# Estadística III 3009137, semestre 01 de 2022

Equipo de Trabajo No. 2 Serie No. 3 Curso: Ma – Ju, Horario: 10:00 – 12.00

## **ANÁLISIS DE SERIES DE TIEMPO. MODELOS DE REGRESIÓN DE ERRORES ESTRUCTURALES ARMA(p, q) ESTACIONARIOS**

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

*Fecha de entrega: 03-06-2022*

1. **INTRODUCCIÓN**

*Bla, bla, bla.*

1. **ANÁLISIS DESCRIPTIVO**

A continuación se realizará un estudio de la serie de tiempo de ventas nominales que es publicada con periodicidad mensual por el DANE iniciando en enero de 2001 y concluyendo, de momento, en noviembre de 2021, para lo cual se tendrá en cuenta no solo las componentes de la tendencia y la estacionalidad, sino también el error estructural, de tal forma que se logre mejorar el ajuste que se realizaría a partir un modelo global únicamente que ignore tal componente del error estructural, y que por tanto no tenga en consideración a la componente cíclica que, como se verá enseguida, es relevante para esta serie temporal. Dicho esto, es importante tener presente que esta serie cuenta con registros hasta el momento, y entonces, para poder entender mejor esta serie de tiempo, se debe comenzar observando la ***figura 1***.

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

***Figura 1.*** Gráficos descriptivos asociados a la serie temporal del índice de ventas nominales mensuales entre enero de 2001 y noviembre de 2021 calculado y publicado por el *Departamento Administrativo Nacional de Estadística, DANE* de Colombia con modificaciones para los meses abarcados por la pandemia de la *COVID-19*. **(a)** Índice de ventas nominales contra el tiempo. Nótese que la escala del índice nominal de ventas está en escala original, mientras que el tiempo es presentado en años calendario. En esta gráfica se nota que hay un aumento en la variabilidad de este índice conforme avanza el tiempo. **(b)** Índice de ventas nominales contra el tiempo. En este caso el índice de ventas nominales se encuentra en escala logarítmica y puede evidenciarse que se ha estabilizado la variabilidad que se da alrededor de la tendencia en la figura *(a)* con la escala original. **(c)** Gráfica de la función de autocorrelación del *logaritmo* del índice de ventas nominales mensuales, la cual refleja la presencia de la componente estacional en la serie

Con ayuda de la figura 1 se puede realizar un análisis descriptivo de esta serie temporal. Así pues, se debe comenzar en la ***figura 1 (a)***, en la cual se refleja que la tendencia de la serie es creciente en tanto el índice de ventas nominales en Colombia tiende a aumentar con el tiempo, y se tiene también que la serie es multiplicativa, ya que la variabilidad alrededor del valor medio de este índice incrementa con el tiempo; además, es posible advertir la presencia de un comportamiento estacional, dado que hay un patrón repetitivo anualmente.

De igual forma, al observar la ***figura 1 (b)*** que realiza una transformación logarítmica a la escala original del índice de ventas nominales, se confirma que la tendencia es creciente y es claro igualmente que esta es determinística, ya que es posible identificar la presencia de efectos permanentes en las componentes estructurales de la serie temporal, y es destacable el hecho que en el rango histórico de la serie la tendencia aparenta ser global, de manera que si se define a como el logaritmo del índice de ventas nominales mensuales de Colombia en un tiempo , entonces pueda ser modelada en la escala logarítmica como , donde los asteriscos indican que se asocian a una transformación logarítmica, donde y serán descritos a continuación.

Es así que resulta posible hacer una representación de la tendencia mediante una curva suave de la forma , donde representa el tiempo de cada uno de los periodos de esta serie (por ejemplo, enero de 2001 está asociado a ), es el valor del componente de la tendencia en la escala logarítmica del índice de ventas nominales para un tiempo y es el grado más alto del polinomio que define a la tendencia; además, a partir de la forma que muestra la tendencia, es razonable plantear que y que es impar, ya que un modelo cuadrático o par mostrará que en algún punto la tendencia comenzará a decrecer, lo cual no se corresponde con lo que se advierte en esta figura. Adicional a lo anterior, es importante tener presente que las cualidades determinística y global de la tendencia son posibles gracias que han sido imputados los índices realmente observados entre marzo de 2020 y noviembre de 2021 por los efectos que tuvo la pandemia de COVID-19 en la economía y por tanto en índices económicos como este; asimismo, esta representación podría tener dificultades por la presencia de ciclos en algunos periodos, fundamentalmente entre los años 2007 y 2015, lo que abre la puerta al planteamiento de modelos locales.

Ahora bien, respecto a la estacionalidad, como se puede ver en las ***figuras 1 (a)*** y ***1 (b)***, existe un comportamiento repetitivo en el índice de ventas nominales dentro de un año calendario, teniendo un aumento progresivo a lo largo del año con algunos picos dentro de cada año, y con esto es posible para los modelos globales considerar funciones trigonométricas, y a través de un periodograma es posible mostrar que esta para la escala logarítmica puede ser modelada como , para , de manera que en la escala original se obtendría que para los antes mencionados. Además, al analizar la ***figura 1 (c),*** se corrobora la presencia de la componente estacional, puesto que para observaciones rezagadas periodos en el tiempo hay una interrupción en el decrecimiento de los valores estimados de la ACF para tener un aumento ligero. Además, nótese que en general se tiene un patrón cola de decaimiento lento, asociado con la presencia de la componente estacional en esta serie temporal, lo cual implica que el índice de ventas nominales no se asocia con un proceso ergódico. Por otro lado, es destacable que , lo cual se debe a la presencia de ciclos en la serie como se describió antes.

A continuación, es posible determinar si esta serie es estacionaria o no, y rápidamente es posible descartar esta posibilidad, puesto que una serie estacionaria demanda que se tenga varianza y media constantes, y si bien la variabilidad se logra estabilizar con la transformación logarítmica de la escala original del índice de ventas nominales, dado a que esta tiene tendencia, se tiene que la media no es constante. Adicional a esto, teniendo la ACF se pueden realizar tests para si el logaritmo del índice de ventas nominales es ruido blanco, para lo que se plantean las siguientes hipótesis para cada :

Y para los cuales el estadístico de prueba es tal que y para un se rechaza si , y observando la ***figura 1 (c),*** la cual demanda que ninguna de las barras verticales supere a las franjas rojas que demarcan los valores de , se concluye que la función de autocorrelación es significativa para lo que implica que la serie no corresponda a un ruido blanco.

Luego, teniendo presente que es posible hacer un ajuste global a esta serie temporal, se va a proceder con esto tomando un modelo exponencial polinomial estacional de grado seis estacional con funciones trigonométricas en cinco frecuencias , cuya ecuación es la **(1)**:

**(1)**

Para el ajuste de este modelo solo van a ser consideradas las primeras observaciones, dejando las últimas doce como parte del periodo *ex post* para poder hacer validación cruzada del ajuste realizado, y con ayuda de ***R*** se realiza el ajuste de esta serie, cuyos coeficientes estimados, así como sus errores estándar y los valores del estadístico calculado y el valor p asociados a la prueba de significancia de cada uno de estos se presenta en la ***tabla 1.***

**Tabla 1.** Parámetros estimados para el modelo global: exponencial polinomial

estacional de grado seis estacional con funciones trigonométricas

en cinco frecuencias

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parámetros | Estimación | Error estándar |  |  |
|  | 3.55 | 3.637×10-2 | 97.608 | < 2×10-16 |
|  | 2.182×10-3 | 3.478×10-3 | 0.627 | 0.531009 |
|  | -2.801×10-4 | 1.107×10-4 | 2.530 | 0.012113 |
|  | -4.481×10-6 | 1.576×10-6 | -2.844 | 0.004876 |
|  | 2.919×10-8 | 1.111×10-8 | 2.628 | 0.009183 |
|  | -8.674×10-11 | 3.794×10-11 | -2.286 | 0.023192 |
|  | 9.781×10-14 | 5.011×10-14 | 1.952 | 0.052203 |
|  | -4.125×10-2 | 3.581×10-3 | -11.519 | < 2 ×10-16 |
|  | 1.473×10-2 | 3.527×10-3 | 4.176 | 4.27×10-5 |
|  | -2.968×10-2 | 3.581×10-3 | -8.289 | 1.09×10-14 |
|  | 1.357×10-2 | 3.506×10-3 | 3.871 | 0.000143 |
|  | -1.925×10-2 | 3.533×10-3 | -5.448 | 1.35×10-7 |
|  | 2.035×10-2 | 3.550×10-3 | 5.732 | 3.22×10-8 |
|  | -1.534×10-2 | 3.572×10-3 | -4.295 | 2.61×10-5 |
|  | 2.346×10-2 | 3.508×10-3 | 6.688 | 1.81×10-10 |
|  | 4.155×10-3 | 3.542×10-3 | 1.173 | 0.242005 |
|  | 2.378×10-2 | 3.530×10-3 | 6.735 | 1.38×10-10 |
|  | | | | |
| |  | | --- | | =2.899 (escala log.); *=9.9001224, =11.52640* | | | | | |

Y así se llega a que la ecuación de ajuste para este modelo es la siguiente:

**XXXX. Tabla de pruebas de hipótesis y comentarios.**

En la ***figura 2*** se presenta un gráfico en el que se contrasta la serie ajustada, en rojo, con la serie original, en negro.Chart

Description automatically generated

**Figura 2.** Contraste de la serie de tiempo del índice de ventas nominales mensuales de Colombia entre enero de 2001 y noviembre de 2021 en negro y la serie ajustada con un modelo exponencial polinomial de grado seis estacional con funciones trigonométricas en las frecuencias en rojo.

De la ***figura 2*** es claro que el modelo logra captar de forma adecuada la tendencia y la estacionalidad de la serie; empero, esta no logra ajustar los ciclos de esta, lo cual se evidencia en una disparidad entre la serie real y la serie ajustada en algunos periodos, siendo esto especialmente evidente en los periodos de los años 2008 y 2009, donde la serie ajustada sigue la tendencia creciente, mientras que los datos reales reflejan una contracción (coincidiendo con la crisis económica mundial de dichos años), si bien es esperable que esto sucediese teniendo presente que la componente cíclica no fue modelada en el modelo global ajustado. Por último, respecto al ajuste es reseñable que la raíz cuadrada del error cuadrático medio es 2.899, el criterio de información de Akaike equivale a 9.9001224 y el criterio de información bayesiano es igual a 11.52640 aproximadamente.

Ahora bien, a la hora de hacer pronósticos luego del periodo se apela a la siguiente ecuación:

Y a partir de esta ecuación se puede construir la tabla de pronósticos para los meses del periodo *ex post*, que se observa en la ***tabla 2.***

**Tabla 2.** Pronósticos para el periodo *ex post* del modelo exponencial polinomial de grado seis.

|  |  |  |  |
| --- | --- | --- | --- |
| **Periodo** | **L** | **Real** | **Pronóstico** |
| 2020 Dic | 1 | 125.95 | 125.6811 |
| 2021 Ene | 2 | 104.49 | 104.9223 |
| 2021 Feb | 3 | 107.87 | 107.3833 |
| 2021 Mar | 4 | 114.16 | 114.9873 |
| 2021 Abr | 5 | 112.05 | 111.4442 |
| 2021 May | 6 | 116.74 | 117.5733 |
| 2021 Jun | 7 | 115.32 | 155.1623 |
| 2021 Jul | 8 | 116.86 | 118.5237 |
| 2021 Ago | 9 | 120.13 | 120.0877 |
| 2021 Sep | 10 | 121.77 | 122.9627 |
| 2021 Oct | 11 | 123.64 | 124.1830 |
| 2021 Nov | 12 | 127.70 | 128.9216 |

Nótese que por ser el modelo global exponencial, entonces no es posible obtener para las estimaciones intervalos de predicción. De forma adicional, como se conocen los valores reales de los índices de ventas nominales para los meses del periodo *ex post,* entonces se pueden calcular medidas de error como el MAE, el MAPE y el RMSE, las cuales son presentadas en la ***tabla 3.***

**Tabla 3.** Precisión de los pronósticos puntuales.

|  |  |
| --- | --- |
| **Medida** | **Valor** |
| RMSE **\*** | 0.8279634 |
| MAE **\*** | 0.6862845 |
| MAPE (%) | 0.5830712 |
| **\*** Unidades en puntos del índice de ventas nominales. | |

De la tabla tres se obtiene que se concluye que el modelo global se equivocó en promedio en cada pronóstico del periodo ex post en 0.8279634 puntos del índice de ventas nominales, mientras que el MAE señala una equivocación en promedio de 0.6862845 puntos; por último, del MAPE se concluye que el modelo global se ha equivocado en promedio para cada pronóstico un 0.4439505 % respecto a cada valor real. De estas métricas pues se puede concluir que se está logrando un ajuste bueno de la serie en tanto los errores cometidos, de acuerdo a estos valores, son pequeños, y como se pudo evidenciar en la figura ***figura 2*** estos errores se dan fundamentalmente no haber incorporado la componente cíclica en este modelo.

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)