

UNIVERSIDAD POLITÉCNICA DE CATALUÑA

ANÁLISIS DE DATOS

GRADO EN CIENCIA E INGENIERÍA DE DATOS

---

# Análisis de la serie temporal de Exportaciones en España

---

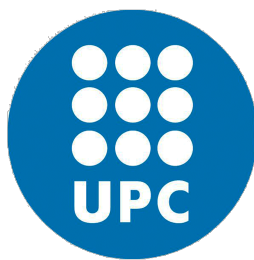
*Autores*

Enric MILLÁN  
y J.P. ZALDIVAR

*Supervisor*

J. Antonio SANCHEZ

3<sup>rd</sup> May, 2023



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

# Contents

<b>1</b>	<b>Introducción</b>	<b>1</b>
1.1	Descripción Serie Temporal . . . . .	1
1.2	Motivación . . . . .	2
1.3	Descripción sintética . . . . .	2
<b>2</b>	<b>Resultados e interpretación</b>	<b>3</b>
2.1	Identificación . . . . .	3
2.1.1	Transformación de la serie . . . . .	3
2.1.2	Serie estacionaria . . . . .	4
2.1.3	Modelos propuestos . . . . .	5
2.2	Estimación . . . . .	6
2.2.1	Estimación modelo 1 . . . . .	6
2.2.2	Estimación modelo 2 . . . . .	7
2.3	Validación . . . . .	7
2.3.1	Varianza constante de residuos . . . . .	7
2.3.2	Normalidad de residuos . . . . .	8
2.3.3	Independencia de residuos . . . . .	9
2.4	Propiedad de causalidad y/o invertibilidad . . . . .	10
2.4.1	Modelo 1 . . . . .	10
2.4.2	Modelo 2 . . . . .	10
2.5	Estabilidad . . . . .	10
2.5.1	Modelo 1 . . . . .	10
2.5.2	Modelo 2 . . . . .	10
2.6	Predicciones . . . . .	10
2.6.1	Modelo 1 . . . . .	11
2.6.2	Modelo 2 . . . . .	11
2.7	Selección . . . . .	11
2.8	Tratamiento de efectos de calendario . . . . .	12
2.8.1	Re-identificación y re-validación tras efectos de calendario . . . . .	12
2.8.2	Predicciones . . . . .	12
2.9	Tratamiento de valores atípicos . . . . .	13
2.9.1	Re-identificación y re-validación tras valores atípicos . . . . .	13
2.9.2	Predicciones . . . . .	14
<b>3</b>	<b>Conclusiones y discusión</b>	<b>14</b>
	<b>Appendices</b>	<b>15</b>
<b>A</b>	<b>Transformación de la serie</b>	<b>15</b>
A.1	Varianza constante . . . . .	15
A.2	Patrón estacional . . . . .	16
A.3	Media constante . . . . .	17
<b>B</b>	<b>Tablas de estimación</b>	<b>17</b>
B.1	Modelo 1 . . . . .	17
B.2	Modelo 2 . . . . .	18

---

<b>C Validación</b>	<b>19</b>
C.1 Varianza constante . . . . .	19
C.2 Normalidad de residuos . . . . .	19
C.3 Independencia de residuos . . . . .	20
<b>D Propiedades</b>	<b>21</b>
<b>E Estabilidad</b>	<b>22</b>
E.1 Modelo 1 . . . . .	22
E.2 Modelo 2 . . . . .	23
<b>F Tratamiento de efectos de calendario</b>	<b>23</b>
F.1 Re-identificación y re-validación tras efectos de calendario . . . . .	23
F.1.1 Varianza constante . . . . .	23
F.1.2 Normalidad de residuos . . . . .	24
F.1.3 Independencia de residuos . . . . .	26
<b>G Tratamiento de atípicos</b>	<b>28</b>
G.1 Re-identificación y re-validación tras tratamiento de atípicos . . . . .	28
G.1.1 Varianza constante . . . . .	28
G.1.2 Normalidad de residuos . . . . .	29
G.1.3 Independencia de residuos . . . . .	31

# 1 Introducción

Este estudio se realiza con el fin de obtener previsiones de una **serie temporal** a partir de las observaciones disponibles, realizado un análisis e informe exhaustivos acerca de la propia serie y sus propiedades.

Tal estudio se compone de una sucesión de apartados en donde se empleará la metodología **Box-Jenkins** mediante modelos **ARIMA** que representen la serie temporal, así como también la incorporación del tratamiento de observaciones atípicas y efectos de calendario.

## 1.1 Descripción Serie Temporal

La serie temporal de elección para el proyecto contiene el **total de exportaciones** (en miles de millones de euros) en España.

Con un total de **240 observaciones** de frecuencia mensual, recorre un periodo de 20 años. Desde el primer mes de 1999 hasta diciembre de 2018. Los valores de las observaciones están registradas en miles de millones de euros. Por simplicidad a la hora de graficar se trabajará con decenas de miles de millones.

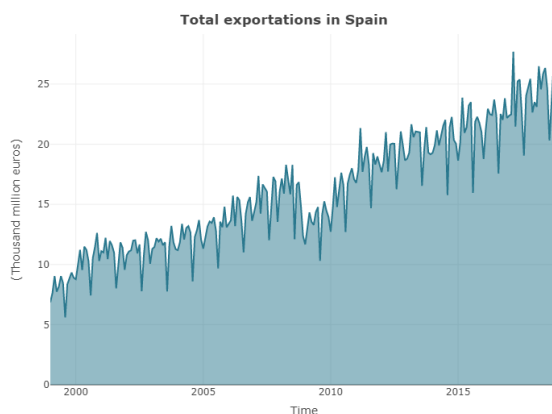


Figure 1: Serie temporal

La serie temporal presenta tendencia creciente a lo largo del tiempo, que mantiene una pendiente casi constante, es decir, no parece haber un aumento repentino del total de exportaciones en España.

Por otro lado, se detecta una caída que supone un cambio estructural (parece que la caída afecta al resto de valores a partir de ese momento) de la serie en lo que corresponde a los años 2008-2009, probablemente debido a la crisis económica. Esto podría generar un valor atípico que se revisará más adelante.

Otro hecho a destacar sobre la serie son un par de observaciones con magnitudes que sobresalen notablemente en el mes de febrero (finales de enero e inicios de febrero) de 2011 y 2017 respectivamente. Para el primer caso (enero/febrero 2011) los bienes de equipo y el sector automóvil fueron los principales impulsores del aumento en los 10 primeros meses del año, por lo que es probable que este pico en febrero esté relacionado. En el segundo caso, las principales contribuciones al crecimiento de las exportaciones provinieron del sector de productos energéticos, alimentación, bebidas y tabaco, sector automóvil y manufacturas de consumo.

Estas explicaciones se realizan en base a investigaciones propias para promover un entendimiento de la causa de tales observaciones. El caso ideal sería contar con un experto en el ámbito de la exportación nacional e incluso internacional para otorgarnos una visión clara sobre dichas observaciones.

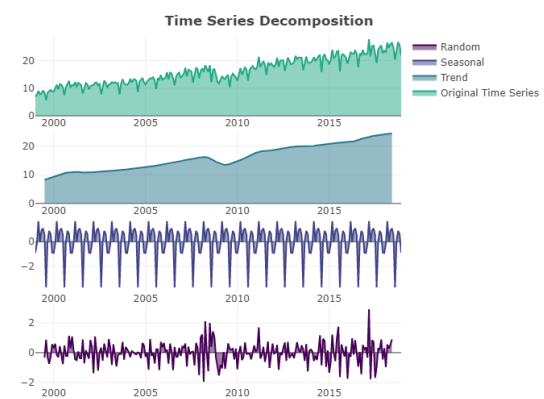


Figure 2: Serie temporal descompuesta en componentes aditivas

La descomposición de la serie muestra la ya

mencionada tendencia creciente, no demasiado pronunciada.

En azul marino, la componente estacional muestra un pico en el número de exportaciones en el primer cuatrimestre del año, mientras que sufre de una disminución considerable durante la temporada de verano, con el mayor descenso en el mes de agosto aproximadamente, que resulta ser el período de mayor descanso laboral.

## 1.2 Motivación

El propósito de este informe es plasmar los conocimientos adquiridos en la asignatura para poder así identificar un modelo que aporte buenas predicciones sobre la serie temporal y se adecue bien a los datos de la serie en cuestión. Es importante recalcar que muchas de las técnicas y gráficos usados a lo largo del informe no se encuentran en este como tal, pero todo este material se encuentra disponible en el RMarkdown adjunto al informe y en el anexo.

## 1.3 Descripción sintética

Como ya se ha mencionado en este proyecto se usará la metodología Box-Jenkins.

Se comenzará con la identificación de la serie. Esto consiste en el análisis de la serie para aplicarle ciertas transformaciones con el objetivo de obtener una serie estacionaria, sobre la cual poder proponer modelos no demasiado complejos y adecuados. La serie será **estacionaria** tras obtener una **varianza constante**, eliminar un

posible **patrón estacional** y conseguir una **media constante**. Con estas condiciones es más sencillo proponer modelos y proseguir a la estimación de sus parámetros.

A continuación, se han de validar los modelos propuestos, que serán 2 de los varios que se identifiquen. Se validarán principalmente analizando sus residuos. En ellos se buscará encontrar varianza constante, normalidad e independencia, siendo esta última la condición más importante. Estas hipótesis se comprobarán con el gráfico de los residuos, el Normal Q-Q plot y su histograma, y el ACF y PACF de estos respectivamente. Estas herramientas se explicarán a lo largo del informe y en el anexo.

Una vez consolidados los modelos propuestos, se estudiarán las propiedades de **estabilidad**, **invertibilidad** y **causalidad** de cada uno y, tras esto, se realizarán las predicciones puntuales y las medidas de capacidad de predicción de cada modelo revisando la adecuación, la exactitud y la precisión de cada modelo.

Con estos resultados se seleccionará el mejor modelo de entre los dos estudiados y se hará un tratamiento de efectos de calendario. En este se revisarán el efecto de la **Semana Santa** y los **Días laborales**. Sobre la nueva versión del modelo se han de volver a realizar los procesos de estimación de parámetros y validación del modelo.

Además, se buscará afinar el resultado al aplicar un tratamiento de atípicos, en el cual se definirá una serie teórica sin estos atípicos para reducir la variabilidad del modelo y mejorar así las cualidades de las predicciones.

## 2 Resultados e interpretación

### 2.1 Identificación

#### 2.1.1 Transformación de la serie

Las transformaciones aplicadas a la serie son una transformación Box-Cox, una diferenciación estacional, y una diferenciación regular. Con los siguientes apartados se justificarán estas decisiones.

**2.1.1.1 Varianza constante** Para comprobar esta condición se hace un gráfico de Media-Varianza y un boxplot por períodos.

En el primer gráfico, la condición de varianza se satisface si la dispersión de los puntos es similar a lo largo de todo el eje horizontal.

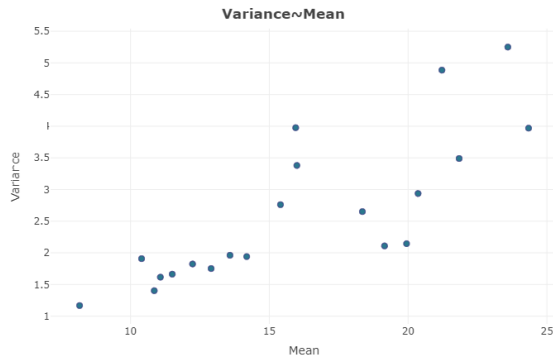


Figure 3: Varianza vs media

En la Figura 3 se observa una tendencia creciente de la varianza con la media, lo que indica debemos aplicar una transformación.

Para evitar la posible influencia de valores atípicos, también se analiza el mencionado boxplot para cada año de la serie, de modo que, idealmente, la altura del IQR de cada caja tiene que ser aproximadamente la misma.

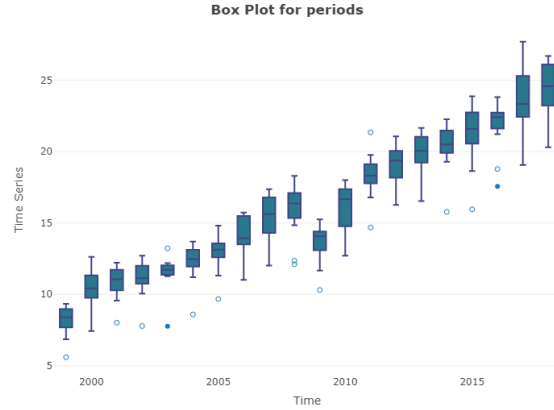


Figure 4: Boxplots de serie temporal

Observando la Figura 4, observamos que las cajas varían considerablemente de tamaño entre años, lo que nos confirma que la serie original no presenta varianza constante.

Aunque la transformación logarítmica es de fácil interpretación, en el anexo se puede comprobar que no mejora la condición de la varianza y no es suficiente, por lo que se ha de encontrar un valor de  $\lambda$  para la transformación Box-Cox ya mencionada, que tiene la siguiente forma:

$$\frac{X_t^\lambda - 1}{\lambda}, \quad \lambda \in [-1, 2], \quad \lambda \neq 0 \quad (1)$$

La función `BoxCox.lambda()` nos permite seleccionar de forma automática la mejor  $\lambda$  posible bajo los criterios que se establezcan. En este caso, usando los métodos "guerrero" y "loglik" se busca, en primer lugar, una  $\lambda$  que minimice el coeficiente de variación para la serie resultante de la transformación y, en segundo lugar, el valor de  $\lambda$  se elige para maximizar la verosimilitud logarítmica de un modelo lineal ajustado a la serie.

Con este método se obtiene un valor de  $\lambda = 0.395$  aproximadamente. Tras la transformación, la varianza para los diferentes grupos de observaciones es similar. Esto se puede comprobar en profundidad en el anexo A, donde se encuentran comparaciones directas entre los gráficos usados antes y después de esta transformación.

**2.1.1.2 Patrón estacional** Para buscar un posible patrón estacional de orden  $s = 12$ , es decir, un patrón que se repita cada año en el mismo momento, se grafican las subseries correspondientes a cada año. Además, el segundo gráfico de la Figura 6 muestra si hay una jerarquía entre los diversos meses pero se ve que no hay ninguno que sobresalga siempre. El tercer gráfico muestra los boxplot de las distribuciones de las observaciones por meses.

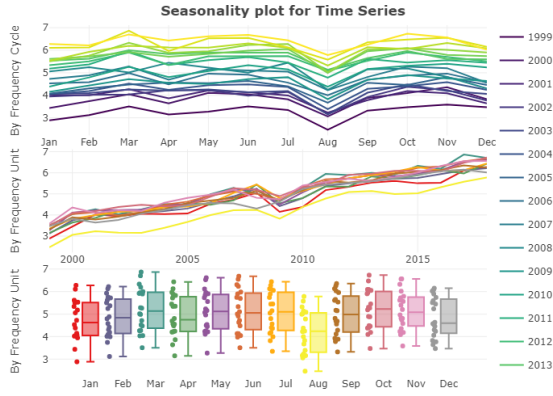


Figure 5: Gráficos de estacionalidad

Observamos un claro patrón estacional en el mes de agosto, donde hay una clara caída de los valores de las observaciones. Esto se confirma observando los boxplots, en los que se aprecia que todos tienen valores similares a excepción agosto, que tiene los valores más bajos, tal y como se había observado al presentar la serie.

Se observa también en el primer gráfico que el tercer mes del año tiene muestras con valores que aumentan considerablemente y seguidamente caen ligeramente en abril o viceversa. Posiblemente este fenómeno esté relacionado con los efectos de calendario, ya que por estos meses se encuentra la festividad de Semana Santa. Esto se tratará más adelante si es necesario.

Como se ha confirmado la presencia de un patrón estacional, se procede a realizar la diferenciación estacional de orden 12. Esto hace que cada observación de la nueva serie sea fruto de la resta entre la observación original y la misma observación del año pasado, eliminando así el patrón estacional.

$$(1 - B^{12})X_t \quad (2)$$

Para corroborar que la transformación sea apropiada basta con observar la descomposición estacional tras dicha transformación para ver que se ha eliminado todo rastro de un patrón. Esta figura se encuentra en el anexo A.

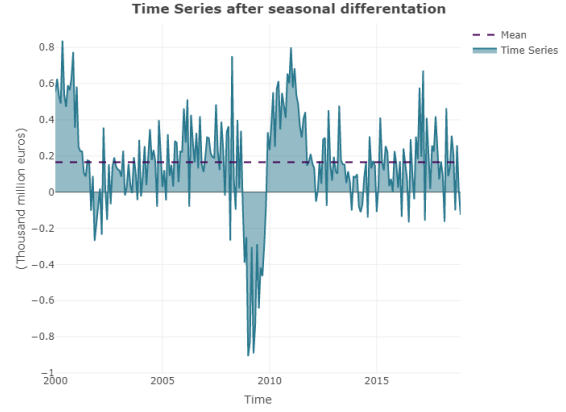


Figure 6: Serie temporal sin patrón estacional

**2.1.1.3 Media constante** Al aplicar la última transformación observamos que en la Figura 6 la serie no presenta signos de algún tipo de tendencia, por lo que podemos considerar ya de entrada que se cumple de forma aceptable la condición de media constante. No obstante, siguiendo la metodología aplicaremos diferenciaciones regulares para mejorar esta condición hasta que la varianza de la serie aumente.

d12bcser	d1d12bcser	d1d1d12bcser
0.0755	0.0657	0.2075

Al aplicar la segunda diferenciación la varianza aumenta, esto se conoce como sobre-diferenciación. Por esta razón se concluye que se aplicará una única diferenciación regular.

## 2.1.2 Serie estacionaria

La que la serie estacionaria  $W_t$  resulta ser:

$$W_t = (1 - B)(1 - B^{12}) \frac{X_t^\lambda - 1}{\lambda}, \quad \lambda \approx 0.395 \quad (3)$$

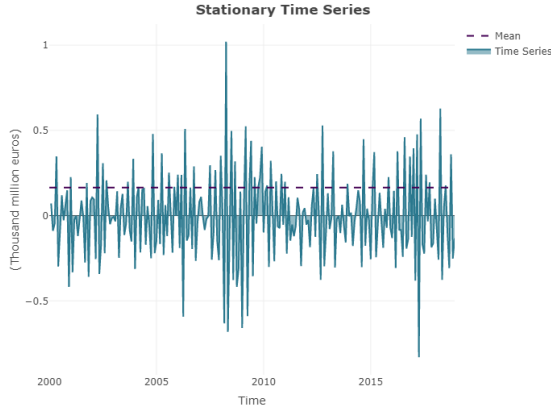


Figure 7: Serie estacionaria

Con valores a los parámetros  $d = 1$  y  $D = 1$  indicando respectivamente la diferenciación regular y la estacionaria que se han aplicado a la serie.

### 2.1.3 Modelos propuestos

Para poder proponer modelos deben analizarse los gráficos del ACF y el PACF de la serie tras haberle aplicado las transformaciones.

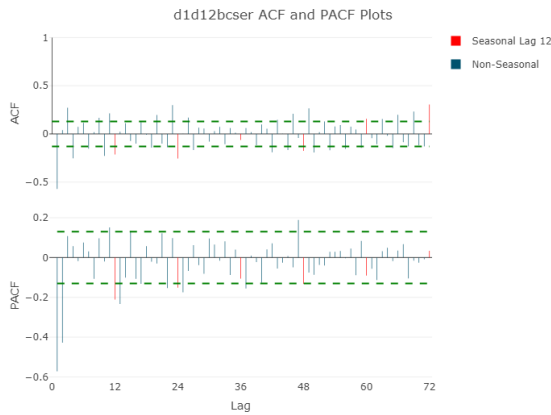


Figure 8: ACF y PACF de serie estacionaria

Para determinar los valores de  $p$ ,  $P$  y  $q$ ,  $Q$  de los modelos ARMA se necesita contar el número de retrasos en los gráficos hasta la última barra significativamente diferente de 0.  $\rho(h) \neq 0$ . Esto significa que la última barra que se considere debe encontrarse fuera de las bandas del intervalo de confianza para poder rechazar la hipótesis nula de no significación de  $\rho(h)$  (por simplicidad del

modelo, las barras más alejadas del origen que presenten significación, se considerarán fruto del azar).

$$\begin{aligned} H_0 : \rho(h) &= 0 \\ H_1 : \rho(h) &\neq 0 \end{aligned} \quad (4)$$

Si bien es cierto que se observa cierta tendencia logarítmica en los retrasos del PACF, característica típica de modelos MA, se considera que es el más adecuado para considerar los retrasos finitos inicialmente.

Comenzando por la parte estacional, se podría considerar ajustar AR(1), AR(2) y AR(4). Pues esta claro que el primer retraso estacional es significativo, mientras que el segundo y el cuarto sobresalen ligeramente por el límite inferior del intervalo de confianza. Es favorable decantarse por un AR(4) ya que se tendría en cuenta un mayor número de relaciones entre las variables del pasado, y si resulta haber parámetros no significativos, estos se pueden eliminar del ajuste más adelante.

Aun sabiendo que los retrasos del ACF no presentan un decrecimiento evidente, se podría considerar una media móvil de parámetro  $q = 6$  MA(6). Esta propuesta incorpora al modelo una mayor complejidad (consta de más parámetros) en comparación con la opción anterior, con lo que nos decantaremos inicialmente por mantener una parte AR(4) o bien proponer un ARMA que explique esta parte estacional, pero no se desechará el MA(6) hasta haber ajustado algunos modelos.

La decisión de presentar un ARMA estacional se debe a que las posibilidades propuestas contienen un número elevado de parámetros. Tras experimentar con los casos que no excedan el número de parámetros mínimo ya presente (4 del AR(4)), se ha llegado a la conclusión, basándose en el AIC (una medida de adecuación de modelos, que cuanto menor sea una mejor adecuación supone) de los modelos que se han ajustado, de que el ARMA(2,2) y el ARMA(2,1) son los más adecuados junto al AR(4) y el MA(6). Todas las opciones son versátiles ya que pueden explicar retrasos que con opciones más simples no se considerasen y además permiten ser reducidos en caso necesario si algún coeficiente no es significativo. Para la parte regular resulta muy evidente que



la mejor opción es un AR(2), debido a que son los únicos dos retrasos con obvia significación en el PACF. El resto de barras que sobresalen están suficientemente lejos del origen como para considerarse como fruto del 5% de azar en las observaciones o se encuentran cerca de desfases estacionales por lo que podrían ser satélites de estos.

Por otro lado, si nos queremos fijar en el ACF, un MA(4) sería suficiente por las mismas razones.

Dado que la última opción cuenta con bastantes parámetros, se estudiará proponer un modelo ARMA para reducir la complejidad del modelo también en la componente regular. Para reconocer cual es el modelo ARMA regular más adecuado, se han estudiado distintas combinaciones teniendo como base ARMA(1,1), ARMA(1,2), ARMA(2,1) y ARMA(2,2). Finalmente, teniendo en cuenta el AIC en primer lugar y la significación de parámetros en segundo lugar, las mejores opciones para la parte regular son el AR(2) y el ARMA(2,1).

Los cuatro mejores modelos identificados (en función principalmente del AIC) han sido los siguientes:

$$\begin{aligned} & \text{ARIMA}(2, 1, 0)(0, 1, 6)_{12} \\ & \text{ARIMA}(2, 1, 1)(4, 1, 0)_{12} \\ & \text{ARIMA}(2, 1, 0)(2, 1, 2)_{12} \\ & \text{ARIMA}(2, 1, 1)(2, 1, 1)_{12} \end{aligned} \quad (5)$$

Todos estos modelos tienen un AIC muy similar, por lo que hay que fijarse también en la significación de sus parámetros. Cabe recalcar que han sido ajustados teniendo como base la serie ya diferenciada, por lo que aparece el coeficiente de la media (intercept), que en ninguno de los casos resulta ser significativo como se verá en el siguiente apartado.

La significación de un parámetro se calcula mediante el  $t$ -ratio, que se explicará en el siguiente apartado, pero en resumen, si el valor del  $t$ -ratio es menor que 2, el parámetro no se considera significativo.

El primer modelo tiene tres parámetros no significativos pertenecientes a la MA (Moving Average) de la parte estacional:  $\Theta_3$ ,  $\Theta_4$  y  $\Theta_5$ , que coinciden con el previo análisis del ACF, donde

se ha visto que el sexto retraso sobresale notablemente pero ni el tercero ni el cuarto ni el quinto lo hacen.

El segundo modelo propuesto solo tiene un parámetro no significativo  $\theta_1$ , perteneciente a la MA regular, que tiene un  $t$ -ratio de aproximadamente 1.64.

A diferencia de los demás, el tercer modelo no presenta ningún parámetro no significativo, pero cabe mencionar que la significación del parámetro  $\Theta_2$ , de la MA estacional, roza el límite con un valor de 2.009.

El cuarto modelo presenta el parámetro  $\theta_1$  no significativo, con un  $t$ -ratio de aproximadamente 1.58, y el parámetro  $\Phi_1$ , de la parte AR (Auto-Regressive) estacional, que consta con un  $t$ -ratio cerca de 1.96 y que podría considerarse como significativo si fuese necesario, en función del cambio en el AIC al suponerlo significativo o no.

Teniendo en cuenta esto, y la comparación de los AIC, hay dos modelos que sobresalen en ambos casos. Estos son el modelo ARIMA(2, 1, 1)(4, 1, 0)<sub>12</sub>, y el modelo ARIMA(2, 1, 0)(2, 1, 2)<sub>12</sub>, con AIC's -190.12 y -189.7 respectivamente. Siendo también los dos modelos con una mejor significación de parámetros. Tales modelos pasarán a la fase de estimación y validación a continuación.

## 2.2 Estimación

### 2.2.1 Estimación modelo 1

El modelo 1 ARIMA(2, 1, 1)(4, 1, 0)<sub>12</sub> es de la forma:

$$\begin{aligned} \phi_2(B)\Phi_4(B^{12})(1-B)(1-B^{12})W_t &= \theta_1(B)Z_t, \\ W_t &= \frac{(X_t - \mu_1)^\lambda - 1}{\lambda} \end{aligned} \quad (6)$$

Se comprueba la significación de la media  $\hat{\mu}_1$  mediante su  $t$ -ratio para determinar si es significativo para el modelo, es decir, si se rechaza o no la hipótesis nula de que la media es igual a 0.

$$\begin{aligned} H_0 : \hat{\mu}_1 &= 0 \\ H_1 : \hat{\mu}_1 &\neq 0 \end{aligned} \quad (7)$$

donde el  $t$ -ratio viene dado por

$$t = \frac{\hat{\mu}_1}{sd_{\hat{\mu}_1}} \sim N(0, 1) \quad (8)$$

asintóticamente normal al tratarse de un estimador calculado a partir del estimador de máxima verosimilitud. Para no poder rechazar la hipótesis nula, si queremos un nivel de confianza del 95%, el valor absoluto del t-ratio tendrá que ser menor que 2.

La tabla de los coeficientes del modelo se encuentra en el anexo B. Al no tener significación la media de la serie  $\hat{\mu}_1$ , se puede aplicar la función ‘arima’ a la serie después de la transformación Box-Cox, aplicándole las diferenciaciones como parámetros a la función, de esta manera se evitan deshacer las transformaciones después y se puede predecir directamente sobre la serie transformada sin diferenciaciones.

El AIC disminuye al eliminar la media, lo que indica la mejoría del ajuste. Otro indicador de que  $\hat{\mu}_1$  no es significativa, es que los valores de los coeficientes estimados del modelo no han variado notablemente, al igual que sus desviaciones estándar.

Bajo la misma hipótesis, se comprobará la significación de los parámetros del modelo. Se observa que el estimador del parámetro  $\hat{\theta}_1$ , tal y como se vio en el apartado anterior, está por debajo del borde del intervalo de confianza, por lo que ajustaremos un modelo eliminándolo. Como se puede comprobar en el anexo, el AIC ha aumentado tras eliminar  $\hat{\theta}_1$ , por lo que se considerará que el modelo inicial es más adecuado.

El resto de parámetros si son significativos, con lo que el modelo estimado es ARIMA(2, 1, 1)(4, 1, 0)<sub>12</sub>.

### 2.2.2 Estimación modelo 2

El modelo 2, ARIMA(2, 1, 0)(2, 1, 2)<sub>12</sub> es de la forma:

$$\begin{aligned} \phi_2(B)\Phi_2(B^{12})W_t &= \Theta_2(B^{12})Z_t \\ W_t &= (1 - B)(1 - B^{12})\frac{(X_t - \mu_2)^\lambda - 1}{\lambda} \end{aligned} \quad (9)$$

Se han repetido los mismos pasos que con el modelo 1, disponibles en el anexo B, y se han

obtenido los siguientes resultados: la media no es significativa y el resto de coeficientes si son significativos.

No se detallan los pasos de la comprobación de significación como se ha realizado con el modelo 1 dada la semejanza y para dar cabida a la información más relevante.

## 2.3 Validación

La validación pasa por revisar que los residuos tengan varianza constante, sigan una distribución normal y sean independientes.

### 2.3.1 Varianza constante de residuos

Con el fin de validar la condición de homocedasticidad (varianza constante) en los residuos, nos apoyaremos en el gráfico de estos. Para un modelo correcto se esperan encontrar gráficos de los residuos similares a un ruido blanco, es decir, sin una gran cantidad de outliers que sobrepasen las bandas de probabilidad, sin una tendencia clara (aspecto que a priori ha sido tratado con la transformación Box-Cox), y sin grupos (clústeres) de volatilidad que anulen esta hipótesis.

**2.3.1.1 Modelo 1** Observamos que los residuos del Modelo 1 presentan dos valores por debajo del límite inferior de las bandas de probabilidad, lo cual respecta a 0.833% de los valores. No es una gran cantidad pero supera el 0.03% que se establece como aceptable, por lo que será revisado en el tratamiento de atípicos si se escoge este modelo.

En el primer gráfico de la Figura 9 vemos una dispersión de los residuos alrededor del 0 sin ningún patrón claro y sin presencia de clústeres de volatilidad. Sin embargo, el segundo gráfico que muestra ajuste local suave de la variabilidad en donde la línea de ajuste muestra una muy leve forma sinusoidal a lo largo de los residuos. Además, los intervalos de confianza de mayor amplitud se encuentran al inicio y hacia el final de las observaciones.

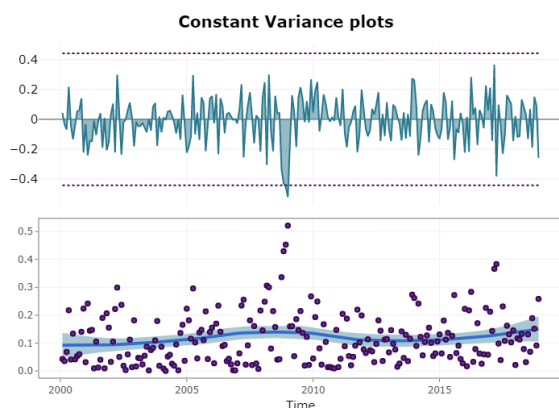


Figure 9: Gráficos para el análisis de varianza constante

Como la tendencia observada es muy leve no debería suponer un gran problema, pero por el momento no se puede considerar como definitiva principalmente debido a la presencia de outliers.

**2.3.1.2 Modelo 2** Como se observa en la Figura 10, para el modelo 2, la presencia de outliers ha aumentado, siendo 3 esta vez. El resto de factores, como la tendencia del ajuste local suave y sus intervalos de confianza tienen un comportamiento muy similar a los del modelo 1.

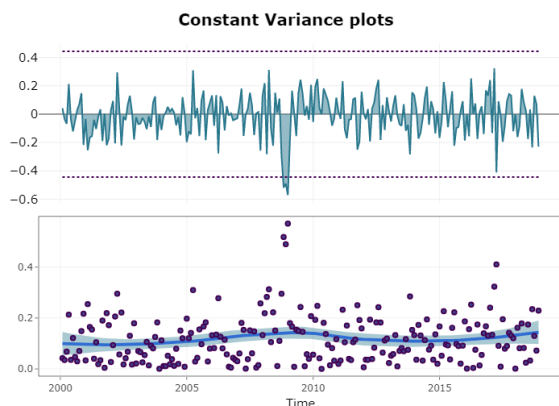


Figure 10: Gráficos para el análisis de varianza constante 2

Así pues, la hipótesis de homocedasticidad tampoco se puede establecer como definitiva en este caso. Puesto a que ambos modelos tienen condiciones similares respecto a la varianza pero el segundo modelo cuenta con más atípicos, se

considera que el primer modelo valida mejor que el segundo la hipótesis de varianza constante.

### 2.3.2 Normalidad de residuos

Para analizar si se cumple esta segunda hipótesis se utilizarán herramientas como el *Normal Q-Q plot*, el histograma de los residuos y el test de *Shapiro-Wilk*. El histograma sirve principalmente para obtener una primera intuición, pero las otras dos herramientas son las más importantes.

**2.3.2.1 Modelo 1** Como se vio en el anterior apartado, hay presencia de posibles outliers. Sin embargo, es razonable decir que la normalidad de los residuos se cumple de forma aceptable, pues las muestras se ajustan adecuadamente a los cuantiles teóricos.

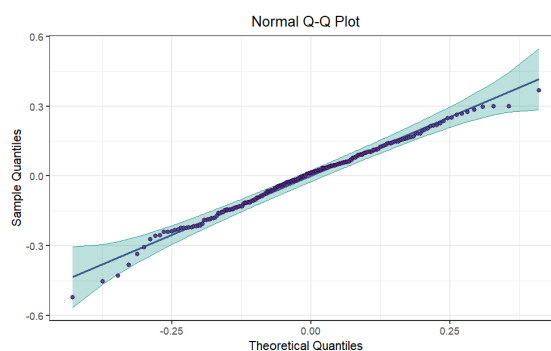


Figure 11: Gráfico para el análisis de normalidad

En el anexo C se puede observar el histograma, que parece seguir de forma aceptable una distribución normal. Tras realizar el test de *Shapiro-Wilk* se observa que el *pvalor* del test es superior al grado de significación, por lo que se concluye como correcta la decisión de asumir la normalidad de los residuos.

**2.3.2.2 Modelo 2** El segundo modelo presenta una presencia de outliers más alta y acentuada que el primer modelo, hecho que ya se había vislumbrado con el análisis de la varianza. A diferencia del primer modelo, hay algunos valores atípicos que quedan fuera de los intervalos de confianza. Si este modelo es escogido, se revisará esta problemática en el tratamiento de los

outliers. Si no se tienen en cuenta dichos atípicos, el ajuste a los cuantiles teóricos es aceptable.

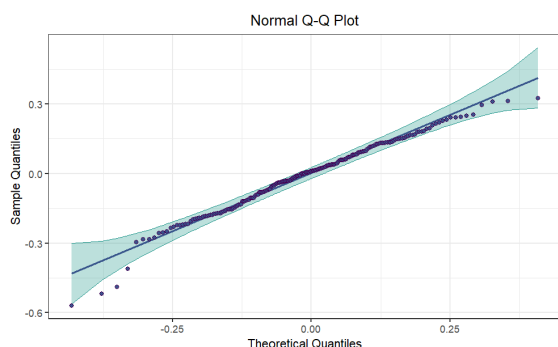


Figure 12: Gráfico para el análisis de normalidad 2

A diferencia del modelo anterior, el histograma C de este flaquea más ya que apenas hay valores entre  $-0.3$  y  $-0.2$  que puedan sostener la hipótesis de normalidad. Al realizar el test de *Shapiro-Wilk* se confirma que la hipótesis de normalidad se debe rechazar al tener un *p*valor de 0.00323. Es importante mencionar que este tipo de tests son muy sensibles a los atípicos, y como ya se ha mencionado este segundo modelo contiene outliers más claros que el primero.

### 2.3.3 Independencia de residuos

La independencia de los residuos se considera la hipótesis más importante a la hora de validar el modelo. Para este caso, se usarán como herramientas el ACF y PACF de los residuos y el test de *Ljung-Box* que permite rechazar o no la hipótesis nula de que las auto-correlaciones de los residuos hasta un desfase  $k$  son conjuntamente iguales a cero; en otras palabras, independientes.

**2.3.3.1 Modelo 1** Analizando la Figura 13, el retraso número 60 sobresale notablemente. Sin embargo, al estar alejado del origen puede considerarse como fruto del azar. Esta suposición se confirma con el test estadístico de *Ljung-Box*, que muestra que el *p*-valor de la hipótesis en ese desfase está por encima del nivel de significación. De hecho, todos los desfases cumplen esto, por lo que se puede afirmar que los residuos siguen una

estructura similar a la de un *White Noise* y son, por tanto, independientes.

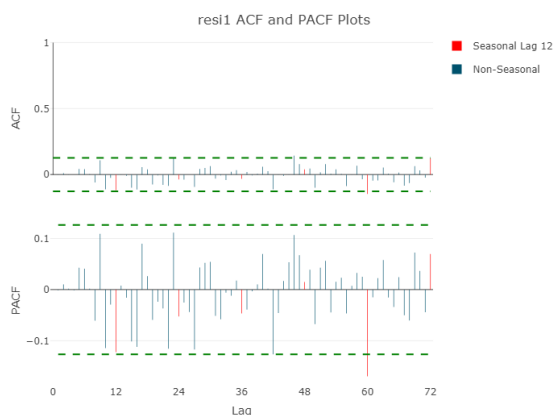


Figure 13: Gráfico para el análisis de independencia

Además de la interpretación visual del ACF y PACF de los residuos, los *p*valores del test de *Ljung-Box* C están por encima en todos los casos del nivel de significación. Por lo que se confirma la hipótesis de independencia de residuos para el modelo 1.

**2.3.3.2 Modelo 2** En el segundo modelo, destaca el retraso número 23 que sobresale de forma notable en ambos gráficos. También se pueden destacar las observaciones 16, 46 y 72 que rozan incluso sobrepasan las bandas.

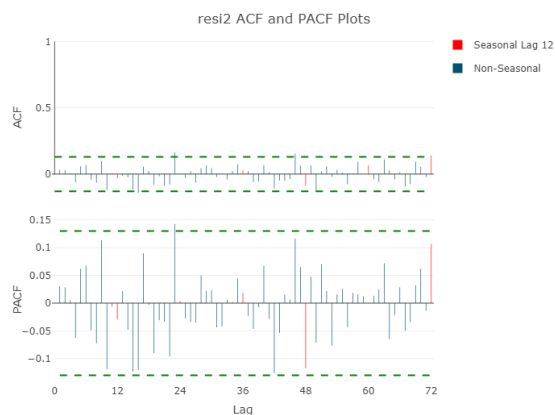


Figure 14: Gráfico para el análisis de independencia 2

Para comprobar los desfases se utilizará nuevamente el test de *Ljung-Box* C. El test muestra

que la observación 23 posee un p-valor superior al nivel de significación. Lo mismo sucede con las demás a excepción de la número 72, que se encuentra por debajo y no puede ser pasada por alto. A pesar de ser una observación lejana, este hecho podría dar indicios de que la hipótesis de independencia no se cumple tan bien como en el primer modelo, en el que para todos los residuos no puede ser rechazada la hipótesis de independencia.

## 2.4 Propiedad de causalidad y/o invertibilidad

Para comprobar la causalidad de un modelo se ha de ver que los pesos  $\psi_i$  de

$$\phi(B) = \frac{\theta_q(B)}{\phi_p(B)} \quad (10)$$

converjan, de modo que se pueda expresar como un  $MA(\infty)$ . Lo cuál será de gran utilidad a la hora del cálculo de la varianza de las predicciones puntuales realizadas en breve.

Para asegurar la invertibilidad de un modelo, se ha de comprobar, de manera análoga, que los pesos  $\pi_i$  de

$$\pi(B) = \frac{\phi_p(B)}{\theta_q(B)}$$

tengan convergencia, de modo que se obtenga una expresión  $AR(\infty)$ . Bajo esta propiedad y con el uso de observaciones pasadas será posible el cálculo de las predicciones puntuales en el apartado de predicciones.

### 2.4.1 Modelo 1

Al ver que todas las raíces del polinomio de grado 50 son mayores que la unidad, podemos confirmar que el Modelo 1 es causal.

Como solo hay una raíz para la componente de media móvil del modelo y tiene módulo mayor que 1, el modelo 1 también cumple con la condición de invertible. Para una aclaración visual mirar D.

### 2.4.2 Modelo 2

Por la misma razón podemos afirmar que el modelo 2 también es causal e invertible, ya que la inversa de todas las raíces caen dentro del círculo unidad.

## 2.5 Estabilidad

Para poder comprobar si los modelos son estables se debe verificar que el modelo propuesto para todas las observaciones de la serie temporal es similar, en cuanto a unas condiciones cualitativas, al modelo de *training*, conformado por la serie temporal sin las 12 últimas muestras. Por lo que se ha de comprobar si, al ajustar el modelo con la serie recortada se mantienen unos coeficientes muy similares con respecto a la magnitud, al signo y a su significación.

### 2.5.1 Modelo 1

Para el modelo 1, se mantienen los signos de los coeficientes y los errores y, además, las magnitudes de estos son muy similares, siendo las variaciones más significativas de apenas una centésima E. Para corroborar finalmente al estabilidad del modelo, hay que fijarse en que la significación de los parámetros apenas ha variado.

### 2.5.2 Modelo 2

Respecto al segundo modelo, a pesar que los signos no cambian, las magnitudes de los coeficientes varían más considerablemente que para el primer modelo. E Lo mismo sucede con las significaciones, que también presentan variaciones más drásticas a pesar de no cambiar el signo. Esto indica que el modelo 1 es más estable que el segundo.

## 2.6 Predicciones

Como el propósito del estudio explica, el objetivo es poder realizar predicciones de futuros valores de la serie temporal, conocida su última observación. Para lo cual, en este caso se realizarán las

predicciones puntuales de las últimas 12 observaciones separadas previamente, para comprobar el ajuste y analizar los intervalos de confianza.

Téngase en cuenta que al haber realizado una transformación **Box-Cox** al inicio de la identificación, será necesario deshacer estas transformaciones a las 12 observaciones predichas, al igual que a los límites del intervalo de confianza.

$$IC(X_{t+h})_{1-\alpha} = \tilde{X}_{t+h} \pm Z_{1-\alpha/2} \sqrt{\text{Var}(\tilde{X}_{t+h})},$$

$$h = \{1, \dots, 12\}$$
(11)

### 2.6.1 Modelo 1

Tras la predicción del modelo de *training* sobre las 12 observaciones de validación se observa que las predicciones de las muestras se alinean con las muestras observadas. Si bien en ocasiones puntuales difieren, los intervalos de confianza parecen ser considerablemente reducidos.

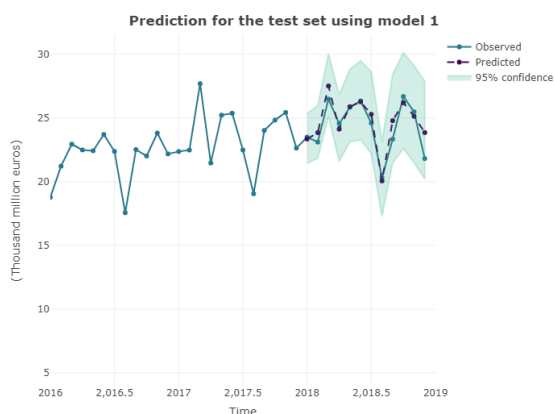


Figure 15: Predicciones modelo 1

Se nota el incremento de la amplitud entre los límites de confianza como es de esperar debido al incremento de la incertidumbre para predicciones puntuales más alejadas de la última observación.

### 2.6.2 Modelo 2

Sobre el segundo modelo se observa una mayor diferencia entre los valores predichos con respecto a los valores observado del último año.

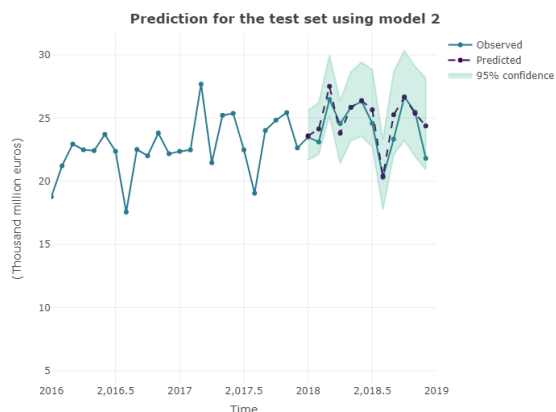


Figure 16: Predicciones modelo 2

Los intervalos de confianza parecen ser a simple vista similares al modelo 1. Con lo que se necesitarán de más recursos o métricas para evaluar los intervalos de confianza en ambos casos.

## 2.7 Selección

Llegados a este punto se seleccionará uno de los dos modelos propuestos para continuar con el desarrollo del reporte. Analizamos la exactitud con dos medidas de error (Root Mean Square Percentage Error y Mean Absolut Percentage Error) y como se ve en la tabla, el modelo 1 es mejor en este aspecto. Pues la calidad de predicción es superior, ya que los valores de los errores son menores.

	Modelo 1	Modelo 2
RMSPE	0.0377	0.0475
MAPE	0.0273	0.0313

Se busca un modelo con error mínimo, pues este indica una mayor exactitud a la hora de realizar las predicciones. En parte por eso se observa en el apartado 2.6 el porque las predicciones del modelo 1 se acercan en mayor medida a los valores reales observados de la serie.

En cuanto a la precisión de las predicciones, del modelo 1 se obtiene una media de 6.02 (en miles de millones) para la anchura del intervalo. Mientras que el segundo modelo cuenta con 5.74 de anchura media. Sin embargo, hay que tener en cuenta el hecho de que no se ha podido validar la varianza constante en ninguno de los dos

modelos, y esto podría afectar la certeza de los intervalos. Con lo que aunque la diferencia entre medidas es de magnitud pequeña, no es una garantía clara a la hora asegurar la precisión de las predicciones.

Fijándose en las medidas de adecuación a los datos, ambos modelos presentan un AIC muy similar, de  $-191.87$  y  $-191.36$  para el modelo 1 y 2 respectivamente. Con lo que la bondad de adecuación no difiere mucho dependiendo del modelo seleccionado. Pero, como se ha visto a lo largo del proyecto, el modelo 1 tiene un parámetro más (no significativo, a pesar de que se ha comprobado que eliminarlo empeora la adecuación del modelo) en comparación del modelo 2, con lo que el BIC de  $-164.47$  es inferior al segundo;  $-167.38$ . Premiando en complejidad, el segundo modelo puede considerarse más adecuado.

Por otro lado, el ambos modelos son causales e invertibles. El modelo 1 presenta una mejor propiedad de estabilidad al haber menos fluctuación. A partir del apartado de validación del modelo, se ha visto que el modelo 1 cumple con todas las condiciones a excepción de la varianza constante por culpa del número de outliers. Mientras que el segundo modelo no satisface de forma clara las condiciones impuestas. La decisión pues, resulta evidente el decantarse por el **modelo 1**.

## 2.8 Tratamiento de efectos de calendario

En este apartado se buscará mejorar las predicciones del modelo 1 mediante un tratamiento para identificar los efectos de Semana Santa y días laborales. A tal fin se ajustaron cuatro modelos de las combinaciones de los efectos para escoger el que presente una mejora en términos de AIC. El modelo escogido incluye ambos coeficientes de los efectos (ambos con significación) y un AIC total de  $-304.86$  F.

En las secciones siguientes, por restricciones de número de páginas, no se detallarán como anteriormente los pasos de re-identificación (ni para los efectos de calendario ni atípicos). Más detalle de gráficas en los apartados del anexo F.

### 2.8.1 Re-identificación y re-validación tras efectos de calendario

Tras linearizar la serie temporal, se ha realizado una re-identificación sobre la serie para constatar si se mantiene el mismo modelo propuesto hasta el momento. La serie linealizada ya se le ha aplicado la transformación Box-Cox, además es necesaria una diferenciación estacional y una regular. Con la que se obtiene una varianza de  $0.0251$  (reducción de aproximadamente  $0.04$  sobre la serie original).

Tras el análisis del ACF y PACF de la serie linearizada, los modelos propuestos son los siguientes:

$$\begin{aligned} \text{ARIMA}(0, 0, 1)(0, 0, 1)_{12} \\ \text{ARIMA}(2, 0, 0)(0, 0, 1)_{12} \end{aligned} \quad (12)$$

Como el segundo modelo presenta mejores cualidades de AIC (y todos los parámetros son significativos), será el seleccionado para pasar a la fase de re-validación. En este caso, respecto a la varianza, el comportamiento es muy similar al del primer modelo teniendo presencia de únicamente dos outliers, por lo que si no los tenemos en cuenta, se puede considerar como constante. En el caso de la hipótesis de normalidad, casi todos los residuos están dentro de los intervalos de confianza y tienen un ajuste casi perfecto a la recta de normalidad a excepción de un par de atípicos. Sin embargo el test de *Shapiro-Wilk* rechaza la hipótesis de normalidad. En cuanto a la independencia de residuos, solamente hay un retraso estacional que roza el límite del intervalo, pero demasiado alejado del origen. Con los pvalores del test de *Ljung-Box* se confirma la independencia de los residuos.

El modelo cumple con las propiedades de causalidad e invertibilidad, además de ser estable.

### 2.8.2 Predicciones

Los intervalos de confianza se han reducido en  $0.62$  unidades (a  $5.41$ ), lo que supone una mejora en la precisión.

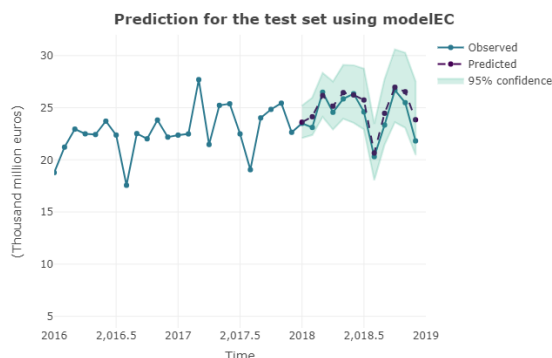


Figure 17: Predicciones con modelo de efectos de calendario

A pesar de presentar una ligera pérdida de exactitud con respecto a las medidas de error ( $RM-SPE = 0.0393$  y  $MAPE = 0.0307$ ), el AIC ha disminuido a  $-310.71$ , indicando una mejor adecuación.

## 2.9 Tratamiento de valores atípicos

Tal y como se ha hecho con los efectos de calendario, se llevará a cabo un tratamiento de atípicos con la intención de obtener la serie linealizada (sin presencia de atípicos) y así poder re-identificar un nuevo modelo que, idealmente, se adecue mejor y nos lleve a unas predicciones más exactas y precisas.

Para esta labor, se usará el programa 'outdetect' brindado por el supervisor de este proyecto. Este programa encontrará los atípicos de los tipos **AO** (*Additive Outlier*), **TC** (*Transitory Change*) y **LS** (*Level Shift*), presentes tras haber tratado ya los efectos de calendario.

Obs	type_detected	W_coeff	ABS_L_Ratio	Fecha	PercVar
<int>	<chr>	<dbl>	<dbl>	<chr>	<dbl>
5	23 TC	0.2464986	3.090202	Nov 2000	127.95374
4	108 AO	-0.2560540	3.344374	Dic 2007	77.41002
2	114 AO	-0.2897433	3.620892	Jun 2008	74.84556
3	118 LS	-0.2674118	3.233383	Oct 2008	76.53578
1	120 LS	-0.5841314	6.732292	Dic 2008	55.75900
6	136 AO	-0.2316833	3.146972	Abr 2010	79.31973
7	171 AO	0.2265319	3.129537	Mar 2013	125.42426

Figure 18: Atípicos identificados por el programa 'outdetect'

Como se observa en la salida del programa 'outdetect', se han identificado un total de 7 outliers, siendo el más importante y el causante un cambio estructural en la serie el Level Shift pro-

ducido en diciembre de 2008, probablemente por la crisis financiera.

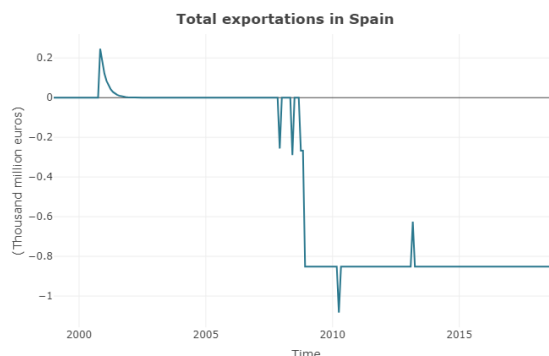


Figure 19: Efecto de los atípicos sobre la serie temporal

A continuación se muestra la serie linealizada sobre la original para observar con mejor detalle los efectos de estos atípicos.

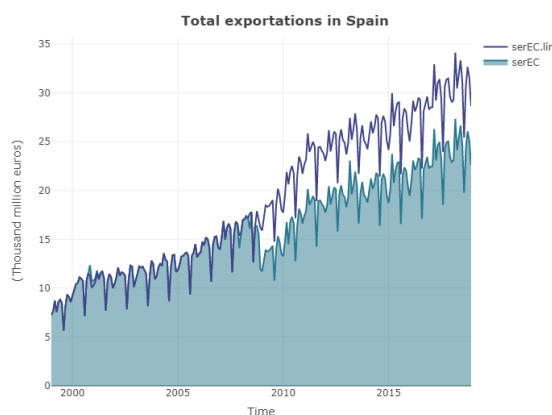


Figure 20: Serie linealizada vs Serie original

### 2.9.1 Re-identificación y re-validación tras valores atípicos

Siguiendo los mismos pasos de la metodología Box-Jenkins que se han visto a lo largo del informe (gráficos y outputs disponibles en el anexo), se ha realizado las mismas transformaciones que a la serie original. Tras eliminar el efecto de los atípicos, se han re-identificado los dos modelos siguientes:

$$\begin{aligned} & \text{ARIMA}(2, 1, 0)(0, 1, 1)_{12} \\ & \text{ARIMA}(0, 1, 1)(0, 1, 1)_{12} \end{aligned} \quad (13)$$



En vista de que el segundo modelo presenta un AIC ligeramente menor al primero y todos sus parámetros son significativos, se seleccionará este modelo como el mejor.

Además, en el proceso de validación se concluye que se cumple la varianza constante, la normalidad y la independencia de los residuos de forma mucho más adecuada que en los previos modelos. También, el nuevo modelo es causal e invertible, y estable.

### 2.9.2 Predicciones

Tras haber validado el modelo y haber comprobado sus propiedades, se procede a hacer predicciones de las últimas 12 muestras con el objetivo de ver si el modelo, además de adecuarse mejor a los datos, obtiene una mayor exactitud y precisión en las predicciones.

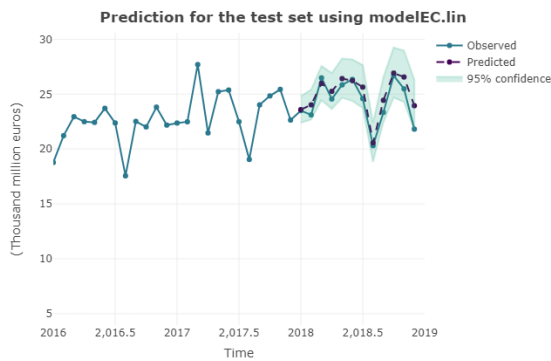


Figure 21: Predicciones con la serie linealizada

Al haber eliminado los atípicos y los efectos de calendario, las medidas de exactitud prácticamente no han cambiado ( $\text{RMSPE} = 0.0397$  y  $\text{MAPE} = 0.0307$ ), pero cabe destacar que el grosor medio de los intervalos de confianza se ha reducido en aproximadamente 1.7 unidades (a 3.70), mejorando nuevamente la precisión de las predicciones. Un punto a destacar es que se ha aumentado muy notablemente la adecuación, con un AIC de  $-416.96$ .

## 3 Conclusiones y discusión

La metodología Box-Jenkins y los diversos procedimientos como el tratamiento de efectos de

calendario o el tratamiento de atípicos nos han permitido estimar nuevos modelos que, si bien no han mejorado la exactitud de los primeros modelos propuestos, si que han aumentado la precisión a la hora de hacer predicciones y también la adecuación del modelo a los datos. Dicho esto, si lo que se busca es obtener las predicciones más exactas a corto plazo, el mejor modelo resulta ser el modelo antes del tratamiento de efectos de calendario y de atípicos. Si, en cambio, se busca una buena adecuación a los datos y además una mayor precisión de las predicciones, perdiendo muy ligeramente exactitud en el proceso a corto plazo, la mejor opción es el modelo tras tratar los efectos de calendario y los atípicos. El modelo tratando únicamente los efectos de calendario no supone ninguna ventaja considerable sobre el modelo habiendo tratado estos y los atípicos, por lo que para las situaciones propuestas no resulta ser el mejor modelo.

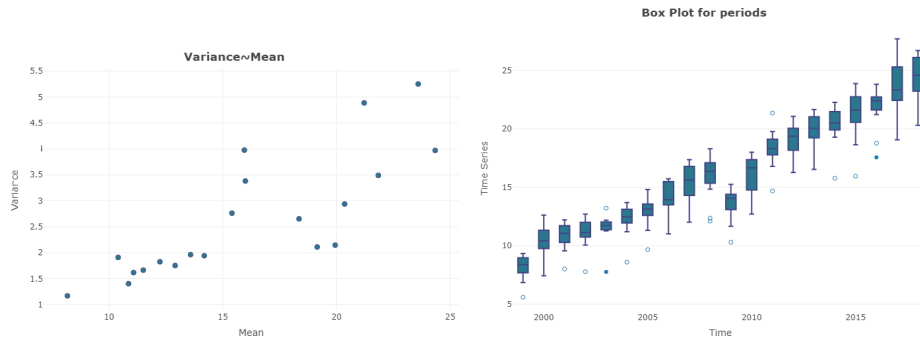
En términos generales, el mejor modelo resulta el que incluye ambos tratamientos.

# Appendices

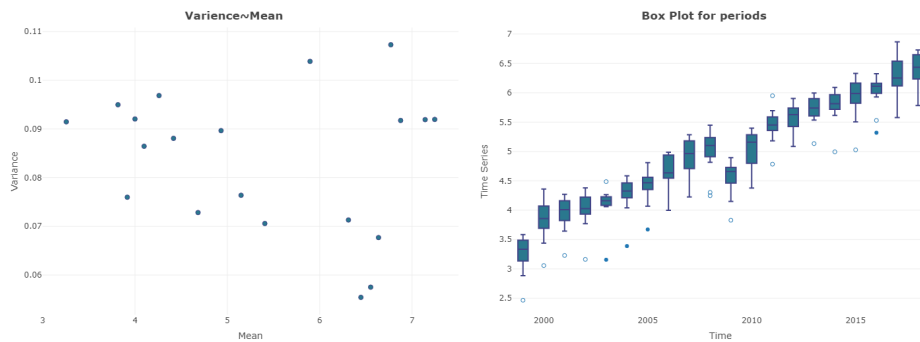
## A Transformación de la serie

### A.1 Varianza constante

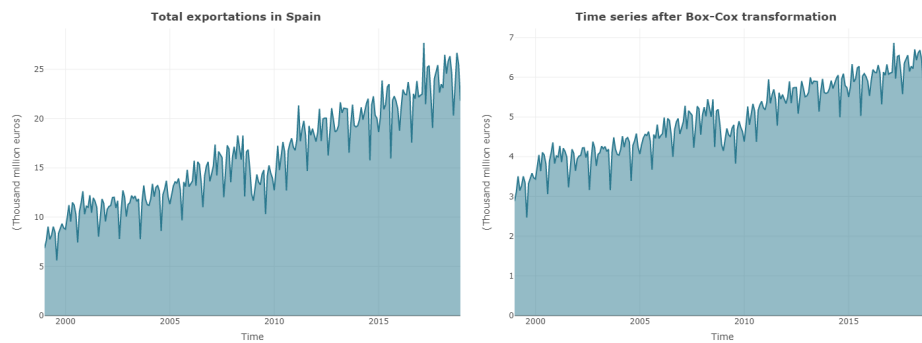
Gráficos de comprobación de variancia constante antes de transformación Box-cox.



Gráficos de comprobación de variancia constante después de transformación Box-Cox.

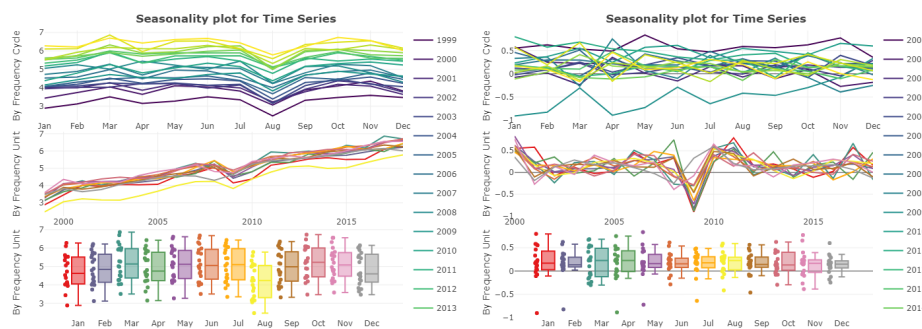


Serie temporal antes y después de aplicar transformación Box-Cox.

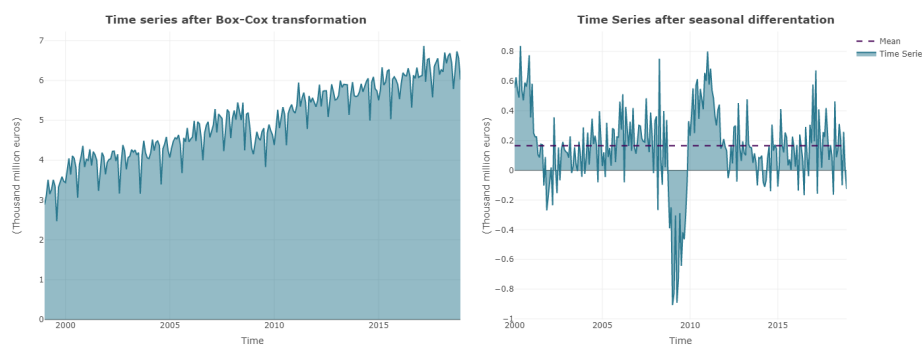


## A.2 Patrón estacional

Gráficos de identificación del patrón estacional. Antes y después de aplicar la diferenciación estacional.

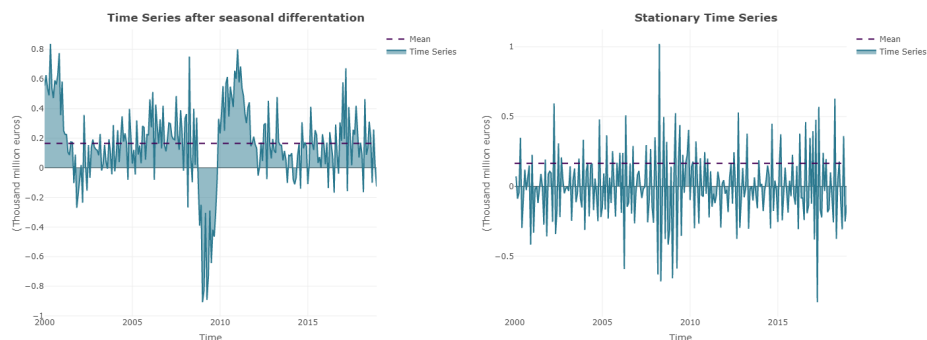


Serie temporal antes y después de aplicar la diferenciación estacional.



### A.3 Media constante

Serie temporal antes y después de aplicar la diferenciación regular.



## B Tablas de estimación

### B.1 Modelo 1

```
Call:
arima(x = d1d12bcser, order = c(2, 0, 1), seasonal = list(order = c(4, 0, 0),
  period = 12))
```

```
Coefficients:
      ar1      ar2      ma1      sar1      sar2      sar3      sar4  intercept
-0.9582 -0.4938  0.2302 -0.6011 -0.6407 -0.4500 -0.4240  -0.0009
s.e.    0.1339  0.0844  0.1499  0.0632  0.0719  0.0699  0.0662   0.0017
```

```
sigma^2 estimated as 0.02175: log likelihood = 104.06, aic = -190.12
```

Figure 22: Modelo 1 con estimación de la media

```
Call:
arima(x = bcser, order = c(2, 1, 1), seasonal = list(order = c(4, 1, 0), period =
12),
  fixed = c(NA, NA, 0, NA, NA, NA, NA))
```

```
Coefficients:
      ar1      ar2      ma1      sar1      sar2      sar3      sar4
-0.7602 -0.3825    0 -0.5954 -0.6411 -0.4560 -0.4251
s.e.    0.0657  0.0671    0  0.0630  0.0712  0.0691  0.0660
```

```
sigma^2 estimated as 0.02197: log likelihood = 102.87, aic = -191.75
```

Figure 23: Modelo 1 eliminando parámetros no significativos

```
Call:
arima(x = bcser, order = c(2, 1, 1), seasonal = list(order = c(4, 1, 0), period =
12))

Coefficients:
      ar1      ar2      ma1      sar1      sar2      sar3      sar4
-0.9588 -0.4938  0.2315 -0.6009 -0.6412 -0.4502 -0.4243
s.e.    0.1338  0.0843  0.1498  0.0632  0.0720  0.0699  0.0662

sigma^2 estimated as 0.02177:  log likelihood = 103.93,  aic = -191.87
```

Figure 24: Modelo 1 final

## B.2 Modelo 2

Se confirma el hecho que ya se había comentado en el apartado 2.2 de Estimación, la media no es significativa.

```
Call:
arima(x = d1d12bcser, order = c(2, 0, 0), seasonal = list(order = c(2, 0, 2),
period = 12))

Coefficients:
      ar1      ar2      sar1      sar2      sma1      sma2  intercept
-0.7863 -0.4274  0.5090 -0.3275 -1.2191  0.3715      -6e-04
s.e.    0.0639  0.0651  0.1744  0.1149  0.1663  0.1849      1e-03

sigma^2 estimated as 0.02189:  log likelihood = 102.85,  aic = -189.7
```

Figure 25: Modelo 2 con estimación de la media

Ahora es correcto ajustar el modelo con las diferenciaciones como parámetros de la función ARIMA:

```
Call:
arima(x = bcser, order = c(2, 1, 0), seasonal = list(order = c(2, 1, 2), period =
12))

Coefficients:
      ar1      ar2      sar1      sar2      sma1      sma2
-0.7855 -0.4268  0.5073 -0.3287 -1.2173  0.3704
s.e.    0.0639  0.0650  0.1745  0.1130  0.1667  0.1840

sigma^2 estimated as 0.02192:  log likelihood = 102.68,  aic = -191.36
```

Figure 26: Modelo 2 final

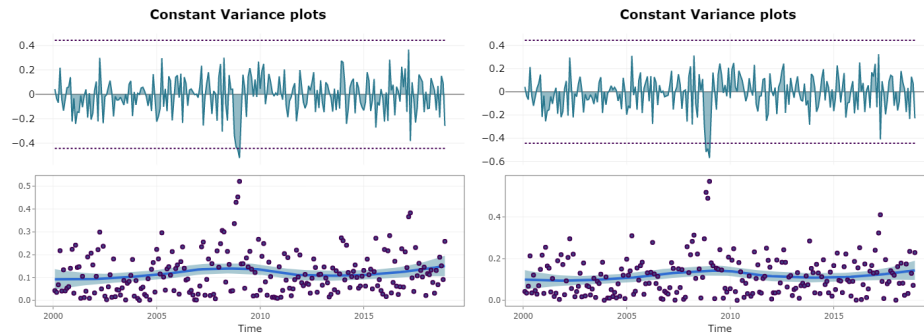
La disminución del AIC muestra que es acertado suponer que la media no es significativa. Esto también se observa en los valores de los coeficientes del modelo, que apenas varían tras la eliminación de la media. El siguiente paso es ver si el resto de los coeficientes son significativos.

Fijándose en los  $t$ -ratios, se contempla que todos tienen un valor absoluto por encima de 2, indicando que el resto de parámetros si tienen significación en el modelo.

## C Validación

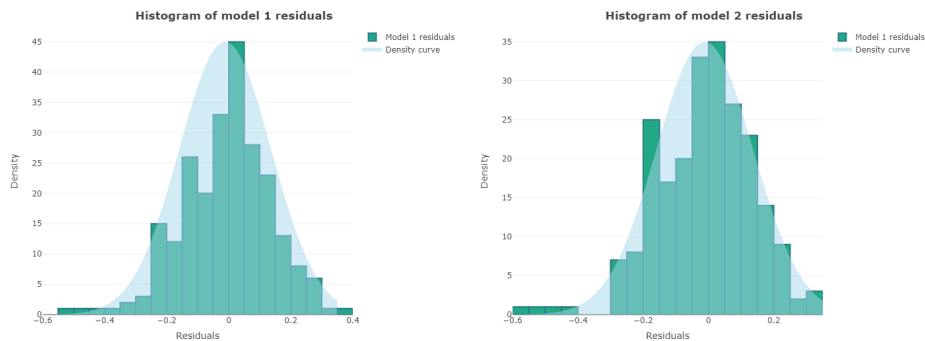
### C.1 Varianza constante

Gráficos de comprobación de varianza constante para los modelos 1 y 2 respectivamente.

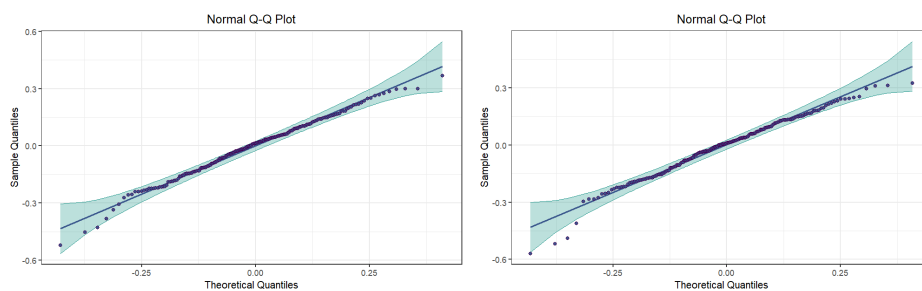


### C.2 Normalidad de residuos

Gráficos de comprobación de normalidad de residuos para los modelos 1 y 2 (izquierda y derecha) respectivamente. Entre los cuales se muestra el histograma de los residuos, el QQ-plot y el test de *Shapiro-Wilk*.



Histogramas de residuos del modelo 1 y 2 (izquierda y derecha)



Q-Q plot de los residuos del modelo 1 y 2 (izquierda y derecha)

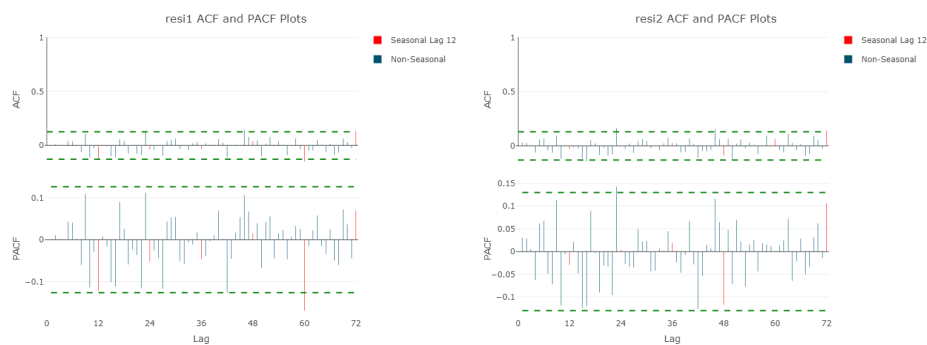
```
Shapiro-Wilk normality test
data: resi1
W = 0.99087, p-value = 0.1657
```

```
Shapiro-Wilk normality test
data: resi2
W = 0.98054, p-value = 0.003228
```

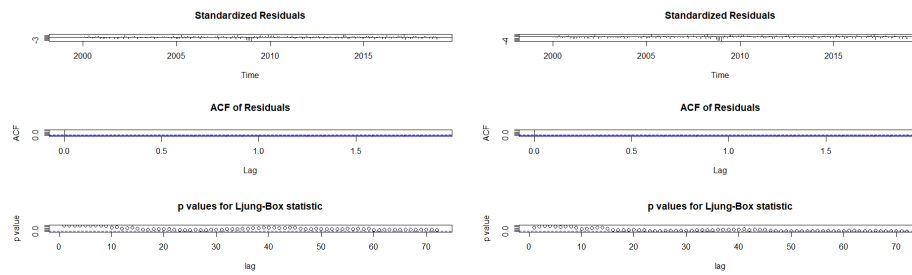
Test de Shapiro-Wilk para el modelo 1 y 2 (izquierda y derecha)

### C.3 Independencia de residuos

Gráficos de comprobación de independencia de residuos para el modelo 1 y 2 (izquierda y derecha) respectivamente. Para los cuales, en primer lugar está el ACF y PACF de cada uno de los residuos y el test estadístico de *Ljung-Box*.



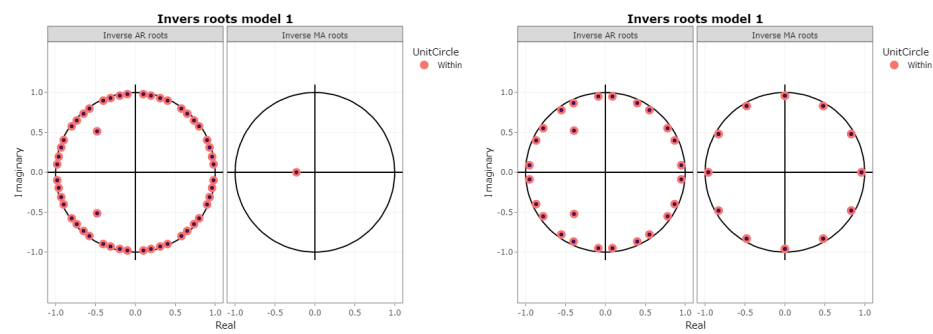
ACF y PACF de residuos del modelo 1 y 2 (izquierda y derecha)



Test de Ljung-Box de residuos del modelo 1 y 2 (izquierda y derecha)

## D Propiedades

Gráficos para visualizar las raíces inversas de los modelos 1 y 2 respectivamente. Todas las raíces están dentro del círculo unitario en ambos modelos.



Raíces inversas del modelo 1 y 2 (izquierda y derecha)



## E Estabilidad

### E.1 Modelo 1

```
Call:
arima(x = bctrain, order = c(2, 1, 1), seasonal = list(order = c(4, 1, 0), period
= 12))
```

Coefficients:

	ar1	ar2	ma1	sar1	sar2	sar3	sar4
	-0.9572	-0.4874	0.2414	-0.5975	-0.6496	-0.4421	-0.4189
s.e.	0.1361	0.0857	0.1524	0.0680	0.0741	0.0723	0.0678

sigma^2 estimated as 0.02203: log likelihood = 96.83, aic = -177.66

```
Call:
arima(x = bcser, order = c(2, 1, 1), seasonal = list(order = c(4, 1, 0), period =
12))
```

Coefficients:

	ar1	ar2	ma1	sar1	sar2	sar3	sar4
	-0.9588	-0.4938	0.2315	-0.6009	-0.6412	-0.4502	-0.4243
s.e.	0.1338	0.0843	0.1498	0.0632	0.0720	0.0699	0.0662

sigma^2 estimated as 0.02177: log likelihood = 103.93, aic = -191.87

Figure 27: Comparación de parámetros modelo 1

Comparación de la significación de los parámetros del modelo train y conjunto para el modelo 1.

```
T-ratios: -7.03 -5.69 1.58 -8.78 -8.77 -6.12 -6.18
T-ratios: -7.17 -5.86 1.55 -9.51 -8.91 -6.44 -6.41
```

Figure 28: Comparación de significación parámetros modelo 1

## E.2 Modelo 2

```
Call:
arima(x = bctrain, order = c(2, 1, 0), seasonal = list(order = c(2, 1, 2), period = 12))
```

Coefficients:

	ar1	ar2	sar1	sar2	sma1	sma2
	-0.7643	-0.3976	0.6071	-0.4351	-1.3039	0.5273
s.e.	0.0658	0.0683	0.1267	0.1078	0.1169	0.1338

sigma^2 estimated as 0.02193: log likelihood = 96.87, aic = -179.73

```
Call:
arima(x = bcser, order = c(2, 1, 0), seasonal = list(order = c(2, 1, 2), period = 12))
```

Coefficients:

	ar1	ar2	sar1	sar2	sma1	sma2
	-0.7855	-0.4268	0.5073	-0.3287	-1.2173	0.3704
s.e.	0.0639	0.0650	0.1745	0.1130	0.1667	0.1840

sigma^2 estimated as 0.02192: log likelihood = 102.68, aic = -191.36

Figure 29: Comparación de parámetros modelo 2

Comparación de la significación de los parámetros del modelo train y conjunto para el modelo 2.

T-ratios: -11.62 -5.82 4.79 -4.04 -11.15 3.94  
 T-ratios: -12.3 -6.56 2.91 -2.91 -7.3 2.01

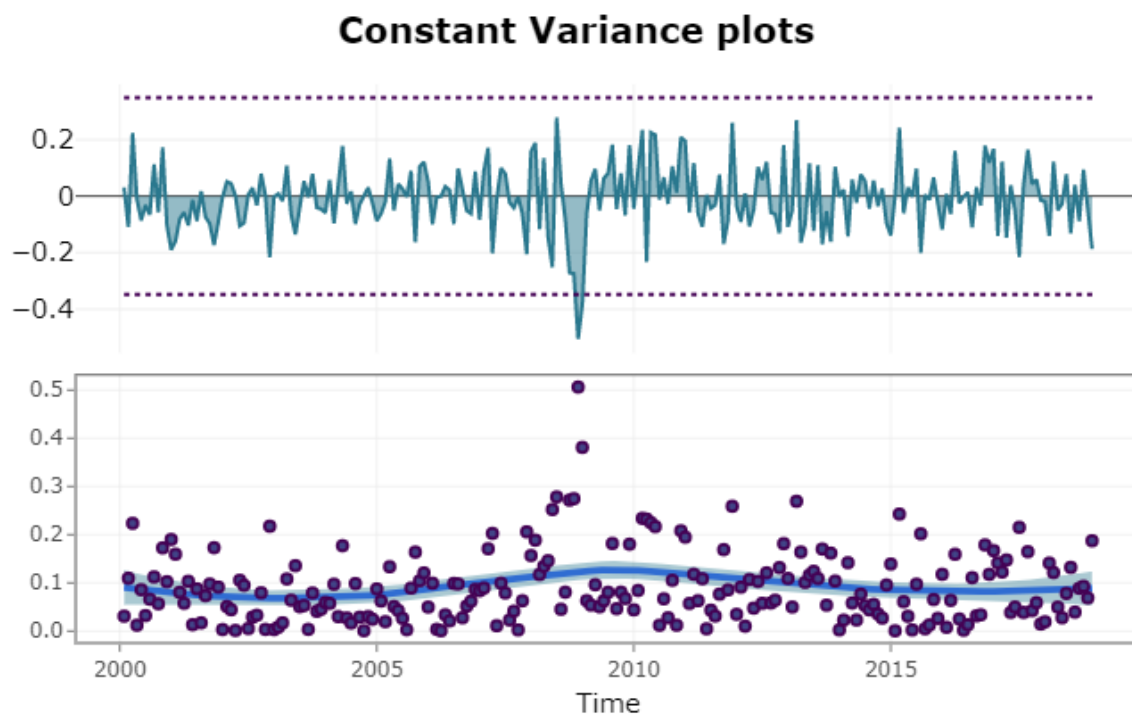
Figure 30: Comparación de significación parámetros modelo 2

## F Tratamiento de efectos de calendario

### F.1 Re-identificación y re-validación tras efectos de calendario

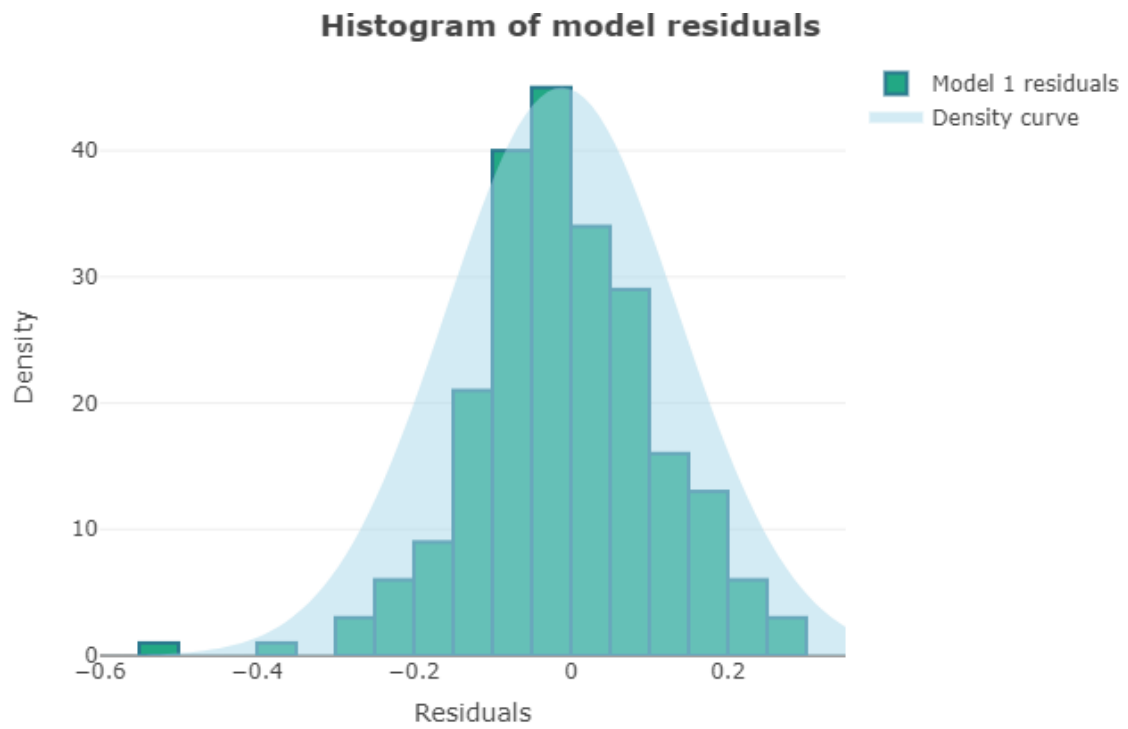
#### F.1.1 Varianza constante

Gráficos de comprobación de varianza constante para el modelo de la serie.

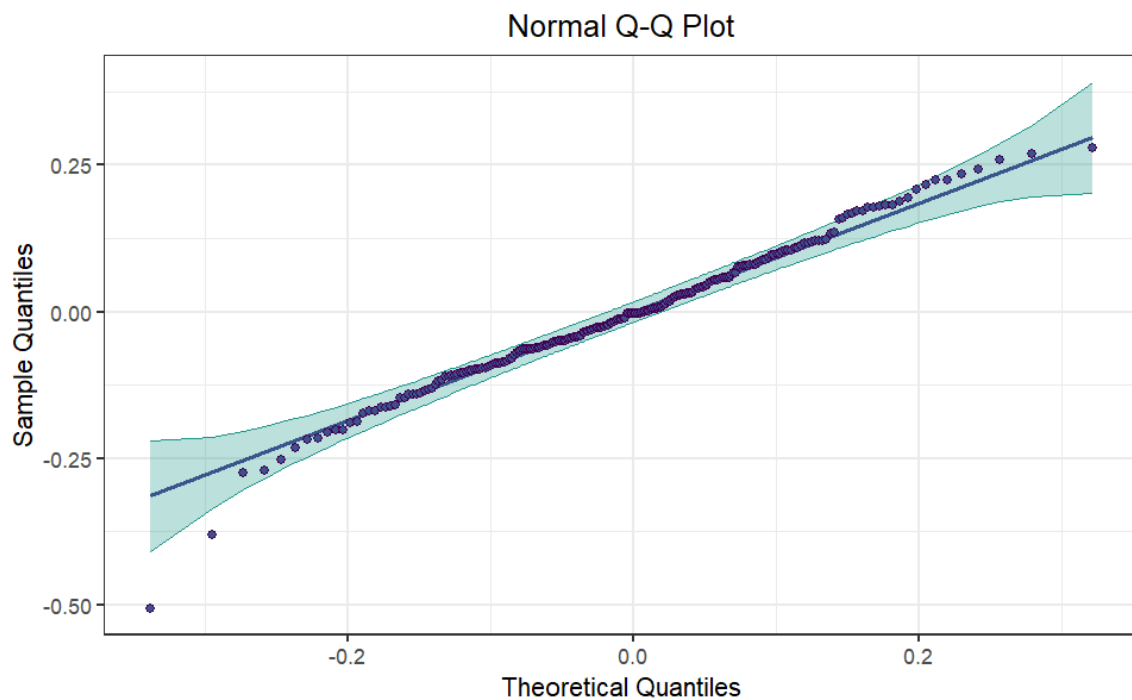


### F.1.2 Normalidad de residuos

Gráficos de comprobación de normalidad de residuos para el modelo de la serie. Entre los cuales se muestra el histograma de los residuos, el QQ-plot y el test de *Shapiro-Wilk*.



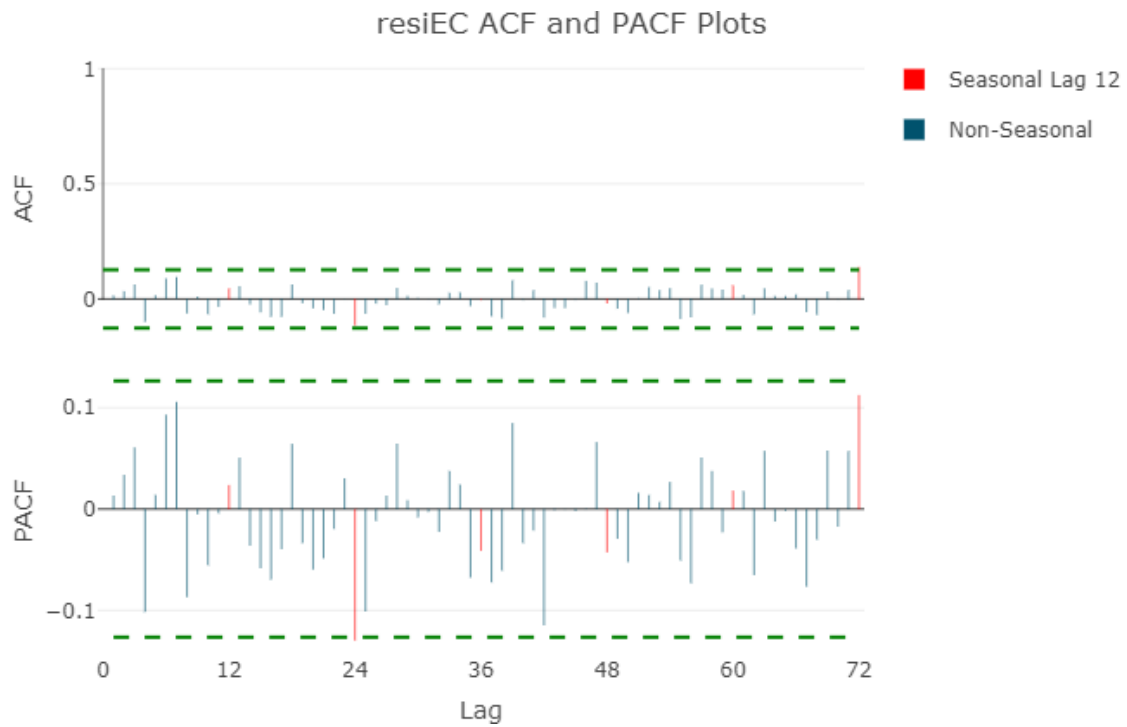
Histogramas de residuos para el modelo de la serie.



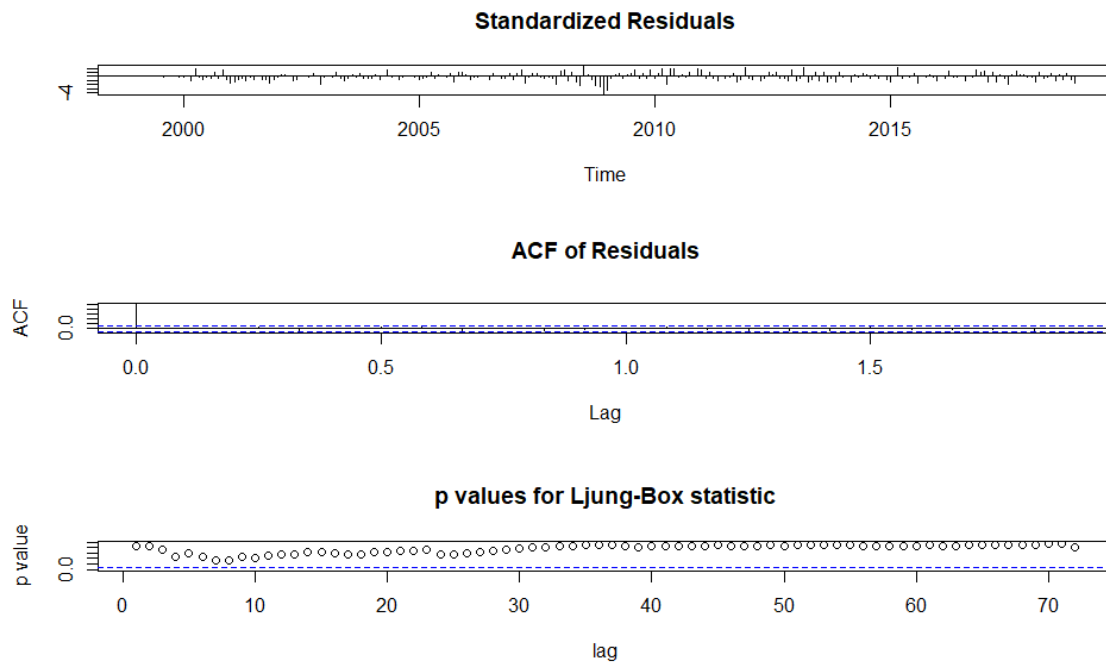
Q-Q plot de los residuos para el modelo de la serie.

### F.1.3 Independencia de residuos

Gráficos de comprobación de independencia de residuos para el modelo de la serie. Para los cuales, en primer lugar está el ACF y PACF de cada uno de los residuos y el test estadístico de *Ljung-Box*.



ACF y PACF de residuos para el modelo de la serie.



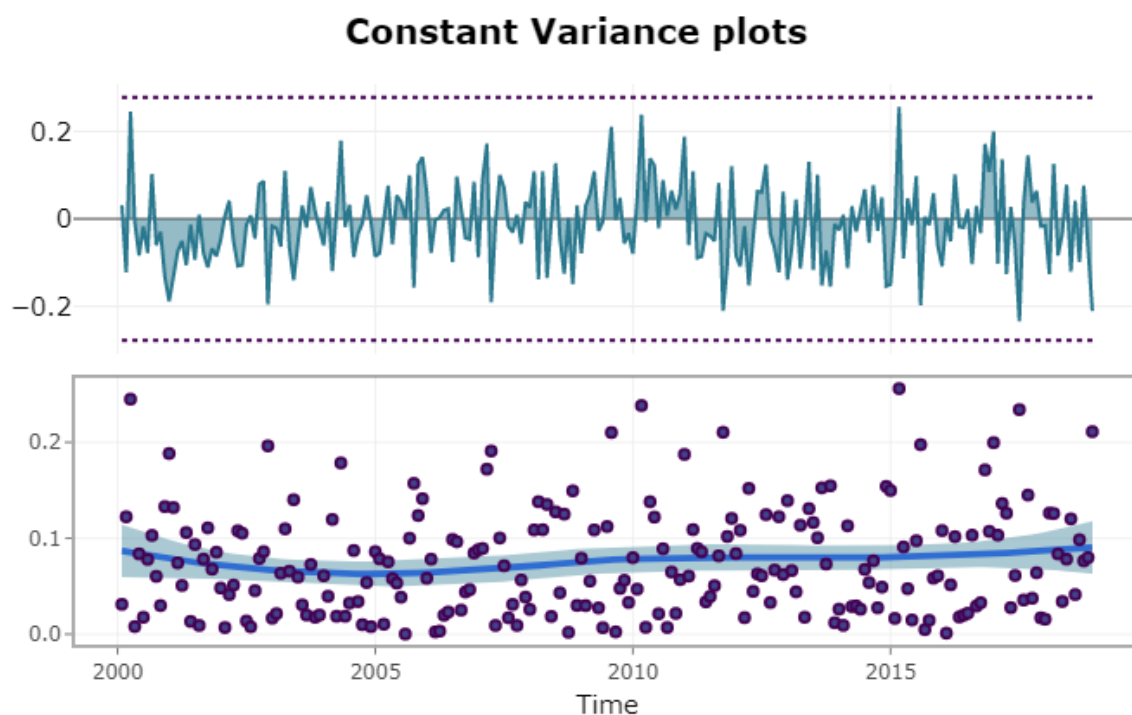
Test de Ljung-Box de residuos para el modelo de la serie.

## G Tratamiento de atípicos

### G.1 Re-identificación y re-validación tras tratamiento de atípicos

#### G.1.1 Varianza constante

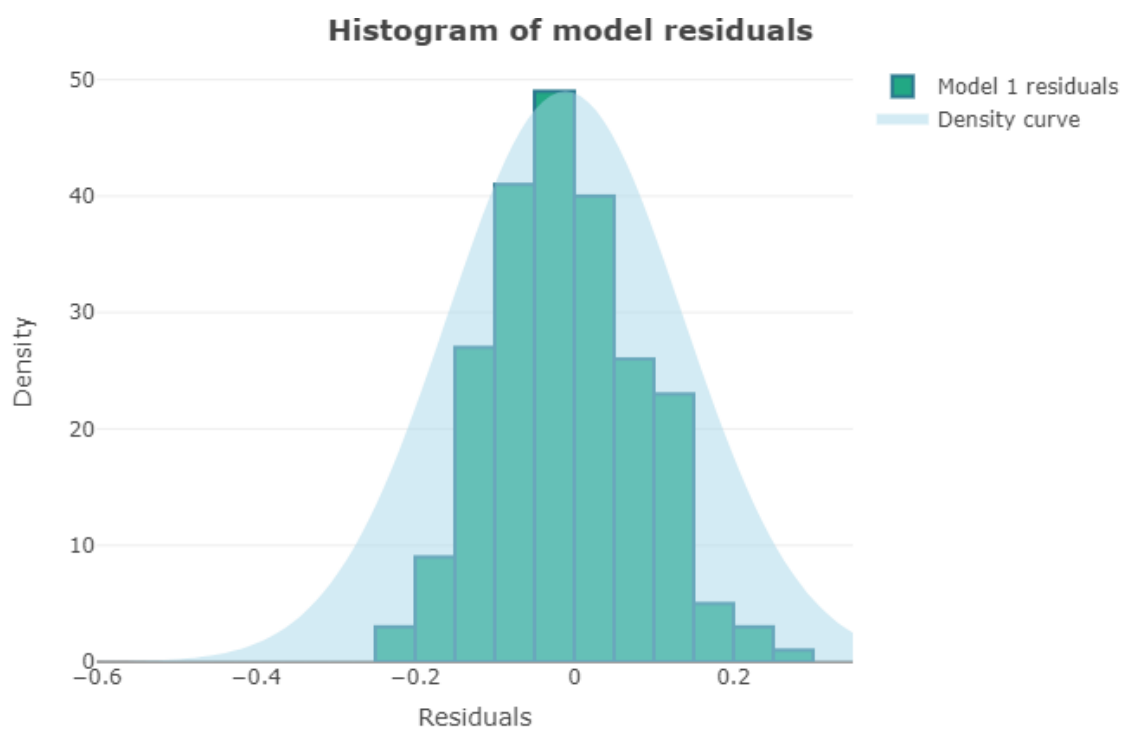
Gráficos de comprobación de varianza constante para el modelo de la serie.



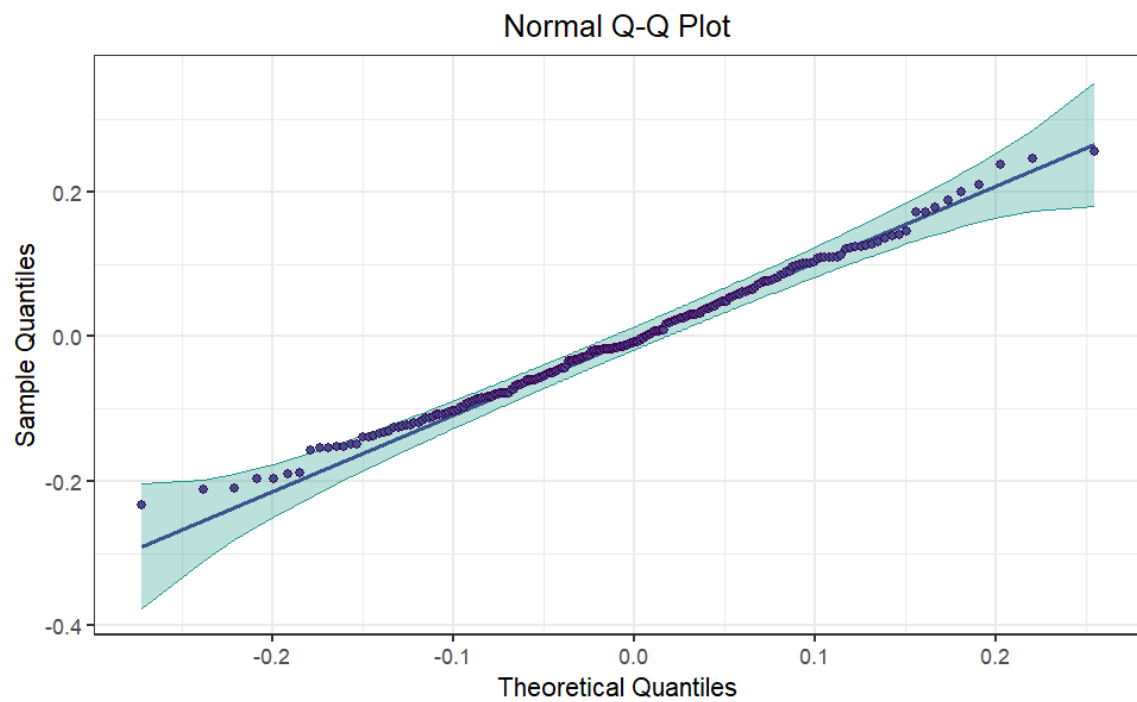
### G.1.2 Normalidad de residuos

Gráficos de comprobación de normalidad de residuos para el modelo de la serie. Entre los cuales se muestra el histograma de los residuos, el QQ-plot y el test de *Shapiro-Wilk*.





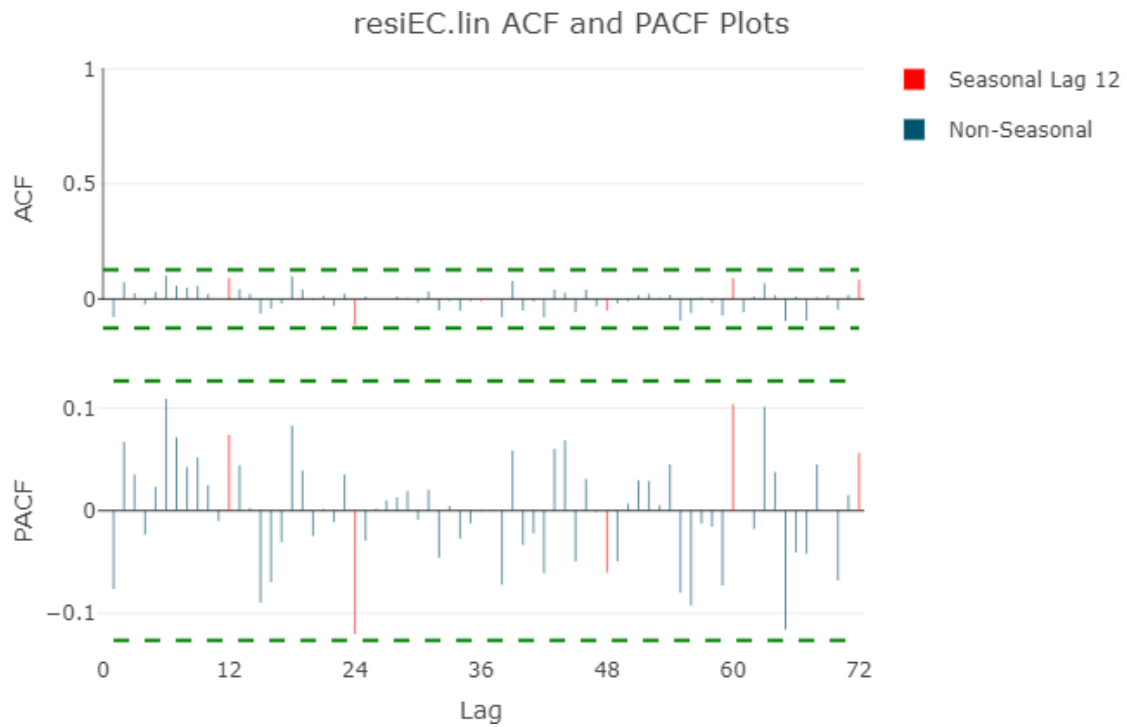
Histogramas de residuos para el modelo de la serie.



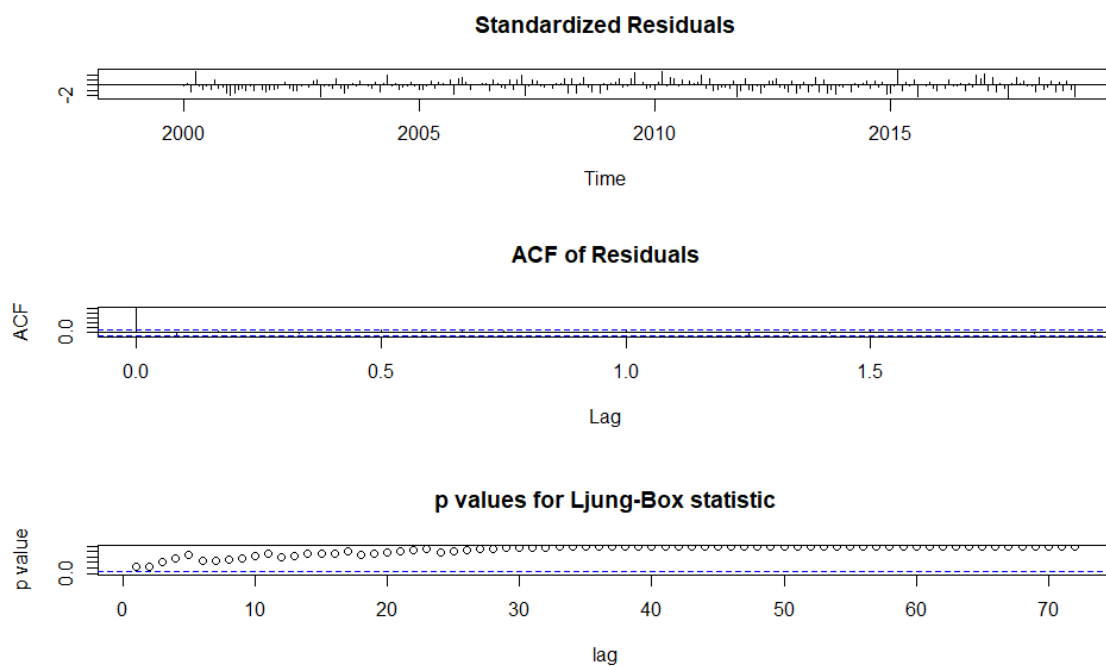
Q-Q plot de los residuos para el modelo de la serie.

### G.1.3 Independencia de residuos

Gráficos de comprobación de independencia de residuos para el modelo de la serie. Para los cuales, en primer lugar está el ACF y PACF de cada uno de los residuos y el test estadístico de *Ljung-Box*.



ACF y PACF de residuos para el modelo de la serie.



Test de Ljung-Box de residuos para el modelo de la serie.