

ANÁLISIS DE SERIES DE TIEMPO: AJUSTE DE TENDENCIA Y ESTACIONALIDAD.

Jessika Julieth Silva Ramos¹, Juan Nicolas Piedrahita Salas², Karolina Tupue Puenayan³, Paulina Ríos Pulgarín⁴, y Yarleny Cataño Posada⁵

Fecha de entrega: 29 – 04 – 2022

1. Introducción

El DANE, con el fin de lograr un mejor estudio y análisis de los fenómenos económicos del país, se vio en la necesidad de clasificar o agrupar los procesos económicos existentes en categorías lo más homogéneas posible respecto de las características que son el objeto de las estadísticas en cuestión, de esta necesidad surgió la CIIU, que tiene por finalidad establecer una clasificación uniforme de las actividades económicas productivas^[1]. Para la EMMET la clasificación se compone de 39 dominios, o agrupaciones de la CIIU Rev.4 A.C (DANE) a nivel nacional y su desagregación por departamentos, áreas metropolitanas y principales ciudades del país^[2].

Con el fin de estudiar y llevar el control en el corto plazo de la evolución de las principales variables económicas del sector fabril colombiano, lleva a cabo la Encuesta Mensual Manufacturera con Enfoque Territorial (EMMET), una investigación que como su nombre lo indica, mide la evolución mensual del sector manufacturero del país basado en las variables de producción, ventas, empleo, sueldos y horas trabajadas, a partir de allí el DANE genera índices, variaciones y contribuciones; además es muy útil también para la elaboración de las estimaciones del Producto Interno Bruto (PIB) del sector industrial que realiza la Dirección Técnica de Síntesis y Cuentas Nacionales. Su población objetivo son las empresas que hacen parte del sector manufacturero colombiano, los establecimientos industriales manufactureras en el país que ocupan diez (10) o más personas y que en suma produjeron el 80% de la producción industrial reportada por la EAM 2017 y concentran el 65% del empleo total; en cada dominio de estudio publicado (dominio para departamento, área metropolitana, ciudad). La población objetivo se delimita a estos establecimientos considerando en primer lugar el tipo de distribución de la producción de la actividad manufacturera. El sector manufacturero tiene una distribución asimétrica al estar concentrada la mayoría de la producción en pocos establecimientos relativamente a la cantidad de establecimientos. El objetivo de la operación estadística es medir la evolución del sector manufacturero; por lo tanto, este grupo de establecimientos a los que se limita la población objetivo es preciso para determinar el comportamiento mensual^[2].

Producción nominal es el valor de los productos elaborados y los subproductos y desechos que resultan de la producción y que se destinan a la venta, valor de los productos manufacturados para terceros que entregan la materia prima, valor de los ingresos por servicios industriales, valorados a precio promedio de venta en fábrica y sin incluir los impuestos indirectos (IVA, consumo, etc.). Y las ventas nominales son el valor de los productos y subproductos elaborados por el establecimiento y vendidos durante el mes de referencia, a precio de venta en fábrica y sin incluir los impuestos indirectos. Incluye también aquellos productos elaborados por otros establecimientos a los cuales se haya suministrado la materia prima para su transformación, el valor cobrado por los productos elaborados para terceros, industriales y no industriales y el valor de los ingresos por servicios industriales prestados. El valor de ventas se precisa según destino, así: ventas realizadas en el país y al exterior. por otro lado las producción y ventas reales es el valor nominal de la producción y las ventas respectivamente, deflactadas por el índice de precios al productor según clase industrial^[2].

Según la CIIU del DANE, se entiende por industria manufacturera a toda unidad dedicada a la transformación física o química de materiales, sustancias y componentes en productos nuevos ya sea manualmente, a través de maquinaria, que elaboren los productos por sí mismos o que contraten a otras unidades para que lo hagan en su lugar^[1]. Para este trabajo se analiza la elaboración de productos de panadería, macarrones, fideos y sus productos, que corresponde a la industria manufacturera para obtener los índices de producción real, que no tienen en cuenta el efecto de la inflación y se trabajan a partir de los valores del año base, a diferencia de los índices de producción nominal, que sí lo hace^[5]. Los datos usados fueron tomados de la EMMET recopilada mensualmente, entre enero de 2001 y noviembre de 2021, y específicamente la variable producción, con 251 observaciones de frecuencia mensual.

¹Estudiante Ingeniería Industrial, Universidad Nacional de Colombia – Sede Medellín.

²Estudiante Ingeniería Sistemas e Informática, Universidad Nacional de Colombia – Sede Medellín.

³Estudiante Ingeniería Industrial, Universidad Nacional de Colombia – Sede Medellín.

⁴Estudiante Ingeniería Industrial, Universidad Nacional de Colombia – Sede Medellín.

⁵Estudiante Ingeniería Industrial, Universidad Nacional de Colombia – Sede Medellín.

Un número índice es una medida estadística que permite observar la evolución, cambios o el estado de uno o varios fenómenos económicos en el tiempo, comparándolos con un periodo base o de referencia que se fija arbitrariamente. Estos índices se clasifican en dos grupos: el primero, son los números índices simples, que proporcionan la variación o evolución que ha sufrido un fenómeno económico o magnitud entre dos periodos o locaciones. El segundo, son los números índices complejos, que proporcionan la evolución conjunta de dos o más fenómenos económicos o magnitudes, aquí se tiene otra clasificación. Los números índices complejos se clasifican en ponderados y sin ponderar, para los primeros se tiene que cada magnitud tiene un peso distinto asignado en función de varios criterios que ayudan a definir su importancia relativa, mientras que para los segundos, cada magnitud posee el mismo peso o importancia que las demás^[3].

Los datos acerca de las características de los establecimientos industriales manufactureros son recopilados con fines estadísticos para agregarlos a nivel de dominios y total del sector. Con esta información se obtiene el índice de producción industrial manufacturera (IPIM), el cual permite conocer la estructura y evolución del sector manufacturero en Colombia. Para estas mediciones, se recolecta información de las características de los establecimientos industriales manufactureros y a partir de esta, se obtiene el **índice de producción industrial manufacturera (IPIM)**, el cual permite conocer la estructura y evolución del sector manufacturero en Colombia. Este es un índice ponderado por la relación de valor agregado sobre la producción bruta de 2017, se presenta en año base 2018. El IPIM para el total nacional es el resultado de sumar la multiplicación de cada uno de los valores de producción de los establecimientos por su respectiva ponderación y dividirlo en el año base, su cálculo se define a continuación^[2]:

1. Valor de la producción total ponderada en el mes de referencia:

$$VPTP^{(t)} = \sum_{j=1}^m \sum_{r=1}^n \sum_{e=1}^l (VP_{ejr}^{(t)} * W_{jr}) \quad (1)$$

$VP_{ejr}^{(t)}$: Valor total de la producción para el establecimiento e en el dominio j en la región r para el periodo de referencia t .

W_{jr} : Ponderador del dominio j en la región r .

2. Valor de la producción total ponderada para el mes t en el año 2018

$$A_{2018}^{(t)} = \sum_{j=1}^m \sum_{r=1}^n \sum_{e=1}^l (VP_{ejr}^{(t)} * W_{jr}) \quad (2)$$

En este caso se calcula para $t = 1, 2, 3, \dots, 12$ donde t recorre los meses del año 2018.

3. Índice de producción industrial manufacturera

Es el resultado de la división de (1) en el promedio de (2)

$$IPIM^{(t)} = \frac{\sum_{j=1}^m \sum_{r=1}^n \sum_{e=1}^l (VP_{ejr}^{(t)} * W_{jr})}{\frac{\sum_{j=1}^m A_{2018}^{(t)}}{12}} \quad (3)$$

Este cálculo aplica tanto para producción nominal, producción real, ventas nominales y ventas reales, que son las variables ponderadas.

4. Valor total de la variable X en el mes de referencia

$$VTX^{(t)} = \sum_{j=1}^m \sum_{r=1}^n \sum_{e=1}^l VX_{ejr}^{(t)} \quad (4)$$

donde $VX_{ejr}^{(t)}$: Valor total de la variable X para el establecimiento e en el dominio j en la región r para el periodo de referencia.

5. Valor total de la variable X para el mes t del año 2018

$$B_{2018}^{(t)} = \sum_{j=1}^m \sum_{r=1}^n \sum_{e=1}^l VX_{ejr}^{(t)} \quad (5)$$

Este es calculado para $t = 1, 2, 3, \dots, 12$ donde t recorre los meses del año 2018

6. Índice de la variable X en el mes de referencia: resultado de división entre (4) y (5)

$$IX^{(t)} = \frac{\sum_{j=1}^m \sum_{r=1}^n \sum_{e=1}^l (VP_{er}^{(t)})}{\sum_{j=1}^m B_{2018}^{(j)}} \quad (6)$$

El índice de producción real es calculado como un índice complejo ponderado, utilizando la fórmula (3) como los datos de la EMMET son recogidos de 6 regiones del país, n es igual a 6 y al haber solo un dominio $m=1$. Teniendo en cuenta que en este caso $VP_{er}^{(t)}$ representa el valor total de la producción real de productos de panadería, macarrones, fideos y sus productos para el establecimiento e en la región r para un periodo t . Es de tipo % ya que las unidades de numerador y denominador se cancelan.

Por ejemplo, teniendo en cuenta el cálculo del índice (3), el valor total de producción real en el mes de enero de 2001 ponderadas según la importancia del dominio: productos de panadería en la región r representa el 70,5% del valor promedio ponderado obtenido para el total de producción real del año base 2018.

Cuando se produce un cambio de metodología en la elaboración de los índices de una serie se genera una ruptura en la continuidad de dicha serie. Desde un punto de vista teórico esta ruptura no admite solución; sin embargo, es necesario disponer de series continuadas, al menos a ciertos niveles de desagregación, que permitan calcular tasas de variación entre diferentes periodos, disponer de series continuadas ofrece a los usuarios la posibilidad de realizar estudios y predicciones acerca de la evolución de los precios^[4].

El Instituto Nacional de Estadística calculará un enlace denominado enlace legal, para mantener en lo posible esta continuidad. Este enlace conserva las tasas de variación anteriores y posteriores al momento de implantación del nuevo sistema, este procedimiento se basa en el cálculo de un coeficiente de enlace y recibe el nombre de enlace legal porque tradicionalmente lo aplica el Instituto Nacional de Estadística en sus certificaciones oficiales.

$$\text{Coeficiente de enlace} = \frac{I_o \text{ nueva}}{I_o \text{ antigua}}$$

$$I_1 \text{ nueva} = I_1 \text{ antiguo} \times \text{Coeficiente de enlace}$$

El coeficiente de enlace se obtiene como cociente entre los índices del mes x de base antigua y base nueva.

2. Análisis descriptivo de la serie y modelos propuestos

2.1 Análisis descriptivo

Serie de datos Y_t = puntos porcentuales del índice de producción real de la elaboración de productos de panadería, macarrones, fideos y sus productos respecto al tiempo. La serie que se ilustra en la figura 1(a), va desde enero de 2001 a noviembre de 2021, con un total de $N = 251$ observaciones. Se desea ajustar e implementar la estrategia de validación cruzada así: ajuste con los primeros $n = 239$ datos (es decir, hasta noviembre de 2020) y pronóstico ex- post para los últimos $m = 12$ datos que corresponden al período comprendido entre diciembre de 2020 hasta noviembre de 2021. En la figura 2 se exhibe la tendencia de la descomposición aditiva del logaritmo natural de la serie, los boxplots comparativos por meses y el periodograma para evaluar la presencia de efectos estacionales.

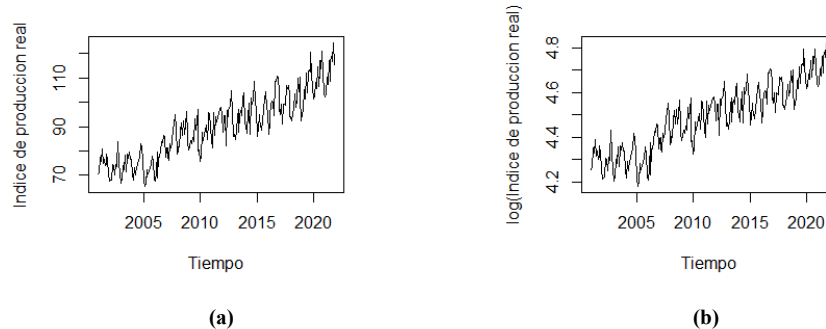


Figura 1. (a) Índice de producción real mensual de la elaboración de productos de panadería, macarrones, fideos y sus productos; **(b)** Logaritmo natural de la serie

En la figura 1(a), puede observarse un rango de variación anual que está creciendo levemente a medida que la tendencia crece, por lo que se sospecha que la serie es de componentes multiplicativas. En la figura 1(b) se analiza la serie transformada en logaritmo natural de la serie para identificar si ésta es de componentes aditivas o multiplicativas. Se observa un patrón de varianza más estable en la gráfica $\log(Y_t)$ vs. t que en la gráfica de la serie sin transformar, y la variabilidad cambia en el sentido creciente de la tendencia, entonces se considera que la serie es de componentes multiplicativas. Por otra parte, se nota claramente un patrón de tendencia creciente en ambas gráficas, 1(a) y 1(b) y se pueden observar patrones estacionales, es decir, es visible un patrón que se repite anualmente y este tiende a ser similar con el paso de los años. Este patrón aumenta desde enero, en el mes de octubre alcanza su pico máximo y en diciembre vuelve a bajar.

Debido a que se considera una serie de componentes multiplicativas, se analizarán las descomposiciones aditivas del logaritmo natural de la serie para determinar qué modelos pueden formularse para la componente de tendencia y estacionalidad en la escala logarítmica. Por esto, y teniendo en cuenta que la serie es de componentes multiplicativas, se percibe evidencia de tendencia creciente y de componente estacional. Las posibles ecuaciones de la serie son:

Caso completamente multiplicativo, $Y_t = T_t \times S_t \times \exp(E_t)$, $E_t \sim iid N(0, \sigma^2)$

$$\log(Y_t) = \log(T_t) + \log(S_t) + E_t, E_t \sim iid N(0, \sigma^2)$$

$$Y_t^* = T_t^* + S_t^* + E_t, E_t \sim iid N(0, \sigma^2)$$

Caso parcialmente multiplicativo, $Y_t = T_t \times S_t + E_t$, $E_t \sim iid N(0, \sigma^2)$

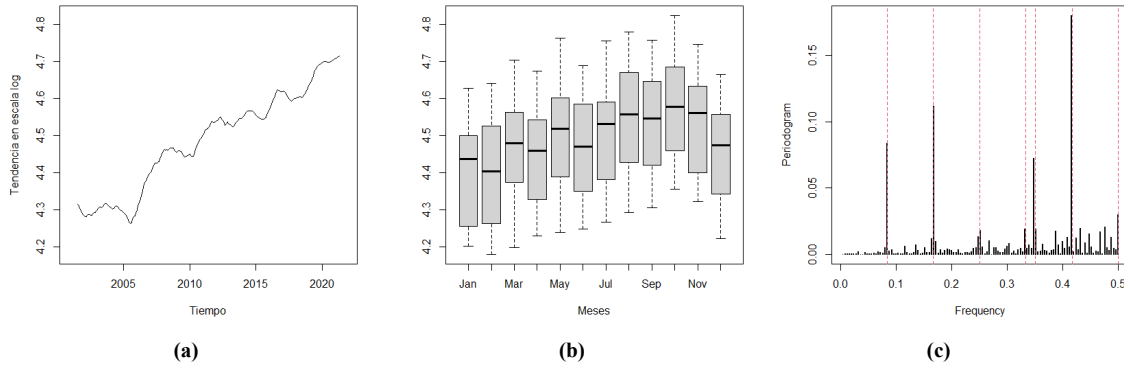


Figura 2. (a) Componente de tendencia para el logaritmo natural; (b) Distribución por meses del logaritmo natural de la serie; (c) Periodograma sobre las diferencias de los logaritmos naturales de la serie.

En la figura 2(a), la estimación de tendencia sobre el logaritmo natural puede caracterizar un patrón creciente y se identifica la presencia de ciclos. Cualitativamente teniendo en cuenta que la serie es la elaboración de productos de panadería, macarrones, fideos y sus productos vemos que los datos históricos muestran que se ha ido incrementando en el tiempo y sabemos también que a mayor población mayor consumo, a la larga la serie debe ser creciente, esto mismo se espera en la escala log, entonces se puede identificar que un posible grado de polinomio que podría explicar la trayectoria a largo plazo sin seguir ciclos en la escala logaritmo natural es $p = 3$. Un posible modelo para la Tendencia del $\log(Y_t) = T^* = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 \rightarrow Y_t = \exp(T_t^*)$.

También se observan cambios estructurales en el año 2005 por lo que es preferible ajustar la tendencia localmente.

En la figura 2(b), en el boxplot se observa que la media de la distribución de la serie en escala logaritmo natural según el periodo del año, no es constante y cambia a lo largo de un año calendario, lo cual nos indica la presencia de la componente estacional, lo que corrobora lo observado en la figura 1(b).

En la figura 2(c) del periodograma sobre las diferencias de los logaritmos naturales de la serie, se muestra una medida que da cuenta de la asociación entre la serie y un fenómeno periódico. Se puede observar que existen 7 frecuencias donde puede existir componente periódica significativa, las cuales son $F1 = \frac{1}{12}$, $F2 = \frac{1}{6}$, $F3 = \frac{1}{4}$, $F4 = \frac{1}{3}$, $F5 = \frac{5}{12}$, $F6 = \frac{1}{2}$, y $F7 = 0.35$. Así, un posible modelo para S_t como la suma de las ondas sinusoidales en frecuencias F1, F2, F3, F4, F5, F6 y F7 sería:

$$S_t = \alpha_1 \sin\left(\frac{\pi t}{6}\right) + \gamma_1 \cos\left(\frac{\pi t}{6}\right) + \alpha_2 \sin\left(\frac{\pi t}{3}\right) + \gamma_2 \cos\left(\frac{\pi t}{3}\right) + \alpha_3 \sin\left(\frac{\pi t}{2}\right) + \gamma_3 \cos\left(\frac{\pi t}{2}\right) + \alpha_4 \sin\left(\frac{2\pi t}{3}\right) + \gamma_4 \cos\left(\frac{2\pi t}{3}\right) + \alpha_5 \sin\left(\frac{5\pi t}{6}\right) + \gamma_5 \cos\left(\frac{5\pi t}{6}\right) + \gamma_6 \cos(\pi t) + \alpha_7 \sin\left(\frac{0.35\pi t}{6}\right) + \gamma_7 \cos\left(\frac{0.35\pi t}{6}\right)$$

2.2 Modelos propuestos.

Se postulan 4 modelos, así: los modelos 1 y 2 son correspondientes a ajustes globales de componentes completamente multiplicativos y parcialmente multiplicativos, considerando un polinomio de grado $p \geq 3$.

Modelo 1: log - polinomial estacional de grado 3 usando funciones trigonométricas en la representación de la componente estacional en frecuencias $F_j = \frac{j}{12}$, $j = 1, \dots, 6$; y $F7 = 0.35$.

Modelo 2: exponencial - polinomial estacional de grado 3 usando funciones trigonométricas en la representación de la componente estacional, en frecuencias $F_j = \frac{j}{12}$, $j = 1, \dots, 6$; y $F7 = 0.35$.

Para estos modelos, el número de parámetros es 17, repartidos de la siguiente forma: 4 en la tendencia y 13 en la componente estacional.

Los modelos 3 y 4 corresponden a:

Modelo 3: Modelo de tendencia local de cambio de nivel y de pendientes que evolucionan lentamente con t , multiplicativa a factor estacional con efectos estacionales que evolucionan lentamente en el tiempo, con método de ajuste Suavizamiento Exponencial Holt -Winters multiplicativo.

Modelo 4: Modelo de tendencia local lineal multiplicativa a factor estacional con efectos estacionales constantes en el tiempo, con método de ajuste descomposición multiplicativa combinado con LOESS lineal, cuyo parámetro de suavizamiento se obtiene usando criterio AICC.

Tabla 1. Ecuaciones de los modelos propuestos

<p>Modelo 1 (Log Polinomial de grado 3 multiplicativo, estacional con trigonometricas)</p> $\log(Y_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \alpha_1 \sin\left(\frac{\pi t}{6}\right) + \gamma_1 \cos\left(\frac{\pi t}{6}\right) + \alpha_2 \sin\left(\frac{\pi t}{3}\right) + \gamma_2 \cos\left(\frac{\pi t}{3}\right) + \alpha_3 \sin\left(\frac{\pi t}{2}\right) + \gamma_3 \cos\left(\frac{\pi t}{2}\right) + \alpha_4 \sin\left(\frac{2\pi t}{3}\right) + \gamma_4 \cos\left(\frac{2\pi t}{3}\right) + \alpha_5 \sin\left(\frac{5\pi t}{6}\right) + \gamma_5 \cos\left(\frac{5\pi t}{6}\right) + \gamma_6 \cos(\pi t) + \alpha_7 \sin\left(\frac{0.35\pi t}{6}\right) + \gamma_7 \cos\left(\frac{0.35\pi t}{6}\right) + E_t, \text{ con } E_t \sim iid N(0, \sigma^2)$
<p>Modelo 2 (Exponencial Polinomial de grado 3 multiplicativo, estacional con trigonometricas)</p> $Y_t = \exp(\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \alpha_1 \sin\left(\frac{\pi t}{6}\right) + \gamma_1 \cos\left(\frac{\pi t}{6}\right) + \alpha_2 \sin\left(\frac{\pi t}{3}\right) + \gamma_2 \cos\left(\frac{\pi t}{3}\right) + \alpha_3 \sin\left(\frac{\pi t}{2}\right) + \gamma_3 \cos\left(\frac{\pi t}{2}\right) + \alpha_4 \sin\left(\frac{2\pi t}{3}\right) + \gamma_4 \cos\left(\frac{2\pi t}{3}\right) + \alpha_5 \sin\left(\frac{5\pi t}{6}\right) + \gamma_5 \cos\left(\frac{5\pi t}{6}\right) + \gamma_6 \cos(\pi t) + \alpha_7 \sin\left(\frac{0.35\pi t}{6}\right) + \gamma_7 \cos\left(\frac{0.35\pi t}{6}\right)) + E_t, \text{ con } E_t \sim iid N(0, \sigma^2)$
<p>Modelo 3 (SEHW multiplicativo)</p> $Y_{t+h} = (\beta_{0,t} + \beta_{1,t} \times h) \times \sum_{i=1}^{12} \delta_{i,t} I_{i,t+h} + E_{t+h}, \text{ con } \sum_{i=1}^{12} \delta_{i,t} = 12, \text{ con } \beta_{0,t}, \beta_{1,t} \text{ y los } \delta_{i,t} \text{ nivel, pendiente y efectos estacionales en } t, \text{ respectivamente, cambiando lentamente en el tiempo y } E_t \sim iid N(0, \sigma^2)$
<p>Modelo 4 (Filtro de la descomposición multiplicativa y Loess lineal)</p> <p>En la vecindad de un tiempo t_k ($t_k = 1, 2, \dots, 239$) $Y_t = (\beta_{0,k} + \beta_{1,k} t) \times \sum_{i=1}^{12} \delta_{i,t} I_{i,t} + E_t \quad \forall t$ en la vecindad de t_k, con $E_t \sim iid N(0, \sigma^2)$ y $\sum_{i=1}^{12} \delta_{i,t} = 12$, con parámetros $\beta: \beta_{0,k}, \beta_{1,k}$ en la vecindad con centro en t_k</p>

3. Ajustes de modelos con validación cruzada.

En la tabla 2, se puede observar el valor estimado de los parámetros, el error estándar de cada uno, el estadístico de prueba T_0 y el valor P para probar la significancia individual de los parámetros de los modelos 1 y 2.

En la tabla 3, se determinan las pruebas de hipótesis para probar la significancia, en primer lugar de los parámetros de los modelos con énfasis en la significancia del parámetro β_3 en la componente de tendencia, en segundo lugar, de la componente estacional para los modelos 1 y 2. Analizando los resultados arrojados por las tablas 2a y 2b, se puede concluir que:

Con un nivel de significancia de $\alpha = 0.05$, se observa que el modelo 1 arroja un valor P para β_3 menor al nivel de significancia, por lo que se rechaza H_0 y se asume que el parámetro es significativo y equivalentemente el polinomio cúbico propuesto también lo es, bajo el supuesto de $E_t \sim iid N(0, \sigma^2)$, sin embargo, es necesario comprobar los supuestos sobre los errores del modelo.

Por el contrario, con un nivel de significancia de $\alpha = 0.05$, el modelo 2 muestra un valor P para β_3 mayor a este nivel de significancia, por lo que no se rechaza H_0 y se asume que el parámetro no es significativo y equivalentemente el polinomio cúbico propuesto tampoco lo es, bajo el supuesto de $E_t \sim iid N(0, \sigma^2)$, se recomendaría hacer otro modelo donde se elimine este parámetro y analizar cómo se comporta en ajustes, residuales y pronósticos, puede que ese modelo que no tiene este parámetro que no le está

aportando, se comporte igual e incluso mejor, en este caso es mejor tomar el nuevo modelo porque es más parsimonioso y logra un desempeño igual o mejor.

Finalmente, para la componente estacional, con un nivel de significancia de $\alpha = 0.05$, se puede concluir que, para ambos modelos (1 y 2) los parámetros $\alpha_1, \gamma_1; \alpha_2, \gamma_2; \alpha_3, \gamma_3; \alpha_5, \gamma_5$, y γ_7 , el valor P es menor al nivel de significancia, por lo que se rechaza H_0 y se asume

que los parámetros son significativos, lo que corresponde a que las ondas sinusoidales con frecuencia $F_j = \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{5}{12}$ y 0.35 respectivamente, también lo son, pues se sabe que, si al menos un parámetro (α o γ) es estadísticamente significativo ($\alpha \neq 0$ o $\gamma \neq 0$) la onda en la respectiva frecuencia también lo es. Esto implica que, habiendo al menos una onda significativa, la componente estacional en el modelo 1 y el modelo 2, cumple la misma condición de significancia. Se sugiere hacer otro modelo donde se eliminen las ondas correspondientes a los parámetros con subíndice $j = 4, 6$ que no aportan a los modelos, y analizar cómo se comporta en ajustes, residuales y pronósticos, puede que estos modelos que no tienen estas ondas, se comporten igual e incluso mejor, en este caso es mejor tomar los nuevos modelos porque son más parsimoniosos y logran un desempeño igual o mejor.

Tabla 2. Parámetros estimados Modelos de regresión global

Tabla 2a. Parámetros estimados en Modelo 1 (Log Polinomial)					Tabla 2b. Parámetros estimados en Modelo 2 (Exponencial Polinomial)				
Parámetros	Estimación	Error Std	T_0	$P(t_{222} > T_0)$	Parámetros	Estimación	Error Std	T_0	$P(t_{222} > T_0)$
β_0	4.274	1.216×10^{-2}	351.541	0.00000	β_0	4.260	1.396×10^{-2}	305.194	0.00000
β_1	9.452×10^{-4}	4.378×10^{-4}	2.159	0.031939	β_1	1.669×10^{-3}	4.653×10^{-4}	3.588	0.00041
β_2	9.583×10^{-6}	4.234×10^{-6}	2.263	0.024594	β_2	2.348×10^{-6}	4.272×10^{-6}	0.550	0.58307
β_3	-2.737×10^{-8}	1.160×10^{-8}	-2.360	0.019141	β_3	-7.813×10^{-9}	1.128×10^{-8}	-0.693	0.48911
α_1	-5.178×10^{-2}	4.223×10^{-3}	-12.261	0.00000	α_1	-5.224×10^{-2}	4.028×10^{-3}	-12.969	0.00000
γ_1	-1.018×10^{-2}	4.232×10^{-3}	-2.407	0.016910	γ_1	-1.030×10^{-2}	4.100×10^{-3}	-2.511	0.01275
α_2	-2.922×10^{-2}	4.215×10^{-3}	-6.932	4.45×10^{-11}	α_2	-2.864×10^{-2}	4.041×10^{-3}	-7.087	1.80×10^{-11}
γ_2	-1.555×10^{-2}	4.232×10^{-3}	-3.674	0.000299	γ_2	-1.629×10^{-2}	4.07×10^{-3}	3.993	8.88×10^{-5}
α_3	-9.523×10^{-3}	4.214×10^{-3}	-2.260	0.024787	α_3	-9.045×10^{-3}	4.045×10^{-3}	-2.247	0.02562
γ_3	-4.906×10^{-3}	4.232×10^{-3}	-1.159	0.247526	γ_3	-4.355×10^{-3}	4.094×10^{-3}	-1.064	0.28859
α_4	-2.881×10^{-3}	4.213×10^{-3}	-0.684	0.494797	α_4	-4.445×10^{-3}	4.042×10^{-3}	-1.100	0.27263
γ_4	-6.385×10^{-3}	4.232×10^{-3}	-1.509	0.132757	γ_4	-6.519×10^{-3}	4.081×10^{-3}	-1.5972	0.1116
α_5	1.568×10^{-2}	4.213×10^{-3}	3.722	0.000250	α_5	1.124×10^{-2}	4.018×10^{-3}	4.120	5.35×10^{-5}
γ_5	1.108×10^{-2}	4.232×10^{-3}	2.619	0.009421	γ_5	1.124×10^{-2}	4.100×10^{-3}	2.742	0.00661
γ_6	-5.447×10^{-3}	2.986×10^{-3}	-1.824	0.069432	γ_6	-5.313×10^{-3}	2.870×10^{-3}	-1.851	0.06551
α_7	-9.746×10^{-4}	4.213×10^{-3}	-0.231	0.817267	α_7	4.047×10^{-4}	4.029×10^{-3}	0.100	0.92007
γ_7	1.156×10^{-2}	4.232×10^{-3}	2.732	0.006811	γ_7	1.211×10^{-2}	4.048×10^{-3}	2.991	0.00310
$\sqrt{MSE} = 0.04615$ (escala log); AIC= 17.03724 BIC= 21.81683					$\sqrt{MSE} = 3.958$, AIC= 16.77638, BIC= 21.48278				

Tabla 3: Formulación de los tests de hipótesis pertinentes en los modelos globales

En la componente de tendencia		
Tests de hipótesis	Estadístico de prueba y su distribución	Criterio de decisión
$H_0: \beta_3 = 0 \text{ vs. } H_1: \beta_3 \neq 0$	$T_0 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} \sim t_{222}$	Rechazar H_0 si $P(t_{222} \geq T_0)$ es pequeño
En la componente estacional		
En las frecuencias $F_j = \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3}, \frac{5}{12}$ y 0.35		
Tests de hipótesis	Estadístico de prueba y su distribución	Criterio de decisión
$H_0: \alpha_j = 0 \text{ vs. } H_1: \alpha_j \neq 0$	$T_0 = \frac{\hat{\alpha}_j}{se(\hat{\alpha}_j)} \sim t_{222}$	Rechazar H_0 si $P(t_{222} \geq T_0)$ es pequeño
$H_0: \gamma_j = 0 \text{ vs. } H_1: \gamma_j \neq 0$	$T_0 = \frac{\hat{\gamma}_j}{se(\hat{\gamma}_j)} \sim t_{222}$	Rechazar H_0 si $P(t_{222} \geq T_0)$ es pequeño
En la frecuencia $F_j = \frac{1}{2}$		
$H_0: \gamma_6 = 0 \text{ vs. } H_1: \gamma_6 \neq 0$	$T_0 = \frac{\hat{\gamma}_6}{se(\hat{\gamma}_6)} \sim t_{222}$	Rechazar H_0 si $P(t_{222} \geq T_0)$ es pequeño

Tabla 4: Parámetros óptimos de suavizamiento y coeficientes estimados, Holt-Winters multiplicativo en $t = n = 239$

Parámetros de suavizamiento	Valor óptimo
α	0.30783757
β	0.01863519
γ	0.22697393

Coeficientes	Valor en $n = 239$
$a = \hat{\beta}_{0,239}$	108.9671617
$b = \hat{\beta}_{1,239}$	0.1875303
$s_1 = \hat{\delta}_{1,239}$	0.9398588
$s_2 = \hat{\delta}_{2,239}$	0.9433335
$s_3 = \hat{\delta}_{3,239}$	0.9951379
$s_4 = \hat{\delta}_{4,239}$	0.9719459
$s_5 = \hat{\delta}_{5,239}$	1.0348849
$s_6 = \hat{\delta}_{6,239}$	0.9808778
$s_7 = \hat{\delta}_{7,239}$	1.0244489
$s_8 = \hat{\delta}_{8,239}$	1.0627160
$s_9 = \hat{\delta}_{9,239}$	1.0506450
$s_{10} = \hat{\delta}_{10,239}$	1.1037679
$s_{11} = \hat{\delta}_{11,239}$	1.0413662
$s_{12} = \hat{\delta}_{12,239}$	0.9758230
AIC = 14.32972 BIC = 17.43498	

Tabla 5: Efectos estacionales estimados en el modelo 4 DLL(AICC)

i	$\hat{\delta}_i$
1	0.9287706
2	0.9307854
3	0.9852414
4	0.9624388
5	1.0157878
6	0.9750618
7	1.0182585
8	1.0344685
9	1.0420417
10	1.0962298
11	1.0402207
12	0.9706949
Suma	12.0000
AIC = 9.86281	
BIC = 13.98342	

Tabla 6: Ecuaciones ajustadas en modelos globales. Ecuaciones de suavizamiento Holt-Winters. Ajuste componente estacional por filtro de descomposición.

<p style="text-align: center;">Modelo 1</p> $\hat{Y}_t \approx \exp(4.274 + 9.452 \times 10^{-4}t + 9.583 \times 10^{-6}t^2 - 2.737 \times 10^{-8}t^3 - 5.178 \times 10^{-2} \sin\left(\frac{\pi t}{6}\right) - 1.018 \times 10^{-2} \cos\left(\frac{\pi t}{6}\right) - 2.922 \times 10^{-2} \sin\left(\frac{\pi t}{3}\right) - 1.555 \times 10^{-2} \cos\left(\frac{\pi t}{3}\right) - 9.523 \times 10^{-3} \sin\left(\frac{\pi t}{2}\right) - 4.906 \times 10^{-3} \cos\left(\frac{\pi t}{2}\right) - 2.881 \times 10^{-3} \sin\left(\frac{2\pi t}{3}\right) - 6.385 \times 10^{-3} \cos\left(\frac{2\pi t}{3}\right) + 1.568 \times 10^{-2} \sin\left(\frac{5\pi t}{6}\right) + 1.108 \times 10^{-2} \cos\left(\frac{5\pi t}{6}\right) - 5.447 \times 10^{-3} \cos(\pi t) - 9.746 \times 10^{-4} \sin\left(\frac{0.35\pi t}{6}\right) + 1.156 \times 10^{-2} \cos\left(\frac{0.35\pi t}{6}\right)) \times \exp\left(-\frac{(0.04615)^2}{2}\right)$
<p style="text-align: center;">Modelo 2</p> $\hat{Y}_t = \exp(4.260 + 1.669 \times 10^{-3}t + 2.348 \times 10^{-6}t^2 - 7.813 \times 10^{-9}t^3 - 5.224 \times 10^{-2} \sin\left(\frac{\pi t}{6}\right) - 1.030 \times 10^{-2} \cos\left(\frac{\pi t}{6}\right) - 2.864 \times 10^{-2} \sin\left(\frac{\pi t}{3}\right) - 1.629 \times 10^{-2} \cos\left(\frac{\pi t}{3}\right) - 9.045 \times 10^{-3} \sin\left(\frac{\pi t}{2}\right) - 4.355 \times 10^{-3} \cos\left(\frac{\pi t}{2}\right) - 4.445 \times 10^{-3} \sin\left(\frac{2\pi t}{3}\right) - 6.519 \times 10^{-3} \cos\left(\frac{2\pi t}{3}\right) + 1.124 \times 10^{-2} \sin\left(\frac{5\pi t}{6}\right) + 1.124 \times 10^{-2} \cos\left(\frac{5\pi t}{6}\right) - 5.313 \times 10^{-3} \cos(\pi t) + 4.047 \times 10^{-4} \sin\left(\frac{0.35\pi t}{6}\right) + 1.211 \times 10^{-2} \cos\left(\frac{0.35\pi t}{6}\right))$
<p style="text-align: center;">Modelo 3</p> $\begin{aligned} \hat{\beta}_{0,t} &= 0.30783757(Y_t / \hat{S}_{t-12}) + 0.69216243(\hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1}) \\ \hat{\beta}_{1,t} &= 0.01863519(\hat{\beta}_{0,t} - \hat{\beta}_{0,t-1}) + 0.98136481\hat{\beta}_{1,t-1} \\ \hat{S}_t &= 0.22697393(Y_t / \hat{\beta}_{0,t}) + 0.77302607\hat{S}_{t-12} \\ \hat{Y}_t &= (\hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1}) \times \hat{S}_{t-12} \text{ donde } \hat{S}_{t-12} = \sum_{i=1}^{12} \hat{\delta}_{i,t-12} I_{i,t} \end{aligned}$
<p style="text-align: center;">Modelo 4</p> $\begin{aligned} \hat{S}_t = \sum_{i=1}^{12} \hat{\delta}_i I_{i,t} &= 0.9287706 I_{1,t} + 0.9307854 I_{2,t} + 0.9852414 I_{3,t} + 0.9624388 I_{4,t} + 1.0157878 I_{5,t} + 0.9750618 I_{6,t} + 1.0182585 I_{7,t} \\ &+ 1.0344685 I_{8,t} + 1.0420417 I_{9,t} + 1.0962298 I_{10,t} + 1.0402207 I_{11,t} + 0.9706949 I_{12,t} \end{aligned}$


```

> modelo3 <- SuavizamientoEstacional(Yt, seasonal = "multiplicative", h = m)

Call:
HoltWinters(x = serie, alpha = alpha, beta = beta, gamma = gamma,
  seasonal = seasonal, optim.start = optim.start, optim.control = optim.control)

Smoothing parameters:
alpha 0.30783757
beta 0.01863519
gamma 0.22697393

Coefficients:
a 108.9671617
b 0.1875303
s1 0.9398588
s2 0.9433335
s3 0.9951379
s4 0.9719459
s5 1.0348849
s6 0.9808778
s7 1.0244489
s8 1.0627160
s9 1.0506450
s10 1.1037679
s11 1.0413662
s12 0.9758230

```

Figura 3. Print-Screen de la ejecución y salidas resumen del ajuste por SEHW multiplicativo, mediante la función de usuario SuavizamientoEstacional(.) del Modelo 3.

```

> modelo4 <- Descomp.Loess(Yt, h=m, tipo.descomp = "multiplicative", grado=1, criterio = "aicc")
Efectos estacionales estimados
s1 0.9287706
s2 0.9307854
s3 0.9852414
s4 0.9624388
s5 1.0157878
s6 0.9750618
s7 1.0182585
s8 1.0344685
s9 1.0420417
s10 1.0962298
s11 1.0402207
s12 0.9706949

Resumen Loess sobre serie desestacionalizada:
Call:
loess(formula = y ~ x, data = data.bind, span = span1, degree = degree,
  family = family)

Number of Observations: 239
Equivalent Number of Parameters: 12.54
Residual Standard Error: 2.955
Trace of smoother matrix: 14.89 (exact)

Control settings:
span : 0.1256679
degree : 1
family : gaussian
surface : interpolate cell = 0.2
normalize: TRUE
parametric: FALSE
drop.square: FALSE

```

Figura 4. Print-Screen de la ejecución y salidas resumen del ajuste del modelo combinando de forma multiplicativa el filtro de la descomposición con Loess, mediante la función de usuario Descomp.Loess(.) del Modelo 4

En la figura 4, se puede ver el valor del parámetro de suavizamiento Loess (α) que indica la proporción de los $n = 239$ datos que son usados en el ajuste local de la serie desestacionalizada del modelo 4, y el número de parámetros equivalentes loess, que indica cuántos parámetros serían necesarios en un modelo polinomial global para alcanzar la misma calidad del correspondiente ajuste loess.

Tabla 7. Parámetro de suavizamiento loess y número de parámetros equivalentes, según resultados en la Figura 4.

Modelo	Parámetro de suavizamiento loess	Número de parámetros equivalentes
Modelo 4	0.1256679	12.54

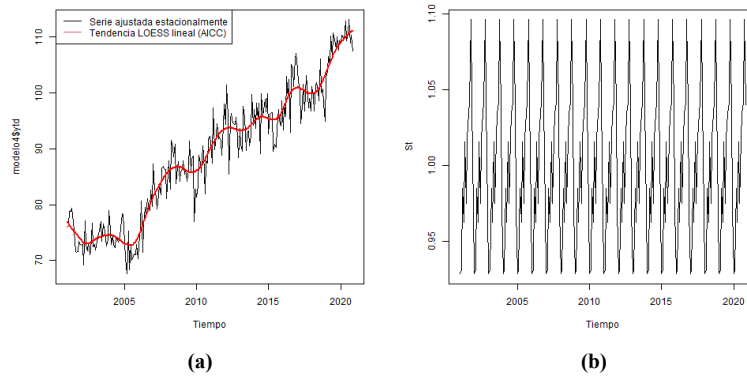


Figura 5. (a) Ajuste de la tendencia por regresión local Loess lineal usando criterio AICC; **(b)** Ajuste componente estacional con el filtro de descomposición multiplicativa.

De la figura 5(a), se observa que, aunque la tendencia se encuentra perturbada por los ciclos, la curva LOESS ajusta bien esta tendencia de la serie desestacionalizada, ya que sigue de manera similar tanto la tendencia como los patrones locales debidos a variaciones cíclicas. También es visible que el ajuste local final en esta misma figura, termina con pendiente negativa. Lo anterior implica que las extrapolaciones de la tendencia en la validación cruzada con el modelo 4 serán con decrecimiento. Por otro lado, de la Tabla 7, es claro que los ajustes locales por loess involucran una baja proporción de los datos ($\alpha = 0.1256679$) y que sería necesario un ajuste global polinomial con 13 parámetros (polinomio de grado 12) para lograr seguir de la misma manera las variaciones observadas sobre la serie desestacionalizada. Por lo que se considera que no es posible este ajuste global.

En la figura 5(b), se concluye que la forma del patrón estacional se observa en la gráfica de la serie, ya que año tras año este patrón inicia creciendo y decreciendo periódicamente, hasta alcanzar su pico máximo generalmente en el mes de octubre y finalmente decrece abruptamente. Este comportamiento no es constante en el tiempo, por el contrario, varía lentamente.

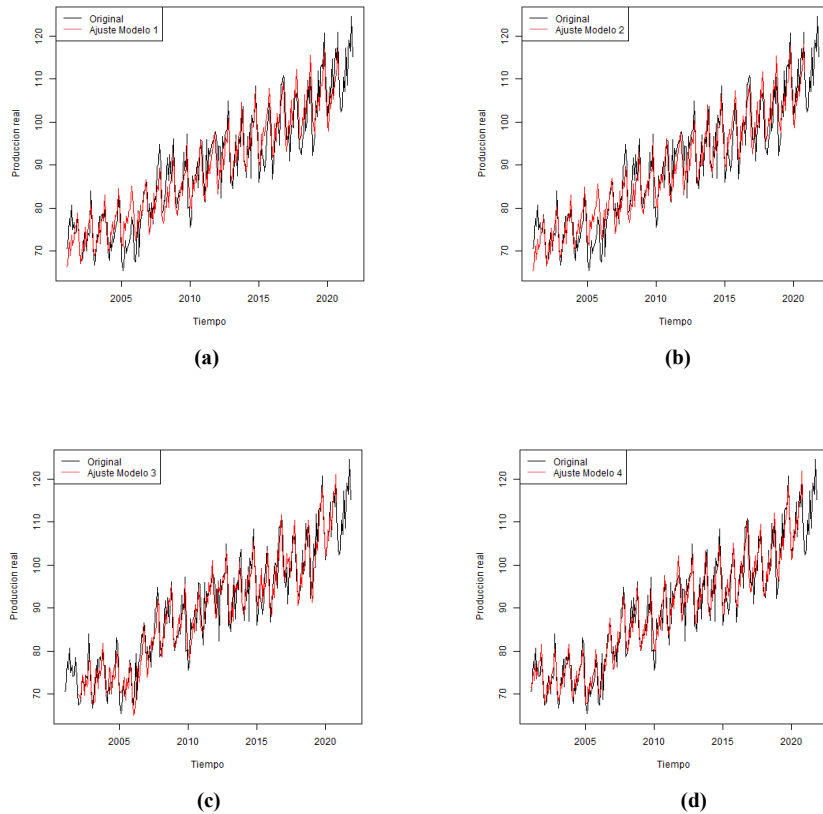


Figura 6. Ajustes de los primeros 239 datos de la serie: **(a)** En modelo 1; **(b)** En modelo 2; **(c)** En modelo 3; **(d)** En modelo 4.

De acuerdo a la figura 6, ajuste de los modelos, se observa que los modelos globales tienen problemas al ajustar los patrones cíclicos que se presentan más notoriamente desde el inicio de la serie y hasta el año 2010, lo que no ocurre con los modelos locales. A partir del año 2010 los modelos globales siguen un comportamiento similar al de los modelos locales, por lo que se puede decir que estos

últimos tienen un mejor ajuste que los modelos globales. Sin embargo entre los modelos locales no es posible decidir cuál ajusta mejor ya que no se notan diferencias significativas entre el ajuste de estos, por lo que se procede a analizar los CI para poder decidir cuál modelo tiene un mejor ajuste. De acuerdo a las tablas 2, 4 y 5, se observan las medidas de bondad de ajuste AIC y BIC, es visible que estas medidas son mejores en los 2 modelos locales como era de esperar. Finalmente, el modelo 4 obtiene los mejores resultados tanto en AIC y BIC con los valores $AIC = 9.86281$ y $BIC = 13.98342$.

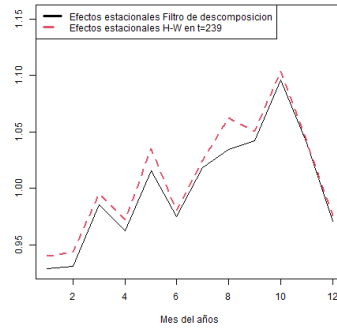


Figura 6. Comparación de las estimaciones de los efectos estacionales por el filtro de la descomposición multiplicativa (en color negro) y Holt-Winters al final del suavizamiento (en color rojo).

En la figura 6, se observa que la forma del patrón estacional que estima al final Holt-Winters y el que estima globalmente el filtro de la descomposición, refleja algunas diferencias para los meses de enero, febrero, marzo, abril, y mayo que no distan mucho, en las que Holt-Winters termina con estimaciones un poco mayores a las del filtro de la descomposición. Entre julio y septiembre se observa una diferencia en la que Holt-Winters termina con estimaciones aparentemente mucho mayores a las del filtro de la descomposición. Se tiene en cuenta que, por la escala de la gráfica esta disparidad se ve mucho más grande, pero realmente es muy pequeña. Se concluye entonces que aunque en los meses descritos anteriormente, los efectos estacionales que estima al final Holt-Winters se muestran por encima del filtro de la descomposición, hay indicios de que el patrón estacional es más o menos estable porque estas diferencias no son muy grandes.

4. Análisis de residuales y validación de supuestos.

En principio se quiere evaluar si no hay evidencia contra los siguientes supuestos sobre los errores de ajuste: media igual a cero, varianza constante, patrón aleatorio alrededor de la media, y que no hay carencia de ajuste.

En la figura 7, se muestran las gráficas de los residuos de ajuste vs. tiempo, mientras que en la figura 8 se exponen las gráficas de los residuos de ajuste vs. los valores ajustados, todas en la respectiva escala de ajuste usada para cada modelo.

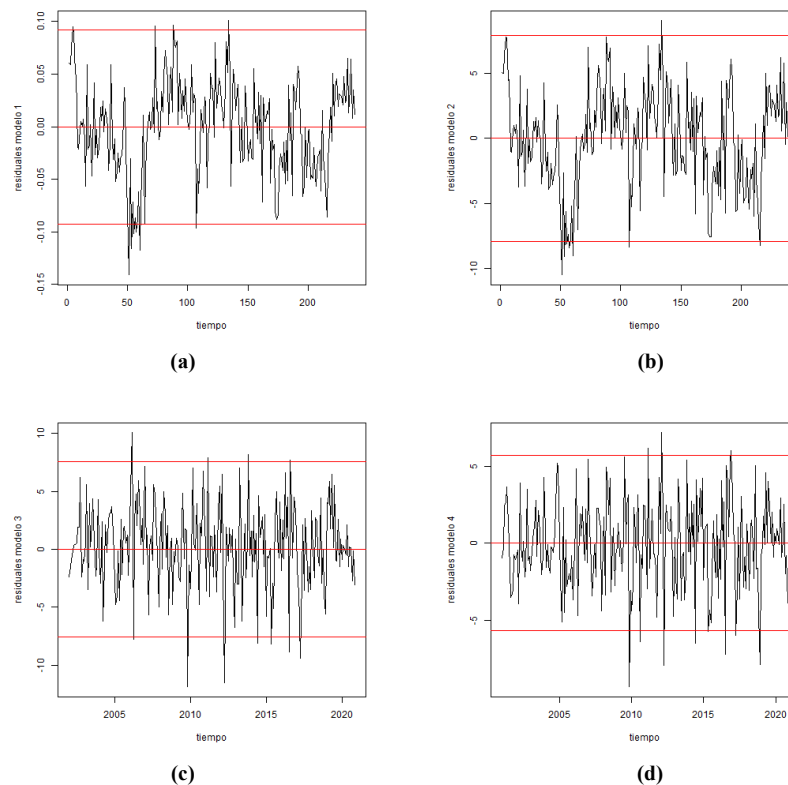


Figura 7. Residuos de ajuste vs. tiempo: (a) En modelo 1; (b) En modelo 2; (c) En modelo 3; (d) En modelo 4

En las gráficas 7(a) y 7(b) de los modelos ajustados globalmente, se puede observar la presencia de ciclos en la serie de tiempo que se mueven en torno a cero, esto indica que la media de la serie es cero. Dichos ciclos indican que existe correlación positiva entre los errores del modelo (E_t), separados al menos un periodo de tiempo, es decir, $corr(E_t, E_{t+1}) > 0$, por tanto, no se cumple el supuesto de la independencia en los errores del modelo. Se debe tener en cuenta que, en caso de no observar ciclos en los residuos, no es evidencia suficiente acerca de la validez del supuesto de independencia, aunque es una condición necesaria. Debido a que no se cumplió el supuesto de independencia, no es posible probar el supuesto de normalidad. Para verificar si existe o no carencia de ajuste, se observan las gráficas 8(a) y 8(b) y se determina que, en ambas gráficas la dispersión de los puntos mostrados no tiene patrones en forma de U o W, lo que indica que no hay carencia de ajuste.

En la figura 2(a), un polinomio de grado 3 que siga la tendencia presentada, no está obligado a seguir los ciclos que se ven reflejados en los residuales, ya que estos ciclos son perturbaciones y no hacen parte de dicha tendencia.

En las gráficas de 7(c) y 7(d) de los modelos ajustados localmente, no se observa componente cíclica de forma notoria ni rachas en signos \pm o \mp , aunque esto no es suficiente para la validación del supuesto de independencia en los errores pero si es necesario, los valores parecen estar bien centrados en torno a cero, lo que indica que la media de los errores en ambas gráficas es cero.

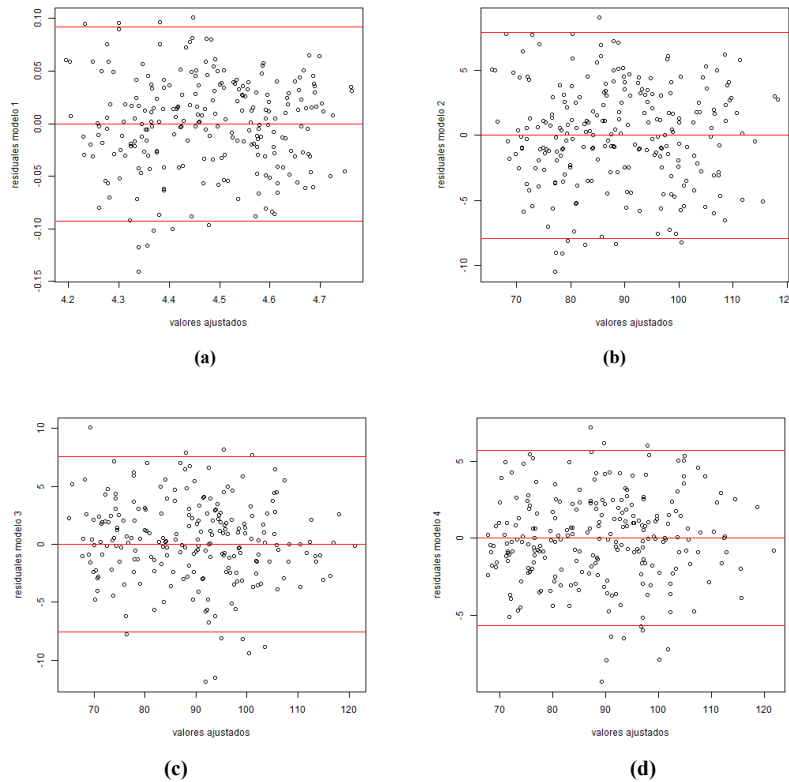


Figura 8. Residuos de ajuste vs. Valores ajustados: (a) Modelo 1; (b) Modelo 2; (c) Modelo 3; (d) Modelo 4

En las gráficas 8(a) y 8(b) de los modelos ajustados globalmente, como se dijo anteriormente, no se observa que exista carencia de ajuste porque no se presentan patrones de U o W que den indicios de dicha carencia. Como no existe carencia de ajuste, el supuesto de varianza constante se prueba observando la dispersión de los puntos alrededor de la media cero y no alrededor de la tendencia en los patrones de U o W. De esto, se tiene que la varianza es constante para estas gráficas.

Para las gráficas 8(c) y 8(d) de los modelos ajustados localmente, se observa el mismo comportamiento de las gráficas 8(a) y 8(b), no presentan carencia de ajuste y tienen varianza constante.

Es apropiado decir entonces, que los modelos globales no son la mejor opción, ya que al poseer componente cíclica, muestran evidencia que contradice el supuesto de independencia que por ende, no posibilita la validación del supuesto de normalidad, y comparados con los modelos locales que aunque no es posible probar la independencia de los errores y por lo tanto, no es posible probar la normalidad, no poseen evidencia que contradiga alguno de los supuestos, se concluye que los modelos locales (modelo 3 y modelo 4) son los más indicados en el análisis de residuales y validación de supuestos.

5. Pronósticos para la validación cruzada.

A continuación, en la tabla 8 se muestran las ecuaciones de los pronósticos de los 4 modelos que se postulan para esta serie con origen en $t = n = 239$, es decir, para los últimos $m = 12$ datos que corresponden al periodo comprendido entre diciembre de 2020 y

noviembre de 2021. La tabla 9, refleja los valores de los pronósticos puntuales y por intervalos de predicción del 95% de confianza para los 4 modelos planteados. En la tabla 10 se presentan los valores de las medidas de precisión de los pronósticos puntuales y de los I.P del 95%. Finalmente, en la figura 9, se ilustra el comportamiento de los pronósticos puntuales con respecto a los datos reales en los tiempos de los pronósticos ex – post.

Tabla 8. Ecuaciones de pronóstico

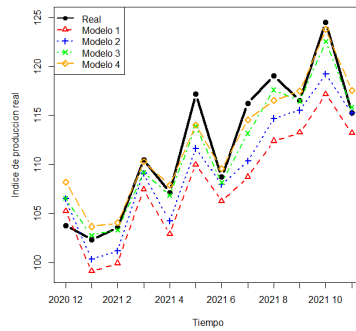
<p>Modelo 1</p> $\hat{Y}_{239}(L) \approx \exp(4.274 + 9.452 \times 10^{-4}(239 + L) + 9.583 \times 10^{-6}(239 + L)^2 - 2.737 \times 10^{-8}(239 + L)^3 - 5.178 \times 10^{-2} \sin\left(\frac{\pi(239+L)}{6}\right) - 1.018 \times 10^{-2} \cos\left(\frac{\pi(239+L)}{6}\right) - 2.922 \times 10^{-2} \sin\left(\frac{\pi(239+L)}{3}\right) - 1.555 \times 10^{-2} \cos\left(\frac{\pi(239+L)}{3}\right) - 9.523 \times 10^{-3} \sin\left(\frac{\pi(239+L)}{2}\right) - 4.906 \times 10^{-3} \cos\left(\frac{\pi(239+L)}{2}\right) - 2.881 \times 10^{-3} \sin\left(\frac{2\pi(239+L)}{3}\right) - 6.385 \times 10^{-3} \cos\left(\frac{2\pi(239+L)}{3}\right) + 1.568 \times 10^{-2} \sin\left(\frac{5\pi(239+L)}{6}\right) + 1.108 \times 10^{-2} \cos\left(\frac{5\pi(239+L)}{6}\right) - 5.447 \times 10^{-3} \cos(\pi(239 + L)) - 9.746 \times 10^{-4} \sin\left(\frac{0.35\pi(239+L)}{6}\right) + 1.156 \times 10^{-2} \cos\left(\frac{0.35\pi(239+L)}{6}\right)) \times \exp\left(-\frac{(0.04615)^2}{2}\right)$
<p>Modelo 2</p> $\hat{Y}_{239}(L) = \exp(4.260 + 1.669 \times 10^{-3}(239 + L) + 2.348 \times 10^{-6}(239 + L)^2 - 7.813 \times 10^{-9}(239 + L)^3 - 5.224 \times 10^{-2} \sin\left(\frac{\pi(239+L)}{6}\right) - 1.030 \times 10^{-2} \cos\left(\frac{\pi(239+L)}{6}\right) - 2.864 \times 10^{-2} \sin\left(\frac{\pi(239+L)}{3}\right) - 1.629 \times 10^{-2} \cos\left(\frac{\pi(239+L)}{3}\right) - 9.045 \times 10^{-3} \sin\left(\frac{\pi(239+L)}{2}\right) - 4.355 \times 10^{-3} \cos\left(\frac{\pi(239+L)}{2}\right) - 4.445 \times 10^{-3} \sin\left(\frac{2\pi(239+L)}{3}\right) - 6.519 \times 10^{-3} \cos\left(\frac{2\pi(239+L)}{3}\right) + 1.124 \times 10^{-2} \sin\left(\frac{5\pi(239+L)}{6}\right) + 1.124 \times 10^{-2} \cos\left(\frac{5\pi(239+L)}{6}\right) - 5.313 \times 10^{-3} \cos(\pi(239 + L)) + 4.047 \times 10^{-4} \sin\left(\frac{0.35\pi(239+L)}{6}\right) + 1.211 \times 10^{-2} \cos\left(\frac{0.35\pi(239+L)}{6}\right))$
<p>Modelo 3</p> $\hat{Y}_{239}(L) = (108.9671617 + 0.1875303 \times L) \times (0.9398588I_{1,239+L} + 0.9433335I_{2,239+L} + 0.9951379I_{3,239+L} + 0.9719459I_{4,239+L} + 1.0348849I_{5,239+L} + 0.9808778I_{6,239+L} + 1.0244489I_{7,239+L} + 1.0627160I_{8,239+L} + 1.0506450I_{9,239+L} + 1.1037679I_{10,239+L} + 1.0413662I_{11,239+L} + 0.9758230I_{12,239+L})$
<p>Modelo 4</p> $\hat{Y}_{239}(L) = T_{239}(L) \times S_{239}(L)$ $\hat{Y}_{239}(L) = (\beta_{0,239} + \beta_{1,239} \times (239 + L)) \times (0.9287706I_{1,239+L} + 0.9307854I_{2,239+L} + 0.9852414I_{3,239+L} + 0.9624388I_{4,239+L} + 1.0157878I_{5,239+L} + 0.9750618I_{6,239+L} + 1.0182585I_{7,239+L} + 1.0344685I_{8,239+L} + 1.0420417I_{9,239+L} + 1.0962298I_{10,239+L} + 1.0402207I_{11,239+L} + 0.9706949I_{12,239+L})$

Tabla 9. Pronósticos puntuales y por I.P del 95% de confianza

	Modelo 1			Modelo 2			Modelo 3			Modelo 4		
Período	Pronóstico	Lim. Inf	Lim. Sup	Pronóstico	Lim. Inf	Lim. Sup	Pronóstico	lim Inf	lim Sup	Pronóstico	Lim. Inf	Lim. Sup
Dec 2020	105.2205 ₁	95.52602	115.8988	106.5068	---	---	106.5157	103.2622 ₃	109.7691	108.1745	---	---
Jan 2021	99.10499	89.96553 ₉	109.1729	100.3354	---	---	102.7663	98.88667	106.6458	103.6598	---	---
Feb 2021	99.93003	90.70104	110.0981	101.1667	---	---	103.3231	98.83971	107.8065	104.0359	---	---
Mar 2021	107.4431 ₈	97.49784	118.4030	109.0732	---	---	109.1838	103.9953 ₆	114.3723	110.2764	---	---
Apr 2021	102.8844 ₅	93.34540	113.3983	104.2139	---	---	106.8215	101.2246 ₁	112.4184	107.8690	---	---
May 2021	109.9680 ₄	99.75370	121.2283	111.6332	---	---	113.9329	107.5767 ₀	120.2891	113.9960	---	---
Jun 2021	106.2516 ₀	96.35734	117.1618	107.9330	---	---	108.1711	101.6404 ₉	114.7017	109.5627	---	---
Jul 2021	108.7277 ₇	98.58447	119.9147	110.3344	---	---	113.1682	105.9540 ₄	120.3824	114.5553	---	---
Aug 2021	112.3949 ₃	101.8860 ₈	123.9877	114.6893	---	---	117.5948	109.7247 ₂	125.4648	116.5160	---	---
Sep 2021	113.2472 ₇	102.6291 ₃	124.9640	115.5311	---	---	116.4561	108.2573 ₇	124.6548	117.5033	---	---
Oct 2021	117.1605 ₇	106.1530 ₇	129.3095	119.2350	---	---	122.5513	113.5767	131.5260	123.7514	---	---
Nov 2021	113.1811 ₇	102.5187 ₉	124.9525	115.2655	---	---	115.8182	95.94189	135.6944	117.5561	---	---

Tabla 10. Precisión de los Pronósticos puntuales y de los I.P del 95%

Medidas	Modelo 1	Modelo 2	Modelo 3	Modelo 4
RMSE	4.833304	3.432064	1.729023	2.013989
MAE	4.343043	2.857168	1.322271	1.614678
MAPE (%)	3.812008	2.513249	1.163709	1.445068
Amplitud. Media I.P	21.13093	NA	14.57040	NA
Cobertura (%) I.P	100%	NA	100%	NA

**Figura 9.** Comparación de los pronósticos puntuales vs. valores reales de los 4 modelos.

Los datos mostrados en la tabla 9, indican el valor que cada modelo ajustado con los primeros $n = 239$ datos, predice que será observado para la serie en cada periodo de los pronósticos ex-post. En esta misma tabla, para los modelos 1 y 3, se muestran los intervalos de predicción del 95% de confianza, que predicen el rango de valores en el cual puede presentarse el valor real de la serie. Por ejemplo, para el modelo 1 en junio de 2021, se pronostica que el índice de producción real de la elaboración de productos de panadería, macarrones, fideos y sus productos, en ese mes será de 106.26160 puntos porcentuales del índice, mientras que para los modelos 2, 3 y 4, pronostican que el índice de producción será 107.9330, 108.1711, 109.5627 puntos porcentuales del índice, respectivamente. Adicionalmente, para los modelos 1 y 3, se tienen intervalos de predicción del 95% de confianza que pronostican el rango de valores que podrá tomar el valor de la serie, por ejemplo, para el mes de febrero de 2021, el modelo 1 predice que, el índice de producción real de la elaboración de productos de panadería, macarrones, fideos y sus productos, en ese mes estará entre 90.70104 y 110.0981 puntos porcentuales del índice; mientras el modelo 3 pronostica que este índice, para el mismo mes, estará entre 98.83971 y 107.8065 puntos porcentuales del índice.

Se concluye que, para el modelo 1, en el mes de febrero de 2021, el valor del pronóstico puntual (99.93003) se encuentra entre los límites de los intervalos de predicción de este mismo mes. Así mismo, para el modelo 3, en el mes de febrero de 2021, el valor del pronóstico puntual (103.3231) está entre los límites de los intervalos de predicción correspondientes a este mes. Esto se cumple para todas las fechas de la tabla 9 en estos modelos. Vemos diferencias importantes tanto en los pronósticos puntuales como por intervalos, por lo que es necesario calcular medidas promedios.

Siguiendo con la tabla 10, se analiza la precisión de los pronósticos puntuales y de los intervalos de predicción de la tabla 9, la cual expone el valor de las medidas del RMSE, MAE, y MAPE de los 4 modelos, y la amplitud y cobertura de los modelos 1 y 3. Se analizarán estas precisiones para los modelos locales, es decir, los modelos 3 y 4, ya que los modelos globales (modelos 1 y 2) presentan las peores medidas de RMSE, MAE y MAPE, además, no cumplen con los supuestos sobre los errores de los modelos. Se puede observar que el valor de RMSE en puntos porcentuales del índice, para los modelos 3 y 4 es 1.729023 y 2.013989, respectivamente; también se observa que el valor del MAE en puntos porcentuales del índice, para estos mismos modelos es 1.322271 y 1.614678, respectivamente; por otro lado, el valor del MAPE en términos porcentuales en relación con cada valor real observado, para estos modelos es 1.163709 y 1.445068, respectivamente. Siguiendo con el valor de la amplitud media del intervalo en puntos porcentuales del índice, para el modelo 3 es 14.57040, y finalmente, el valor de la cobertura en términos porcentuales para el modelo 3 es 100%.

Con lo anterior, se puede concluir que, con base en la medida de RMSE, el modelo 3 es el mejor, ya que estima que en promedio, en cada uno de los 12 periodos del horizonte de pronósticos, se equivocará en ± 1.729023 puntos porcentuales del índice; mientras que el modelo 4, estima que en promedio, en cada uno de los 12 periodos del horizonte de pronósticos, se equivocará en ± 2.013989 puntos porcentuales del índice.

Con base en la medida de MAE, el modelo 3 sigue siendo el mejor, ya que en promedio estima que, en cada uno de los 12 periodos del horizonte de pronósticos, se equivocará en ± 1.322271 puntos porcentuales del índice; mientras que el modelo 4, estima que en promedio, en cada uno de los 12 periodos del horizonte de pronósticos, se equivocará en ± 1.614678 puntos porcentuales del índice.

De acuerdo con la medida de MAPE, el modelo 3 siguen apuntando a ser el mejor, ya que en cada uno de los 12 periodos del horizonte de pronósticos, estima que se equivocará en 1.163709% con relación al valor real, mientras que el modelo 4 estima que se equivocará en 1.445068% con relación al valor real.

En resumen, según el valor de las medidas analizadas anteriormente, el mejor modelo es el modelo 3 ya que tiene los menores valores en cada una de estas medidas.

Por último, en la figura 9 se observa que, al inicio todos los modelos sobre estiman el valor real, pero rápidamente estos modelos pasan a sub estimar los siguientes valores reales a lo largo de todo el pronóstico. Sin embargo, en promedio los pronósticos del modelo 3 están más próximos a los valores reales. Por el análisis anterior, se reafirma que el modelo 3 sigue siendo el mejor.

6. Conclusiones.

Teniendo en cuenta el análisis de residuales representado en las figuras 7 y 8 como primer criterio de selección para elegir el mejor modelo propuesto, se tiene que, los modelos globales (1 y 2) se descartan como candidatos a mejores modelos porque no cumplen con los supuestos del error: media cero, varianza constante e independencia de los errores. Por otro lado, los modelos locales (3 y 4) son candidatos a mejores modelos, debido a que no se presenta evidencia en contra del cumplimiento de los supuestos del error, además de que no se nota la presencia de ciclos que indican carencia de ajuste.

Continuando con los resultados de pronósticos expuestos en la tabla 10 y figura 9 como segundo criterio de selección, y considerando que los modelos globales (1 y 2) fueron descartados anteriormente, se concluye que el mejor modelo es el 3, debido a que posee los menores valores que estiman el error que en promedio se puede estar cometiendo en cada pronóstico realizado en el horizonte de $m = 12$ periodos.

Adicionalmente, con los resultados obtenidos en los criterios de información (AIC y BIC) usados en el ajuste como tercer y último criterio de selección, se determina que el mejor modelo es el 4, ya que tiene los menores valores en AIC y BIC, seguido por el modelo 3.

Finalmente, sabiendo que, en el análisis de residuales los modelos 3 y 4 están empatados, se tiene en cuenta que, los resultados de pronóstico tienen mayor prioridad sobre los resultados de ajuste a la hora de elegir el mejor modelo, y se concluye que el modelo más indicado que logró ajustar en gran medida a la serie, tanto en la componente de tendencia como en la componente estacional, entre los 4 modelos propuestos es el modelo 3, que corresponde al modelo de tendencia local de cambio de nivel y de pendientes que evolucionan lentamente con t , multiplicativa a factor estacional con efectos estacionales que evolucionan lentamente en el tiempo, con método de ajuste Suavizamiento Exponencial Holt -Winters multiplicativo.

Referencias

- [1] DANE, «Clasificación Industrial Internacional Uniforme de todas las actividades económicas,» DANE, 2020.
- [2] DANE, «Encuesta Mensual Manufacturera con Enfoque Territorial - EMMET,» DANE, Bogotá, 2021.
- [3] S. de la Fuente Fernández, «Números Indices,» [En línea]. Available: <https://www.estadistica.net/PAU2/numeros-indices.pdf>.
- [4] Instituto Nacional de Estadística, «INE,» [En línea]. Available: <https://www.ine.es/ipc01/enlaces.htm>.
- [5] Y. Montero Castellanos, «Economipedia - Diferencia entre nominal y real,» 2012. [En línea]. Available: <https://economipedia.com/definiciones/nominal-y-real.html>.