |  | 10 |  |
| 2 |  | 12 |  |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Parte AR*** | | ***Parte MA*** | |
| **Celda sombreada** | **Parámetros identificados** | **Celda sombreada** | **Parámetros identificados** |
| 1 |  | 10 |  |
| 2 |  | 12 |  |
| 9 |  |  |  |

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)