

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 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 las últimas $m = 12$ observaciones como parte del periodo *ex post* para verificar la calidad del pronóstico mediante la estrategia de validación cruzada.

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

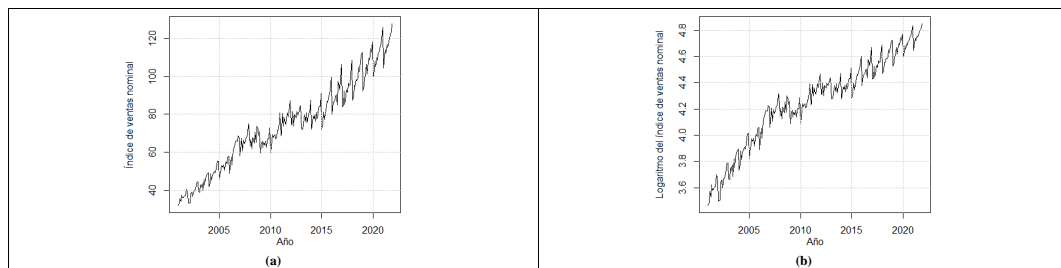


Figura 1. Serie temporal de *índice de ventas del sector manufacturero colombiano en pesos nominales* desde enero de 2001 a noviembre de 2021. (a) Serie con el índice de ventas en pesos nominales en escala original. (b). Serie temporal con el índice de ventas en pesos nominales en escala logarítmica.

En la *figura 1 (a)* se encuentra el índice de ventas del sector manufacturero colombiano en pesos nominales presentado en escala original. En esta se puede observar que la serie muestra tendencia creciente, ya que el índice bajo estudio tiende a aumentar con el tiempo; asimismo, se puede evidenciar cómo este índice va presentando una mayor varianza conforme pasa el tiempo, lo cual señala que esta serie es multiplicativa; de igual manera se aprecia que la tendencia es global en tanto es posible emplear una función suave del tiempo para modelarla, y además, se puede decir que la tendencia es determinística, puesto que su evolución se puede calificar como predecible. Por otro lado, es clara la presencia de la componente estacional en esta serie, debido a que existe un comportamiento repetitivo en cada año de la serie, y se puede decir que esta es aproximadamente exacta con $s = 12$ 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 brusca al comenzar enero del siguiente año y comenzar así un nuevo periodo anual, y por lo anterior es también razonable decir que la estacionalidad es determinística. No obstante, vale la pena señalar que en algunos años este comportamiento estacional no tiene la misma apariencia en comparación con la mayoría de los años, según como se acabó de describir, esto gracias a la presencia de ciclos en la serie, como ocurre por ejemplo entre los años 2008 a 2011, lo cual implica que una modelación global de la estacionalidad pueda tener inconvenientes en contraste con un modelo local que puede tener mejores resultados en este respecto. Para finalizar esta descripción inicial, se debe destacar que en la *figura 1 (b)* se aprecia como la variabilidad alrededor de la tendencia media de largo plazo se estabiliza al hacer una transformación logarítmica en la escala del índice de ventas del sector manufacturero colombiano en pesos nominales.

A continuación, es necesario observar la *figura 2*, para la que se considera la serie recortada con las primeras $n = 239$ observaciones, 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.

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

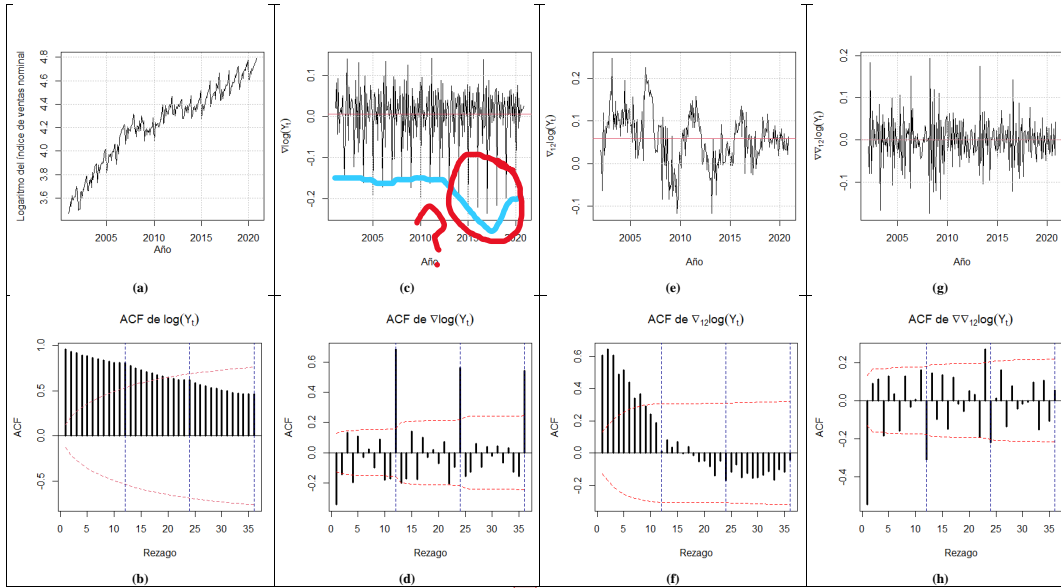


Figura 2. (a). Serie recortada; (b) ACF muestral de la serie recortada; (c) primera diferencia regular de la serie recortada; (d) ACF de la primera diferencia regular de la serie recortada; (e) primera diferencia estacional serie recortada; (f) ACF muestral de la primera diferencia estacional de la serie recortada; (g) primera diferencia regular y estacional de la serie recortada; (h) ACF muestral de la serie recortada diferenciada por tendencia y estacionalidad. Nótese que los rezagos asociados a la estacionalidad ($k = 12, 24, 36$) han sido delineados con una línea punteada azul oscura vertical.

Se comienza entonces discutiendo la serie temporal transformada mediante logaritmo natural y sin diferenciar por tendencia ni estacionalidad, graficada en la **figura 2 (a)**. En esta salta a la vista inicialmente que la media no es constante gracias a la componente estructural de la tendencia, que hace que la media sea creciente, pero también gracias a la estacionalidad, que provoca que disminuya al comenzar cada año respecto al último periodo del año anterior y aumente hasta finalizar el año. No obstante, parece que la varianza es constante, lo cual se debe a la transformación logarítmica. Luego, estudiando la función de autocorrelación muestral (en adelante **ACF**) asociada a esta serie e ilustrada en la **figura 2 (b)**, definida en este caso como $\hat{\rho}(k) = \text{Corr}(\log Y_t, \log Y_{t+k})$, se observa para primeros seis rezagos que la función de autocorrelación toma valores muy próximos a uno, lo que hace factible la existencia de raíz unitaria regular en esta serie, lo cual hace necesario diferenciar por tendencia a la serie, y de igual manera, se observa que los valores de la función de autocorrelación disminuyen muy lentamente, lo cual implica que la parte regular no es un proceso ergódico; después, al analizar la parte estacional en $k = 12, 24, 36$, se evidencia una cola positiva que decae de forma lenta, lo cual indica que también sea muy probable la existencia de una raíz unitaria estacional, por lo que resultará también conveniente hallar la primera diferencia estacional para el logaritmo del índice de ventas en pesos nominales del sector manufacturero, y como también se observa que estos valores se acercan a cero de forma lenta, se concluye que la parte estacional no es ergódica. Con todo lo anterior, se concluye que este primer proceso **no es estacionario**.

Después, tomando la primera diferencia del logaritmo del índice de interés cuya gráfica se tiene en la **figura 2 (c)**, se puede evidenciar que el nivel se ha estabilizado pero la media no es constante ya que persiste el patrón periódico en la serie, reflejado en los picos que se observan a lo largo de la serie y que ocurren justo en el comienzo y final de cada año (nótese los picos que se ciernen sobre las líneas punteadas verticales de los años 2005, 2010 y 2015); la varianza por su parte parece ser constante. Luego, pasando a la ACF de esta serie, definida como $\hat{\rho}(k) = \text{Corr}(\nabla \log Y_t, \nabla \log Y_{t+k})$ y retratada en la **figura 2 (d)**, y comenzando con la parte regular, se puede evidenciar que la ACF, para los primeros seis rezagos, toma valores que son cercanos a cero, por lo que podría ser un proceso ergódico; ahora, respecto a la parte estacional, se puede notar que la ACF toma sus valores más altos en $k = 12, 24, 36$ y con disminuciones progresivas muy pequeñas, lo que es señal de que el proceso, en su parte estacional, no es ergódico. Así, se concluye que el proceso diferenciado por tendencia **no es estacionario en sentido débil**. No obstante, es importante notar que basta con diferenciar una vez la serie por tendencia, ya que la tendencia es estabilizada y la parte regular del proceso es ergódico.

A continuación, se va a analizar el proceso diferenciado por estacionalidad, graficado en la **figura 2 (e)**, en la cual se puede que ya no hay un patrón periódico exacto, si bien el nivel no es estable, siendo claro entonces que la media no es constante; de la misma manera, se evidencia que la variación del proceso diferenciado por estacionalidad no es igual para todos los periodos (evidenciado por ejemplo el primer quinquenio con una alta variabilidad en contraste con el último), pero es importante señalar que ya no se observa un patrón periódico en la serie. Así pues, avanzando con la ACF muestral, definida en este caso como $\hat{\rho}(k) = \text{Corr}(\nabla_{12} \log Y_t, \nabla_{12} \log Y_{t+k})$, resalta para la parte regular con los primeros cinco rezagos que estos toman valores muy cercanos a uno

Commented [SCR1]: -La media es ligeramente positiva. ¿Se debe hacer algún comentario relevante respecto a este hecho?
-¿Es normal que la escala se reduzca tanto?

y con un decaimiento **lento**, por lo que se concluye que la parte regular no es ergódica y, además, parece que su **patrón es tipo cola sinusoidal exponencial**. Luego, en cuanto a la parte regular, se observa que siempre toma valores que pueden ser considerados estadísticamente iguales a cero, por lo que se puede tomar como ergódico. Con todo esto se concluye que la serie diferenciada por primera vez por estacionalidad **no es estacionaria en covarianza**, aunque se debe señalar que es suficiente con diferenciarla una vez por estacionalidad en tanto esto hace desaparecer las componentes periódicas exactas o aproximadamente exactas y la parte estacional es ergódica, eliminando entonces la raíz unitaria estacional.

Commented [SCR2]: Es más lento que con $\log Y_t$, pero no parece tan rápido como en el ejemplo de cemento Portland. ¿Es adecuado decir que es lento?

Por último, al revisar el proceso asociado a la serie temporal del índice de ventas del sector manufacturero colombiano en pesos nominales tomando tanto la primera diferencia regular como la primera diferencia estacional, y cuya gráfica se tiene en la **figura 2 (g)**, se observa que el nivel ha sido estabilizado y que ya no se aprecian patrones periódicos exactos o casi exactos (nótese que ya no hay picos separados entre sí por la misma cantidad de periodos), y además, se puede juzgar al proceso como homocedástico en tanto la varianza es aproximadamente igual a lo largo del proceso. Después, analizando la **figura 2 (h)** que grafica la función de autocorrelación muestral para este proceso, que se define como $\hat{\rho}(k) = \widehat{Corr}(\nabla_{12} \log Y_t, \nabla_{12} \log Y_{t+k})$, se tiene para la parte regular que para el primer y el cuarto rezago la función de autocorrelación es estadísticamente diferente de cero, aunque se debe resaltar que para el primer rezago toma un valor que no es próximo a uno y este es el mayor de todos entre los rezagos analizados para la parte regular ($k = 1, 2, 3, 4, 5, 6$), siendo finalmente la función de autocorrelación muestral muy próxima a cero (y estadísticamente igual a cero al no superar los límites de Bartlett señalados en líneas rojas punteadas), por lo que se puede asumir que la parte regular es ergódica; dicho esto, y analizando la parte estacional, se observa que para $k = 12, 24$ la función de autocorrelación muestral es estadísticamente significativa al tomar valores que superan los límites de Bartlett, pero este no es el caso para $k = 36$, donde no solo no se superan tales límites sino que se tiene un valor muy próximo a cero, de tal suerte que se puede considerar que el proceso en su parte estacional es ergódico, y cumpliéndose la ergodicidad para la parte estacional y regular, entonces se concluye que el proceso es ergódico, lo cual muestra además que ya no hay evidencia de existencia de raíces unitarias regulares y estacionales, por lo que no es necesario diferenciar más por tendencia o estacionalidad, por lo que los modelos SARIMA que van ser planteados considerarán $d = D = 1$. En definitiva, cumpliéndose los supuestos de media constante, homocedasticidad y ergodicidad, se concluye que este proceso **sí es estacionario en sentido débil**.

Commented [SCR3]: En los análisis que se hacen tanto de la primera diferencia regular como de la primera diferencia estacional se menciona que no es necesario diferenciar más. ¿Eso se puede hacer, o este análisis se restringe a la última gráfica?

Finalmente, se tiene la **tabla 1** que resume todas las conclusiones argumentadas anteriormente.

Tabla 1. Resumen de los análisis realizados sobre cada una de las cuatro series. Con \checkmark se indica el cumplimiento del supuesto y con \times el no cumplimiento del supuesto.

Característica	Serie uno $\log Y_t$	Serie dos $\nabla \log Y_t$	Serie tres $\nabla_{12} \log Y_t$	Serie cuatro $\nabla \nabla_{12} \log Y_t$
Media constante	\times	\times	\times	\checkmark
Homocedasticidad	\checkmark	\checkmark	\times	\checkmark
Ergodicidad	\times	\times	\times	\checkmark
Ergodicidad en la tendencia	\times	\checkmark	\times	\checkmark
Ergodicidad en la estacionalidad	\times	\times	\checkmark	\checkmark
Estacionariedad en covarianza	\times	\times	\times	\checkmark

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:

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

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

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

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:

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

Así pues, el juego de hipótesis para las pruebas del test HEGY y sus interpretaciones son las presentadas en la segunda columna de la **tabla 2**. 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 las últimas tres columnas de 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 hay evidencia muestral suficiente para sugerir que el proceso asociado a la serie temporal de interés **presenta raíces unitarias regulares y estacionales**, a saber: cuatrimestral, semestral, con cinco ciclos al año y anual,

Commented [SCR4]: - ¿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?

con una significancia de $\alpha = 0.10$, lo cual es un indicador de que $\{\log Y_t\}_{t \in \mathbb{Z}^+}$ no es estacionario y es necesario aplicarle filtro de diferencia mixto, como el resultado en la **figura 2 (g)**, que como se analizó recientemente, al tomar la primera diferencia regular y la primera diferencia estacional, se llega a un proceso estacionario en covarianza.

Tabla 2. Juego de hipótesis para las pruebas del test HEGY y sus resultados calculados con el paquete estadístico **R**. En la última columna se usa “✓” para indicar que se rechaza la hipótesis nula.

Test	Hipótesis	Frecuencia angular (θ)	Ciclos por año (j)	Estadístico de prueba calculado	Valor p (V_p)	¿Rechazo? $\alpha = 0.10$
1	$\{H_0: \pi_1 = 0$ $H_1: \pi_1 < 0$	0	0	2.9135488	0.10000000	
2	$\{H_0: \pi_2 = 0$ $H_1: \pi_2 < 0$	π	6	-2.8854438	0.01000000	✓
3	$\{H_0: \pi_3 = \pi_4 = 0$ $H_1: \pi_3 < 0 \vee \pi_4 \neq 0$	$\pm \frac{1}{2}\pi$	3	0.8104567	0.10000000	
4	$\{H_0: \pi_5 = \pi_6 = 0$ $H_1: \pi_5 < 0 \vee \pi_6 \neq 0$	$\pm \frac{2}{3}\pi$	4	2.4693601	0.09062608	✓
5	$\{H_0: \pi_7 = \pi_8 = 0$ $H_1: \pi_7 < 0 \vee \pi_8 \neq 0$	$\pm \frac{1}{3}\pi$	2	1.7683136	0.10000000	
6	$\{H_0: \pi_9 = \pi_{10} = 0$ $H_1: \pi_9 < 0 \vee \pi_{10} \neq 0$	$\pm \frac{5}{6}\pi$	5	2.2239858	0.10000000	
7	$\{H_0: \pi_{11} = \pi_{12} = 0$ $H_1: \pi_{11} < 0 \vee \pi_{12} \neq 0$	$\pm \frac{1}{6}\pi$	1	1.6332640	0.10000000	

Commented [SCR5]: ¿Qué tratamiento se le debe dar a los valores p iguales a 0.1?

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 colombianos 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, cuya gráfica se presentó en la **figura 2 (h)** y con la PACF, que se presenta en la **figura 3**.

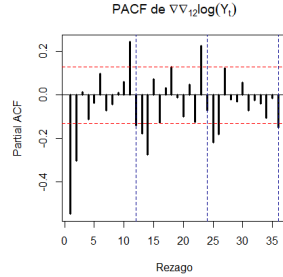


Figura 3. Función de autocorrelación parcial muestral del logaritmo del índice de ventas del sector manufacturero colombiano en pesos nominales tomando su primera diferencia regular y su primera diferencia estacional.

Así pues, comenzando con el análisis de la parte regular, se observa en la gráfica de la ACF que esta tiene un patrón tipo **cola**, mientras que en la PACF se identifica un patrón tipo corte, donde el último rezago estadísticamente diferente de cero es el segundo, por lo que para la parte regular se identifica un **AR**(2), pero teniendo presente que se realiza este análisis a partir de la serie diferenciada por tendencia y estacionalidad, entonces se tiene un **ARMA**(2, 1, 0). Después, pasando a la parte estacional, se puede observar en la función de autocorrelación parcial que los dos primeros rezagos estacionales son significativamente diferentes de cero, mientras que el primero no lo es, por lo que se puede pensar que la ACF tiene un patrón tipo corte con último rezago estacional significativo en $k = 24$, mientras que con la PACF se observa que esta es significativa para $k = 12, 36$ y que tiene un patrón cola, por lo que se debe cumplir que la parte estacional se pueda modelar como un **MA**(q)[s], donde $q = 24/_{12} = 2$, es decir, un **MA**(2)[12], pero teniendo presente la diferencia estacional empleada nuevamente, entonces se llega a un **ARMA**(0, 1, 2)[12], y al unir los análisis de la parte estacional y regular se concluye que un primer modelo adecuado es un **ARIMA**(2, 1, 0)(0, 1, 2)[12], coincidiendo entonces con el **modelo uno** propuesto en el enunciado, y cuya ecuación teórica está dada por:

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

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**.

```

> 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
s.e. -0.7143  0.1770  0.548  0.2990
      drift
s.e.  0.0049
s.e.  0.0059

sigma^2 = 0.001605: log likelihood = 423.53
AIC=-835.06 AICc=-834.69 BIC=-814.22

(a)

> auto.arima(lny,ic="bic",seasonal.test="ocsb")
Series: lny
ARIMA(1,1,1)(2,0,0)[12]

Coefficients:
      ar1      mal      sar1      sar2
s.e. -0.2556 -0.4531  0.5544  0.2918
      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

(d)

> 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
s.e. -0.7143  0.1770  0.548  0.2990
      drift
s.e.  0.0049
s.e.  0.0059

sigma^2 = 0.001605: log likelihood = 423.53
AIC=-835.06 AICc=-834.69 BIC=-814.22

(b)

> auto.arima(lny,ic="bic",seasonal.test="ch")
Series: lny
ARIMA(1,1,1)(2,0,0)[12]

Coefficients:
      ar1      mal      sar1      sar2
s.e. -0.2556 -0.4531  0.5544  0.2918
      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

(e)

> auto.arima(lny,ic="aic",seasonal.test="seas")
Series: lny
ARIMA(4,1,0)(1,1,2)[12]

Coefficients:
      ar1      ar2      ar3      ar4
s.e. -0.6849 -0.2928  0.0235 -0.0695
      0.0672  0.0013  0.0811  0.0665
      sar1      sma1      sma2
s.e. -0.6689 -0.0910 -0.7050
      0.2634  0.2475  0.1922

sigma^2 = 0.001141: log likelihood = 439.51
AIC=-863.02 AICc=-862.36 BIC=-835.65

(c)

> auto.arima(lny,ic="bic",seasonal.test="seas")
Series: lny
ARIMA(0,1,2)(0,1,1)[12]

Coefficients:
      mal      ma2      sma1
s.e. -0.7286  0.2237 -0.0661
      0.0632  0.0675  0.0572

sigma^2 = 0.001197: log likelihood = 432.59
AIC=-857.18 AICc=-857 BIC=-843.5

(f)

```

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

Así, se presenta en la **tabla 3** 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 los modelos no se da esta situación.

Tabla 3. Ecuaciones teóricas de los modelos identificados de forma automática con la función `auto.arima()`.

<p>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. $(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B) \log Y_t = \delta + (1 + \theta_1 B + \theta_2 B^2)E_t$, con E_t un R.B. $\sim N(0, \sigma^2)$</p>
<p>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. $(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B) \log Y_t = \delta + (1 + \theta_1 B + \theta_2 B^2)E_t$, con E_t un R.B. $\sim N(0, \sigma^2)$</p>
<p>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. $(1 - \Phi_1 B - \Phi_2 B^2 - \Phi_3 B^3 - \Phi_4 B^4)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12}) \log Y_t = (1 + \theta_1 B^{12} + \theta_2 B^{24})E_t$, con E_t un R.B. $\sim N(0, \sigma^2)$</p>
<p>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. $(1 - \Phi_1 B)(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B) \log Y_t = (1 + \theta_1 B)E_t$, con E_t un R.B. $\sim N(0, \sigma^2)$</p>
<p>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. $(1 - \Phi_1 B)(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B) \log Y_t = (1 + \theta_1 B)E_t$, con E_t un R.B. $\sim N(0, \sigma^2)$</p>
<p>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. $(1 - B)(1 - B^{12}) \log Y_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \theta_3 B^{12})E_t$, con E_t un R.B. $\sim N(0, \sigma^2)$</p>

De este modo, 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, hecho que también fue verificado al observar la ACF de la serie con transformación logarítmica y sin transformar en la **figura 2 (b)**, 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 bien, para el modelo **seis** se debe notar que para la parte regular sugiere un proceso de medias móviles, lo cual contradice lo identificado con la ACF y la PACF de $\nabla_{12} \log Y_t$, que en el sentido de que debería tratarse de un proceso autorregresivo, por lo que se descarta esta opción también, quedando únicamente la **tres**, que de hecho coincide con el modelo **dos** del enunciado y con el que se trabajará enseguida.

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** y otro tablero 18 × 18, presentado en la **figura 6**, y 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 $d = D = 1$.

Así, 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(2, 1, 10)(0, 1, 1)[12]** que solo incluye a los términos $\phi_1, \phi_2, \theta_{10}$ y θ_1 , como se puede apreciar en la **tabla 4**:

Commented [SCR6]: ¿Es necesario cambiar el valor de los coeficientes por los presentados en las salidas de **R**? ¿O se debe dejar así tal cual, indicado?

Commented [SCR7]: En este caso se tiene a la primera diferencia regular, pero no a la primera diferencia estacional. ¿Esto también lo explica la deriva en el caso de los modelos uno y dos, o debe descartarse por estar asociado con la tendencia y no con la estacionalidad?

Commented [SCR8]: ¿Correcto?: como se está trabajando con un tablero 12 x 12, entonces no se considera la parte estacional, y por tanto, no se considera tampoco la primera diferencia estacional.

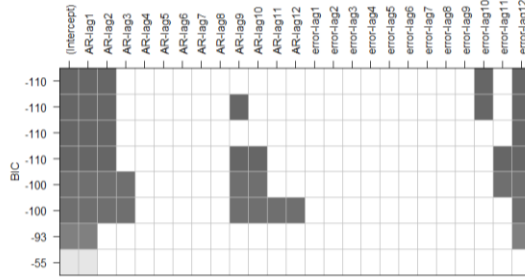


Figura 5. Gráfico 12 × 12 arrojado por *armasubsets* sobre $\nabla \nabla_{12} \log Y_t$ usando el método “ols”.

Tabla 4. Parámetros identificados a partir del primer renglón del tablero 12 × 12 arrojado por *armasubsets* sobre $\nabla \nabla_{12} \log Y_t$ usando el método “ols”.

Parte AR		Parte MA	
Celda sombreada	Parámetros identificados	Celda sombreada	Parámetros identificados
1	ϕ_1	10	θ_{10}
2	ϕ_2	12	θ_1

Así las cosas, la ecuación teórica del modelo identificado es la **ecuación 3**:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^{12}) \log Y_t = (1 + \theta_{10} B^{10})(1 + \theta_1 B^{12}) E_t, \text{ con } \{E_t\}_{t \in \mathbb{Z}^+} \text{ un R.B.} \sim N(0, \sigma^2) \quad [3]$$

Sin embargo, si se incorpora el término ϕ_6 , si bien continuaría siendo un *ARIMA*(2, 1, 10)(0, 1, 1)[12], la ecuación sería como se muestra en la **ecuación 4**:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_6 B^6)(1 - B)(1 - B^{12}) \log Y_t = (1 + \theta_{10} B^{10})(1 + \theta_1 B^{12}) E_t, \text{ con } \{E_t\}_{t \in \mathbb{Z}^+} \text{ un R.B.} \sim N(0, \sigma^2) \quad [4]$$

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 $\phi_1, \phi_2, \phi_9, \theta_{10}$ y θ_1 , tal y como se ilustra en la **tabla 5**.

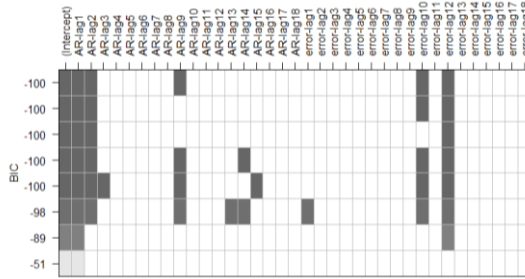


Figura 6. Gráfico 18 × 18 arrojado por *armasubsets* sobre $\nabla \nabla_{12} \log Y_t$ usando el método “ols”.

Tabla 4. Parámetros identificados a partir del primer renglón del tablero 12 × 12 arrojado por *armasubsets* sobre $\nabla \nabla_{12} \log Y_t$ usando el método “ols”.

Parte AR		Parte MA	
Celda sombreada	Parámetros identificados	Celda sombreada	Parámetros identificados
1	ϕ_1	10	θ_{10}
2	ϕ_2	12	θ_1
9	ϕ_9		

Y se llega entonces a la **ecuación 6**:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_9 B^9)(1 - B)(1 - B^{12}) \log Y_t = (1 + \theta_{10} B^{10})(1 + \theta_1 B^{12}) E_t, \quad E_t \text{ un R.B.} \sim N(0, \sigma^2) \quad [6]$$

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