TRABAJO 1
ESTADÍSTICA 3

Grupo 2

Jennifer Salazar Galvis
Juan Esteban Sánchez Pulgarín
Miguel Ángel Londoño Ciceros
Cristina Mercedes Ortega Benavides

CONTENIDO

CONTENIDO

INTRODUCCIÓN

Según el DANE la industria manufacturera es aquella que "abarca la transformación física o química de materiales, sustancias o componentes en productos nuevos. Está dividida en 21 grandes grupos que se pueden estudiar gracias a la EMMET, una investigación estadística que busca explicar la estructura y evolución del sector manufacturero en Colombia por medio de indicadores como: el IPIM (Índice de producción manufacturero) o el PIB (Producto Interno Bruto) del sector industrial. Estos valores son calculados con ayuda de variables de producción, ventas, empleo, sueldos y horas trabajadas.

Cambios en la EMMET

1

Debido a la transformación del sector en el que se desarrolla su cambio ha sido constante. Desde 1962 el DANE ha diseñado varias encuestas cuyos objetivos son estimar la evolución del sector industrial. Sin embargo, al observar que la naturaleza misma de la información, no se permitía utilizarla para formular planes a corto plazo, se consideró realizar una nueva encuesta donde se encuentren reportes mensuales por parte de los establecimientos industriales.

2

A finales de 1999, debido a los cambios que estaba presentando el sector tanto a nivel nacional como internacional, se conformó un marco muestral a partir de la información de Encuesta Anual Manufacturera (EAM) de 1997. Para el año 2014, después de evaluaciones y con el fin de actualizar la metodología de la investigación y contar con un directorio de establecimientos actualizado y completo, se rediseña nuevamente la Muestra Mensual Manufacturera – MMM, que en adelante se denominará Encuesta Mensual Manufacturera – EMM.

3

Finalmente, en el continuo proceso de mejoramiento del DANE y buscando que los diferentes usuarios contarán con la desagregación de información de la EMM pero adicionalmente, presentar información por departamentos, áreas metropolitanas y ciudades con periodicidad mensual creó la EMMET, cuya población objetivo son los establecimientos industriales manufactureros en el país que ocupan diez o más personas donde se analiza tanto la producción y ventas nominales, que muestran el valor de los productos elaborados por el establecimiento y su venta sin incluir los impuestos; como los valores reales de los mismos donde se tiene en cuenta la deflactación por el índice de precios al productor según clase industrial.

En este caso se estudia la serie mensual de índices de producción nominal de otros productos químicos; estas actividades abarcan la fabricación de una amplia gama de productos, como: plaguicidas, pinturas, tintas, jabones y preparados para limpiar, perfumes y preparados de tocador; explosivos y productos pirotécnicos, preparados químicos para usos fotográficos; gelatinas; preparados compuestos para diagnóstico, etc.

La 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.).

Por medio del número índice, que es una medida estadística con la cual se recoge la evolución relativa en el periodo t de una magnitud económica (precios, producciones, etc) de un conjunto de bienes o productos respecto de un periodo base o de referencia, es decir, es el cambio porcentual en el crecimiento o decrecimiento de una actividad con respecto a un periodo de referencia.

El número índice es una expresión numérica que acumula las variaciones porcentuales observadas.

Fórmulas

Valor de la producción total ponderada en el mes de referencia

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

Valor de la producción total ponderada para el mes t del año 2018

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

Índice de Producción Industrial Manufacturera en el mes de referencia:

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

Con el fin de ofrecer una serie de datos que permita a los usuarios contar con información histórica del sector manufacturero, el DANE enlazó las series estadísticas por medio de las variaciones anuales para los dominios comparables y para el total nacional con las anteriores operaciones MMM (Muestra Mensual Manufacturera) y EMM (Encuesta Mensual Manufacturera).

El enlace legal consiste en aplicar las tasas de variación anuales calculadas con la base 2001 y 2014, al nivel de la serie determinado por el nuevo año base (2018), garantizando las variaciones anuales de las anteriores muestras.

CONTENIDO

ANÁLISIS DESCRIPTIVO

Se consideran los datos sobre el índice de producción nominal de la clase industrial otros productos químicos sector de manufactura, de enero de 2001 a diciembre de 2019 (N=228), año base 2018. Ver Figura 1(a).

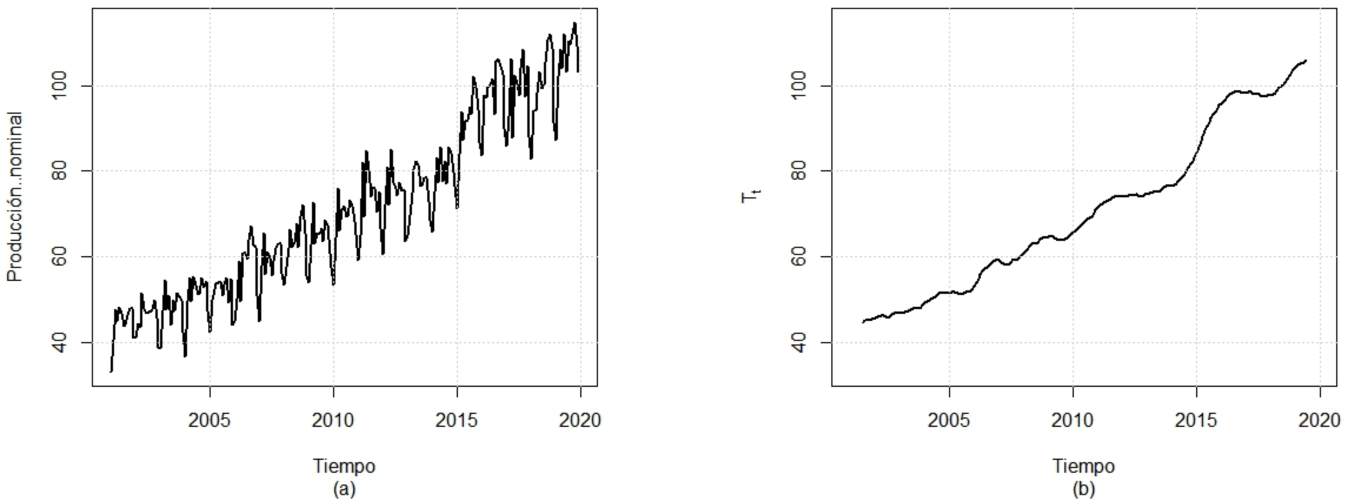
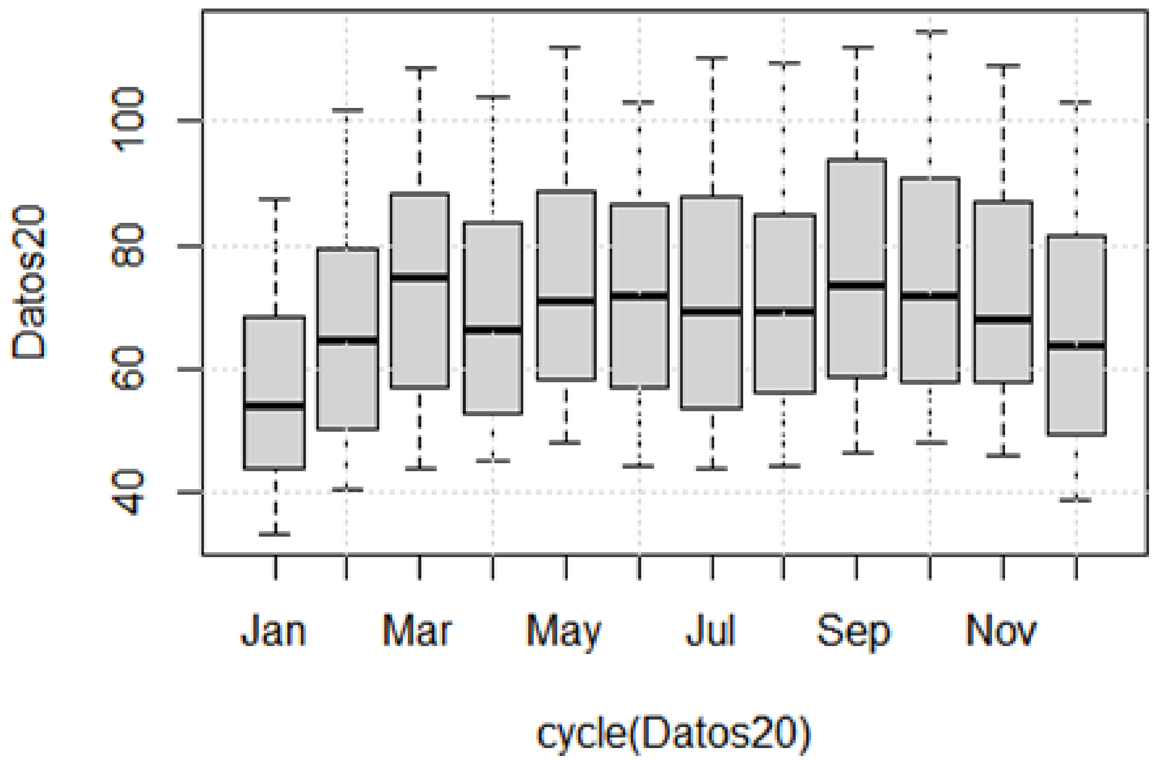


Figura 1. (a) Índice de producción nominal del sector manufactura(Colombia), Clase industrial: otros productos químicos. Enero 2001 - diciembre 2019; (b) Tendencia según el filtro de descomposición aditiva.

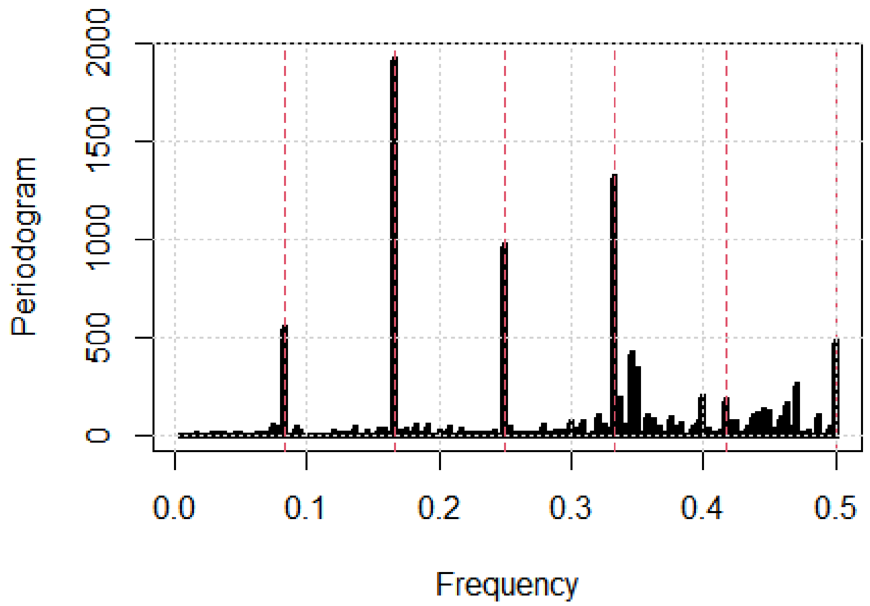
De la Figura 1(a) se deduce la presencia de estacionalidad y tendencia aditivas, es aditiva dado que su varianza es constante alrededor de su trayectoria de largo plazo(tendencia). En la Figura 1(b) es visible que la serie es de tendencia creciente y que existen ciclos.

Gráfico de boxplots comparativos de la distribución de la serie versus periodos del año calendario



en los boxplots comparativos se ve cómo la media (o mediana) de la distribución de la serie según el periodo del año, cambia a lo largo de un año calendario siendo menor en los meses de enero, febrero, abril y diciembre comparados con el resto del año. Tiene un crecimiento de enero a marzo y luego vuelve y baja en abril para volver a subir en mayo manteniéndose relativamente al mismo nivel en los siguientes meses hasta noviembre, para luego decaer en diciembre y retomar el siguiente año; se confirma que existe un patrón periódico anual con una forma constante en el tiempo.

Periodograma



El periodograma muestra la asociación de la serie con cinco componentes periódicas (valor alto del periodograma en esas frecuencias): en las frecuencias 1/12, 2/12, 3/12, 4/12, 6/12. El periodograma reafirma la existencia de componente estacional.

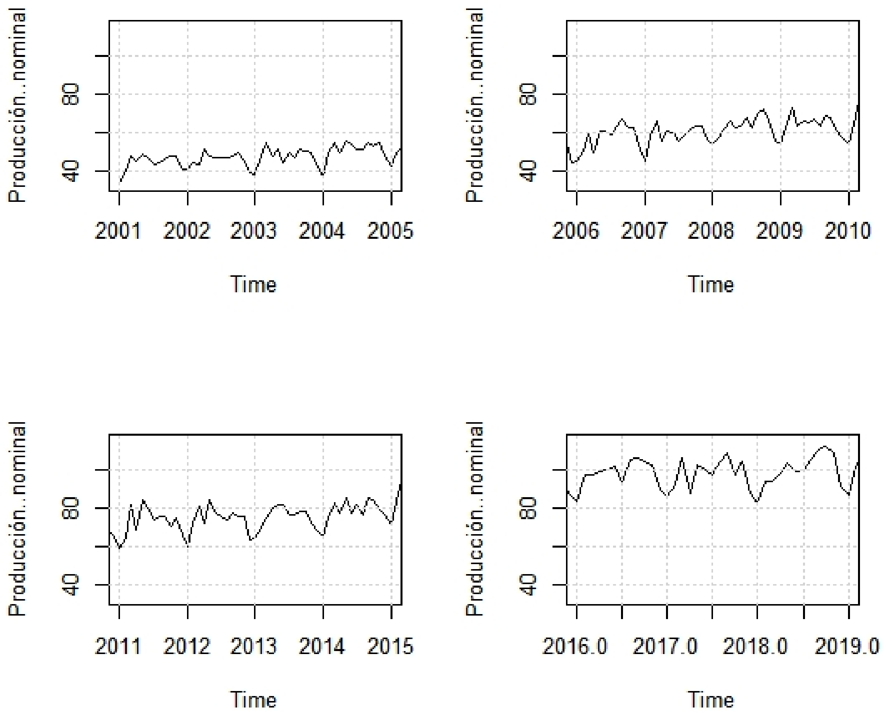
Por qué existe componente estacional y si su forma es constante o no en el tiempo

- Mediante el gráfico de la serie vs el tiempo se ve un patrón que se repite cada año; pues en cada año se observa que en los meses del inicio y del final el índice de producción nominal es bajo y tiene aproximadamente 2 picos de decaimiento en el transcurso del año.

- Dado que en el gráfico de boxplot al menos una de las medias de los meses cambia respecto al resto de los meses del año calendario, lo cual tiene incidencia sobre el valor medio de la serie; por eso, se dice que existe la componente estacional.

- Mediante el periodograma se resalta que hay una asociación de la serie con fenómenos periódicos, indicando la existencia de una componente estacional.

Es la estacionalidad constante o no en el tiempo:



Las gráficas anteriores muestran la serie dividida en periodos de 5 años, para observar si la estacionalidad cambia a través del tiempo, se puede notar que en todos los tramos tiene una forma similar, que realmente lo que cambia es el nivel, por lo tanto se considera la componente estacional como constante en el tiempo.

La tendencia se puede ajustar globalmente o si es local.

- La tendencia de la serie es global ya que se puede ajustar una curva suave creciente en todos los tiempos de la serie, También se puede ajustar modelos locales para comparar los ajustes y pronósticos respecto a la ajuste global.

Identificación de posibles ciclos y cambios estructurales.

- Hay presencia de ciclos en la serie a través del tiempo ya que al aplicar la descomposición aditiva y obtener la tendencia, no se observa una curva suave que describa la tendencia de la serie.

- Hay aproximadamente entre 4 y 5 ciclos en la serie a través del tiempo (montañas).

- Parece haber un posible cambio estructural en la serie en el 2015 ya que el nivel de la serie cambia un poco a partir de ese año.

CONTENIDO

MODELOS PROPUESTOS

Por el análisis descriptivo; para, los modelos globales la tendencia puede ser ajustada por un polinomio de grado $p = 2,3$ y la componente estacional puede representarse como un factor y por tanto usar variables indicadoras. Se proponen modelos de regresión, donde t es el índice de tiempo y $I_{i,t}$ es la indicadora del mes i en el tiempo t. Por otro lado, asumiendo que las componentes estructurales, es decir la tendencia y la estacionalidad de la serie, pueden cambiar en el tiempo, se proponen modelos locales.

Tabla 1. Ecuación de modelos propuestos

Modelo 1: Modelo cuadrático estacional con indicadoras, mes de referencia diciembre $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{i=1}^{11} \delta_i I_{i,t} + E_t, \quad E_t \sim iidN(0, \sigma^2)$
Modelo 2: Modelo cúbico estacional con indicadoras, mes de referencia diciembre $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \sum_{i=1}^{11} \delta_i I_{i,t} + E_t, \quad E_t \sim iidN(0, \sigma^2)$
Modelos 3: Descomposición aditiva & Loess cuadrático En la vecindad de un tiempo t_k donde se quiere ajustar $Y_t = \beta_{0,k} + \beta_{1,k} t + \beta_{2,k} t^2 + \sum_{i=1}^{12} \delta_i I_{i,t} + E_t, \ E_t \sim iidN(0, \sigma^2)$ para todo t vecino a t_k, con $\sum_{i=1}^{12} \delta_i = 0, \quad \beta_{0,k}, \ \beta_{1,k} \ y \ \beta_{2,k}$ los parámetros de la parábola local en la vecindad de t_k
Modelo 4: Suavizamiento exponencial Holt-Winters aditivo $Y_{t+h} = \beta_{0,t} + \beta_{1,t} \times h + \sum_{i=1}^{12} \delta_{i,t} I_{i,t+h} + E_{t+h}, \ E_t \sim iidN(0, \sigma^2)$, con $\sum_{i=1}^{12} \delta_{i,t} = 0, \ \beta_{0,t} \ y \ \beta_{1,t} \ y \ \delta_{i,t}$, el nivel en t, la pendiente en t y los efectos estacionales en t, respectivamente, cambiando lentamente en el tiempo.

CONTENIDO

AJUSTE DE LOS MODELOS

Se visualiza como se ajustaron los modelos mediante el software R y sus respectivas salidas, posteriormente se comentan cada uno de sus resultados.

Ajuste modelo 1.

```
> mod1 <- lm(yt~t+I(t^2)+I1+I2+I3+I4+I5+I6+I7+I8+I9+I10+I11)
> summary(mod1)

Call:
lm(formula = yt ~ t + I(t^2) + I1 + I2 + I3 + I4 + I5 + I6 +
    I7 + I8 + I9 + I10 + I11)

Residuals:
     Min       1Q   Median       3Q      Max
-11.2796  -2.4057  -0.1672   2.2769  11.0489

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.582e+01  1.225e+00  29.243  < 2e-16 ***
t            1.320e-01  1.745e-02   7.564 1.35e-12 ***
I(t^2)       7.051e-04  7.787e-05   9.055  < 2e-16 ***
I1          -3.182e+00  1.327e+00  -2.397   0.0174 *
I2           5.307e+00  1.327e+00   3.999 8.93e-05 ***
I3           1.202e+01  1.327e+00   9.059  < 2e-16 ***
I4           7.269e+00  1.327e+00   5.478 1.27e-07 ***
I5           1.211e+01  1.327e+00   9.130  < 2e-16 ***
I6           9.826e+00  1.327e+00   7.406 3.47e-12 ***
I7           8.429e+00  1.327e+00   6.354 1.37e-09 ***
I8           9.093e+00  1.327e+00   6.854 8.50e-11 ***
I9           1.252e+01  1.327e+00   9.435  < 2e-16 ***
I10          1.055e+01  1.327e+00   7.953 1.28e-13 ***
I11          8.948e+00  1.326e+00   6.745 1.57e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.979 on 202 degrees of freedom
Multiple R-squared:  0.9589,    Adjusted R-squared:  0.9562
F-statistic: 362.3 on 13 and 202 DF,  p-value: < 2.2e-16
```

Ajuste modelo 2.

```
> mod2 <- lm(yt~t+I(t^2)+I(t^3)+I1+I2+I3+I4+I5+I6+I7+I8+I9+I10+I11)
> summary(mod2)

Call:
lm(formula = yt ~ t + I(t^2) + I(t^3) + I1 + I2 + I3 + I4 + I5 +
    I6 + I7 + I8 + I9 + I10 + I11)

Residuals:
     Min       1Q   Median       3Q      Max
-11.6505  -2.3126  -0.1508   2.1501  11.1269

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.511e+01  1.452e+00  24.184  < 2e-16 ***
t            1.690e-01  4.405e-02   3.836 0.000168 ***
I(t^2)       2.800e-04  4.712e-04   0.594 0.552936
I(t^3)       1.306e-06  1.428e-06   0.915 0.361432
I1          -3.115e+00  1.330e+00  -2.342 0.020160 *
I2           5.368e+00  1.329e+00   4.038 7.67e-05 ***
I3           1.208e+01  1.329e+00   9.088  < 2e-16 ***
I4           7.318e+00  1.329e+00   5.508 1.10e-07 ***
I5           1.216e+01  1.328e+00   9.153  < 2e-16 ***
I6           9.862e+00  1.328e+00   7.427 3.10e-12 ***
I7           8.460e+00  1.328e+00   6.372 1.25e-09 ***
I8           9.118e+00  1.327e+00   6.869 7.93e-11 ***
I9           1.253e+01  1.327e+00   9.444  < 2e-16 ***
I10          1.056e+01  1.327e+00   7.958 1.25e-13 ***
I11          8.954e+00  1.327e+00   6.747 1.57e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.981 on 201 degrees of freedom
Multiple R-squared:  0.959,     Adjusted R-squared:  0.9562
F-statistic: 336.2 on 14 and 201 DF,  p-value: < 2.2e-16
```

Ajuste modelo 3

```
> ytd=yt-St
> mod3=loess.as(t,ytd,degree=2,criterion="aicc",family="gaussian",plot=F)
> summary(mod3)
Call:
loess(formula = y ~ x, data = data.bind, span = span1, degree = degree,
    family = family)

Number of Observations: 216
Equivalent Number of Parameters: 13.21
Residual Standard Error: 3.11
Trace of smoother matrix: 14.61  (exact)

Control settings:
  span     :  0.2244775
  degree   :  2
  family   :  gaussian
  surface  :  interpolate          cell = 0.2
  normalize:  TRUE
 parametric:  FALSE
drop.square:  FALSE
```

En los 4 modelos se implementa la estrategia de validación cruzada excluyendo del ajuste los últimos m = 12 datos, que comprende de enero de 2001 a diciembre de 2018 por tanto, la validación cruzada se hará con los pronósticos ex-post de estos últimos.

Ajuste modelo 4

```
> mod4=HoltWinters(yt,seasonal="additive") #Suavizamiento con valores Optimos en parametros
> mod4
Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:
HoltWinters(x = yt, seasonal = "additive")

Smoothing parameters:
 alpha: 0.2279994
 beta : 0.010042
 gamma: 0.397323

Coefficients:
           [,1]
a   101.9644720
b     0.2949128
s1  -12.3938476
s2   -2.0117762
s3    3.6989236
s4    0.1017724
s5    6.8427850
s6    4.2348284
s7    2.7412904
s8    6.6698440
s9   10.1574426
s10   6.3504048
s11   4.5463245
s12  -9.2424032
>
```

Tabla. Parámetros ajustados en los modelos globales

Parametro	*Estimación*	*Error Estándar*	T_0	$P(\lvert T_{202}\rvert > \lvert T_0\rvert)$
β_0	35.8242551	1.2250562	29.242949	0.0000000
β_1	0.1319807	0.0174478	7.564309	0.0000000
β_2	0.0007051	0.0000779	9.055350	0.0000000
δ_1	−3.1817818	1.3273503	−2.397093	0.0174368
δ_2	5.3069465	1.3272009	3.998601	0.0000893
δ_3	12.0220423	1.3270660	9.059114	0.0000000
δ_4	7.2690612	1.3269454	5.478041	0.0000001
δ_5	12.1146698	1.3268389	9.130475	0.0000000
δ_6	9.8255350	1.3267465	7.405736	0.0000000
δ_7	8.4294343	1.3266681	6.353838	0.0000000
δ_8	9.0930346	1.3266037	6.854372	0.0000000
δ_9	12.5163357	1.3265533	9.435230	0.0000000
δ_{10}	10.5493377	1.3265171	7.952659	0.0000000
δ_{11}	8.9475962	1.3264952	6.745291	0.0000000

Parámetro	*Estimación*	*Error Estándar*	T_0	$P(\lvert T_{201}\rvert > \lvert T_0\rvert)$
β_0	35.1122117	1.4518673	24.1841735	0.0000000
β_1	0.1689800	0.0440540	3.8357445	0.0001675
β_2	0.0002800	0.0004712	0.5943651	0.5529364
β_3	0.0000013	0.0000014	0.9147258	0.3614316
δ_1	−3.1146344	1.3299145	−2.3419809	0.0201603
δ_2	5.3679308	1.3294103	4.0378283	0.0000767
δ_3	12.0768988	1.3289564	9.0875055	0.0000000
δ_4	7.3178173	1.3285516	5.5081165	0.0000001
δ_5	12.1573451	1.3281951	9.1532824	0.0000000
δ_6	9.8621412	1.3278862	7.4269474	0.0000000
δ_7	8.4599754	1.3276245	6.3722651	0.0000000
δ_8	9.1175066	1.3274098	6.8686450	0.0000000
δ_9	12.5347270	1.3272421	9.4441906	0.0000000
δ_10	10.5616286	1.3271216	7.9582979	0.0000000
δ_11	8.9537593	1.3270487	6.7471217	0.0000000

Modelo 1

Debido a que el valor ajustado para el coeficiente correspondiente al grado 2 del polinomio tiene un p-valor muy pequeño(menor a 0.05), se dice que es significativo y por lo tanto la estructura del polinomio cuadrático es significativa. Por otro lado, ya que al menos uno de los coeficientes estimados asociados a las variables indicadoras, son significativos(p-valor menor a 0.05), en este caso todos, se concluye que la componente estacional es significativa.

Modelo 2

Debido a que el valor ajustado para el coeficiente correspondiente al grado 3 del polinomio tiene un p-valor grande(mayor a 0.05), se dice que no es significativo y por lo tanto la estructura del polinomio cúbico no es significativa. Por otro lado, ya que al menos uno de los coeficientes estimados asociados a las variables indicadoras, son significativos(p-valor menor a 0.05), en este caso todos, se concluye que la componente estacional es significativa.

Tests sobre significancia de parámetros importantes en modelos globales

Prueba de hipótesis	Estadístico de la prueba
$H_0 : \beta_p = 0$ vs. $H_1 : \beta_p \neq 0$	$T_0 = \frac{\widehat{\beta}_p}{\text{s.e}(\widehat{\beta}_p)} \sim t_\nu$
Tests individuales para δ_i, $i = 1, 2, \ldots, 12$ $H_0 : \delta_i = 0$ vs. $H_1 : \delta_i \neq 0$	$T_0 = \frac{\widehat{\delta}_i}{\text{s.e}(\widehat{\delta}_i)} \sim t_\nu$
En todos estos tests se rechaza H_0 si $P(\lvert t_\nu \rvert > \lvert T_0 \rvert)$ es pequeño, donde $\nu = 216 - k$, con k el total de parámetros en cada modelo. p es el grado del polinomio.	

Conclusión:

Desarrollando los tests de significancia de los parámetros, con base en resultados, se concluye que el modelo 1 global es significativo el respectivo polinomio propuesto (aunque esto no implica que el modelo es correcto), en cambio el modelo 2 no es significativo respecto al polinomio propuesto. Para la componente estacional, a un nivel de significancia de 0.05, en ambos modelos se concluye que todos los δ_i son estadísticamente significativos, incluso siendo solo uno de ellos significativo es suficiente para afirmar que la componente estacional modelada globalmente como un factor, es estadísticamente significativa.

Tabla. Parámetros óptimos de suavizamiento y coeficientes estimados, Holt Winters aditivos en t = n = 216

Parámetros de suavizamiento:	
$alpha : \alpha$	0.2279994
$beta : \beta$	0.010042
$gamma : \gamma$	0.397323
Coeficientes:	
$a : \hat{\beta}_{0,216}$	101.9644720
$b : \hat{\beta}_{1,216}$	0.2949128
$s1 : \hat{\delta}_{1,216}$	−12.3938476
$s2 : \hat{\delta}_{2,216}$	−2.0117762
$s3 : \hat{\delta}_{3,216}$	3.6989236
$s4 : \hat{\delta}_{4,216}$	0.1017724
$s5 : \hat{\delta}_{5,216}$	6.8427850
$s6 : \hat{\delta}_{6,216}$	4.2348284
$s7 : \hat{\delta}_{7,216}$	2.7412904
$s8 : \hat{\delta}_{8,216}$	6.6698440
$s9 : \hat{\delta}_{9,216}$	10.1574426
$s10 : \hat{\delta}_{10,216}$	6.3504048
$s11 : \hat{\delta}_{11,216}$	4.5463245
$s12 : \hat{\delta}_{12,216}$	−9.2424032

Tabla. Efectos estacionales estimados en el modelo DLC

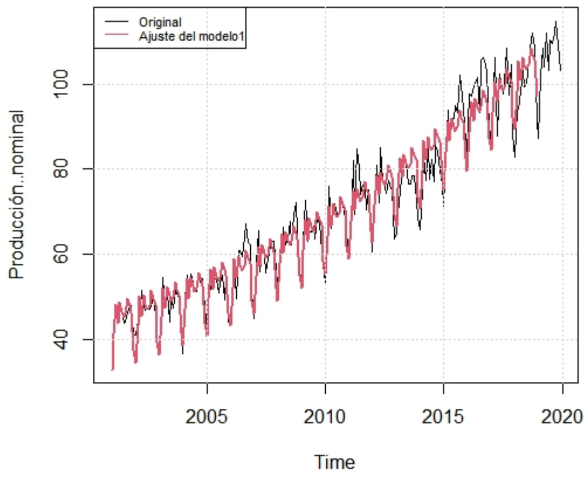
i	$\hat{\delta}_i$
1	−10.8780433
2	−2.3143178
3	4.3650940
4	−0.5319649
5	4.3979371
6	2.0783292
7	0.7969567
8	1.2373979
9	4.6148489
10	2.5229371
11	1.0633783
12	−7.3525531
$suma$	0

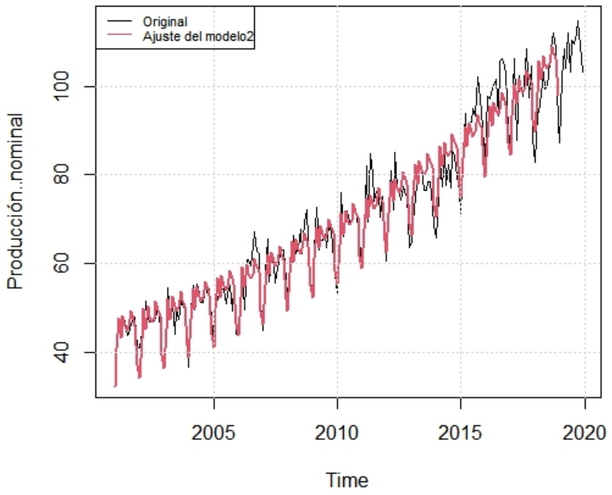
A partir de los modelos ajustados salen las siguientes ecuaciones de ajustes:

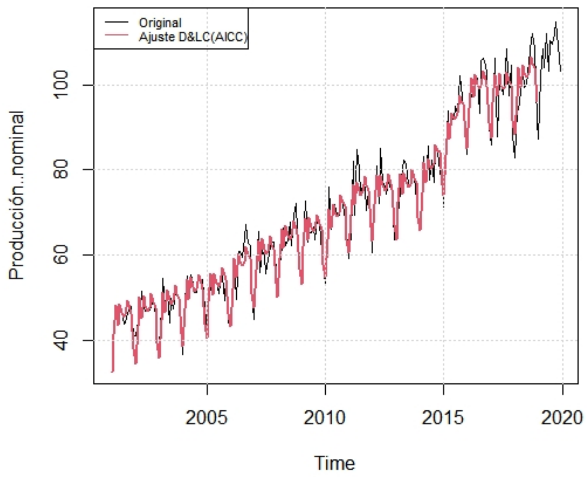
Tabla 2. Ecuaciones de los modelos ajustados

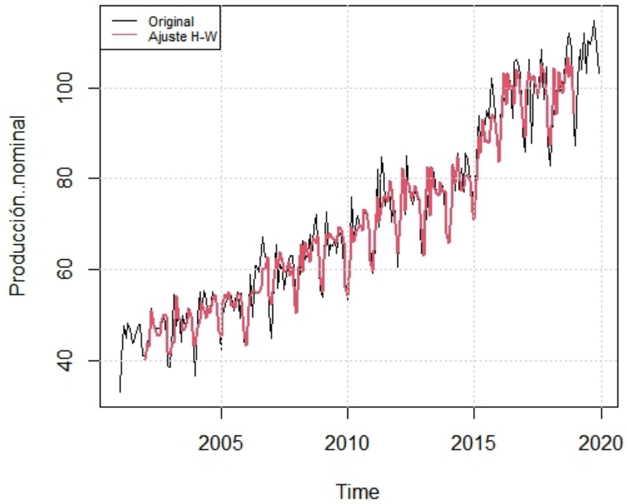
Modelo	*Ecuación*
1	$\hat{Y}_t \approx 35.8242551 + 0.1319807t + 0.0007051t^2 - 3.1817818I_{1,t} + 5.3069465I_{2,t} + 12.0220423I_{3,t} + 7.2690612I_{4,t} +$ $12.1146698I_{5,t} + 9.8255350I_{6,t} + 8.4294343I_{7,t} + 9.0930346I_{8,t} + 12.5163357I_{9,t} + 10.5493377I_{10,t} + 8.9475962I_{11,t}$
2	$\hat{Y}_t \approx 35.1122117 + 0.1689800t + 0.0002800t^2 + 0.0000013t^3 - 3.1146344I_{1,t} + 5.3679308I_{2,t} + 12.0768988I_{3,t} +$ $7.3178173I_{4,t} + 12.1573451I_{5,t} + 9.8621412I_{6,t} + 8.4599754I_{7,t} + 9.1175066I_{8,t} + 12.5347270I_{9,t} + 10.5616286I_{10,t} + 8.9537593I_{11,t}$
3	$\hat{S}_t = \sum_{i=1}^{11} \hat{\delta}_i I_{i,t} = -10.8780433I_{1,t} - 2.3143178I_{2,t} + 4.3650940I_{3,t} - 0.5319649I_{4,t} + 4.3979371$ $I_{5,t} + 2.0783292I_{6,t} + 0.7969567I_{7,t} + 1.2373979I_{8,t} + 4.6148489I_{9,t} + 2.5229371I_{10,t} + 1.0633783I_{11,t} - 7.3525531I_{12,t}$
4	$\beta_{0,t} = 0.2279994(Y_t - \hat{S}_{t-12}) + 0.7720006(\hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1})$ $\hat{\beta}_{1,t} = 0.010042(\hat{\beta}_{0,t} - \hat{\beta}_{0,t-1}) + 0.989958\hat{\beta}_{1,t-1}$ $\hat{S}_t = 0.397323(Y_t - \hat{\beta}_{0,t}) + 0.602677\hat{S}_{t-12}$ $\hat{Y}_t = (\beta_{0,t-1} + \hat{\beta}_{1,t-1}) + \hat{S}_{t-12}$

Gráficos de ajuste de cada uno de los modelos:









$Modelo$	p	AIC	BIC
1	14	16.85948	20.98234
2	15	17.01631	21.51105
3	24	11.18417	16.27338
4	13	16.87146	20.84412

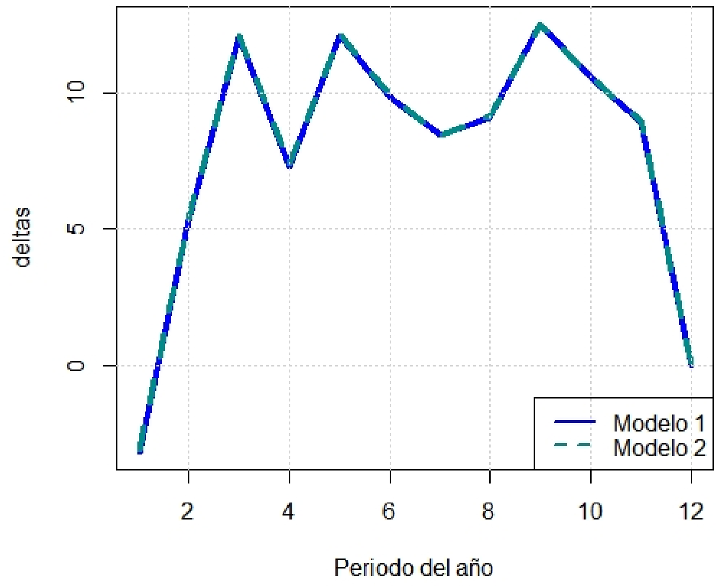
Con relación a la calidad de ajuste, vemos un ajuste muy similar entre los modelos, todos consiguen seguir muy bien los patrones de tendencia, estacionalidad pero los modelos locales siguen mejor los ciclos. Por otro lado, los valores de AIC y BIC también son similares aunque numéricamente resulta menor el del modelo 3, tanto AIC y BIC coinciden en el mismo modelo. El modelo de regresión cúbico con indicadoras tiene el mayor AIC y BIC (aunque como ya se indicó, el grado del polinomio no es significativo).

Entre los modelos globales se recomienda el modelo cuadrático con indicadoras ya que con un grado menos del polinomio ajusta igual de bien que el cúbico, por tanto se prefiere por ser un modelo más simple (parsimonioso) además que en los ajustes se noto que el grado 3 del polinomio del segundo modelo no es significativo.

Ya que la serie es aditiva, los δ_i es la diferencia entre la media de la serie en el mes i del año menos la media en el periodo de referencia, es decir, diciembre.

$\delta\ i$	*Modelo* 1	*Modelo* 2
δ_1	-3.181782	-3.114634
δ_2	5.306946	5.367931
δ_3	12.022042	12.076899
δ_4	7.269061	7.317817
δ_5	12.114670	12.157345
δ_6	9.825535	9.862141
δ_7	8.429434	8.459975
δ_8	9.093035	9.117507
δ_9	12.516336	12.534727
δ_{10}	10.549338	10.561629
δ_{11}	8.947596	8.953759

Componente estacional en los modelos globales:



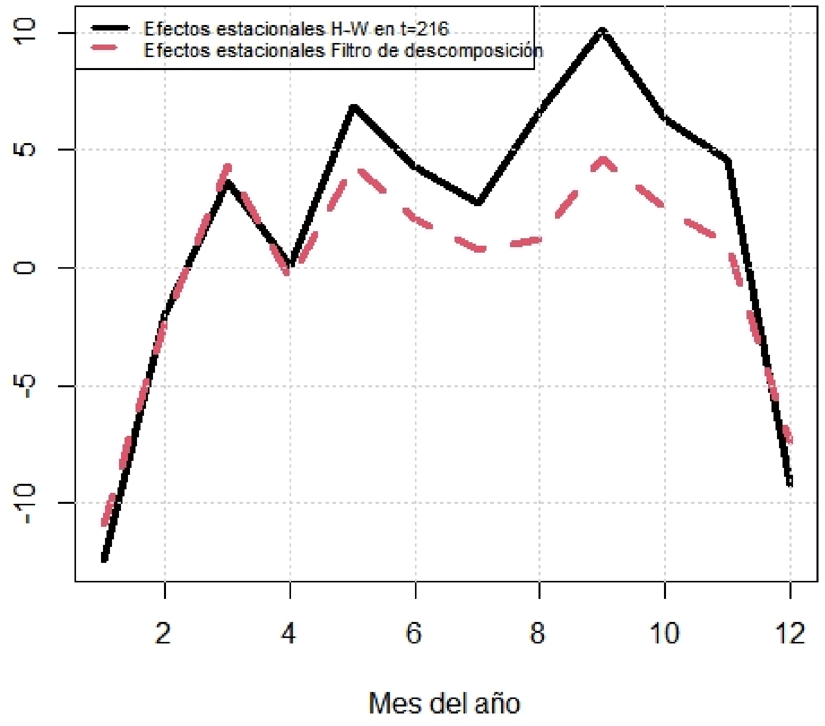
Vemos que el modelo 1 y el modelo 2 estiman la misma forma para S_t además de que sus valores son muy similares para los efectos estacionales (no son iguales pero tienen diferencias muy pequeñas en las estimaciones de los δ_i)

- Mediante el gráfico se puede notar que las estimación del δ_1 qué es el del mes de enero es la única con valor negativo, es decir es el único mes donde la producción nominal tuvo una disminución respecto al mes de diciembre que es el de referencia, por lo tanto para el resto de meses de febrero a noviembre tuvo un aumento en comparación con el mes de diciembre (Este es un comportamiento dentro de cada año).

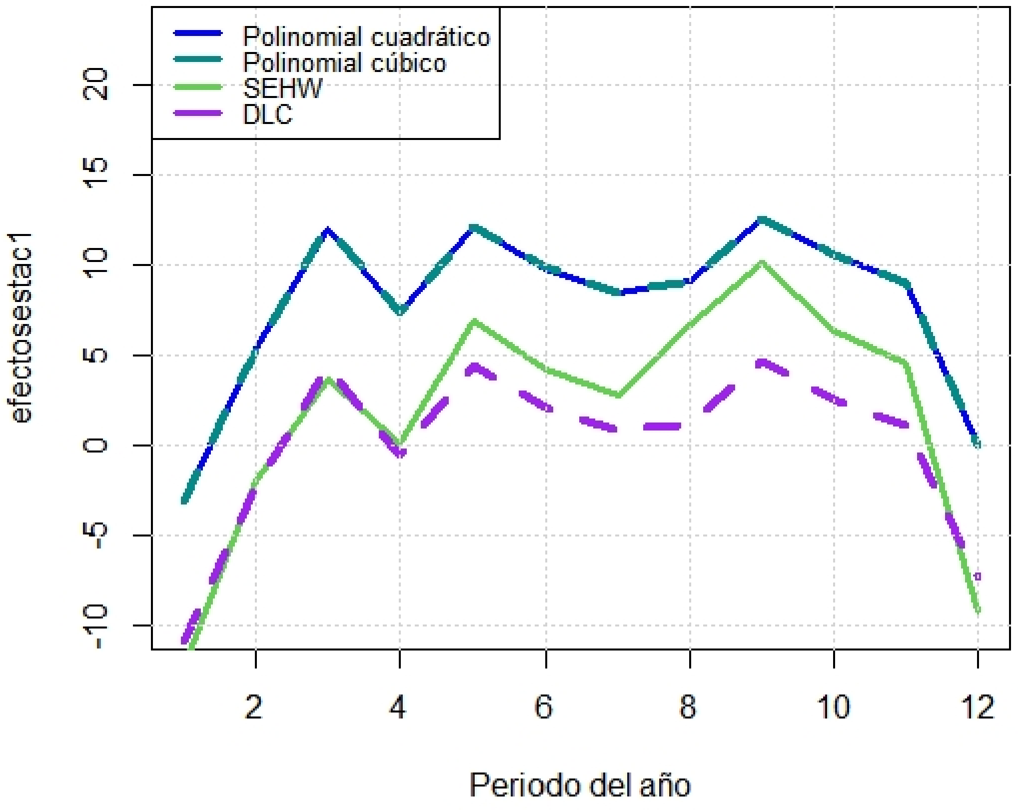
- Se puede visualizar mediante la gráfica que la estimación del δ
- para el mes 12 es decir de diciembre toma el valor de cero, debido a que es el mes de referencia.

Interpretación del δ_1 :

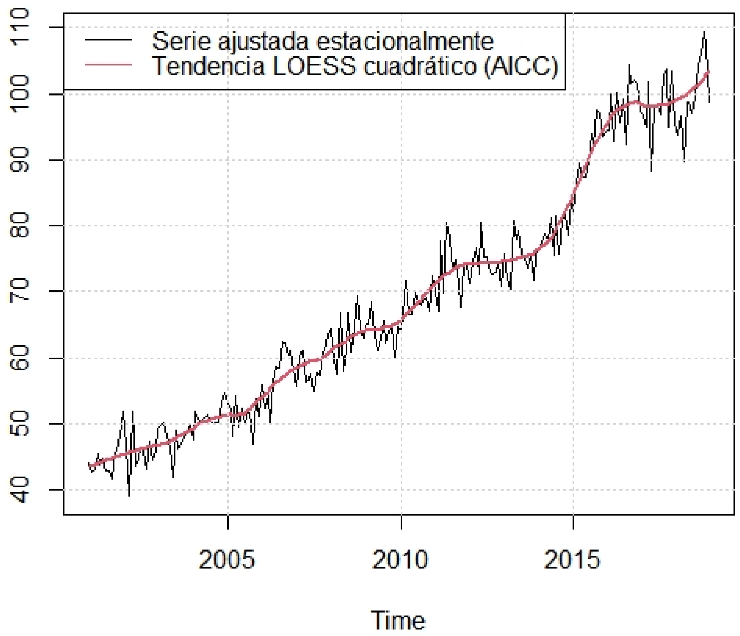
* El modelo 1 estima que en promedio del índice de producción nominal en el mes de enero disminuyó en 3.18178 unidades en comparación con el del mes de diciembre en cada año, en cambio para el modelo 2 la disminución fue de 3.1146.



Vemos que los modelos globales respecto a los locales estiman la misma forma de S_t pero con valores diferentes de los efectos estacionales al no coincidir los δ_i



En esta figura vemos que la forma del patrón estacional que estima al final Holt-Winters y el que estima globalmente el filtro de la descomposición, es muy similar aunque se observan algunas diferencias que destacan para los meses de mayo a noviembre, periodos del año en los que Holt-Winters termina con estimaciones mayores a las del filtro. Podemos decir que dado lo anterior, la forma del patrón estacional es más o menos estable y que en esta componente las proyecciones en los pronósticos ex–post del filtro y de Holt-Winters, no distan mucho.




```
> ytd=yt-St
> mod3=loess.as(t,ytd,degree=2,criterion="aicc",family="gaussian",plot=F)
> summary(mod3)
Call:
loess(formula = y ~ x, data = data.bind, span = span1, degree = degree,
    family = family)

Number of Observations: 216
Equivalent Number of Parameters: 13.21
Residual Standard Error: 3.11
Trace of smoother matrix: 14.61  (exact)

Control settings:
  span     :  0.2244775
  degree   :  2
  family   :  gaussian
  surface  :  interpolate         cell = 0.2
  normalize:  TRUE
 parametric:  FALSE
drop.square:  FALSE
```

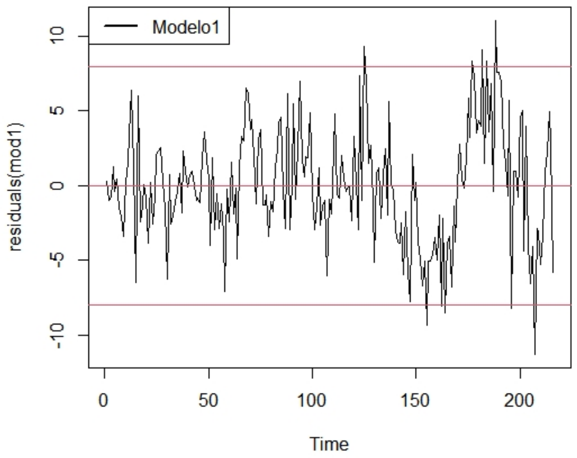
- Se observa que los valores ajustados se acercan a los valores que toma la serie; la componente cíclica que toma la serie está presente ya que se puede observar que la tendencia no es una curva completamente suave.

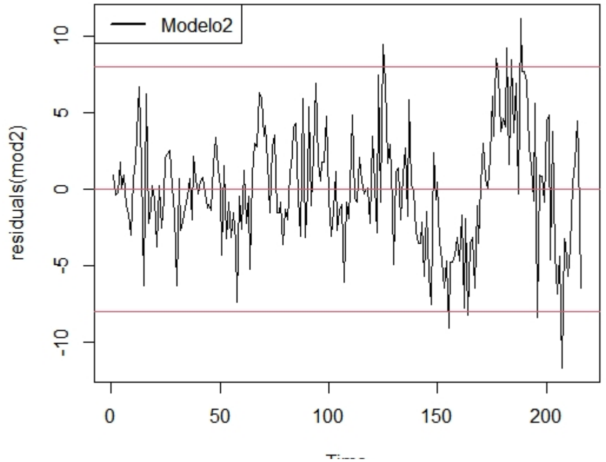
- Por otra parte, el ajuste que hace LOESS al final de la serie es creciente, aunque el valor real muestra una tendencia decreciente.

- Para un ajuste global polinomial alcanzando el mismo ajuste que LOESS, el modelo global debería ser un polinomial con aproximadamente 13 parámetros.

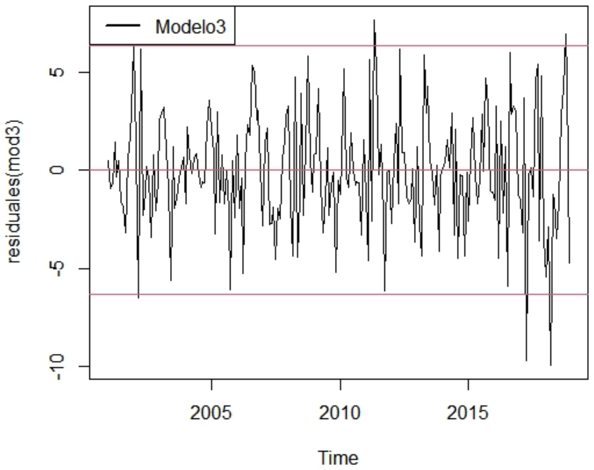
CONTENIDO

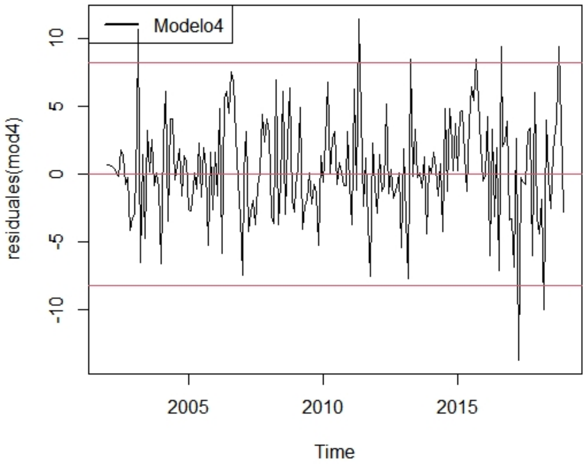
ANÁLISIS DE RESIDUALES





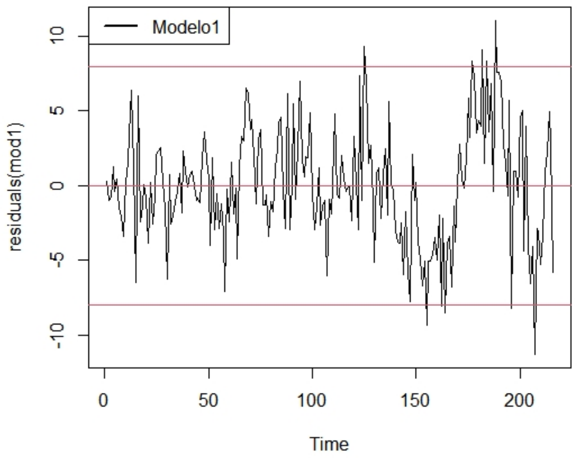
- En las figuras presentadas en las gráficas de residuos vs. tiempo y vs. valores ajustados, no se observa evidencia contra el supuesto de que los errores tienen media cero (a pesar de la variación cíclica alrededor de cero, que se observa en las series de tiempo de los residuos de los modelos globales), pues en todos los casos los residuos parecen bien centrados en cero.

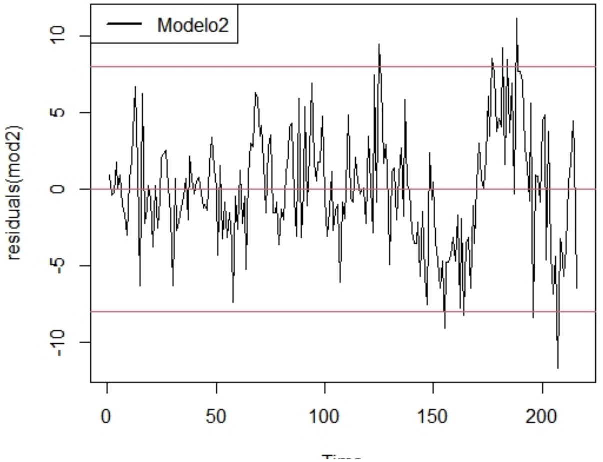


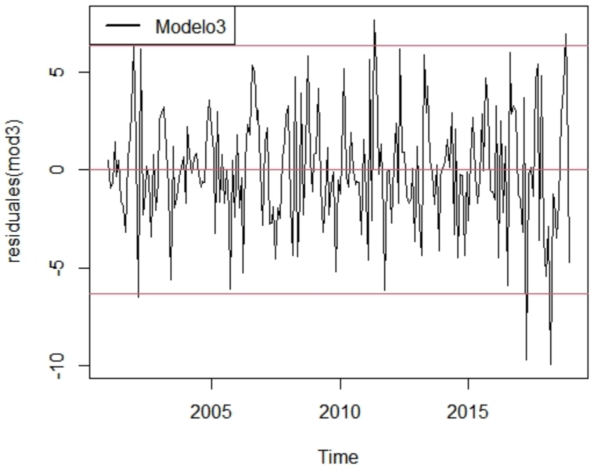


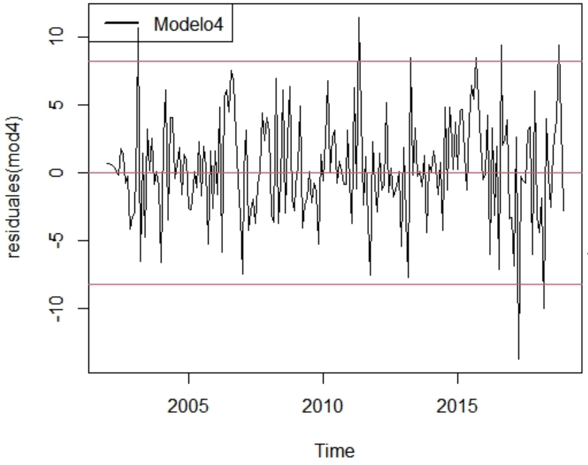
- Se observa una distribución relativamente homogénea de los residuos alrededor de cero, es decir, no hay evidencias contrariando el supuesto de varianza constante para los errores de ajuste.

ANÁLISIS DE RESIDUALES







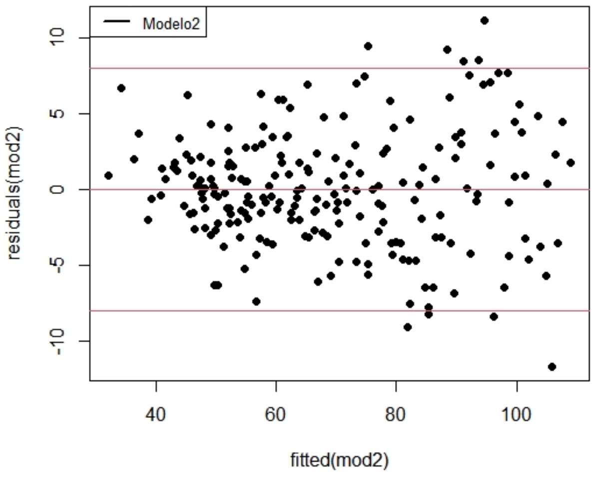


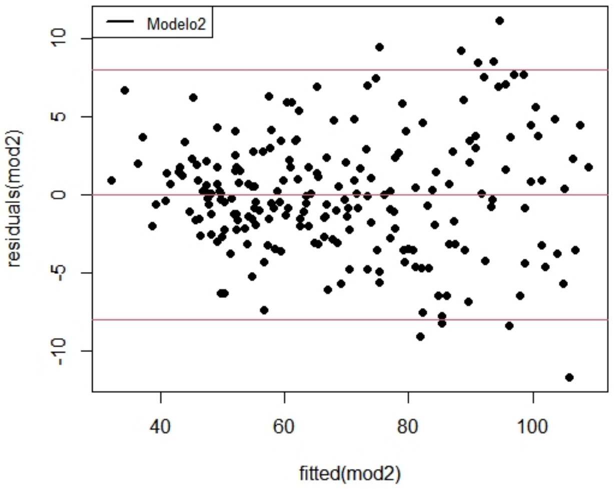
- En las series de tiempo de los residuos de los modelos globales hay evidencia de ciclos no explicados (los cuales se dan centrados en cero), esto implica que los errores en los modelos globales, separados un periodo en el tiempo, están positivamente correlacionados, es decir, $Corr(E_t, E_{t+1}) > 0$, por lo cual ya no es válido el supuesto de independencia.

  Por el contrario, en las series de tiempo de los residuos de los dos modelos locales no son observables a simple vista patrones cíclicos, aunque esto no es una garantía suficiente para afirmar la validez de la independencia entre los errores de ajuste de estos dos modelos, es un punto su favor.

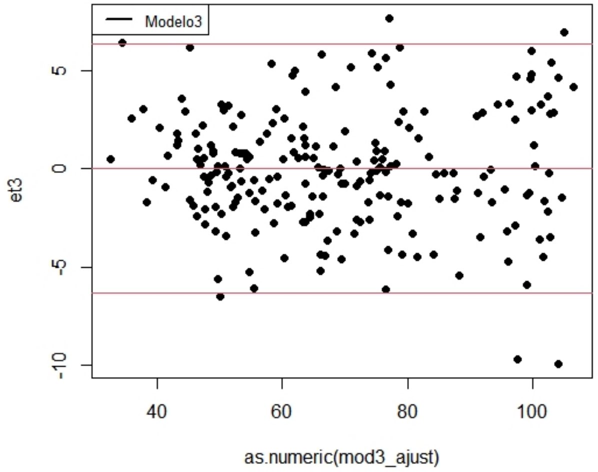
- Los métodos locales logran seguir ciclos, por eso son mejores que los modelos globales.

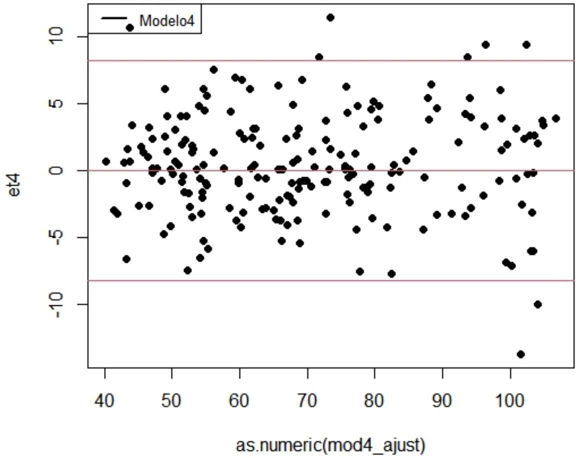
ANÁLISIS DE RESIDUALES





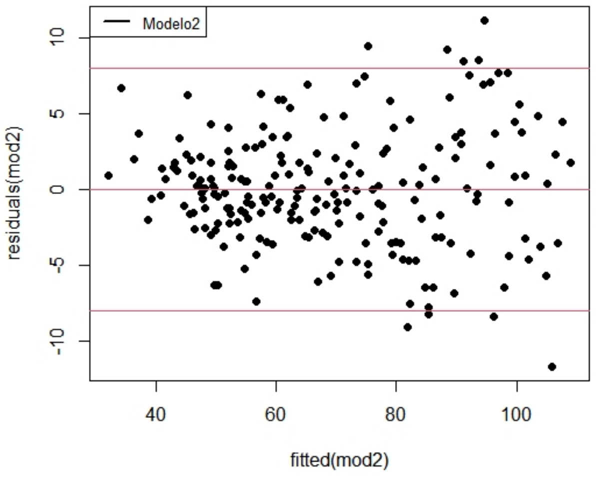
- No hay evidencia de carencia de ajuste en las componentes estructurales (o sea en tendencia y estacionalidad), pues en los residuos vs. valores ajustados no se observan patrones claros con forma de U o de W que indiquen mal ajuste de la tendencia, ni patrones periódicos en las gráficas de las series de tiempo de residuos de ajuste que indiquen mal ajuste de la estacionalidad.

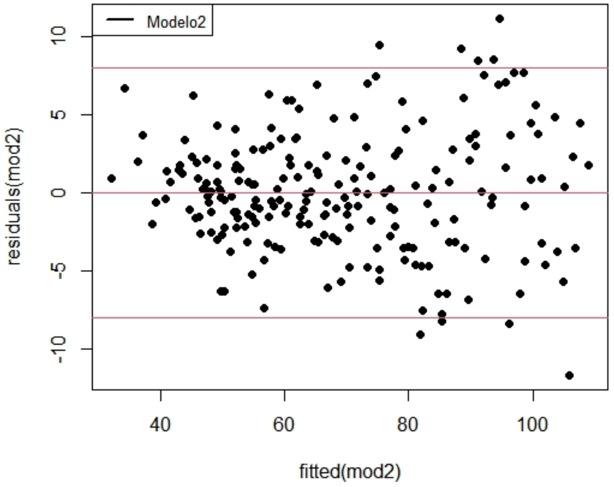


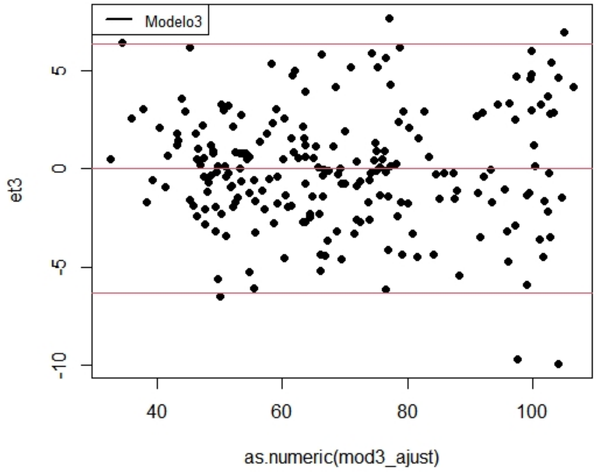


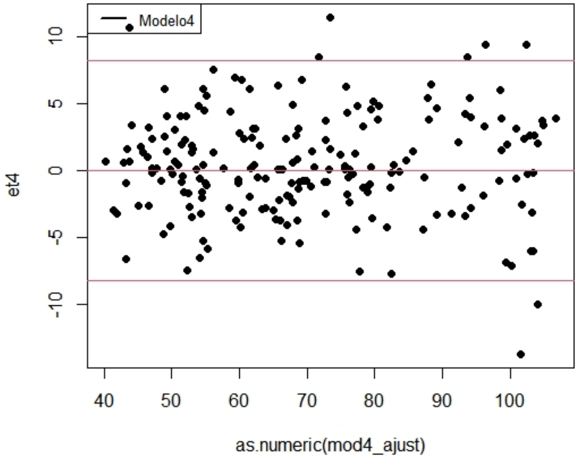
- En general todos los modelos parecen no tener evidencia en contra de los supuestos de media cero y varianza constante, por lo que esto es positivo pero los métodos locales logran seguir ciclos, por eso son mejores que los modelos globales.

ANÁLISIS DE RESIDUALES









Con todo lo antes dicho, de los 4 modelos, se considera que los modelos de ajuste local son mejores, ya que en ellos no hay patrones cíclicos y por tanto dan indicios de no incumplir el supuesto de independencia.

Si es necesario elegir entre los locales, hasta el momento se prefiere el modelo 3 ya que le fue bien los supuestos del modelo y fue el mejor en los ajustes, solo queda mirar en los pronósticos para garantizar de que es el mejor.

CONTENIDO

PRONÓSTICOS PARA LA VALIDACIÓN CRUZADA

Tabla ecuaciones de pronóstico

Modelo	*Ecuaciones de Pronósticos*
1	$\hat{Y}_{216}(L) \approx 35.8242551 + 0.1319807(216+L) + 0.0007051(216+L)^2 - 3.1817818 I_{1,216+L} + 5.3069465 I_{2,216+L} + 12.0220423 I_{3,216+L} + 7.2690612 I_{4,216+L} + 12.1146698 I_{5,216+L} + 9.8255350 I_{6,216+L} + 8.4294343 I_{7,216+L} + 9.0930346 I_{8,216+L} + 12.5163357 I_{9,216+L} + 10.5493377 I_{10,216+L} + 8.9475962 I_{11,216+L}$
2	$\hat{Y}_{216}(L) \approx 35.1122117 + 0.1689800(216+L) + 0.0002800(216+L)^2 + 0.0000013(216+L)^3 - 3.1146344 I_{1,216+L} + 5.3679308 I_{2,216+L} + 12.0768988 I_{3,216+L} + 7.3178173 I_{4,216+L} + 12.1573451 I_{5,216+L} + 9.8621412 I_{6,216+L} + 8.4599754 I_{7,216+L} + 9.1175066 I_{8,216+L} + 12.5347270 I_{9,216+L} + 10.5616286 I_{10,216+L} + 8.9537593 I_{11,216+L}$
3	$\hat{Y}_{216}(L) = \left(\hat{\beta}_{0,216} + \hat{\beta}_{1,216}(216+L) + \hat{\beta}_{2,216}(216+L)^2\right) + (-10.8780433 I_{1,216+L} - 2.3143178 I_{2,216+L} + 4.3650940 I_{3,216+L} - 0.5319649 I_{4,216+L} + 4.3979371 I_{5,216+L} + 2.0783292 I_{6,216+L} + 0.7969567 I_{7,216+L} + 1.2373979 I_{8,216+L} + 4.6148489 I_{9,216+L} + 2.5229371 I_{10,216+L} + 1.0633783 I_{11,216+L} - 7.3525531 I_{12,216+L})$
4	$\hat{Y}_{216}(L) = (101.9644720 + 0.2949128 \times L) + (-12.3938476 I_{1,216+L} - 2.0117762 I_{2,216+L} + 3.6989236 I_{3,216+L} + 0.1017724 I_{4,216+L} + 6.8427850 I_{5,216+L} + 4.2348284 I_{6,216+L} + 2.7412904 I_{7,216+L} + 6.6698440 I_{8,216+L} + 10.1574426 I_{9,216+L} + 6.3504048 I_{10,216+L} + 4.5463245 I_{11,216+L} - 9.2424032 I_{12,216+L})$

Pronósticos puntuales y por intervalos de los modelos globales:

Valores reales:

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
87.3	102.1	108.5	104.0	112.0	103.3	110.3	109.4	112.1	114.7	109.2	103.3

Modelo 1:

Periodo	*L*	*Pronósticos*	*Lim. Inf*	*Lim. Sup*
Ene 2019	1	94.4852	86.27520	102.6952
Feb 2019	2	103.4126	95.19741	111.6279
Mar 2019	3	110.5678	102.34725	118.7884
Abr 2019	4	106.2564	98.03028	114.4825
May 2019	5	111.5449	103.31316	119.7767
Jun 2019	6	109.7001	101.46257	117.9377
Jul 2019	7	108.7498	100.50627	116.9933
Agos 2019	8	109.8606	101.61094	118.1102
Sept 2019	9	113.7324	105.47658	121.9883
Oct 2019	10	112.2154	103.95319	120.4777
Nov 2019	11	111.0651	102.79632	119.3338
Dic 2019	12	102.5703	94.29486	110.8457

Modelo 2:

Periodo	*L*	*Pronósticos*	*Lim. Inf*	*Lim. Sup*
Ene 2019	1	95.19725	86.84148	103.5530
Feb 2019	2	104.15594	95.78241	112.5295
Mar 2019	3	111.34331	102.95097	119.7356
Abr 2019	4	107.06490	98.65269	115.4771
May 2019	5	112.38738	103.95418	120.8206
Jun 2019	6	110.57743	102.12208	119.0328
Jul 2019	7	109.66281	101.18411	118.1415
Ago 2019	8	110.81020	102.30691	119.3135
Sep 2019	9	114.71959	106.19043	123.2487
Oct 2019	10	113.24098	104.68463	121.7973
Nov 2019	11	112.12994	103.54501	120.7149
Dic 2019	12	103.67534	95.06044	112.2903

Pronósticos puntuales y por intervalos de los modelos locales:

Valores reales:

Ene	*Feb*	*Mar*	*Abr*	*May*	*Jun*	*Jul*	*Ago*	*Sep*	*Oct*	*Nov*	*Dic*
87.3	102.1	108.5	104.0	112.0	103.3	110.3	109.4	112.1	114.7	109.2	103.3

Modelo 3:

Fecha	L	$\hat{T}_{216}(L)$	$\hat{S}_{216}(L)$	$\hat{Y}_{216}(L) = \hat{T}_{216}(L) + \hat{S}_{216}(L)$
Ene 2019	1	103.9982	−10.8780433	93.12013
Feb 2019	2	104.6239	−2.3143178	102.30959
Mar 2019	3	105.2877	4.3650940	109.65277
Abr 2019	4	105.9900	−0.5319649	105.45802
May 2019	5	106.7313	4.3979371	111.12925
Jun 2019	6	107.5121	2.0783292	109.59040
Jul 2019	7	108.3326	0.7969567	109.12959
Ago 2019	8	109.1933	1.2373979	110.43074
Sep 2019	9	110.0945	4.6148489	114.70934
Oct 2019	10	111.0364	2.5229371	113.55930
Nov 2019	11	112.0192	1.0633783	113.08259
Dic 2019	12	113.0433	−7.3525531	105.69070

Modelo 4:

Periodo	L	*Pronósticos*	$Lim.Sup$	$Lim.Inf$
Ene 2019	1	89.86554	97.40692	82.32415
Feb 2019	2	100.54252	108.28129	92.80375
Mar 2019	3	106.54813	114.48318	98.61309
Abr 2019	4	103.24590	111.37622	95.11557
May 2019	5	110.28182	118.60654	101.95710
Jun 2019	6	107.96878	116.48710	99.45046
Jul 2019	7	106.77015	115.48136	98.05894
Ago 2019	8	110.99362	119.89709	102.09014
Sep 2019	9	114.77613	123.87132	105.68094
Oct 2019	10	111.26400	120.55041	101.97759
Nov 2019	11	109.75484	119.23204	100.27763
Dic 2019	12	96.26102	105.92865	86.59339

Interpretación de los resultados anteriores

Los pronósticos puntuales, indican el valor que cada modelo ajustado con los primeros n = 216 datos, predice que será observado para la serie en cada periodo de los pronósticos ex-posts.

Los pronósticos por intervalos, indican que el valor real estará en el intervalo de cada uno de los modelos con una confianza del 95%.

La producción nominal de la categoría otros productos químicos para el mes de Junio de 2019, comparado con los meses del año 2018 es de:

- En el modelo global cuadrático : 109.59040 % es decir, durante este periodo aumentó un 9.59%. Además, con un 95% de confianza el valor real del índice se encuentra entre 101.46257 y 117.9377.

- En el modelo global cúbico : 110.57743% es decir, durante este periodo aumentó aproximadamente un 10.57%. Además, con un 95% de confianza el valor real del índice se encuentra entre 102.12208 y 119.0328 .

- En el modelo local con Descomposición aditiva & Loess cuadrático (DLC) : 109.12959 es decir, durante este periodo aumentó aproximadamente un 9.13%.

- En el modelo local aditivo con SEHW: 107.96878% es decir, durante este periodo aumentó aproximadamente un 7.9%. Además, con un 95% de confianza el valor real del índice se encuentra entre 99.45046 y 116.48710

Medidas MAE, MAPE y RMSE en todos los modelos

$RMSE$	MAE	$MAPE$	$Amplitud$	$Cobertura(\%)$
3.153708	2.366654	2.338060	16.48278	100
3.620468	2.746457	2.711682	16.94829	100
3.014298	2.335510	2.283224	–	–
3.187973	2.670446	2.533667	17.22162	100

Según este criterio todos los modelos que tienen intervalo de predicción tienen una cobertura del 100%, es decir todos los valores reales caen en el intervalo.

El modelo 1 es el de menor amplitud, por lo tanto según este criterio sería el mejor modelo.

Modelo 1:

En promedio en cada pronóstico se estima un error de $\pm$ 3.153708 puntos del índice de producción nominal según el RMSE y de $\pm$ 2.366654 puntos del índice de producción nominal según MAE, mientras que MAPE estima que en promedio cada pronóstico se comete un error de $\pm$ 2.338% con relación al valor real del índice de producción nominal, todo esto en el modelo de regresión cuadrático con variables indicadores

Modelo 2:

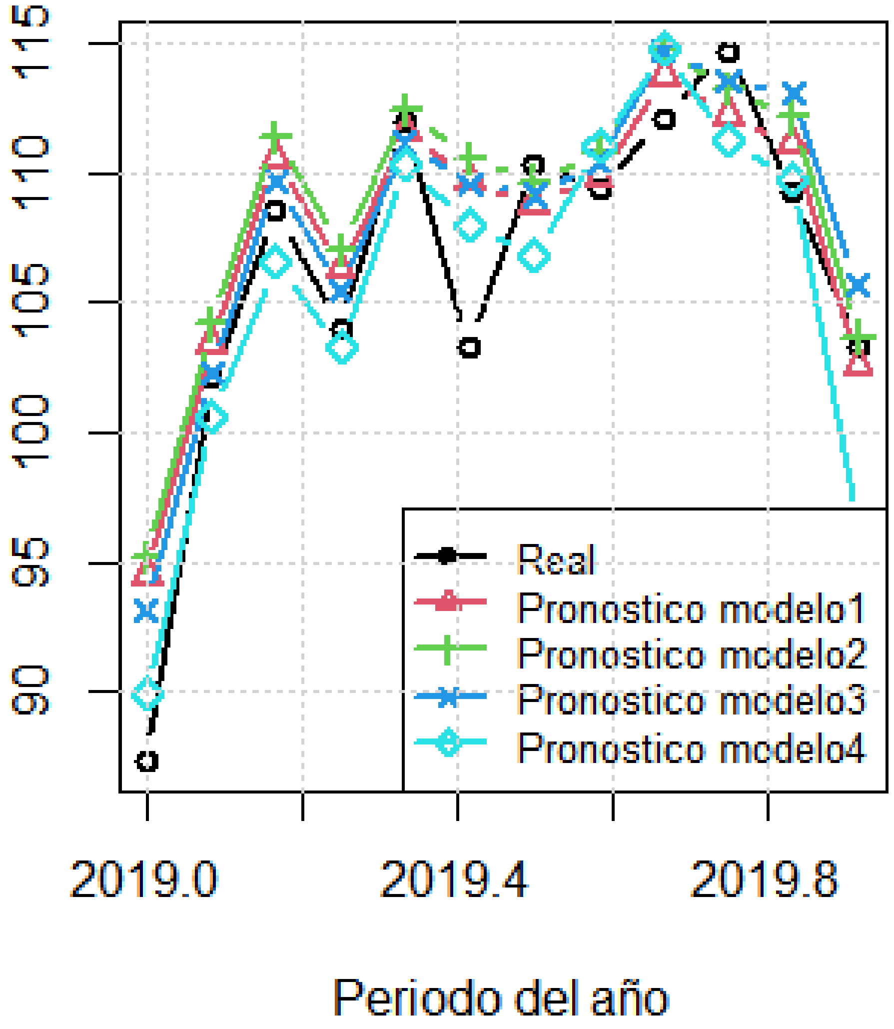
En promedio en cada pronóstico se estima un error de $\pm$ 3.620468 puntos del índice de producción nominal según el RMSE y de $\pm$ 2.746457 puntos del índice de producción nominal según MAE, mientras que MAPE estima que en promedio cada pronóstico se comete un error de $\pm$ 2.71% con relación al valor real del índice de producción nominal, todo esto en el modelo de regresión cúbico con variables indicadores.

Modelo 3:

En promedio en cada pronóstico se estima un error de $\pm$ 3.014298 puntos del índice de producción nominal según el RMSE y de $\pm$ 2.33551 puntos del índice de producción nominal según MAE, mientras que MAPE estima que en promedio cada pronóstico se comete un error de $\pm$ 2.28% con relación al valor real del índice de producción nominal, todo esto en el modelo de Filtro de descomposición combinado Loess cuadrático.

Modelo 4:

En promedio en cada pronóstico se estima un error de $\pm$ 3.187973 puntos del índice de producción nominal según el RMSE y de $\pm$ 2.670446 puntos del índice de producción nominal según MAE, mientras que MAPE estima que en promedio cada pronóstico se comete un error de $\pm$ 2.53% con relación al valor real del índice de producción nominal, todo esto en el modelo de SEHW.



- En base a la gráfica de pronósticos se escoge como mejor modelo el modelo 3.

CONTENIDO

CONCLUSIONES

Teniendo como primer criterio de selección los resultados en el análisis de residuales, en segundo lugar los resultados de pronósticos y por último los de ajustes, se concluye que el mejor modelo por el momento, es el modelo 3.

Este no mostró evidencia en contra de los supuestos (aunque aún no se ha probado el supuesto de independencia ni el de normalidad), además capturó muy bien la dinámica de la serie, este tiene las mejores medidas de pronóstico y ajuste, además de que sigue bien la tendencia, estacionalidad y variaciones cíclicas, aunque en algunos pronósticos no parece ser tan realista y confiable debido a que en la mayoría tiende a subestimar los valores y al final de los pronósticos es el modelo más alejado del valor real, por tanto se cree que podría cometer errores en pronósticos ex-ante.

Se recomienda, debido a un cambio de nivel muy leve que presenta la serie aproximadamente en el año 2015, analizar con más detalle qué pasó en este año en el sector de los productos químicos y así determinar con mayor certeza si corresponde a una intervención, que por ende requiera del ajuste de modelos con los datos posteriores a dicha fecha, para así explicar mejor el comportamiento de la serie.