

## **Análisis de Serie de Tiempo:**

### **Modelos ARMA-SARMA para los Residuos Estructurales**

Maria Isabel Arango Acevedo y David Fernando Rivera Olarte

*29 Octubre de 2017*

---

#### **Resumen**

Este documento presenta los análisis realizados a la serie de tiempo relacionada con el turismo en Alemania respecto a las estadías de una sola noche, así como un comparativo entre el modelo escogido para pronosticar el último año de la serie y los datos observados de la serie para el mismo año. En esta entrega se profundiza en el análisis de los residuos estructurales y su ajuste con análisis de modelos ARMA y S-ARMA.

**Palabras claves:** Serie de tiempo, AIC, BIC, AR, MA, ARMA, S-ARMA.

#### **Abstract**

This paper presents the analyzes of the time series related to tourism in Germany with regard to single-night stays, as well as a comparison between the model chosen to forecast the last year of the series and the observed data from the series for the same year. In this delivery, we will deepen the analysis of the structural residues and their adjustment with analysis of ARMA and S-ARMA models.

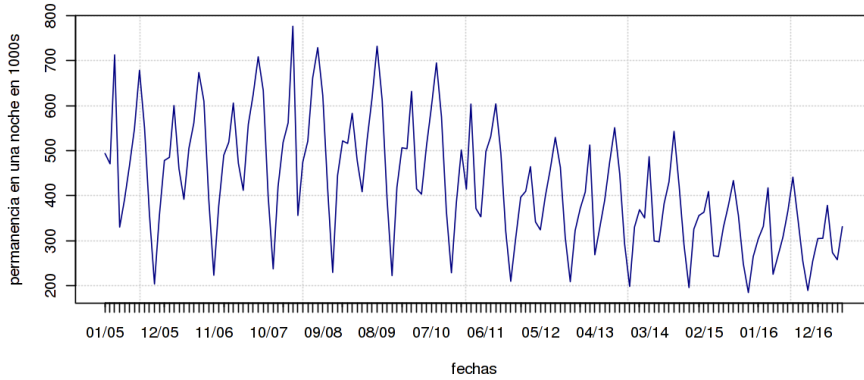
**Key words:** Time series, AIC, BIC, AR, MA, ARMA, S-ARMA.

---

## **1 Introducción**

La serie de datos trabajada corresponde a “permanencia en una noche” (overnight stay), que corresponde a turismo en Alemania. Según la lectura de datos, se está analizando el período comprendido entre 01/01/2005 y

01/06/2017, y se leen en miles de estadías en hoteles, moteles, casas de huéspedes, hospitales/clínicas y spas.



*Figura 1: Estadías de una noche - Turismo Suiza.<sup>1</sup>*

Dentro de los modelos analizados en el entregable anterior, se tiene que el modelo que mejor se ajustó respecto a los datos observados y en pronósticos era una ecuación del tipo exponencial cúbica con estacionalidad marcada con indicadoras:

$$Y_t = \exp(\beta_0 + \sum_{j=1}^3 \beta_j t^j + \sum_{i=1}^{11} \delta_i I_i(t)) + \epsilon_t$$

(1)

Con este modelo, se obtuvieron los siguientes resultados de datos observados sobre los ajustados:

---

<sup>1</sup> Christoph Sax, “Switzerland’s data series in one place”, 2017 <<http://www.dataseries.org/>> [consultado 17 septiembre 2017].

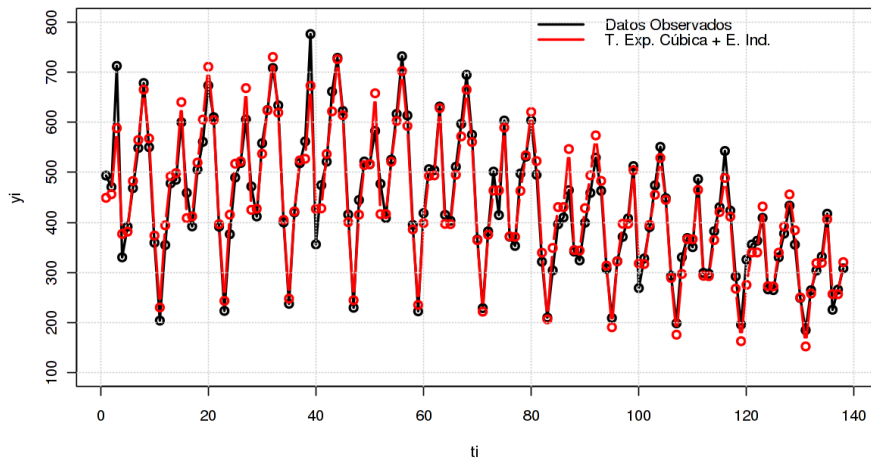


Figura 2- Datos Observados vs Datos Ajustados.

Y de acuerdo con el modelo, el pronóstico de los últimos 12 datos de la serie de tiempo:

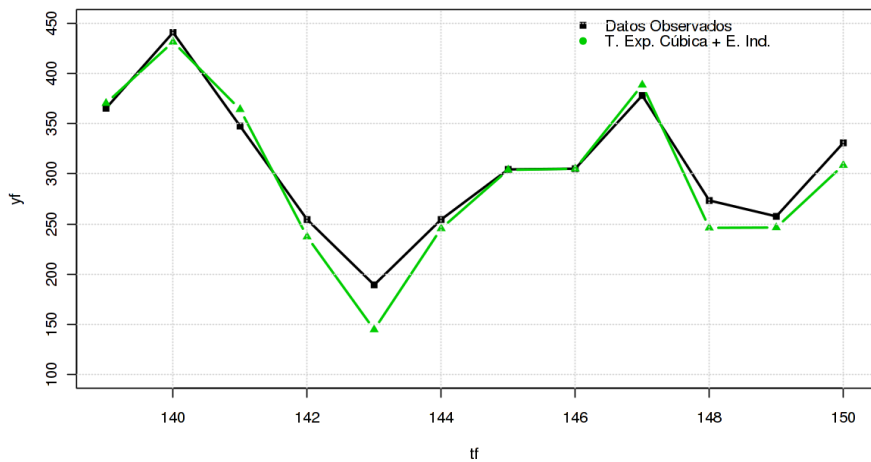


Figura 3- Pronóstico de la Serie

Con estos resultados, es el objetivo del presente entregable hacer el modelamiento de los residuos estructurales  $\epsilon_t$  a través de modelos ARMA o

S-ARMA, de acuerdo con criterios de ajuste mediante estadísticos como AIC, BIC y MAPE para los resultados en los pronósticos.

## 2 Pruebas de Incorrelación

Como primera medida, se determina si los residuos estructurales  $\epsilon_t$  son ruido blanco. A continuación, se expone el comportamiento de esta componente en el tiempo, la gráfica de función de autocorrelación y la gráfica de función de autocorrelación parcial:

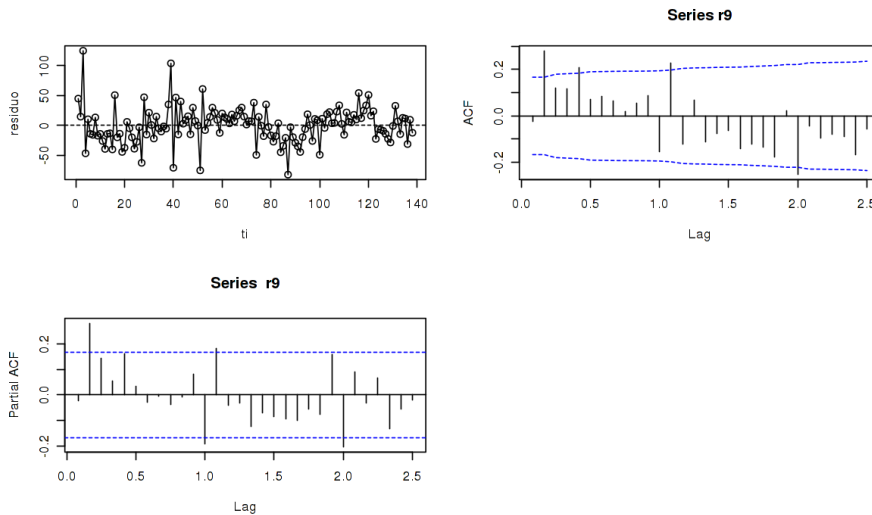


Figura 4- ACF, PACF de los residuos estructurales.

Para los datos en cuestión, se aplican las pruebas Ljung-Box, obteniendo:

Tabla 1- Prueba Ljung-Box residuos estructurales

Rezagos	$\chi^2$	p-value
5	22	0,0005948
15	40	0,0003984
25	69	0,000005491

Y, por último, se aplica la prueba Durbin-Watson generalizada, obteniendo:

*Tabla 2- Prueba Durbin Watson Generalizada*

lag	Autocorrelation	D-W Statistic	p-value
1	-0.02645121	2.033.631	0.954
2	0.27874281	1.420.686	<b>0.000</b>
3	0.11643249	1.605.692	<b>0.020</b>
4	0.11450047	1.592.067	<b>0.026</b>
5	0.20484650	1.409.328	<b>0.000</b>
6	0.06643548	1.682.556	0.134
7	0.07913294	1.655.257	0.102
8	0.06130401	1.681.354	0.176
9	0.01538833	1.771.043	0.472
10	0.05284372	1.687.167	0.256
11	0.08517716	1.612.608	0.150
12	-0.15715372	2.083.252	0.158
13	0.22455291	1.317.485	<b>0.000</b>
14	-0.12341638	2.011.580	0.256
15	0.06463627	1.622.502	0.256

Concluyendo de lo anterior:

- Determinación de componentes AR o MA a partir de las gráficas ACF y PACF.
  - Los residuos no se encuentran contenidos dentro de las bandas den Bartlett en la gráfica ACF, sugiriendo que no se trata de un ruido blanco y posibilitando los análisis de presencia de modelos AR o MA.
  - MA(q): Dado que la gráfica de ACF no termina abruptamente<sup>2</sup>, no es posible determinar que la se trate de una serie tipo MA(q)
  - De la PACF se pueden observar los rezagos 2, 5, 12, 13 y 24 superando de las bandas de Bartlett, sugiriendo un posible modelo AR o AR con estacionalidad, ya que, desde ese punto en adelante, los rezagos se ven contenidos dentro de las bandas.

---

<sup>2</sup> Norman Giraldo, Notas de Clase Series de Tiempo con R, 2006 p112.

- Criterios para aceptación de la prueba Ljung-Box
  - Si  $p\text{-value} < 0.05$ : Se puede rechazar la hipótesis nula suponiendo un 5% de posibilidades de cometer un error<sup>3</sup>. Entonces se puede asumir que sus rezagos muestran dependencia el uno del otro.
- Criterios para aceptación de la prueba Durbin-Watson
  - Se aplica DWG ya que se tiene un modelo diferente a un  $AR(1)$ , de acuerdo a lo encontrado en las evaluaciones de las gráficas ACF, y de la tabla se tiene que con el estadístico  $d < 2$  indicando posible autocorrelación positiva en la serie<sup>4</sup>. Para la mayoría de los valores de los lag evaluados, este valor es menor a 2, indicando que se presenta autocorrelación positiva en esos rezagos de la serie. Además, se confirma que efectivamente se tienen otros rezagos diferentes a  $k = 1$  y son los resaltados en la tabla basados en el valor  $p$ .

De acuerdo con los puntos expuestos se determina que el residuo estructural trabajado no corresponde a ruido blanco y se debe proceder a modelarlos, de acuerdo al objetivo propuesto, como ARMA o S-ARMA.

### 3 Identificación

#### 3.1 Armasubsets

Por medio del uso de la función `armasubsets` de la librería TSA, se determinarán los posibles modelos ARMA o S-ARMA presentes en la serie de datos analizados.

---

<sup>3</sup> Norman Giraldo, Notas de Clase Series de Tiempo con R, 2006 p.94.

<sup>4</sup> Norman Giraldo, Notas de Clase Series de Tiempo con R, 2006 p.99.

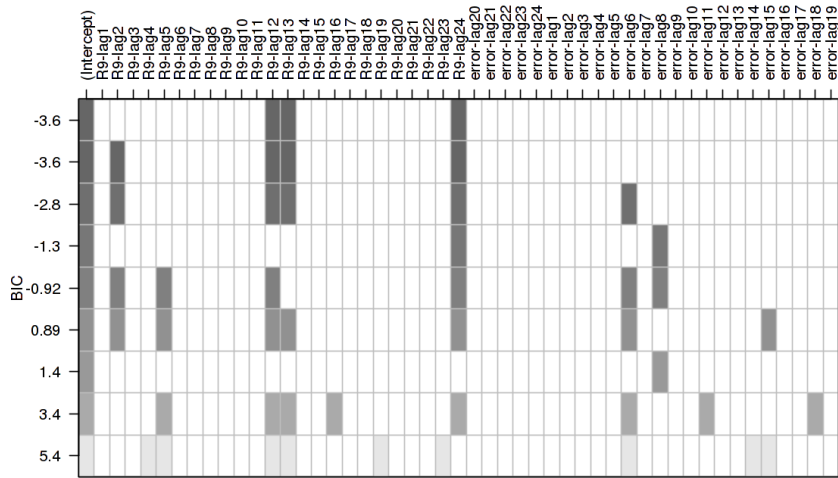


Figura 5- Resultados armasubsets.

Considerando el menor valor del BIC, las componentes MA respecto a las AR estarían descartadas del modelo<sup>5</sup>. Así, se observa un rezago en la posición 12 y otra en la posición 24, dando a entender que se puede tratar de un S-ARMA con  $ps=2$ , dado que la periodicidad de la serie es 12. Observando el segundo menor valor del BIC, dentro de los componentes AR está el rezago 2 y siguen apareciendo el 12 y el 24, por lo tanto, el rezago 2 debería tenerse en cuenta para el modelo sugerido por esta herramienta.

Modelo sugerido:

S-ARMA(2,0)(2,0).

### 3.2 SelectModel-FitAr

Con la aplicación de estas funciones, se trata de encontrar modelos AR(p) que se ajusten al comportamiento de los residuos, bajo el criterio del modelo que arroje mejor AIC, obteniéndose el modelo **AR(5)**, siendo este uno de los rezagos presentes en la gráfica del *armasubsets*, para otros valores más altos del BIC.

<sup>5</sup> Norman Giraldo, Notas de Clase Series de Tiempo con R, 2006 p.141.

### 3.3 Auto.arima

Para esta función se debe colocar la parametrización de detección de componentes estacionales y estacionarias, con rezagos máximos de 24 períodos con el fin de detectar componentes altas.

Con esta configuración se obtiene un modelo **S-ARMA (3,1)(3,1)[12]**.

### 3.4 Autosmarfit

A través de esta función, se trata de buscar un posible ARMA – S-ARMA, reportando el siguiente modelo: **ARMA(2,1)**.

## 4 Estimación

De acuerdo con los modelos resaltados en los anteriores puntos, se realizan los cálculos de AIC para cada uno de ellos:

*Tabla 3- Comparación AIC para los modelos ARMA - S-ARMA*

Modelo	AIC
ARIMA(2,0,0)(2,0,0)[12]	1.311,83
ARIMA(5,0,0)	1.322,28
ARIMA(3,0,1)(3,0,1)[12]	1.300,84
ARIMA(2,0,1)	1.320,61

De aquí se observa que los modelos **ARIMA(3,0,1)(3,0,1)[12]** y **ARIMA(2,0,0)(2,0,0)[12]** presentan los dos menores valores de AIC y con ellos se procede a realizar las validaciones respectivas de curvas ACF, PACF, qqplot, densidad, periodograma acumulado y la prueba Ljung-Box.



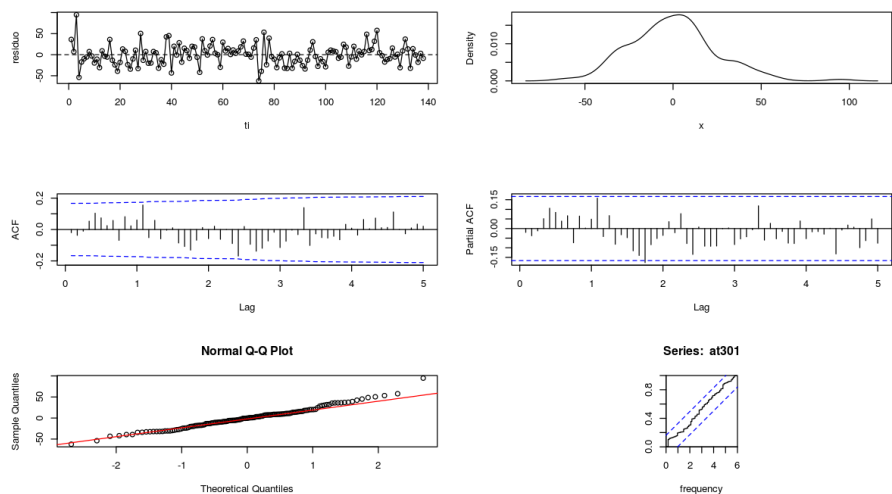


Figura 6 - Validaciones de RB - ARIMA(3,0,1)(3,0,1)[12]

Para este modelo de residuales, se observa que no siguen un cierto comportamiento sino que por el contrario se percibe aleatoriedad, también se tiene una distribución de densidad con cierta simetría y abultamiento al lado izquierdo, así como puntos extendidos en la parte derecha de la gráfica, comprobándose en el QQplot. Para la gráfica ACF se observa que todos los rezagos están contenidos dentro de las bandas de Bartlett, siendo el comportamiento esperado para una serie ruido blanco.

Dentro de las pruebas realizadas de Ljung-Box, se tiene un valor de p de 0.815, que, según los criterios de la prueba, indica que no se rechaza  $H_0$ , que establece que la serie de datos analizada es ruido blanco<sup>6</sup>.

Tabla 4- Coeficientes estimados del modelo ARIMA(3,0,1)(3,0,1)[12]

	Estimate	Std. Error	z value	Pr(> z )
<b>ar1</b>	-0,8271121	0,09110125	-9,07904232	1,09534E-13
<b>ar2</b>	0,3468928	0,10662714	3,25332614	1,14063E+03
<b>ar3</b>	0,3241265	0,08286323	3,91158346	9,16930E+01
<b>ma1</b>	0,955587	0,03350065	28,52443021	5,83220E-173
<b>sar1</b>	-0,7718477	0,16384701	-4,71078268	2,46767E+00
<b>sar2</b>	-0,3331601	0,12582205	-2,64786768	8,10012E+03

<sup>6</sup> Norman Giraldo, Notas de Clase Series de Tiempo con R, 2006. p.94.

<b>sar3</b>	0,1361702	0,12335783	1,10386309	2,69653E+05
<b>sma1</b>	0,6583279	0,15075505	4,36687147	1,26039E+01
<b>intercept</b>	0,2126149	2,94912394	0,07209427	9,42527E+05

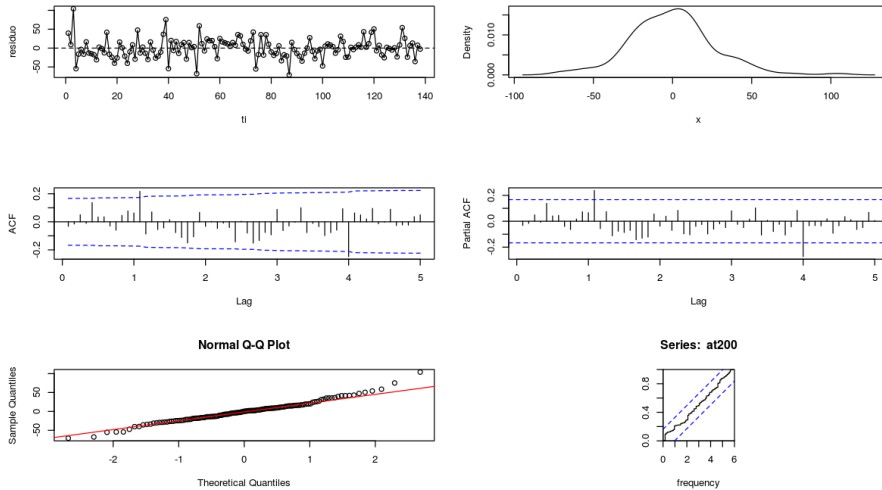


Figura 7- Validaciones de RB - ARIMA(2,0,0)(2,0,0)[12]

Con el modelo ARIMA(2,0,0)(2,0,0)[12], se tiene una distribución en su histograma ciertamente simétrica con abultamiento al lado izquierdo y mayor distribución de puntos al lado de derecho. En la gráfica de QQplot, esta última característica se ve reflejada en varios puntos alejándose de la recta de simetría. Sin embargo, con respecto a la curva de ACF, no se presenta tanta ‘pureza’ como con el modelo anterior, ya que se observan dos rezagos saliéndose de las bandas de Bartlett<sup>7</sup>, el primero en el rezago 12, siendo la frecuencia de la serie de tiempo.

Para la prueba Ljung-Box, se tiene un valor p igual a 0.42, confirmando ruido blanco.

Tabla 5- Coeficientes estimados del modelo ARIMA(2,0,0)(2,0,0)[12]

	Estimate	Std. Error	z value	Pr(> z )
<b>ar1</b>	0,067031490	0,089222190	0,751287200	0,452479837
<b>ar2</b>	0,244109430	0,083426500	2,926042000	0,003433046

<sup>7</sup> Norman Giraldo, Notas de Clase Series de Tiempo con R, 2006. p.91.

<b>sar1</b>	-0,270112200	0,095717190	-2,821982200	0,004772782
<b>sar2</b>	-0,324412810	0,089367340	-3,630104900	0,000283306
<b>intercept</b>	0,327066790	2,154664040	0,151794800	0,879348785

## 5 Pronósticos

Con los dos modelos tratados, ambos se evalúan para la generación de los pronósticos con las últimas 12 muestras de la serie, y se tiene:

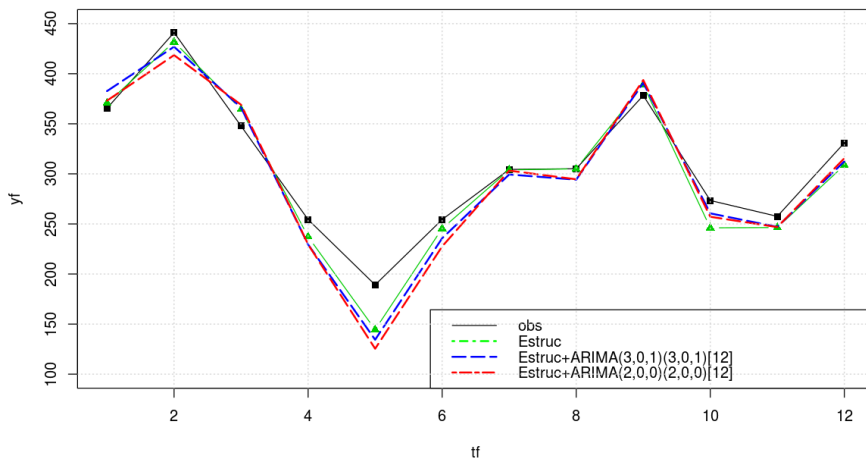


Figura 8- Pronósticos con Modelo Estructural y Modelos S-ARMA

Es evidente que los tres modelos presentan un comportamiento similar, pero visualmente el modelo estructural, calculado con componentes de tendencia exponencial cúbica y estacionalidad de indicadores, tiene ‘menor distancia’ respecto a los datos observados. En los modelos S-ARMA, se puede observar que el modelo en azul correspondiente al ARIMA(3,0,1)(3,0,1)[12], obtenido a través de la función de *auto.arima*, presenta un mejor comportamiento de ajuste, pero no ayuda a la componente estructural a ‘acercarse’ más hacia los datos observados.

Ahora, la cuantificación de estas observaciones se realiza a través de los cálculos del Error Porcentual Absoluto Medio (MAPE):

Tabla 6- Cálculo de Errores de los datos pronosticados

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Estructural	9,11307	18,80009	14,48104	4,02717	<b>5,51843</b>	0,00762	0,29730
Estructural + ARIMA(3,0,1)(3,0,1)[12]	10,18019	21,84885	18,22402	4,63148	6,85572	0,05910	0,35460
Estructural + ARIMA(2,0,0)(2,0,0)[12]	12,33323	24,81367	19,70250	5,46008	7,50217	0,03722	0,41750

Comprobándose lo establecido en la idea anterior, el modelo estructural puro tiene mejor estadístico MAPE con respecto a los demás modelos planteados como estructurales y con modelo de residuos S-ARMA.

## 6 Conclusiones

- El modelo estructural que mejor predice la serie de datos de estadía de una noche en Alemania es el descrito en la ecuación 1: exponencial cúbico con indicadoras. Los residuos de este modelo no cumplen con el criterio de ruido blanco y se pudo evidenciar en las pruebas Ljung-Box y Durbin-Watson Generalizada aplicadas.
- De los modelos S-ARMA para residuos evaluados, el de mejor valoración por el estadístico AIC se trató del modelo S-ARMA(3,0,1)(3,0,1)[12], obtenido mediante la función *auto.arima*. Un segundo modelo escogido para estimación y posteriormente para pronosticar fue el que se realizó mediante evaluación ‘humana’ a través de la función de *armasubsets*, el S-ARMA(2,0,0)(2,0,0)[12].
- En los pronósticos, se observa que ninguno de los dos modelos adicionales introducidos en esta entrega, mejora los estadísticos anteriores (con pronóstico estructural) del MAPE y por el contrario ‘alejan’ la estimación realizada de los puntos de observación originales. Se concluye que para fines de pronóstico da mejor resultado continuar con el modelo estructural y no adicionar los modelos S-ARMA calculados.
- Para el modelado y pronóstico de la serie estudiada, un modelo de componentes estructurales se prefiere en lugar de modelos de componentes estructurales que contengan modelados sus residuos estructurales.
- Seguir profundizando en el estudio de las técnicas de pronóstico, para encontrar metodologías alternativas que puedan mejorar el modelo estructural que hasta ahora se ha identificado como el que mejor describe el comportamiento histórico y futuro de la serie en cuestión.