

ANÁLISIS DE SERIES DE TIEMPO. AJUSTE DE TENDENCIA Y ESTACIONALIDAD

Sofía Cuartas García¹, Simón Cuartas Rendón² y Deivid Zhang Figueroa³
Fecha de entrega: 25 – 06 – 2022

1. ANÁLISIS DESCRIPTIVO Y TEST HEGY DE RAÍCES UNITARIAS

En este trabajo se van a plantear modelos SARIMA para ajustar la serie temporal del **índice de ventas del sector manufacturero colombiano en pesos nominales** que publica el Departamento Administrativo Nacional de Estadística (DANE) cada mes, con el objetivo de poder realizar pronósticos para este índice de cara al futuro, y a su vez, se van a contrastar estos modelos con modelos ARMA y modelos globales y locales que no realizan ningún ajuste a los errores y que han sido estudiados anteriormente. Para ello, es importante tener presente que la serie temporal de índice de ventas en pesos nominales del sector manufacturero colombiano con la que se trabajará va desde enero de 2001 hasta noviembre de 2021, lo cual implica que se tienen $N = 251$ observaciones para esta serie de tiempo (teniendo presente que aquellos datos asociados a los meses de marzo de 2020 en adelante fueron imputados para ignorar los efectos que tuvo la pandemia por la *COVID-19* en este índice, entre otros indicadores económicos), si bien para el ajuste de los todos los modelos fueron empleadas los valores del índice de ventas en pesos nominales para los primeros $n = 239$, dejando los últimas doce observaciones como parte del periodo *ex post* para poder realizar validación cruzada.

Mencionado lo anterior, se comienza revisando las características de la serie de tiempo a trabajar en la **figura 1** considerando las $N = 251$ observaciones de esta.

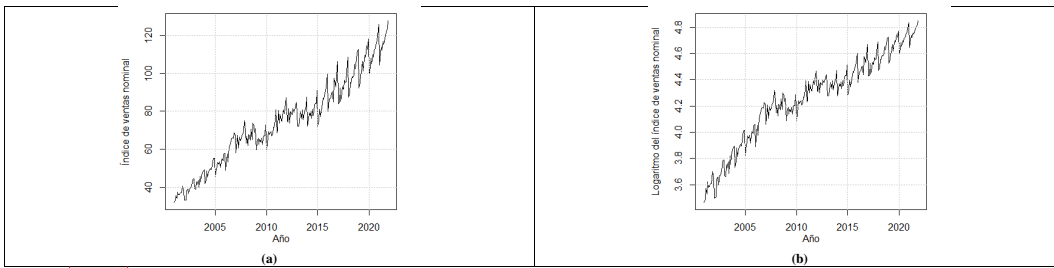


Figura 1. Serie temporal de **índice de ventas del sector manufacturero colombiano en pesos nominales** desde enero de 2001 a noviembre de 2021. (a) Serie con el índice de ventas en pesos nominales en escala original. (b). Serie temporal con el índice de ventas en pesos nominales en escala logarítmica.

En la **figura 1 (a)** el índice de ventas del sector manufacturero colombiano en pesos nominales se presenta en escala original. En esta se puede observar que la serie presenta tendencia creciente, ya que el índice bajo estudio tiende a aumentar con el tiempo; asimismo, se puede evidenciar como este índice va presentando una mayor varianza conforme para el tiempo, lo cual señala que esta serie es multiplicativa; de igual manera se aprecia que la tendencia es global en tanto es posible emplear una función suave del tiempo para modelarla, y además, se puede decir que la tendencia es determinística, puesto que la **evolución sucede de forma perfectamente predecible**. Por otro lado, es clara la presencia de la componente estacional en esta serie, debido a que existe un comportamiento repetitivo en cada año de la serie, y se puede decir que esta es exacta con $s = 12$ meses, pues en general se observa que el índice de ventas en pesos nominales inicia con un mínimo en los meses de enero, y luego muestra un incremento rápido hasta alcanzar sus máximos anuales en los meses de diciembre, para luego mostrar una reducción al comenzar enero del siguiente año y comenzar así un nuevo periodo anual, y por lo anterior es también razonable decir que la estacionalidad es determinística. No obstante, vale la pena señalar que en algunos años este comportamiento estacional no tiene la misma apariencia que la mayoría de los años, según como se acabó de describir, por la presencia de ciclos en la serie, como ocurre por ejemplo entre los años 2008 a 2011. Para finalizar esta descripción inicial, se debe destacar que en la **figura 1 (b)** se aprecia como la variabilidad alrededor de la tendencia media de largo plazo se estabiliza. **XXXX ¿estacionalidad global?**

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

Commented [SCR1]: ¿Se debe presentar también la serie del logaritmo natural de la serie completa? Más abajo se presenta la gráfica de esta serie pero recortada.

Commented [SCR2]: ¿Se puede dar el caso de que solo la tendencia o solo la estacionalidad sea global y la otra no?

¹ Estudiante de Estadística, Universidad Nacional de Colombia – Sede Medellín.

² Estudiante de Estadística, Universidad Nacional de Colombia – Sede Medellín

³ Estudiante de Estadística, Universidad Nacional de Colombia – Sede Medellín

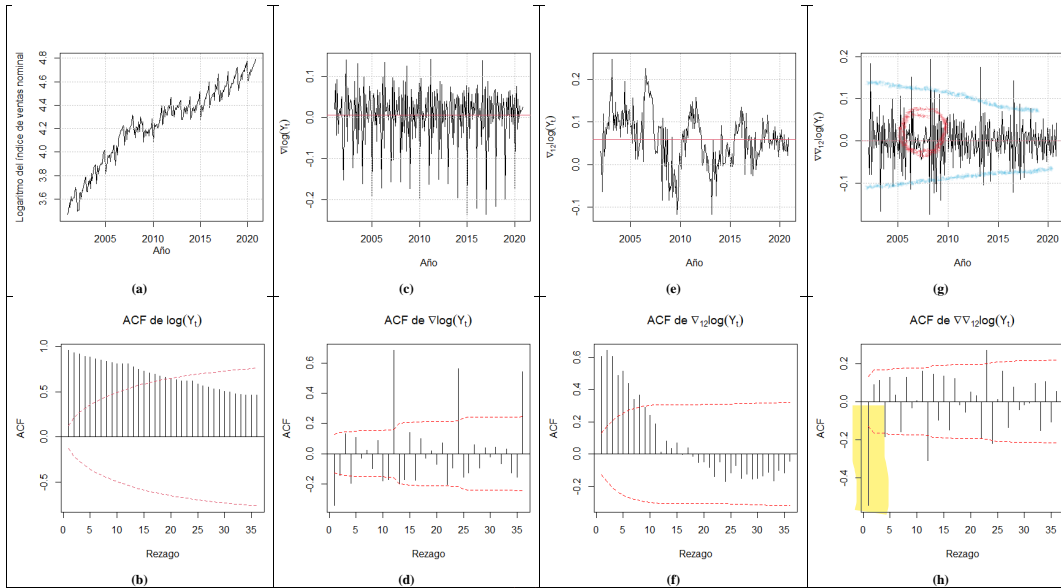


Figura 2. (a). Serie recortada, (b) ACF muestral de la serie recortada; (c) primera diferencia regular de la serie recortada, (d) ACF de la primera diferencia regular de la serie recortada; (e) primera diferencia estacional serie recortada, (f) ACF muestral de la primera diferencia estacional de la serie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

Comenzando con la serie temporal recortada y transformada a escala logarítmica de la **figura 2 (a)** se evidencia que la media no es constante, sino que tiende a crecer con el tiempo en tanto **la tendencia es positiva**, y dentro de cada año también se evidencia un incremento de enero a diciembre, como se analizó previamente; empero, esta transformación permitió estabilizar la varianza del índice de ventas en pesos nominales, y al pasar a su ACF muestral en la **figura 2 (b)** se concluye rápidamente que el proceso no es ergódico, pues la ACF disminuye muy lentamente tanto para k consecutivos en el caso de la parte regular, como para k múltiplos de doce ($k = 12, 24, 36$), que es el número de periodos en cada año. Además, debe destacarse que $\hat{\rho}(k) = \text{Corr}(\log Y_t, \log Y_{t+1}) > 0$, lo que corrobora que esta serie se ve afectada por ciclos. Por lo anterior, se tiene que el logaritmo del proceso asociado al índice de ventas del sector manufacturero colombiano en pesos nominales **no es estacionario**.

Respecto a la primera diferencia de la serie temporal recortada en escala logarítmica, visible en la **figura 2 (c)**, es claro que en este caso la media es constante y es aproximadamente nula, pero en cuanto a la varianza, no es muy claro que esta sea constante, ya que entre el 2015 y el 2020 parece que esta incrementa ligeramente, pero no de forma significativa respecto al resto de periodos ~~XXXX~~. Luego, analizando la **figura 2 (d)**, donde se observa su ACF muestral, donde refleja para la parte regular que esta es ergódica en tanto la autocorrelación muestral toma valores muy cercanos a cero rápidamente para los primeros valores de rezago; en cambio, la parte estacional se observa un patrón de cola positiva, si bien no puede ser considerada ergódica porque para $k = 12$ se puede ver que la función de autocorrelación muestral toma un valor superior a 0.6, para $k = 24$ la ACF es aproximadamente 0.5 y para $k = 36$ la ACF es nuevamente cercana a 0.5. De esta forma, se puede afirmar que la primera diferencia regular del logaritmo del índice de ventas del sector manufacturero en pesos nominales **no es estacionario**.

Después, considerando la primera diferencia estacional de la serie recortada, cuya gráfica se halla en la **figura 2 (e)**, se puede ver que la media no es constante, dado que existen ciclos en esta primera diferencia regular, pero se observa que los valores fluctúan dentro de un rango de tamaño muy semejante en todos los periodos, por lo que la varianza es constante. En cuanto a su función de autocorrelación muestral, visible en la **figura 2 (f)**, se tiene para su parte regular se tiene que esta no es ergódica, ya que la ACF en el segundo rezago es mayor que para el primero, pero a continuación comienza a disminuir hasta $k = 4$, para presentar luego un crecimiento en $k = 5$ y luego volver a decrecer en $k = 6$, y en cualquier caso, cuando decrece no lo hace lo suficientemente rápido como para sugerir que la ACF de la primera diferencia estacional en su parte estacional es ergódica. Respecto a la componente estacional, se tiene que tampoco se cumple el supuesto de ergodicidad, ya que si bien para $k = 12$ la autocorrelación muestral es muy próxima a cero, esta se incrementa para $k = 24, 36$. Así, se concluye que la primera diferencia estacional del índice de ventas del sector manufacturero colombiano en pesos nominales **no es estacionario**.

Finalmente, se estudiará la serie temporal diferenciada por tendencia y estacionalidad, esto es, a $\nabla \nabla_{12} \log Y_t$, cuyo gráfico se presenta en la **figura 2 (g)** y en la que se muestra que la media es nula y que la **varianza es aproximadamente constante** para toda la serie temporal. También resulta relevante analizar su función de autocorrelación muestral, exhibida en la **figura 2 (h)**, en la cual se muestra que para la parte regular se tiene un patrón tipo de **cola exponencial sinusoidal** que se acerca a cero rápidamente en $k = 6$.

Commented [SCR3]: -La media es ligeramente positiva. ¿Se debe hacer algún comentario relevante respecto a este hecho? -¿Es normal que la escala se reduzca tanto?

Commented [SCR4]: ¿Sí es razonable plantear que la varianza es aproximadamente constante? En 2007 se ve una disminución importante y luego un incremento fuerte para luego presentar una varianza igual a la del inicio de la serie.

Commented [SCR5]: ¿Sí se puede decir que sea cola exponencial sinusoidal? Muestra valores positivos y negativos que no cambian consecutivamente y termina en $k = 6$ con un valor muy pequeño.

Respecto a la parte estacional, se puede ver que la ACF muestral toma inicialmente valores grandes pero para $k = 36$ este es próximo a cero. Además, resulta también útil analizar la función de autocorrelación parcial, PACF, de este proceso, observado en la **figura 3**, en la cual se observa, **para la parte regular**, que esta parece ser una cola exponencial sinusoidal que a partir de $k = 3$ se conserva dentro de los límites de Bartlett, por lo que estas autocorrelaciones parciales se toman como significativas y **concluye que se cumple el supuesto de ergodicidad**. De igual forma, para la parte estacional, se tiene que esta toma valores cercanos a -0.15 para $k = 12, 24, 36$, pero se debe tener presente que para rezagos grandes las estimaciones de la PACF son menos confiables, por lo que se puede asumir que estas no son significativas y por tanto se concluye, tanto por la parte estacional como para la regular, que hay ergodicidad, coincidiendo con la conclusión derivada de la ACF en la **figura 2 (h)**.

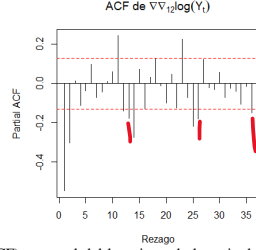


Figura 3. Función de autocorrelación parcial (PACF) muestral del logaritmo de la serie del índice de ventas del sector manufacturero colombiano en pesos nominales recortada y diferenciada por estacionalidad y tendencia.

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

Ahora, se va a realizar el **test HEGY** para esta serie, el cual es un test tipo *Dickey Fuller aumentado*, *ADF*, con la salvedad de que la serie temporal será representando como un proceso autorregresivo infinito que supone que este es invertible. Así, se tiene que un proceso ARMA tendrá raíz unitaria estacional de periodo doce si su polinomio AR tiene como raíces una o más raíces periódicas del polinomio $1 - x^{12}$, las cuales se expresan como c_k , y de este modo, si se asume que el proceso dado por logaritmo del índice de ventas del sector manufacturero colombiano en pesos nominales es invertible, entonces tendría representación $AR(\infty)$, donde el polinomio autorregresivo será $\varphi(B)$, de manera que la representación AR de la serie será dada por:

$$\varphi(B)Y_t = E_t, \quad E_t \text{ un R.B.} \sim N(0, \sigma^2), \quad \varphi(B) = 1 - \sum_{j=1}^{\infty} \phi_j B^j, \quad \text{con } \sum_{j=1}^{\infty} \phi_j^2 < \infty$$

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

$$\nabla_{12} \log Y_t = \pi_1 X_{1,t-1} + \pi_2 X_{2,t-1} + \pi_3 X_{3,t-1} + \pi_4 X_{3,t-2} + \pi_5 X_{4,t-1} + \pi_6 X_{4,t-2} + \pi_7 X_{5,t-1} + \pi_8 X_{5,t-2} + \pi_9 X_{6,t-1} + \pi_{10} X_{6,t-2} + \pi_{11} X_{7,t-1} + \pi_{12} X_{7,t-2} + \sum_{i=1}^{\infty} b_i \nabla_{12} \log Y_{t-i} + E_t, \quad E_t \text{ un R.B.} \sim N(0, \sigma^2) \quad [1]$$

Donde las variables $X_{j,t}$, $j = 1, 2, 3, 4, 5, 6, 7$ resultan de aplicar filtros tales que los coeficientes de regresión π_i estén asociadas a las raíces c_k , y tales variables son las siguientes:

- $X_{1,t} = (1 + B)(1 + B^2)(1 + B^4 + B^8)Y_t$
- $X_{2,t} = -(1 - B)(1 + B^2)(1 + B^4 + B^8)Y_t$
- $X_{3,t} = -(1 - B^2)(1 + B^4 + B^8)Y_t$
- $X_{4,t} = -(1 - B^4)(1 - \sqrt{3} + B^2)(1 + B^4 + B^8)Y_t$
- $X_{5,t} = -(1 - B^4)(1 + \sqrt{3}B + B^2)(1 + B^4 + B^8)Y_t$
- $X_{6,t} = -(1 - B^4)(1 - B^2 + B^4)(1 - B + B^2)Y_t$
- $X_{7,t} = -(1 - B^4)(1 - B^2 + B^4)(1 + B + B^2)Y_t$

Así pues, el juego de hipótesis para las pruebas del test HEGY son las presentadas en la **tabla 1**. Para cada una de las siete pruebas presentadas en tal prueba se tiene que sus estadísticos de prueba bajo H_0 poseen la distribución de un proceso estocástico complejo. Con esto presente, se acude a **R** para poder realizar estas pruebas y los resultados son presentados en la **tabla 2**, y como se puede observar, no se rechazó la hipótesis nula para el primer, tercer, quinto, sexto y séptimo test, lo cual indica que el proceso asociado a la serie temporal de interés **presenta raíces unitarias regulares y estacionales**, con una significancia de $\alpha = 0.10$, lo cual es un indicador de que es necesario aplicarle diferencia mixto $\nabla \nabla_{12}$, como el resultado en la **figura 2 (g)**, que como se pudo observar en su momento, resultó en una serie estacionaria en sentido débil.

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

Realizados los análisis anteriores, ahora se van a identificar modelos **SARIMA(p, d, q)(P, D, Q)[s]**, esto partiendo del proceso derivado de la serie temporal del índice de ventas del sector manufacturero colombiano con precios en pesos nominales que ha sido diferenciada por tendencia y estacionalidad, de manera que ya se sabe que para los modelos a presentar se cumple que $D = d = 1$. Mencionado esto, se partirá de las evaluaciones que se pueden realizar con la ACF y la PACF de las figuras 2 (g) y 3 respectivamente.

Commented [SCR6]: - ¿El orden de p (en rojo) es entonces infinito? ¿La función de R no arroja el valor de p que calcula con AIC?
- ¿El polinomio $AR(\infty)$ corresponde a la suma hacia el final?
- En algunos documentos no se presenta E_t como un proceso sino como una simple v.a. ¿Está bien así?

Commented [SCR7]: En R se arrojan algunos warnings. ¿Esto tiene algún problema? Además los valores p está en el límite.

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

Test	Hipótesis	Frecuencia angular (θ)	Ciclos por año (j)	Estadístico de prueba calculado	Valor p (V_p)	¿Rechazo? $\alpha = 0.10$	
1	$\begin{cases} H_0: \pi_1 = 0 \\ H_1: \pi_1 < 0 \end{cases}$	Hay raíz unitaria regular o no estacional No hay raíz unitaria regular o no estacional	0	0	2.9135488	0.10000000	
2	$\begin{cases} H_0: \pi_2 = 0 \\ H_1: \pi_2 < 0 \end{cases}$	Hay raíz unitaria estacional bimensual No hay raíz unitaria estacional bimensual	π	6	-2.8854438	0.01000000	X
3	$\begin{cases} H_0: \pi_3 = \pi_4 = 0 \\ H_1: \pi_3 < 0 \vee \pi_4 \neq 0 \end{cases}$	Hay raíz unitaria estacional cuatrimestral No hay raíz unitaria estacional cuatrimestral	$\pm \frac{1}{2}\pi$	3	0.8104567	0.10000000	
4	$\begin{cases} H_0: \pi_5 = \pi_6 = 0 \\ H_1: \pi_5 < 0 \vee \pi_6 \neq 0 \end{cases}$	Hay raíz unitaria estacional trimestral No hay raíz unitaria estacional trimestral	$\mp \frac{2}{3}\pi$	4	2.4693601	0.09062608	X
5	$\begin{cases} H_0: \pi_7 = \pi_8 = 0 \\ H_1: \pi_7 < 0 \vee \pi_8 \neq 0 \end{cases}$	Hay raíz unitaria estacional semestral No hay raíz unitaria estacional semestral	$\pm \frac{1}{3}\pi$	2	1.7683136	0.10000000	
6	$\begin{cases} H_0: \pi_9 = \pi_{10} = 0 \\ H_1: \pi_9 < 0 \vee \pi_{10} \neq 0 \end{cases}$	Hay raíz unitaria estacional de cinco ciclos al año No hay raíz unitaria estacional de cinco ciclos al año	$\mp \frac{5}{6}\pi$	5	2.2239858	0.10000000	
7	$\begin{cases} H_0: \pi_{11} = \pi_{12} = 0 \\ H_1: \pi_{11} < 0 \vee \pi_{12} \neq 0 \end{cases}$	Hay raíz unitaria estacional anual No hay raíz unitaria estacional anual	$\pm \frac{1}{6}\pi$	1	2.4880179	0.10000000	

Commented [SCR8]: ¿Qué tratamiento se le debe dar a los valores p iguales a 0.1?

Así pues, iniciando con la parte regular y chequeando únicamente $k = 1, 2, 3, 4, 5, 6$ para ambas funciones, se puede distinguir en la ACF que se da un patrón tipo cola, que podría ser de tipo **exponencial sinusoidal** bajo las justificaciones dadas en la primera sección, mientras que con la PACF se aprecia un patrón tipo corte, donde el primer rezago para el cual se anula estadísticamente la autocorrelación parcial es $k = 3$; así, teniendo un patrón cola en la ACF y corte en la PACF, se obtiene que la parte regular de la serie temporal de interés diferenciada por estacionalidad y tendencia se puede modelar como un proceso de medias móviles con $q = 2$, por lo que de momento se tiene un **ARIMA(2, 1, 0)**.

Por otro lado, para la parte estacional, evaluada para los rezagos $k = 12, 24, 36$, se ve un patrón contrario al caso regular, es decir, la ACF muestra un patrón corte, donde para $k = 12, 24$ se puede considerar que la ACF es estadísticamente significativa, mientras que para $k = 36$ no lo es, toda vez que en la PACF se observa un patrón tipo cola negativa. Por lo anterior, se tiene que la parte estacional puede ser modelada como un proceso de medias móviles con $Q = 24/s = 24/12 = 2$, esto es, un **ARIMA(0, 1, 2)[12]**. Entonces, al unir estos dos análisis, se obtiene que el modelo *completo*, por así decirlo, sería un **ARIMA(2, 1, 0)(0, 1, 2)[12]**. La ecuación de este modelo es la **ecuación 2**:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^{12}) \log Y_t = (1 + \theta_1 B^{12} + \theta_2 B^{24}) E_t, \quad \text{con } E_t \text{ un R.B.} \sim N(0, \sigma^2) \quad [2]$$

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

<pre>> auto.arima(lny, ic="aic", seasonal.test="ocsb") Series: lny ARIMA(0,1,2)(2,0,0)[12] with drift Coefficients: mal ma2 sar1 sar2 -0.7143 0.1770 0.548 0.2990 s.e. 0.0635 0.0677 0.062 0.0663 drift 0.0049 s.e. 0.0059 sigma^2 = 0.001605: log likelihood = 423.53 AIC=-835.06 AICc=-834.69 BIC=-814.22</pre> <p>(a)</p>	<pre>> auto.arima(lny, ic="aic", seasonal.test="ch") Series: lny ARIMA(0,1,2)(2,0,0)[12] with drift Coefficients: mal ma2 sar1 sar2 -0.7143 0.1770 0.548 0.2990 s.e. 0.0635 0.0677 0.062 0.0663 drift 0.0049 s.e. 0.0059 sigma^2 = 0.001605: log likelihood = 423.53 AIC=-835.06 AICc=-834.69 BIC=-814.22</pre> <p>(b)</p>	<pre>> auto.arima(lny, ic="aic", seasonal.test="seas") Series: lny ARIMA(4,1,0)(1,1,2)[12] Coefficients: ar1 ar2 ar3 ar4 -0.6849 -0.2928 0.0235 -0.0695 s.e. 0.0673 0.0813 0.0811 0.0665 sar1 smar sma2 -0.6689 -0.0910 -0.7050 s.e. 0.2634 0.2475 0.1922 sigma^2 = 0.001141: log likelihood = 439.51 AIC=-863.02 AICc=-862.36 BIC=-835.65</pre> <p>(c)</p>
<pre>> auto.arima(lny, ic="bic", seasonal.test="ocsb") Series: lny ARIMA(1,1,1)(2,0,0)[12] Coefficients: ar1 mal sar1 sar2 -0.2556 -0.4531 0.5544 0.2918 s.e. 0.0966 0.0850 0.0620 0.0654 sigma^2 = 0.001608: log likelihood = 422.79 AIC=-835.58 AICc=-835.32 BIC=-818.22</pre> <p>(d)</p>	<pre>> auto.arima(lny, ic="bic", seasonal.test="ch") Series: lny ARIMA(1,1,1)(2,0,0)[12] Coefficients: ar1 mal sar1 sar2 -0.2556 -0.4531 0.5544 0.2918 s.e. 0.0966 0.0850 0.0620 0.0654 sigma^2 = 0.001608: log likelihood = 422.79 AIC=-835.58 AICc=-835.32 BIC=-818.22</pre> <p>(e)</p>	<pre>> auto.arima(lny, ic="bic", seasonal.test="seas") Series: lny ARIMA(0,1,2)(0,1,1)[12] Coefficients: mal ma2 smar -0.7286 0.2237 -0.8661 s.e. 0.0632 0.0675 0.0572 sigma^2 = 0.001197: log likelihood = 432.59 AIC=-857.18 AICc=-857 BIC=-843.5</pre> <p>(f)</p>

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

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

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

Modelo uno. Usando el criterio de información AIC y método de selección de diferencias estacionales <i>ocsb</i> . ARIMA(0, 1, 2)(2, 0, 0)[12] con deriva. $(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B) \log Y_t = \delta + (1 + \theta_1 B + \theta_2 B^2) E_t, \text{ con } E_t \text{ un R.B.} \sim N(0, \sigma^2)$
Modelo dos. Usando el criterio de información AIC y método de selección de diferencias estacionales <i>ch</i> . ARIMA(0, 1, 2)(2, 0, 0)[12] con deriva. $(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B) \log Y_t = \delta + (1 + \theta_1 B + \theta_2 B^2) E_t, \text{ con } E_t \text{ un R.B.} \sim N(0, \sigma^2)$
Modelo tres. Usando el criterio de información AIC y método de selección de diferencias estacionales <i>seas</i> . ARIMA(4, 1, 0)(1, 1, 2)[12] sin deriva. $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)(1 - B)(1 - B^{12}) \log Y_t = (1 + \theta_1 B^{12} + \theta_2 B^{24}) E_t, \text{ con } E_t \text{ un R.B.} \sim N(0, \sigma^2)$
Modelo cuatro. Usando el criterio de información BIC y método de selección de diferencias estacionales <i>ocsb</i> . ARIMA(1, 1, 1)(2, 0, 0)[12] sin deriva. $(1 - \phi_1 B)(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B) \log Y_t = (1 + \theta_1 B) E_t, \text{ con } E_t \text{ un R.B.} \sim N(0, \sigma^2)$
Modelo cinco. Usando el criterio de información BIC y método de selección de diferencias estacionales <i>ch</i> . ARIMA(1, 1, 1)(2, 0, 0)[12] sin deriva. $(1 - \phi_1 B)(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B) \log Y_t = (1 + \theta_1 B) E_t, \text{ con } E_t \text{ un R.B.} \sim N(0, \sigma^2)$
Modelo seis. Usando el criterio de información BIC y método de selección de diferencias estacionales <i>seas</i> . ARIMA(0, 1, 2)(0, 1, 1)[12] sin deriva. $(1 - B)(1 - B^{12}) \log Y_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \theta_3 B^{12}) E_t, \text{ con } E_t \text{ un R.B.} \sim N(0, \sigma^2)$

Commented [SCR9]: ¿Es necesario cambiar el valor de los coeficientes por los presentados en la salidas de *R*? ¿O se debe dejar así tal cual, indicado?

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

Ahora, se va a utilizar otra función, que es `armasubsets`, de la librería TSA (versión 1.3), y la cual genera un tablero en el que se resaltan los parámetros estadísticamente significativos que, de acuerdo con esta función, deben ser incorporados en el modelo. Esto se hará de dos modos: con un tablero 12 × 12, visible en la **figura 5** (por lo que se ignora la parte estacional) y otro tablero 18 × 18 (que sí incorpora términos asociados a la parte estacional de la serie), presentado en la **figura 6**.

Figura 5. Gráfico 12 × 12 arrojado por `armasubsets` sobre $\nabla \nabla_{12} \log Y_t$ usando el método “*ols*”.

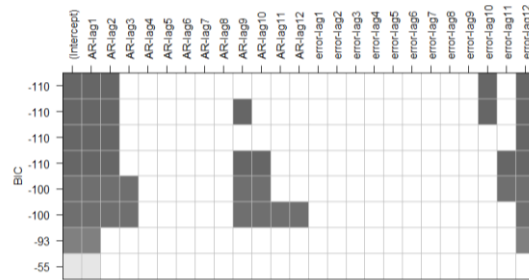
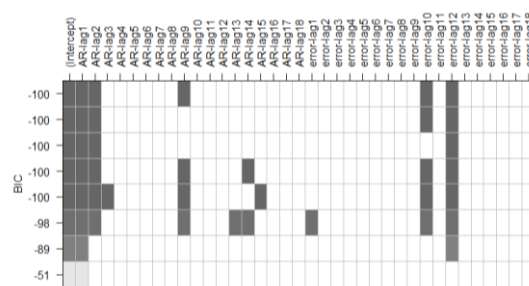


Figura 6. Gráfico 18 × 18 arrojado por `armasubsets` sobre $\nabla \nabla_{12} \log Y_t$ usando el método “*ols*”.



Así pues, teniendo en cuenta que estos tableros han sido obtenidos con la serie diferenciada por estacionalidad y tendencia, por lo que se obtendrán modelos ARIMA con $d = D = 1$, se tiene que, a partir del tablero 12×12 de *armasubsets* de la *figura 5*, y al ceñirse únicamente el primer renglón, se obtiene un modelo *ARIMA(9, 1, 12)* que solo incluye a los términos $\phi_1, \phi_2, \theta_{10}$ y θ_{12} , por lo que su ecuación teórica es la *ecuación 3*:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B) \log Y_t = (1 + \theta_{10} B^{10} + \theta_{12} B^{12}) E_t, \quad E_t \text{ un R.B.} \sim N(0, \sigma^2) \quad [3]$$

Sin embargo, si se incorpora el término ϕ_6 , si bien continuaría siendo un *ARIMA(9, 1, 12)*, la ecuación sería como se muestra en la *ecuación 4*:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_6 B^6)(1 - B) \log Y_t = (1 + \theta_{10} B^{10} + \theta_{12} B^{12}) E_t, \quad E_t \text{ un R.B.} \sim N(0, \sigma^2) \quad [4]$$

Por otro lado, con el tablero 18×18 de *armasubsets* de la *figura 6*, se identifica un modelo *ARIMA(9, 1, 10)(0, 1, 1)[12]* que solo emplea los términos $\phi_1, \phi_2, \phi_9, \theta_1$ y Θ_1 , por lo que ecuación sería la *ecuación 5*:

$$(1 - \phi_1 B - \phi_2 B - \phi_9 B^9)(1 - B)(1 - B^{12}) \log Y_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12}) E_t, \quad E_t \text{ un R.B.} \sim N(0, \sigma^2) \quad [5]$$

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