

Modelos de componente versus Modelos ARIMA-SARIMA

Katerine Marín¹, Julián Mira², Angélica Restrepo³, Joan Rincón⁴

Fecha de entrega: 25-05 – 2016

Resumen

En este trabajo se continúa el análisis de la serie ingresos del sector de suministro de electricidad, gas y agua en miles de millones de pesos, entre el periodo de enero del año 2000 hasta septiembre de 2015 consultada en el departamento administrativo nacional de estadística (DANE), se tratará de establecer definitivamente el mejor modelo estadístico para el ajuste y pronóstico de la misma. En esta ocasión se identifican los modelos SARIMA con los cuales se ajustan y se pronostica la tendencia y la estacionalidad. Para ello se realizará la prueba de test HEGY y un análisis gráfico para identificar si es necesario diferenciar la serie y así convertirla en una serie estacionaria, debido a que la serie es estacional y se verificará por medio del software estadístico R si estos modelos son apropiados para pronosticar y si se cumplen los supuestos de normalidad, independencia, varianza constante.

Palabras claves: Raíces unitarias, SARIMA, estacionariedad, Test HEGY

INTRODUCCION ;

¹Estudiante Ingeniería Administrativa, Universidad Nacional de Colombia – Sede Medellín

²Estudiante Ingeniería Industrial, Universidad Nacional de Colombia – Sede Medellín

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

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

2. Análisis descriptivo y test HEGY de raíces unitarias estacionales

2.a Análisis descriptivo

La serie de suministro de electricidad, gas y agua es una serie que tiene patrón de tendencia creciente, patrón estacional definido y además es de componentes aditivas, pues su varianza es aproximadamente constante en el tiempo a lo largo de la tendencia, la serie también presenta patrones cíclicos que son la componente estocástica vista en los trabajos anteriores.

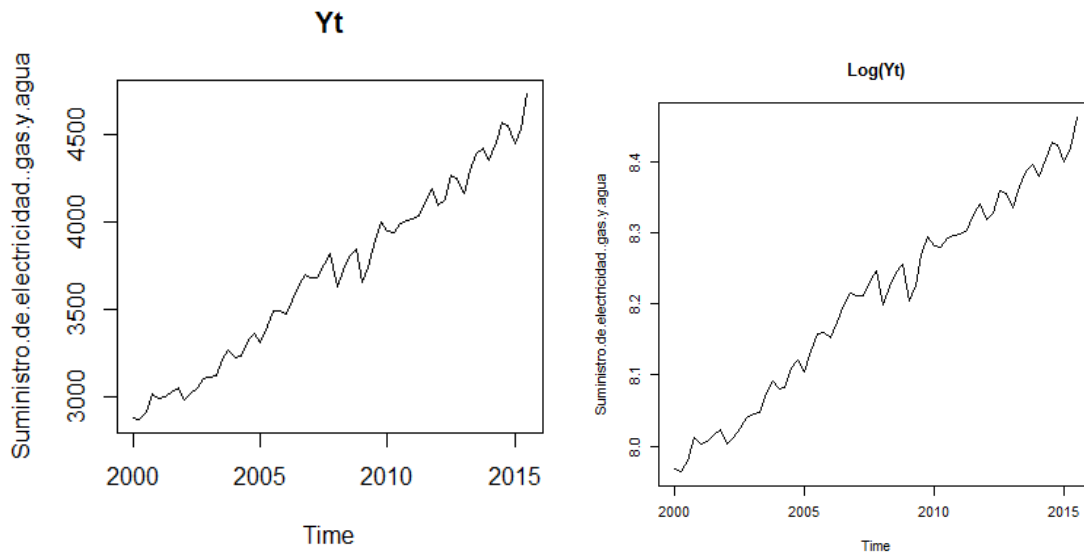


Figura 1: serie de tiempo (izquierda) y logaritmo de la serie de tiempo (derecha)

En la figura 1 se puede notar que tanto la serie como su logaritmo con $N=63$ presentan similitud en cuanto a su varianza por cual se concluyo que las componentes de la serie se comportan de manera aditiva.

En el patrón de tendencia lo se observa que es creciente y que no se puede modelar mediante una curva suave, por lo cual es estocástica; en cuanto al patrón estacional no es constante en el tiempo por lo cual se decidió modelarlo por medio de funciones trigonométricas (y daba mejores resultados que con indicadores) en trabajos anteriores.

También se observan los ciclos y los cambios de nivel como pequeños saltos que tiene la serie en ciertos periodos de tiempo, especialmente entre los años 2008 a 2010.

Desde este punto se puede afirmar que la serie de tiempo no es estacionaria en covarianza debido a que la media de la serie depende de una tendencia y una estacionalidad descritos anteriormente para re afirmarlo se presenta en la figura 2, la gráfica de la serie recortada y su ACF para comprobar la ergodicidad y por lo definido anteriormente en esta grafica posiblemente se observarían los patrones de tendencia y estacionalidad.

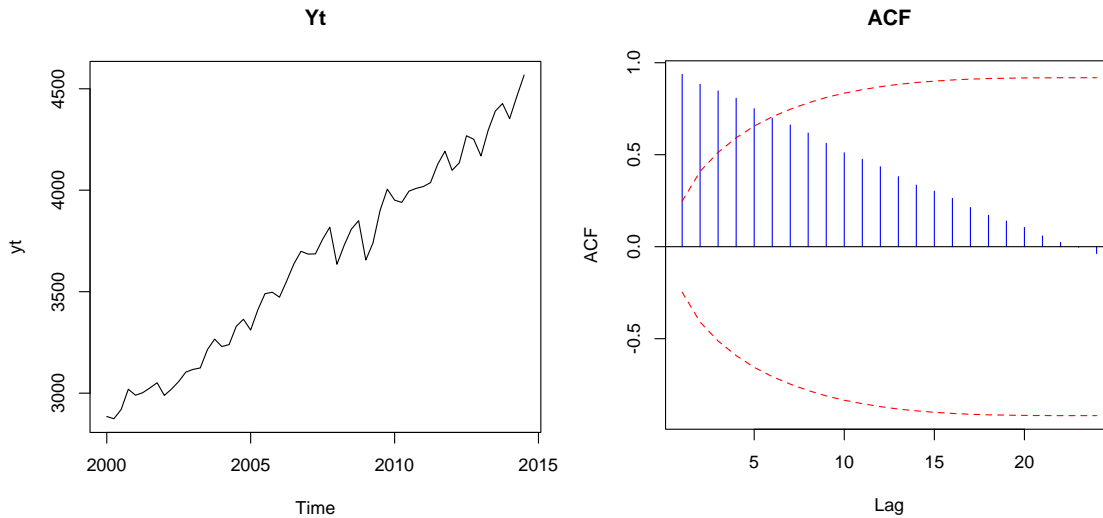


Figura 2: la serie de tiempo recortada junto con su ACF.

De los resultados de la figura 2 se puede ver que la serie no es estacionaria en su ACF porque no es ergódica, debido al patrón de tendencia observado en la figura 1 la cual hace que la ACF caiga lentamente tal como se había dicho anteriormente y por esta razón se analizara la diferencia de forma regular, el patrón estacional no se observa en $k=4,8,12,16,20,24$ y si está presente este estaría cayendo rápidamente aunque se sabe que existe un patrón estacional estocástico por el análisis hecho en trabajos anteriores y a pesar de este comportamiento se estudiara la diferencia estacional. Se analiza cada diferencia por separado, en la figura 3 se presenta la primera diferencia regular con su ACF. Y en la figura 4 se presenta la primera diferencia estacional con su ACF.

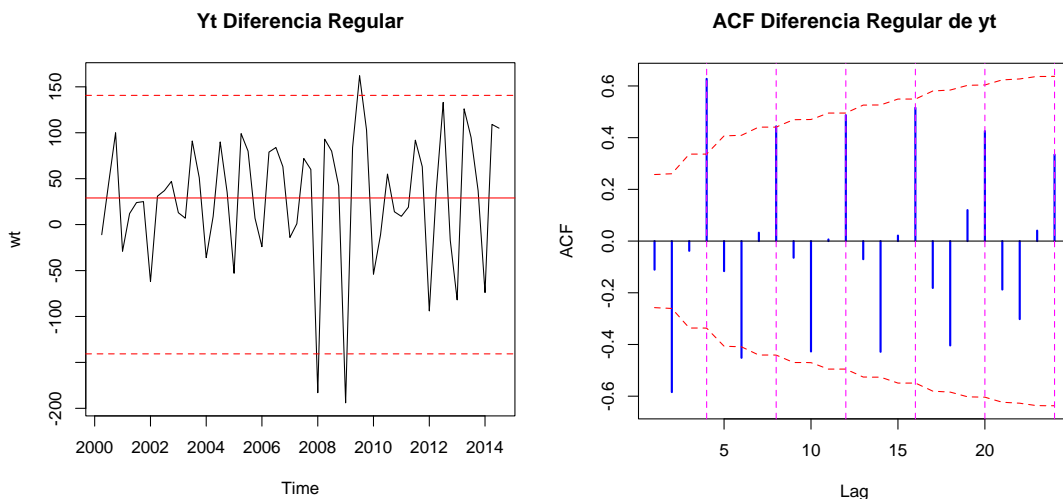


Figura 3: Diferencia regular y su ACF.

$$\nabla Y_t = Y_t - Y_{t-1}$$

De los resultados de la figura 3 se puede ver que al aplicar la diferencia regular la tendencia desaparece quedando un patrón estacional de la serie que se repite año tras año y se observa que la media es aproximadamente constante pero depende del patrón estacional, respecto a la varianza se puede decir que es aproximadamente constante aunque presenta algunos puntos atípicos, en la gráfica de la izquierda se observa que hay un patrón periódico, aunque ya no exista tendencia si existe un patrón estacional y en la ACF se observa que cada múltiplo de s es decir 4,8,12 etc... no hay ergodicidad y es debido al patrón estacional, por tanto este proceso obtenido al aplicar la primera diferencia regular no es estacionario y debe aplicarse la diferencia estacional con $D=1$. Vamos a proceder a examinar la diferencia estacional en la figura 4.

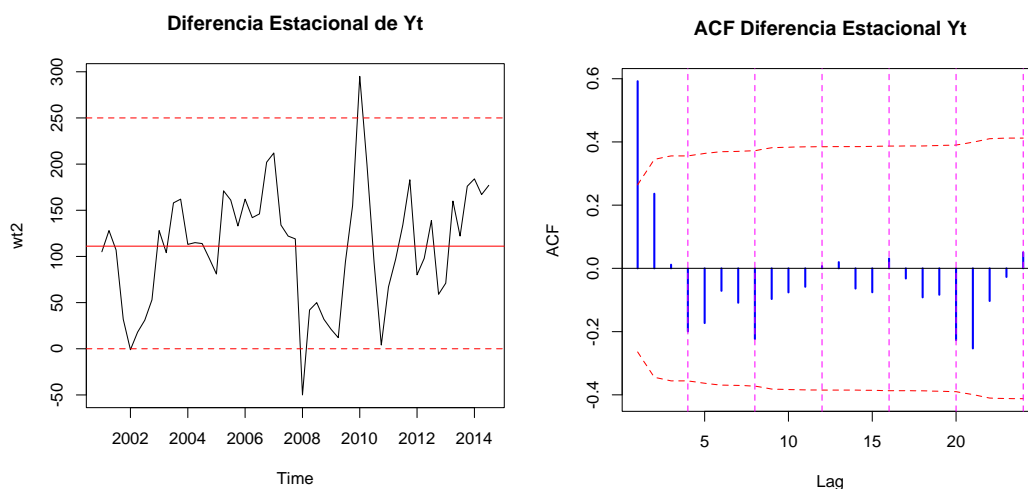


Figura 4: Differencia estacional y su ACF.

$$\nabla_4 Y_t = Y_t - Y_{t-4}$$

De los resultados de la figura 4 se puede ver en la gráfica izquierda que la serie aún tiene un patrón de tendencia, debido a que se presentan cambios de nivel de la media, esto estaría diciendo que aún podría no ser estacionaria y además deja una fuerte autocorrelación entre valores consecutivos, lo que se refleja en el patrón cíclico y como se observa en su ACF se puede ver un comportamiento ergodico en la parte estacional, y debido a que la media depende de una tendencia esta no es constante, además la varianza varia en el tiempo, aunque estas correlaciones decaen rápidamente y parece ser ergodico no es estacionario y debido a que no tiene nivel constante se necesita diferenciar regularmente con $d=1$.

Por lo tanto se necesita diferenciar de forma regular como estacional.

En la figura 5 se aplica la diferencia combinada para la serie junto con su ACF y PACF.

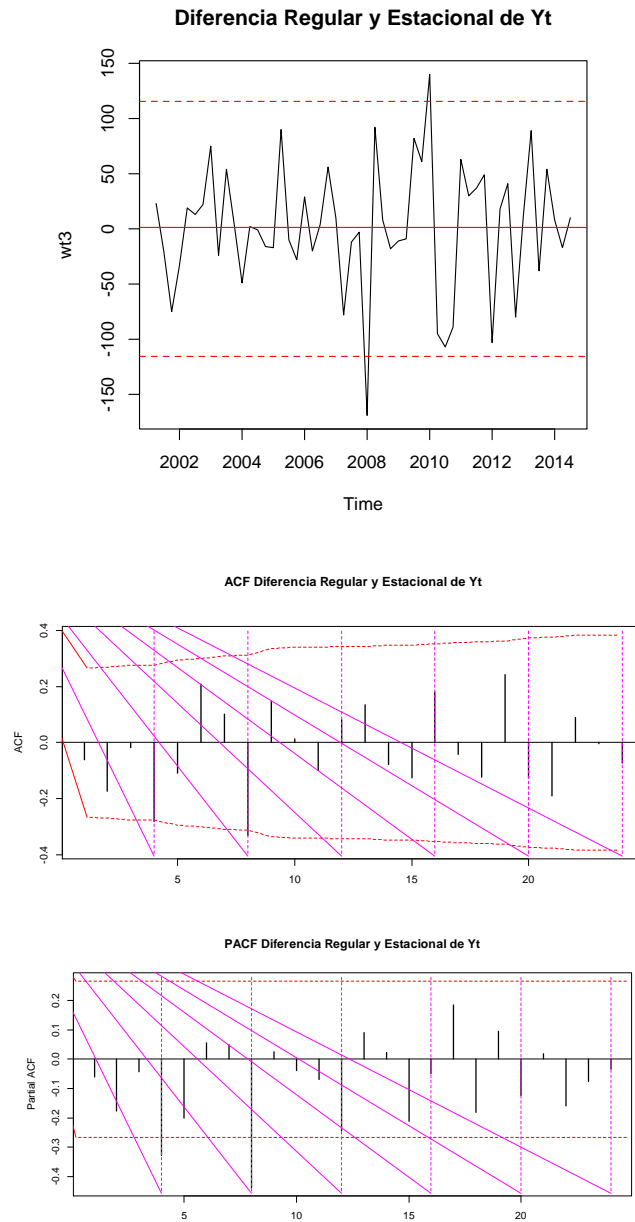


Figura 5: Gráfica de la primera diferencia regular y estacional y su ACF y PACF.

$$\nabla \nabla_4 Y_t = Y_t - Y_{t-1} - Y_{t-4} + Y_{t-5}$$

De los resultados de la figura 5 si se ve claramente que la serie tiene una media constante, varianza aproximadamente constante aunque muestra algunos puntos atípicos, no se observan patrones periódicos exactos, además en la función ACF ya se ve fácilmente que si es proceso estacionario porque decae rápidamente tanto en la parte regular como estacional en esta serie diferenciada de ambas formas, se tiene certeza de que es un proceso estacionario en covarianza. Además en la ACF se observa corte claro justamente en $k=4$ y $k=8$ múltiplos de s y en la PACF también se observan cortes y especialmente en múltiplos de s en $k=4$ y $k=8$, y se observa una estructura MA. La identificación de estos modelos por este método se presentará en la sección posterior.

2.b Test HEGY de raíces unitarias estacionales

Este test trata de identificar si hay raíces unitarias en un proceso ARMA en la parte de su polinomio AR y si tendrá una o más raíces unitarias de $1-X^s=0$.

Si C_k es una raíz del polinomio $1-X^s$ tal que $\varphi(k)=0$ Y_t tiene raíz unitaria y el objetivo del test es determinar si es necesario diferenciar o no la serie y el tipo de diferenciación para identificar un proceso estacionario.

En el test se parte de la representación de Y_t como un AR (∞) es decir, un proceso invertible.

Se plantea el siguiente modelo:

$$\nabla_4(Y_t) = \pi_1 X_{1,t-1} + \pi_2 X_{2,t-1} + \pi_4 X_{3,t-1} + \pi_3 X_{3,t-2} + \sum_{i=1}^{p-1} b_i \nabla_4(Y_{t-i}) + E_t$$

$$\text{con } E_t \text{ un } RB \sim (0, \sigma^2)$$

Con este modelo se evalúa si la serie de tiempo tiene raíces unitarias estacionales y plantea una hipótesis que tiene asociada una frecuencia angular y periodo de la raíz unitaria a examinar. Las pruebas de hipótesis se plantean en el test a continuación.

Tabla 1: Test de HEGY de raíces unitarias estacionales

Test	Hipótesis a probar	Estadístico	Valor p	Frecuencia	Periodo
1	$H_0: \pi_1 = 0 \text{ vs } H_1: \pi_1 < 0$	1.968	0.1	0	0
2	$H_0: \pi_2 = 0 \text{ vs } H_1: \pi_2 < 0$	-0.7389	0.1	π	Semestral
3	$H_0: \pi_3 = \pi_4 = 0 \text{ vs } H_1: \pi_3 < 0 \text{ ó } \pi_4 \neq 0$	0.0264	0.1	$\pm \frac{\pi}{2}$	Anual

Con los resultados de los estadísticos pruebas, valores p de la tabla 1 con sus frecuencias y periodos se hacen los siguientes análisis:

- **Por el test 1:** se observa que no se rechaza la hipótesis nula de este test puesto que el valor p es de 0.1 es alto y con ello se concluye que hay raíz unitaria regular o no estacional en la frecuencia angular 0, con periodo 0, ósea $\varphi(1) = 0$ este test recomienda que se debe diferenciar al menos una vez la serie de tiempo de forma regular con $d=1$.
- **Por el test 2:** se observa que no se rechaza la hipótesis nula de este test puesto que el valor p es de 0.1 es alto y con ello se concluye que hay raíz unitaria estacional semestral en la frecuencia angular π , ósea $\varphi(-1)=0$ y este test recomienda que se debe diferenciar al menos una vez la serie de tiempo de forma estacional con $D=1$.
- **Por el test 3:** se observa que no se rechaza la hipótesis nula de este test puesto que el valor p es de 0.1 es alto y con ello se concluye que hay raíz unitaria estacional anual en la frecuencia angular $\pm \frac{\pi}{2}$, ósea $|\varphi(i)| = 0$ y según este test recomienda que se debe diferenciar al menos una vez la serie de tiempo de forma estacional. Con esto y lo anterior

se recomienda finalmente que se debería diferenciar de forma estacional con $D=1$ y de forma regular con $d=1$.

En conclusión según el test de HEGY encuentra raíz unitaria regular y estacional y se recomienda diferenciar regular y estacionalmente la serie de tiempo, es algo que también se analizó y aplicó en el inciso a) según análisis de las ACF's y las graficas de las respectivas diferencias a la serie, y por tanto se prefiere el proceso $\nabla_4 \nabla(Y_t)$ ya que con seguridad este proceso si es estacionario en covarianza.

3. Postulación de modelos ARIMA estacionales

3.a Identificación de modelos preliminares

Para la serie diferenciada de forma regular y estacional $\nabla_4 \nabla Y_t$ con las gráficas de ACF y PACF de la figura 5.

NOTA: la media de la $\nabla_4 \nabla Y_t$ es diferente a cero por esto se incluirá la deriva para Y_t en los modelos identificados por ACF-PACF y por `ARMASUBSETS()`.

Para observar la parte regular en la ACF solo se debe observar hasta $k=\frac{s}{2}$, en donde se observa un patrón de corte que termina en $k=0$, y para observar la parte regular en la PACF, también se observa hasta $k=\frac{s}{2}$, en donde se analiza un patrón de corte que termina en $k=0$; por lo tanto el orden p y q , de la parte regular son iguales a 0, luego se obtiene un $ARIMA(0,1,0)$ para Y_t .

Para observar la parte estacional en la ACF solo se debe mirar cada $k=m*4$ con $m=1, \dots, 6$. En donde se puede ver claramente un patrón corte, que termina en $K=8$ lo que quiere decir en $m=2$. Y para observar la parte estacional en la PACF solo se debe mirar cada $k=m*4$ con $m=1, \dots, 6$. En donde se puede ver claramente un patrón cola negativa que decrece cada s , por lo tanto el orden $P=0$ y $Q=2$ de la parte estacional.

Se concluye que por el comportamiento de la parte regular y estacional de las ACF y PACF se obtiene un modelo:

$$Y_t \sim ARIMA(0,1,0)(0,1,2)_4 \text{ con } \delta \text{ deriva}$$

La ecuación de este modelo es: $\nabla \nabla_4 Y_t = \delta + \theta_2(B^4) E_t$, $E_t \text{ un RB } \sim N(0, \sigma^2)$
 $Y_t = \delta + Y_{t-1} + Y_{t-4} - Y_{t-5} + E_t + \theta_1 E_{t-4} + \theta_2 E_{t-8}$, $E_t \text{ un RB } \sim N(0, \sigma^2)$

Aplicando la función `auto.arima()` variando tanto el argumento AIC como BIC y cambiando los argumentos de `seasonal.test` se obtiene los siguientes resultados extraídos de R para la identificación de cada modelo, estos se presentan a continuación:

```

> auto.arima(yt,ic="aic",seasonal.test="ocsb")
Series: yt
ARIMA(1,0,0)(2,1,1)[4] with drift

Coefficients:
      ar1      sar1      sar2      sma1      drift
    0.6514  0.1553  -0.2805  -0.7361  27.9280
s.e.  0.1128  0.1851  0.1513  0.1797  1.2003

sigma^2 estimated as 1896: log likelihood=-285.19
AIC=582.38  AICc=584.13  BIC=594.42

> auto.arima(yt,ic="aic",seasonal.test="ch")
Series: yt
ARIMA(0,1,0)(1,0,0)[4] with drift

Coefficients:
      sar1      drift
    0.6602  30.3195
s.e.  0.0956  17.8536

sigma^2 estimated as 2802: log likelihood=-312.63
AIC=631.26  AICc=631.7  BIC=637.44
\ |

> auto.arima(yt,ic="bic",seasonal.test="ocsb")
Series: yt
ARIMA(1,0,0)(0,1,1)[4] with drift

Coefficients:
      ar1      sma1      drift
    0.6598  -0.7592  27.7507
s.e.  0.1093  0.1304  1.3213

sigma^2 estimated as 2030: log likelihood=-287.77
AIC=583.54  AICc=584.34  BIC=591.57

> auto.arima(yt,ic="bic",seasonal.test="ch")
Series: yt
ARIMA(0,1,0)(1,0,0)[4]

Coefficients:
      sar1
    0.7111
s.e.  0.0889

sigma^2 estimated as 2848: log likelihood=-313.88
AIC=631.76  AICc=631.97  BIC=635.88
. |

```

Figura 6: Resultados R de la función auto.arima

Con los resultados de R se identifican cuatro modelos diferentes, estos se presentan en el orden que fueron mostrados los resultados. Primero con el criterio del AIC variando el de seasonal y luego con el criterio del BIC.

El primer modelo identificado es:

$$Y_t \sim ARIMA(1,0,0)(2,1,1)_4 \text{ Con } \delta \text{ deriva}$$

La ecuación de este modelo es: $\varphi_1(B)\Phi_2(B^4)\nabla_4 Y_t = \delta + \Theta_1(B^4)E_t$ con E_t un $RB \sim (0, \sigma^2)$

$$Y_t = \delta + \varphi_1 Y_{t-1} + (1 + \Phi_1)Y_{t-4} - (\varphi_1 + \varphi_1 \Phi_1)Y_{t-5} + (\Phi_2 - \Phi_1)Y_{t-8} + (\varphi_1 \Phi_1 - \varphi_1 \Phi_2)Y_{t-9} \\ - \Phi_2 Y_{t-12} + \varphi_1 \Phi_2 Y_{t-13} + E_t + \Theta_1 E_{t-4} \quad \text{con } E_t \text{ un } RB \sim (0, \sigma^2)$$

El segundo modelo identificado es:

$$Y_t \sim ARIMA(0,1,0)(1,0,0)_4 \text{ con } \delta \text{ deriva}$$

La ecuación de este modelo es: $\Phi_1(B^4)\nabla Y_t = \delta + E_t$ con E_t un $RB \sim (0, \sigma^2)$

$$Y_t = \delta + Y_{t-1} + \Phi_1 Y_{t-4} - \Phi_1 Y_{t-5} + E_t \quad \text{con } E_t \text{ un } RB \sim (0, \sigma^2)$$

El tercer modelo identificado es:

$$Y_t \sim ARIMA(1,0,0)(0,1,1)_4 \text{ con } \delta \text{ deriva}$$

La ecuación de este modelo es: $\varphi_1(B)\nabla_4 Y_t = \delta + \Theta_1(B^4)E_t$ con E_t un $RB \sim (0, \sigma^2)$

$$Y_t = \delta + \varphi_1 Y_{t-1} + Y_{t-4} - \varphi_1 Y_{t-5} + E_t + \Theta_1 E_{t-4} \quad \text{con } E_t \text{ un } RB \sim (0, \sigma^2)$$

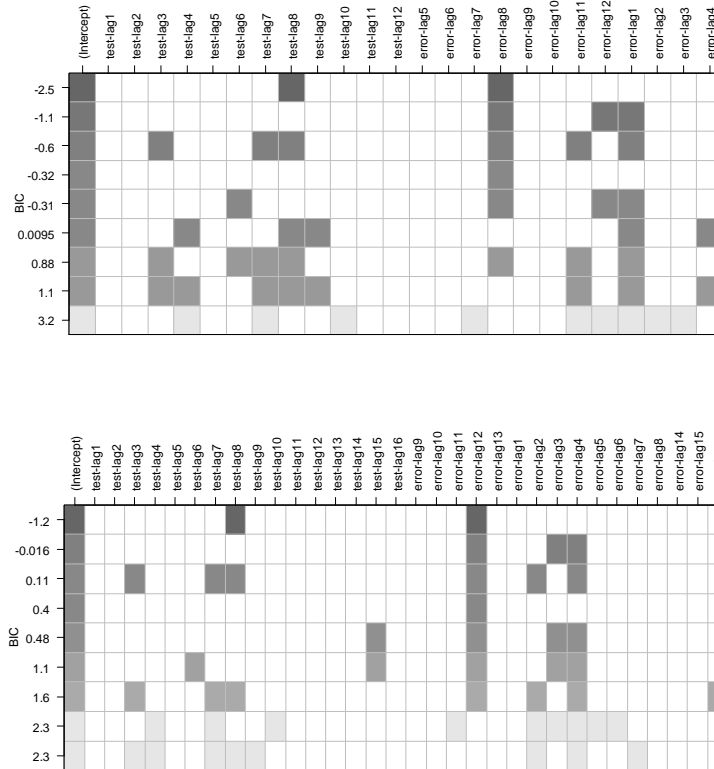
El cuarto modelo identificado es:

$$Y_t \sim ARIMA(0,1,0)(1,0,0)_4 \text{ sin } \delta \text{ deriva}$$

La ecuación de este modelo es: $\Phi_1(B^4)\nabla Y_t = E_t$ con E_t un $RB \sim (0, \sigma^2)$

$$Y_t = Y_{t-1} + \Phi_1 Y_{t-4} - \Phi_1 Y_{t-5} + E_t \quad \text{con } E_t \text{ un } RB \sim (0, \sigma^2)$$

Con la función `armasubsets` se identifican modelos preliminares que puedan examinarse en el supuesto de ruido blanco posteriormente o sino ser modificados por tanto se identifican los mejores de acuerdo al criterio del BIC. Esto se hace para una longitud de la parte AR y MA de 12, 16, y 20 las tres gráficas con estas longitudes se presentan a continuación en la figura 6.



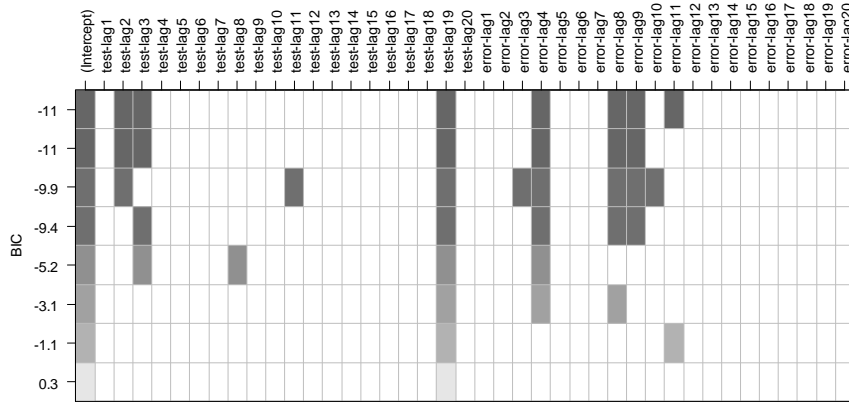


Figura 7: Armasubsets para nar y nma =12 , 16 y 20 respectivamente

De acuerdo a las gráficas de la figura 7 se identifican los siguientes modelos:

De la figura 7 arriba, con longitud 12 el modelo identificado con el mejor BIC eligiendo el primer renglón por tener parámetros en la parte estacional y porque no hay mucha diferencia en el BIC es:

$$\Phi_2(B^4)\nabla\nabla 4Y_t = \Theta_2(B^4)E_t, \text{ con } E_t \text{ un } RB \sim (0, \sigma^2) \text{ con deriva y solo con los parametros } \Phi_2, \Theta_2$$

La ecuación de este modelo es:

$$Y_t = \delta + Y_{t-1} + Y_{t-4} - Y_{t-5} + \Phi_2 Y_{t-8} - \Phi_2 Y_{t-9} - \Phi_2 Y_{t-12} + \Phi_2 Y_{t-13} + E_t \\ + \Theta_2 E_{t-8} \text{ con } E_t \text{ un } RB \sim (0, \sigma^2)$$

De la figura 7 medio, con longitud 16 el modelo identificado con el mejor BIC eligiendo el primer renglón por tener parámetros en la parte estacional y porque no hay mucha diferencia en el BIC es:

$$\Phi_2(B^4)\nabla\nabla 4 Y_t = \Theta_3(B^4)E_t, \text{ con } E_t \text{ un } RB \sim (0, \sigma^2) \text{ con deriva y solo con los parametros } \Phi_2, \Theta_3$$

La ecuación de este modelo es:

$$Y_t = \delta + Y_{t-1} + Y_{t-4} - Y_{t-5} + \Phi_2 Y_{t-8} - \Phi_2 Y_{t-9} - \Phi_2 Y_{t-12} + \Phi_2 Y_{t-13} + E_t \\ + \Theta_3 E_{t-12} \text{ con } E_t \text{ un } RB \sim (0, \sigma^2)$$

De la figura 7 abajo, con longitud 20 el modelo identificado con el mejor BIC eligiendo el primer renglón por tener parámetros en la parte estacional y porque no hay mucha diferencia en el BIC es:

$$\varphi_3(B)\Phi_4(B^4)\nabla\nabla 4Y_t = \theta_3(B)\Theta_2(B^4)E_t, \text{ con } E_t \text{ un } RB \sim (0, \sigma^2) \text{ con deriva y solo con los parámetros } \varphi_2, \varphi_3, \Phi_4, \theta_1, \theta_3, \Theta_1 \text{ y } \Theta_2$$

$$\begin{aligned}
Y_t = & \delta + Y_{t-1} + \varphi_2 Y_{t-2} + (\varphi_3 - \varphi_2) Y_{t-3} + (1 - \varphi_3) Y_{t-4} - Y_{t-5} \\
& - \varphi_2 Y_{t-6} + (\varphi_2 - \varphi_3) Y_{t-7} + \varphi_3 Y_{t-8} + \Phi_4 Y_{t-16} - \Phi_4 Y_{t-17} - \varphi_2 \Phi_4 Y_{t-18} \\
& + (\varphi_2 \Phi_4 - \varphi_3 \Phi_4) Y_{t-19} + (\varphi_3 \Phi_4 - \Phi_4) Y_{t-20} + \Phi_4 Y_{t-21} + \varphi_2 \Phi_4 Y_{t-22} \\
& + (\varphi_3 \Phi_4 - \varphi_2 \Phi_4) Y_{t-23} - \varphi_3 \Phi_4 Y_{t-24} + E_t + \theta_1 E_{t-1} + \theta_3 E_{t-3} + \theta_1 E_{t-4} \\
& + \theta_1 E_{t-5} + \theta_1 \theta_1 E_{t-7} + \theta_1 \theta_2 E_{t-9} + \theta_3 \theta_2 E_{t-11} \quad \text{con } E_t \text{ un } RB \sim (0, \sigma^2)
\end{aligned}$$

3.b Modelos finales propuestos

(ver tabla de la página siguiente)

El modelo 1 será el mejor SARIMA estacional identificado por armasubsets y es el **ARIMA(3,1,3)(4,1,2)_[4]** con δ deriva y solo con $\varphi_2, \varphi_3, \Phi_4, \theta_1, \theta_3, \theta_1$ y θ_2 debido a que es el único que cumple el ruido blanco sobre los errores y se puede apreciar que su error de ajuste según AIC y BIC es muy pequeño y su error de pronóstico según los criterios observados en la tabla también son bajos con respecto a los otros modelos pero es el elegido principalmente por cumplir con el supuesto de ruido blanco. Cuya ecuación se muestra a continuación:

$$\begin{aligned}
Y_t = & \delta + Y_{t-1} + \varphi_2 Y_{t-2} + (\varphi_3 - \varphi_2) Y_{t-3} + (1 - \varphi_3) Y_{t-4} - Y_{t-5} \\
& - \varphi_2 Y_{t-6} + (\varphi_2 - \varphi_3) Y_{t-7} + \varphi_3 Y_{t-8} + \Phi_4 Y_{t-16} - \Phi_4 Y_{t-17} - \varphi_2 \Phi_4 Y_{t-18} \\
& + (\varphi_2 \Phi_4 - \varphi_3 \Phi_4) Y_{t-19} + (\varphi_3 \Phi_4 - \Phi_4) Y_{t-20} + \Phi_4 Y_{t-21} + \varphi_2 \Phi_4 Y_{t-22} \\
& + (\varphi_3 \Phi_4 - \varphi_2 \Phi_4) Y_{t-23} - \varphi_3 \Phi_4 Y_{t-24} + E_t + \theta_1 E_{t-1} + \theta_3 E_{t-3} + \theta_1 E_{t-4} \\
& + \theta_1 E_{t-5} + \theta_1 \theta_1 E_{t-7} + \theta_1 \theta_2 E_{t-9} + \theta_3 \theta_2 E_{t-11} \quad \text{con } E_t \text{ un } RB \sim (0, \sigma^2)
\end{aligned}$$

El modelo 2 será el mejor ARIMA estacional entre los identificados con ACF-PACF y auto.arima y como se puede observar en la tabla solo compararemos las medidas de ajuste y pronóstico con los que cumplan con el supuesto de ruido blanco, y como se puede ver en calidad de ajuste elegiremos el más parsimonioso que nos lo dará según BIC y en calidad de pronóstico es claro que hay uno que se destaca frente a los demás por cometer el menor error en casi todas las pruebas y por esto el modelo elegido será **ARIMA(1,0,0)(0,1,1)_[4]**

$$Y_t = \delta + \varphi_1 Y_{t-1} + Y_{t-4} - \varphi_1 Y_{t-5} + E_t + \theta_1 E_{t-4} \quad \text{con } E_t \text{ un } RB \sim (0, \sigma^2)$$

4. Ajuste de modelos con validación cruzada

Para todos los modelos se escogerá la cantidad de datos necesaria para modelar su comportamiento y los demás datos se usaran para comparar los pronósticos del modelo ajustado con los reales.

Los datos seleccionados a pronosticar serán los últimos 4 trimestres de la serie que empezaría en el cuarto trimestre de 2014 y finalizaría en el tercer trimestre de 2015, es decir se toma un año Para pronosticar, con lo que se concluye que la nueva serie con la que se ajustarán los modelos estará conformada por 59 datos suministrados trimestralmente, empezando en primer trimestre (Enero, Febrero y Marzo) de 2000 y finalizando en tercer trimestre (Julio, Agosto y Septiembre) de 2014. Se usan 4 datos para pronosticar ya que se podría incurrir en un mayor error de pronóstico si toma una muestra más grande para el pronóstico. Se considera que un año es suficiente para poder analizar la capacidad de pronóstico de los modelos SARIMA.

- N: 63 Cantidad total de datos de la serie.
- m: 4 Cantidad de datos de la muestra.
- n: 59 Cantidad de datos para ajustar de los modelos.

4.1 Ajuste Modelo 1: SARIMA (3,1,3) (4,1,2) [4] escogido por Armasubsets

A continuación en la tabla 2 se muestran los parámetros estimados del modelo 1 y además se muestra su ecuación de ajuste

Tabla 2 : Parámetros estimados del modelo 1 (con $Z \sim N(0,1)$)

Parámetro	Estimación	Error Estándar	Valor Z	$P(> Z)$
ϕ_2	-0,2950721	0,0767632	-3,8439	0,0001211
ϕ_3	0,7876478	0,0599702	13,1340	$< 2,2 * 10^{-16}$
θ_1	-0,1398483	0,1268124	-1,1028	0,2701156
θ_3	-1,0268418	0,1482096	-6,9283	$4,26 * 10^{-12}$
Φ_4	-0,0073899	0,1630383	-0,0453	0.9638472
Θ_1	-0,6123549	0,1641537	-3,7304	0,001912
Θ_2	-0,1599894	0,1575963	-1,0152	0,3100177
δ	98,5757189	NA	NA	NA

Ecuación ajustada:

$$\begin{aligned} \hat{Y}_t = & 98,5757189 + Y_{t-1} - 0.2950721 Y_{t-2} + 1.0827199 Y_{t-3} + 0.2123522 Y_{t-4} - Y_{t-5} \\ & - 0.2950721 Y_{t-6} - 1.0827199 Y_{t-7} + 0.7876478 Y_{t-8} - 0.0073899 Y_{t-16} \\ & + 0.0073899 Y_{t-17} - 0.00218055 Y_{t-18} + 0.008001192 Y_{t-19} \\ & + 0.001569262 Y_{t-20} - 0.0073899 Y_{t-21} + 0.002180553 Y_{t-22} \\ & - 0.008001192 Y_{t-23} + -0.005820638 Y_{t-24} \\ & - 0.1398483 \hat{E}_{t-1} - 1.0268418 \hat{E}_{t-3} - 0.6123549 \hat{E}_{t-4} - 0.6123549 \hat{E}_{t-5} \\ & + 0.085636792 \hat{E}_{t-7} + 0.022374245 \hat{E}_{t-9} + 0.164283803 \hat{E}_{t-11} \end{aligned}$$

Significancia de los parámetros:

Se procede a evaluar la significancia de los parámetros del Modelo 1:

SARIMA (3,1,3) (4,1,2) [4], solo interesa evaluar significancia en los parámetros de mayor orden asociados a la parte AR y MA tanto de la parte regular como de la parte estacional, en este caso los parámetros a evaluar son ϕ_3, θ_3, Φ_4 y Θ_2 . En ciertos casos, el parámetro de mayor orden no es significativo pero contribuye a que los errores del modelo sean ruido blanco, en estos casos no se recomienda quitar dicho parámetro. Para evaluar la significancia de dichos parámetros se usa el siguiente test de hipótesis:

Test de Hipótesis (1):

$H_0: \phi_3 = 0,$	$No\ es\ significativo$	vs	$H_a: \phi_3 \neq 0,$	$Es\ significativo$
$H_0: \theta_3 = 0,$	$No\ es\ significativo$	vs	$H_a: \theta_3 \neq 0,$	$Es\ significativo$
$H_0: \Phi_4 = 0,$	$No\ es\ significativo$	vs	$H_a: \Phi_4 \neq 0,$	$Es\ significativo$
$H_0: \Theta_2 = 0,$	$No\ es\ significativo$	vs	$H_a: \Theta_2 \neq 0,$	$Es\ significativo$

Estadístico de prueba

$$t_o = \frac{\widehat{\phi}_i}{SE(\widehat{\phi}_i)} \underset{\sim}{\sim}^{bajo\ H_0} t_{(n-p)} \quad t_o = \frac{\widehat{\theta}_i}{SE(\widehat{\theta}_i)} \underset{\sim}{\sim}^{bajo\ H_0} t_{(n-p)}$$

$$t_o = \frac{\widehat{\Phi}_i}{SE(\widehat{\Phi}_i)} \underset{\sim}{\sim}^{bajo\ H_0} t_{(n-p)} \quad t_o = \frac{\widehat{\Theta}_i}{SE(\widehat{\Theta}_i)} \underset{\sim}{\sim}^{bajo\ H_0} t_{(n-p)}$$

Criterios de rechazo de la hipótesis nula

Rechazo si $V_p = P(|t_{(n-p)}| > |t_0|)$ es pequeño

Haciendo un análisis para el Modelo 1 se observa que el parámetro asociado a la parte AR regular, ϕ_3 , y el parámetro asociado a la parte MA regular, θ_3 , son significativos; los otros dos parámetros, Φ_4 y Θ_2 , asociados a la parte AR estacional y MA estacional respectivamente son no significativos

4.2 Ajuste Modelo 2: ARIMA (1,0,0) (0,1,1) [4] de auto.arima según BIC y test ocsb

A continuación en la tabla 3 se muestran los parámetros estimados del modelo 2 y además se muestra su ecuación de ajuste

Tabla 3 : Parámetros estimados del modelo 2 (con $Z \sim N(0,1)$)

Parámetros	Estimación	Error Estándar	Valor Z	$P(> Z)$
ϕ_1	0,65979	0,10934	6,0342	$1,598 \cdot 10^{-9}$
Θ_1	-0,75919	0,13036	-5,8236	$5,759 \cdot 10^{-9}$
δ	27,75071	1,32133	21,002	$< 2.210^{-16}$

La ecuación para este modelo es:

$$\hat{Y}_t = 27,75071 + 0,65979Y_{t-1} + Y_{t-4} - 0,65979Y_{t-5} - 0,75919\hat{E}_{t-4}$$

Significancia de los parámetros:

Se procede a evaluar la significancia de los parámetros del Modelo 2:

ARIMA (1,0,0) (0,1,1) [4], solo interesa evaluar significancia en los parámetros de mayor orden asociados a la parte AR regular y MA de la parte estacional, en este caso los parámetros a evaluar son ϕ_1 y θ_1 , . En ciertos casos, el parámetro de mayor orden no es significativo pero contribuye a que los errores del modelo sean ruido blanco, en estos casos no se recomienda quitar dicho parámetro. Para evaluar la significancia de dichos parámetros se usa el siguiente test de hipótesis:

Test de Hipótesis (1):

$H_0: \phi_1 = 0$, *No es significativo* *vs* $H_a: \phi_1 \neq 0$, *Es significativo*

$H_0: \theta_1 = 0$, *No es significativo* *vs* $H_a: \theta_1 \neq 0$, *Es significativo*

$H_0: \delta = 0$, *No es significativo* *vs* $H_a: \delta \neq 0$, *Es significativo*

Estadístico de prueba

$$z = \frac{\hat{\phi}_i}{SE(\hat{\phi}_i)} \underset{bajo H_0}{\sim} N_{(0,1)} \quad z = \frac{\hat{\delta}}{SE(\hat{\delta})} \underset{bajo H_0}{\sim} N_{(0,1)}$$

$$z = \frac{\hat{\theta}_i}{SE(\hat{\theta}_i)} \underset{bajo H_0}{\sim} N_{(0,1)}$$

Criterios de rechazo de la hipótesis nula

Rechazo si $V_p = P(> |Z|)$ es pequeño

Haciendo un análisis para el Modelo 2 se observa que el parámetro asociado a la parte AR regular, ϕ_1 , y el parámetro asociado a la parte MA estacionaria, θ_1 , son significativos .

4.3 Gráficos de ajuste y medidas de calidad de ajuste

A continuación se presentan los gráficos de ajuste para los modelos 1 y 2. (Figura 8)

Se puede ver gráficamente como los 2 modelos propuestos en el presente trabajo, logran ajustarse bien a la serie real y logran seguir su componente cíclica, sin embargo el Modelo 1 parece tener un mejor ajuste.

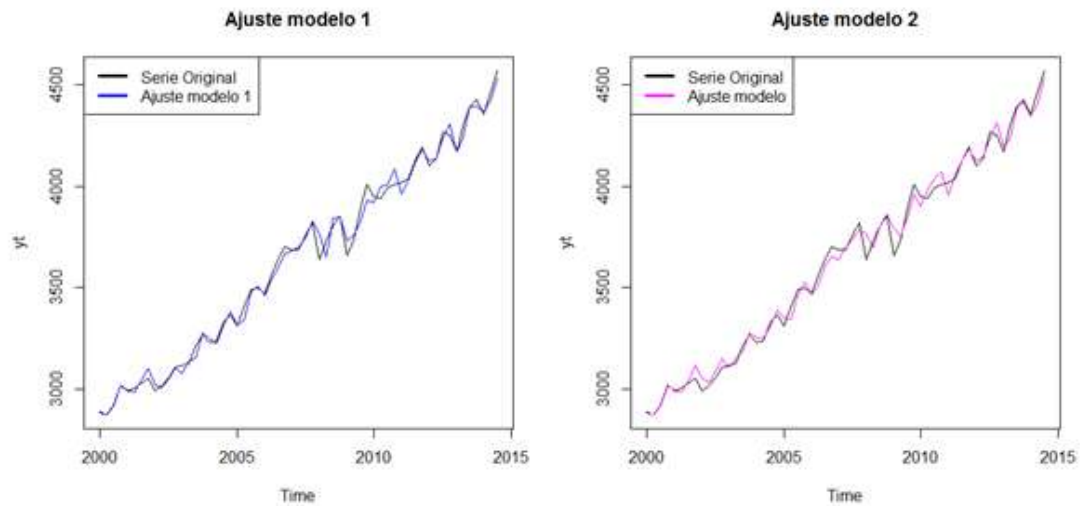


Figura 8 : Ajuste de modelo 1 y 2 con la serie real

Tabla 4: Medidas de calidad de ajuste para Modelos 1 y 2

	AIC	BIC
Modelo 1	1887,914	2502,206
Modelo 2	1980,55	2201,22

Se observa de la tabla 4 que el mejor AIC de los dos modelos propuestos es el del Modelo 1, además de esto el Modelo 2 tiene menor BIC, por lo que se escoge el modelo2 como el de mejor ajuste, solo mirando el criterio del BIC ya que este criterio tiene más relevancia en esta prueba por escoger modelos más parsimoniosos.

5. ANÁLISIS DE RESIDUALES Y VALIDACIÓN DE SUPUESTOS

-Modelo 1: SARIMA (3,1,3) (4,1,2)_[4] con δ deriva y solo con los parámetros $\varphi_2, \varphi_3, \Phi_4, \theta_1, \theta_3, \theta_1$ y θ_2 .

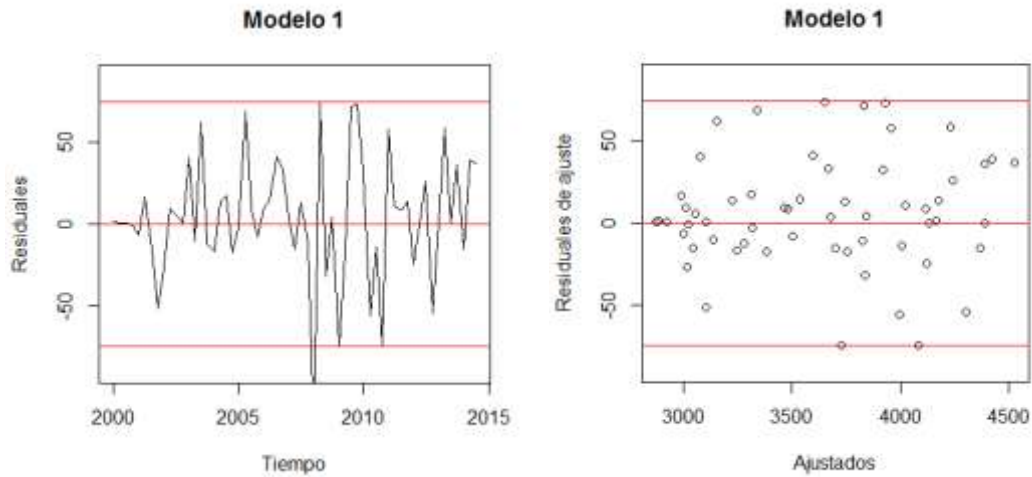


Figura 9. Izq. Gráfico de residuales del modelo $(Y_t \sim \text{SARIMA}(3,1,3) (4,1,2)_{[4]})$ vs tiempo.
Der. Gráfico de residuales del modelo vs valores ajustados.

Según la **figura 9**, en el gráfico de residuales del modelo $(Y_t \sim \text{SARIMA}(3,1,3) (4,1,2)_{[4]})$ (Der.) se logra observar que la varianza es aproximadamente constante, eliminando algunos puntos atípicos, en la **figura 9** (Izq.) residuales vs tiempo no se observan ciclos, y su media esta aproximadamente constante en cero, por lo que se puede concluir que hasta el momento no hay evidencia en contra de que los errores de este modelo sean independientes.

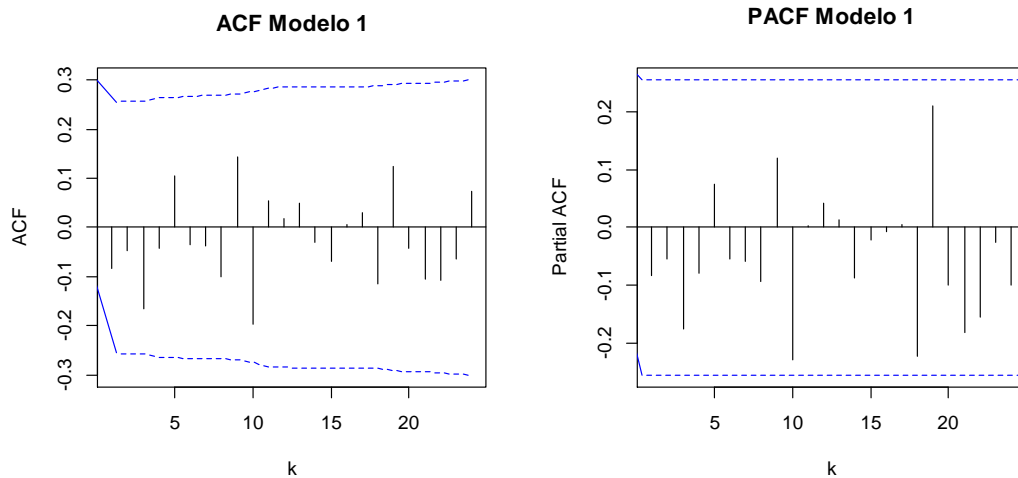


Figura 10. Izq: Gráfica ACF. Der: Gráfica PACF del modelo $(Y_t \sim \text{SARIMA}(3,1,3) (4,1,2)_{[4]})$.

Tabla 5. Estadístico de prueba, prueba de hipótesis y criterio de rechazo para ACF y PACF.

	<i>Test con ACF</i>	<i>Test con PACF</i>
	$\rho(k) = \text{corr}(E_t, E_{t+k})$	$\varphi_{kk} = \text{corr}\langle E_t, E_{t+k} E_{t+1}, \dots, E_{t+k-1} \rangle$
Hipótesis	$H_0: \rho(k) = 0 \quad H_1: \rho(k) \neq 0$	$H_0: \varphi_{kk} = 0 \quad H_1: \varphi_{kk} \neq 0$
Estadístico	$\hat{\rho}(k) = \sum_{t=1}^{59-k} \frac{\hat{E}_t \hat{E}_{t-k}}{\sum_{t=1}^{59} \hat{E}_t^2} \sim_{\text{aprox}} N(0, 1/59)$	$\hat{\varphi}_{kk}$ (construido a partir de la regla de crammer) $\hat{\varphi}_{kk} \sim_{\text{aprox}} N(0, 1/59)$
Criterio de rechazo para H_0	$ \hat{\rho}(k) \geq \frac{2}{\sqrt{59}}$	$ \hat{\varphi}_{kk} \geq \frac{2}{\sqrt{59}}$

Se realizan 24 pruebas independientes.

En la **figura 10**, en la ACF del modelo (Izq) no se observan cortes, y en la PACF (**figura 10 Der**) tampoco se observan cortes, por lo que no se rechaza la hipótesis nula en ningún test y se dice que $\rho(k) = 0$ para todo k; y que $\varphi_{kk} = 0$ para todo k, con $k=1, \dots, 24$. Y como media era constante y la varianza constante por tanto, no hay evidencia en contra de que los errores de este modelo sean Ruido blanco.

Tabla 6: Test Ljung-Box para residuos del modelo SARIMA (3,1,3) (4,1,2)_[4].

Ljung-box			
M	Q _{LB}	gl	P(X _m ² > Q _{LB})
6	3.279716	6	0.7730087
12	8.642353	12	0.7331365
18	10.548474	18	0.9125265
24	15.248537	24	0.9132752

Para el test Ljung-Box se establece la siguiente prueba de hipótesis:

$$H_0: \rho_E(1) = \rho_E(2) = \dots = \rho_E(m) = 0 \quad \text{vs} \quad H_1: \text{Algún } \rho_E(k) \neq 0$$

Con un estadístico de prueba: $Q_{LB} = 59(61) \sum_{k=1}^m \frac{\hat{\rho}^2(k)}{59-k} \approx \text{aprox } X_m^2$

Región de rechazo: $Vp = P(X_m^2 \geq Q_{LB})$ “sea pequeño”.

Como el valor p del test para todos los valores de m son grandes, es decir, la probabilidad de equivocarse si se rechazar H_0 es alta, entonces no rechazo la hipótesis nula (los $\rho_E(k)$ son iguales a cero), por lo tanto los errores del modelo SARIMA (3,1,3) (4,1,2)_[4] son Ruido blanco según LB.

Ahora se procede a evaluar normalidad para este modelo:

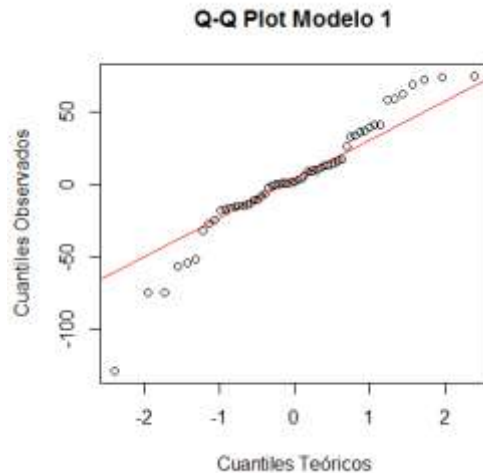


Figura 11: Grafico de normalidad residuos del modelo SARIMA (3,1,3) (4,1,2)_[4].

Tabla 7 :est Shapiro-Wilk
Test de normalidad Shapiro- Wilk residuals(modelo1)
W = 0.94652, p-value = 0.01164

Para la prueba de normalidad en el E_t la prueba de hipótesis es:

$H_0: E_t \sim \text{normal}$ vs $H_1: E_t$ no se distribuye normal

El estadístico de prueba es:

$$W = \frac{(\sum_{i=1}^{59} a_i x_{(i)})^2}{(\sum_{i=1}^{59} x_i - \bar{x})^2}$$

Criterio de rechazo: VP test Shapiro Wilk “pequeño”.

Según la información arrojada por el test de Shapiro-Wilk, el vp es pequeño, por tanto se rechaza la hipótesis nula y se niega la normalidad del error del modelo, en cuanto al análisis grafico, en la **figura 11** se puede ver que la gráfica tiene observaciones extremas en sus colas (muy pronunciadas) las cuales no se alinean a lo largo de la recta ;y por tanto, se concluye que el error del modelo SARIMA (3,1,3) (4,1,2)_[4] no distribuye normal según la gráfica de normalidad.

-Modelo 2: ARIMA(1,0,0)(0,1,1)_[4] con deriva

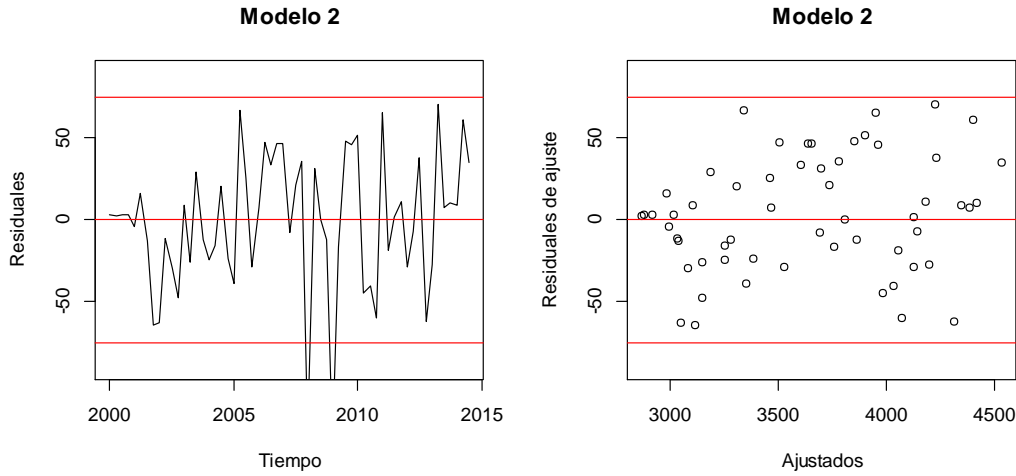


Figura 12 : Izq. Gráfico de residuales del modelo $(Y_t \sim \text{ARIMA}(1,0,0)(0,1,1)_{[4]})$ vs tiempo.
Der. Gráfico de residuales del modelo vs valores ajustados.

Según la **figura 12**, en el gráfico de residuales del modelo $(Y_t \sim \text{ARIMA}(1,0,0)(0,1,1)_{[4]})$ (Der.) se logra observar que la varianza es aproximadamente constante, eliminando algunos puntos atípicos, en la **figura 12** (Izq.) residuales vs tiempo no se observan ciclos, y su media esta aproximadamente constante en cero, por lo que se puede concluir que hasta el momento no hay evidencia en contra de que los errores de este modelo sean independientes.

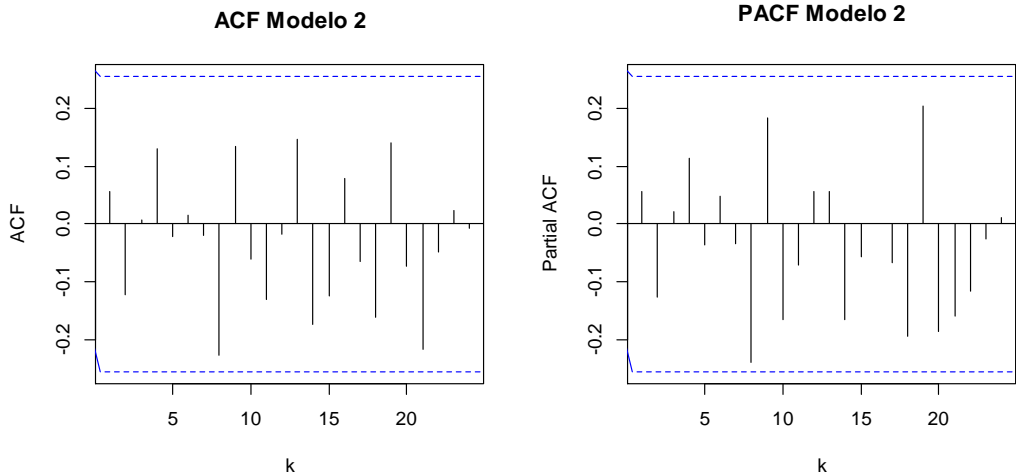


Figura 13 : Izq: Gráfica ACF. Der: Gráfica PACF del modelo $(Y_t \sim \text{ARIMA}(1,0,0)(0,1,1)_{[4]})$

Tabla 8 : Estadístico de prueba, prueba de hipótesis y criterio de rechazo para ACF y PACF.

	<i>Test con ACF</i>	<i>Test con PACF</i>
	$\rho(k) = \text{corr}(E_t, E_{t+k})$	$\varphi_{kk} = \text{corr}\langle E_t, E_{t+k} E_{t+1}, \dots, E_{t+k-1} \rangle$
Hipótesis	$H_0: \rho(k) = 0 \quad H_1: \rho(k) \neq 0$	$H_0: \varphi_{kk} = 0 \quad H_1: \varphi_{kk} \neq 0$
Estadístico	$\hat{\rho}(k) = \sum_{t=1}^{59-k} \frac{\hat{E}_t \hat{E}_{t-k}}{\sum_{t=1}^{59} \hat{E}_t^2} \sim_{\text{aprox}} N(0, 1/59)$	$\hat{\varphi}_{kk}$ (construido a partir de la regla de crammer) $\hat{\varphi}_{kk} \sim_{\text{aprox}} N(0, 1/59)$

Criterio de rechazo para H_0	$ \hat{\rho}(k) \geq \frac{2}{\sqrt{59}}$	$ \hat{\varphi}_{kk} \geq \frac{2}{\sqrt{59}}$
--------------------------------	--	---

Se realizan 24 pruebas independientes.

En la **figura 13**, en la ACF del modelo (Izq) no se observan cortes, y en la PACF (**figura 13 Der**) tampoco se observan cortes, por lo que no se rechaza la hipótesis nula en ningún test y se dice que $\rho(k) = 0$ para todo k; y que $\varphi_{kk} = 0$ para todo k, con $k=1,\dots,24$. Y como media y varianza son constantes, por tanto, no hay evidencia en contra de que los errores de este modelo sean Ruido blanco.

Tabla 9: Test Ljung-Box para residuos del modelo ARIMA(1,0,0)(0,1,1)_[4].

Ljung-box			
M	Q _{LB}	gl	P(X ² _m >Q _{LB})
6	2.295408	6	0.8906254
12	8.817729	12	0.7184149
18	17.372929	18	0.4976209
24	24.368707	24	0.4406833

Para el test Ljung-Box se establece la siguiente prueba de hipótesis:

$$H_0: \rho_E(1)=\rho_E(2)=\dots=\rho_E(m)=0 \quad \text{vs} \quad H_1: \text{Algún } \rho_E(k) \neq 0$$

Con un estadístico de prueba: $Q_{LB} = 59(61) \sum_{k=1}^m \frac{\hat{\rho}^2(k)}{59-k} \approx \chi^2_m$

Región de rechazo: $Vp = P(X^2_m \geq Q_{LB})$ “pequeño”.

Como el valor p del test para todos los valores m son grandes, es decir, la probabilidad de equivocarse si se rechaza H_0 es alta, entonces no rechazo la hipótesis nula (los $\rho_E(k)$ son iguales a cero), por lo tanto los errores del modelo ARIMA(1,0,0)(0,1,1)_[4] son Ruido blanco.

Ahora se procede a evaluar normalidad para este modelo:

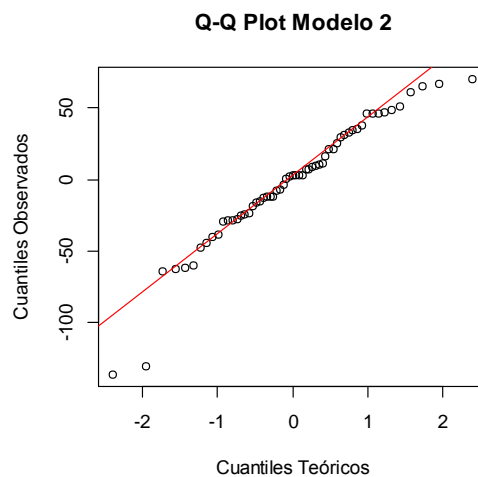


Figura 14 : Grafico de normalidad residuos del modelo ARIMA(1,0,0)(0,1,1)_[4]

Tabla 10 : Test Shapiro-Wilk
Test de normalidad Shapiro- Wilk residuals(modelo1)
W = 0.94903, p-value = 0.01517

Para la prueba de normalidad en el E_t la prueba de hipótesis es:

$$H_0: E_t \sim \text{normal} \quad \text{vs} \quad H_1: E_t \text{ no se distribuye normal}$$

El estadístico de prueba es:

$$W = \frac{(\sum_{i=1}^{59} a_i x_{(i)})^2}{(\sum_{i=1}^{59} x_i - \bar{x})^2}$$

Criterio de rechazo: VP test Shapiro Wilk “pequeño”.

Según la información arrojada por el test de Shapiro-Wilk, el vp es pequeño, por tanto se rechaza la hipótesis nula y se niega la normalidad del error del modelo, pero el análisis gráfico tiene mayor peso, por lo cual se tendrá más en cuenta a la hora de tomar la decisión. En la **figura 14** se puede ver que la gráfica tiene observaciones extremas en sus colas (probablemente por este acontecimiento el test S-W rechazo), pero estas siguen un patrón de escalón a lo largo de toda la recta, y por tanto, se concluye que el error del modelo $\text{ARIMA}(1,0,0)(0,1,1)_{[4]}$ distribuye normal según la gráfica de normalidad.

Ambos modelos cumplen supuestos de ser RB, pero solo el modelo dos cumple el supuesto de normalidad con media cero y varianza aproximadamente constante, por lo cual es el único modelo válido para explicar a Y_t .

6. PRONÓSTICOS PARA LA VALIDACIÓN CRUZADA

-Modelo SARIMA(3,1,3) (4,1,2)_[4]

Tabla 11. Pronostico para el Modelo $Y_t \sim \text{SARIMA}(3,1,3) (4,1,2)_{[4]}$					
Periodo	L	Real	Pronostico	Lim. Inferior	Lim. Superior
2014Q4	1	4555	4613.882	4520.112	4707.653
2015Q1	2	4449	4533.414	4402.736	4664.093
2015Q2	3	4537	4577.724	4428.167	4727.281
2015Q3	4	4736	4682.435	4520.459	4844.410
Cobertura: 100%					
Amplitud promedio: 267.9909					

Tabla 12. Precisión pronostico para el Modelo $Y_t \sim \text{SARIMA}(3,1,3) (4,1,2)_{[4]}$

Medida	RMSE	MAE	MAPE
Resultado	61.48291	59.39649	1.304674

La ecuación de pronóstico para el modelo es:

$$\widehat{Y}_{59}(L-j) = \begin{cases} \text{pronostico } L-j \text{ periodos despues de } t = 59 & \text{si } L-j > 0 \\ \text{observacion } \hat{Y}_{59+L-j} & \text{si } L-j \leq 0 \end{cases}$$

$$\hat{E}_{59}(L-j) = \begin{cases} 0 & \text{si } L-j > 0 \\ \text{residuo } \hat{E}_{59+L-j} & \text{si } L-j \leq 0 \end{cases}$$

$$\begin{aligned} \widehat{Y}_{59}(L) = & 98,5757189 + Y_{59}(L-1) - 0.2950721 Y_{59}(L-2) + 1.0827199 Y_{59}(L-3) \\ & + 0.2123522 Y_{59}(L-4) - Y_{59}(L-5) \\ & - 0.2950721 Y_{59}(L-6) - 1.0827199 Y_{59}(L-7) + 0.7876478 Y_{59}(L-8) \\ & - 0.0073899 Y_{59}(L-16) + 0.0073899 Y_{59}(L-17) - 0.00218055 Y_{59}(L-18) \\ & + 0.008001192 Y_{59}(L-19) + 0.001569262 Y_{59}(L-20) - 0.0073899 Y_{59}(L-21) \\ & + 0.002180553 Y_{59}(L-22) - 0.008001192 Y_{59}(L-23) \\ & + -0.005820638 Y_{59}(L-24) - 0.1398483 \hat{E}_{59}(L-1) - 1.0268418 \hat{E}_{59}(L-3) \\ & - 0.6123549 \hat{E}_{59}(L-4) - 0.6123549 \hat{E}_{59}(L-5) + 0.085636792 \hat{E}_{59}(L-7) \\ & + 0.022374245 \hat{E}_{59}(L-9) + 0.164283803 \hat{E}_{59}(L-11) \end{aligned}$$

La cobertura promedio de los IP es muy buena (100%), por lo que el modelo $Y_t \sim \text{SARIMA}(3,1,3)(4,1,2)_{[4]}$ permite pronosticar de manera acertada la serie.

El MAE indica que se incurre en un error promedio absoluto de pronóstico alrededor de 59.39649 mil millones de pesos. Análogamente el MAPE evidencia que este modelo posee un error medio (absoluto) de cerca de 130,4674%.

-Modelo 2: ARIMA(1,0,0)(0,1,1)_[4]

Tabla 13. Pronostico para el Modelo $Y_t \sim \text{ARIMA}(1,0,0)(0,1,1)_{[4]}$

Periodo	L	Real	Pronostico	Lim. Inferior	Lim. Superior
2014Q4	1	4555	4573.418	4485.106	4661.729
2015Q1	2	4449	4480.342	4374.543	4586.140
2015Q2	3	4537	4537.366	4424.801	4649.931
2015Q3	4	4736	4628.112	4512.726	4743.499
Cobertura: 100%					
Amplitud promedio: 211.031					

Tabla 14. Precisión pronostico para el Modelo $Y_t \sim \text{ARIMA}(1,0,0)(0,1,1)_{[4]}$

Medida	RMSE	MAE	MAPE
Resultado	56.92398	39.50315	0.8487243

La ecuación de pronóstico para el modelo es:

$$\widehat{Y}_{59}(L-j) = \begin{cases} \text{pronostico } L-j \text{ periodos despues de } t = 59 & \text{si } L-j > 0 \\ \text{observacion } \hat{Y}_{59+L-j} & \text{si } L-j \leq 0 \end{cases}$$

$$\hat{E}_{59}(L-j) = \begin{cases} 0 & \text{si } L-j > 0 \\ \text{residuo } \hat{E}_{59+L-j} & \text{si } L-j \leq 0 \end{cases}$$

$$\widehat{Y}_n(L) = 27,75071 + 0,65979Y_{59}(L-1) + Y_{59}(L-4) - 0,65979Y_{59}(L-5) + \hat{E}_{59} - 0,75919\hat{E}_{59}(L-4)$$

La cobertura promedio de los IP es muy buena (100%), por lo que el modelo $Y_t \sim \text{ARIMA}(1,0,0)(0,1,1)_{[4]}$ permite pronosticar de manera acertada la serie.

El MAE indica que se incurre en un error promedio absoluto de pronóstico alrededor de 39.50315 mil millones de pesos. Análogamente el MAPE evidencia que este modelo posee un error medio (absoluto) de cerca de 84,87%.

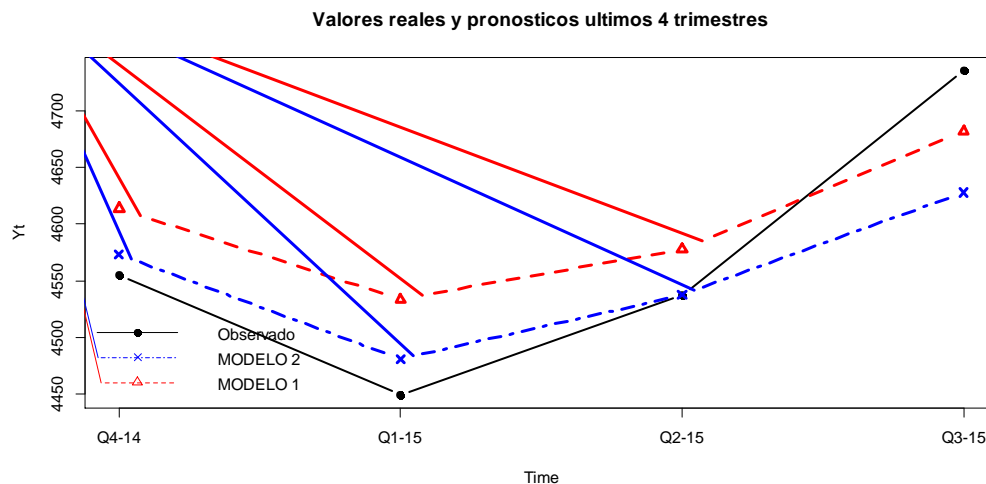


Figura 15. Gráfica de comparación de pronóstico de los modelos

En la **figura 15** se observa claramente que el modelo 2 es muy superior al modelo 1 ya que sigue de muy buena forma los valores reales y de hecho acierta en un punto de pronóstico en Q2-15, aunque preocupa mucho en el último trimestre la diferencia con los valores reales es muy grande, la superioridad grafica del modelo 2 seguramente se verá reflejada en las medidas de pronóstico.

Ambos modelos tienen una cobertura del 100%, siendo mejor (mas pequeña) la del modelo 2, ahora en cuanto a medidas de pronóstico tanto RMSE, MAE y MAPE tienen valores de menor magnitud para el modelo 2, con un cambio significativo, más aún si se tiene en cuenta la naturaleza de los datos.

Por tanto se concluye que el modelo 2 $Y_t \sim \text{ARIMA}(1,0,0)(0,1,1)_{[4]}$ es quien mejor pronostica a la serie.

Por lo tanto el modelo 2 es quien cumple con los supuestos de ruido blanco y normalidad, tiene la mejor calidad de ajuste y la mejor calidad de pronóstico por tanto es el mejor modelo identificado en este trabajo para poder explicar el comportamiento de las componentes de nuestra serie de tiempo.

7) Conclusiones finales

En esta sección Se presentaran los mejores modelos de todos los trabajos, sus ecuaciones teóricas, sus comparaciones y la conclusión final explicando cómo debe modelarse el comportamiento de Y_t .

Modelo 2: $Y_t \sim ARIMA(1,0,0)(0,1,1)_{[4]}$ con δ deriva.

Este fue el mejor modelo en cuanto a calidad de ajuste y pronóstico que se escogió en el trabajo 3.

$$Y_t = \delta + \varphi_1 Y_{t-1} + Y_{t-4} - \varphi_1 Y_{t-5} + E_t + \Theta_1 E_{t-4} \quad \text{con } E_t \sim RB(0, \sigma^2)$$

Modelo 3: Y_t como un Modelo cubico Global Estacional con componentes trigonométricos.

Este fue el mejor modelo global en cuanto a calidad de ajuste y pronóstico que se escogió en el trabajo 1 pero como se ha definido a lo largo de todos los trabajos no se puede considerar la tendencia y la estacionalidad de nuestra serie de tiempo como de tipo global pues estas siguen un proceso estocástico y este modelo no sería el apropiado para modelar el comportamiento de Y_t .

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \alpha_1 \sin\left(\frac{\pi}{2} t\right) + \gamma_1 \cos\left(\frac{\pi}{2} t\right) + E_t, \quad E_t \sim^{iid} N(0, \sigma^2)$$

Modelo 4: Y_t como un Modelo Local por Descomposición aditiva & LOESS Lineal
Óptimo por criterio aicc

Este fue el mejor modelo LOESS escogido ya que contaba con la segunda mejor calidad de ajuste y el menor error de pronóstico el cual se escogió en el trabajo 1.

$$\text{En la vecindad de un tiempo } t_k, (t_k = 1, 2, \dots, n) \\ Y_t = (\beta_{0,k} + \beta_{1,k} t) + \sum_{i=1}^4 (\delta_i I_{i,t}) + E_t, \quad \forall t \text{ en la vecindad de } t_k \text{ con } E_t \sim^{iid} N(0, \sigma^2) \text{ y } \sum_{i=1}^4 (\delta_i) = 0$$

Modelo 5: Y_t como un Modelo Local Global por Suavizamiento Exponencial Holt-Winters Aditivo

Este fue el mejor modelo escogido en el trabajo 1 y 2.

$$Y_{t+h} = (\beta_{0,t} + \beta_{1,t} * h) + \sum_{i=1}^s (\delta_i I_{i,t+h}) + E_{t+h} \text{ con } E_t \sim^{iid} N(0, \sigma^2) \text{ y } \sum_{i=1}^s (\delta_i) = 0$$

Modelo 6: Y_t como un Modelo cubico Global Estacional con componentes trigonométricos con estructura en los errores ARMA (5,0).

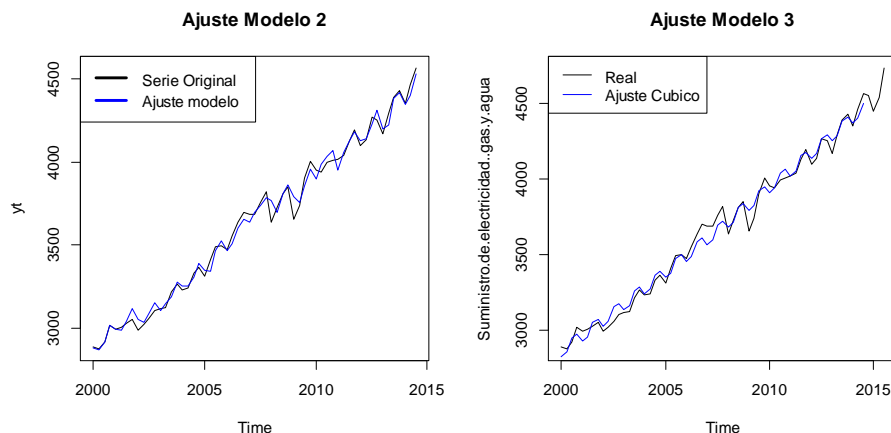
Este fue el mejor modelo con estructura en los errores ARMA en cuanto a calidad de ajuste y pronóstico que se escogió en el trabajo 2 pero como se ha definido a lo largo de todos los trabajos no se puede considerar la tendencia y la estacionalidad de nuestra serie de tiempo como de tipo global lo que se concluye que por más que se insista en explicar los ciclos modelando globalmente la serie no se lograra ajustar y pronosticar bien debido a que

nuestra serie en sus componentes siguen un proceso estocástico y este modelo no sería el apropiado para modelar el comportamiento de Y_t .

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \alpha_1 \sin\left(\frac{\pi}{2} t\right) + \gamma_1 \cos\left(\frac{\pi}{2} t\right) + E_t$$

$$E_t = \phi_1 E_{t-1} + \phi_2 E_{t-2} + \phi_3 E_{t-3} + \phi_4 E_{t-4} + \phi_5 E_{t-5} + a_t, \quad a_t \text{ un RB } \sim N(0, \sigma_a^2)$$

Se ajustaron todos estos cinco modelos identificados por diferentes métodos y modelos determinísticos y estocásticos, para ver qué tan bueno ajustaron estos modelos se presenta las gráficas de ajuste de los 5 modelos mencionados, se observa que el modelo 4, el modelo 5 y 6 no tiene un ajuste tan perfecto o bueno como el modelo 2 en este modelo se logra ver que aproximadamente es el que mejor modela esa posible estacionalidad estocástica presente en la serie de tiempo entre 2007-2009 y logra obtener mejores medidas a simple vista en ajuste, además cabe notar que el modelo 3 no es válido porque esos ciclos se fueron directos al error y por tanto el error no podría llegar a ser ruido blanco. Para analizar con más detalle el ajuste se presentan las medidas de ajuste AIC y BIC. Para estos cinco modelos en la tabla 11.



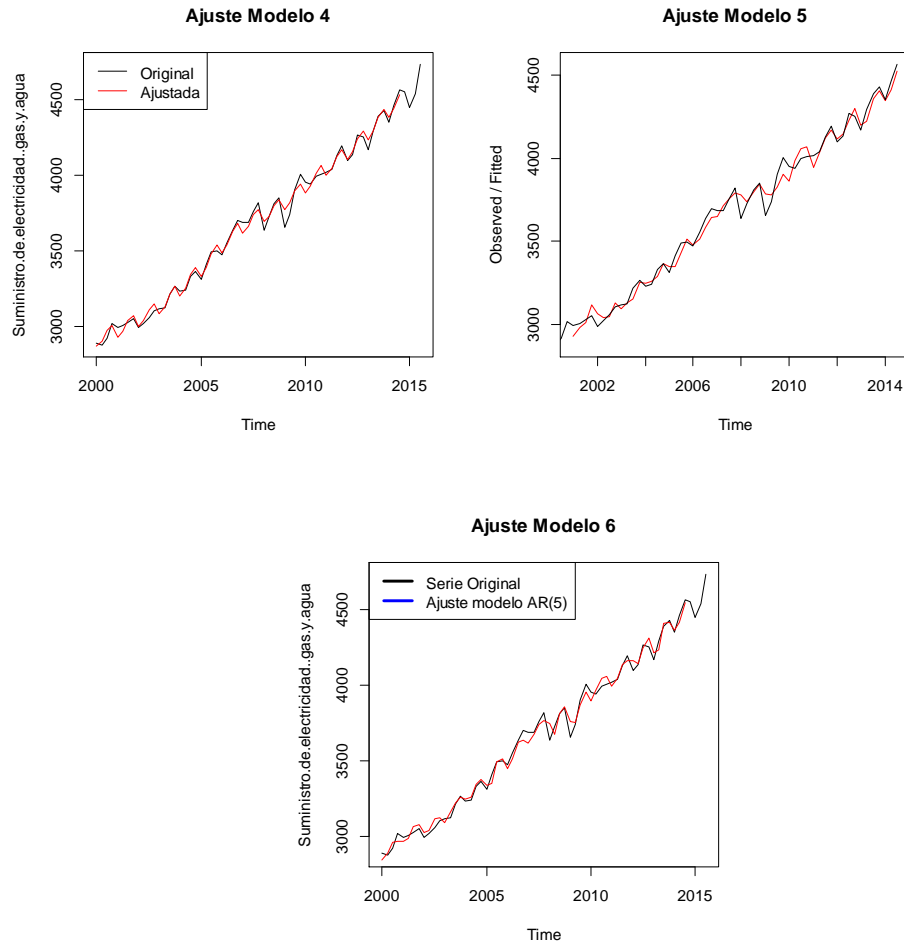


Figura 16. Gráfica de comparación de ajuste de los cinco modelos.

Tabla 15: Medidas AIC y BIC para los Modelos

Medidas	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
<i>AIC</i>	1980.55	3141.746	1713.764	2942.814	2320.942
<i>BIC</i>	2201.22	3880.848	2352.797	3531.953	3418.871

Con las medidas de ajuste se observa según AIC que el modelo 4 parece ser el mejor porque tiene menores valores pero según BIC este modelo es castigado ya que los modelos locales toman muchos parámetros y el de peor ajuste que era de esperarse el del modelo 3 el global. Este quedó de último en ajuste porque no estaba explicando las componentes de ciclos y por eso tampoco es válido, se puede considerar que en ajuste el mejor es el modelo 2 por el criterio del BIC debido a que es más parsimonioso.

NOTA: los modelos 2, 4, 5, 6 cumplieron con el supuesto de ruido blanco y normalidad y el modelo 3 tiene una enorme carencia de ajuste y también no es válido para hacer inferencias debido a que había evidencia en contra del supuesto de independencia lo que posiblemente la $\text{corr}(E_t, E_{t+1}) > 0$.

Para examinar la capacidad de pronóstico de cada uno de los modelos se presentan las gráficas comparativas de pronósticos puntuales en la **figura 16**.

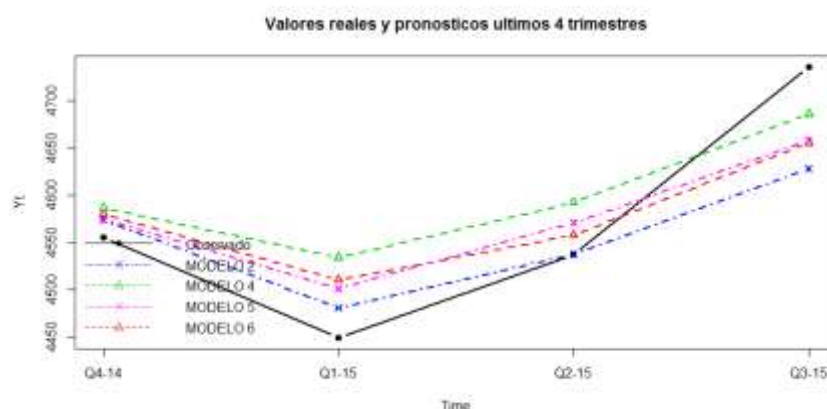


Figura 16 . Gráfica de comparación de pronóstico de los modelos.

De la **figura 16** podemos ver cómo se comporta los pronósticos entre los tres primeros trimestres (Q4 a Q2) el que mejor sigue a los valores reales es el modelo 2 y entre el tercer y cuarto trimestre (Q2 a Q3) el que mejor se comporta es el modelo 4 y 5 el Holt-Winters, pero para mayor certeza en la elección del mejor modelo para pronósticos se presenta la tabla de comparación de errores de pronóstico. En la **tabla 16** se presentan los resultados.

Tabla 16: Medidas de pronóstico para los cuatro mejores modelos

Medidas	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
RMSE	56.92	66.34	58.63	50.509	53.065
MAE	39.5	52.4	55.43	45.68	46.817
MAPE	0.848	1.129	1.32	0.99	1.0194
Amplitud	205.1633	261.99	-	227.7499	180.039
Cobertura	100%	100%	100%	100%	100%

No se analizara el modelo 3 porque no cumple con el supuesto de ruido blanco, Se observa que la cobertura de los pronósticos puntuales sobre los intervalos de predicción es del 100% en todos los modelos, también se observa que el modelo 2 y 5 tienen un error de MAPE menor al 100% , el modelo 5 es el que tiene un RMSE menor y el modelo 2 es el del MAE menor y comparando solo entre el modelo 2 y el 5 el Holt-Winters solo gana en RMSE, pero le daremos más fuerza al MAPE debido a que nos da los resultados en porcentaje y además porque el modelo 5 el Holt-Winters tenía una calidad de ajuste pésima. Así que los dos candidatos para ser el mejor modelo recomendado para la serie de tiempo estarían entre los modelos 2 y 5, y entre ellos el mejor es el modelo 2 debido a que tiene el mejor ajuste y un buen pronóstico y además cumple con los supuestos.

Para el modelo ARIMA estacional se interpreta:

Según RMSE en promedio en un horizonte de pronóstico de 4 trimestres se comete un error de pronóstico de más o menos 56.92 mil millones de pesos

Según MAE en promedio en un horizonte de pronóstico de 4 trimestres se comete un error de pronóstico de más o menos 39.5 mil millones de pesos

Según MAPE en promedio en un horizonte de pronóstico de 4 trimestres el modelo 2 se equivoca más o menos 84,8%.

CONCLUSIONES FINALES

-
- D
-
- Mediante los criterios de información y medidas de pronóstico expuestas en este trabajo se concluye que el modelo 2 es el de mejor ajuste ya que sigue de buena forma los patrones cíclicos de mediano y largo plazo, y es el mejor en pronóstico ya que es un modelo que logra pronosticar mejor el mediano y largo plazo y seguir de buena forma los valores reales.
- Se concluye que el mejor modelo que puede explicar y pronosticar mejor el comportamiento de la serie es $Y_t \sim ARIMA(1,0,0)(0,1,1)_{[4]}$ con δ deriva ya que cumple con el supuesto de ruido blanco y normalidad y es el que menos se equivoca en el ajuste y pronóstico.
-