

Número de pasajeros en el Puerto de Barcelona (2012-2019)

Segona pràctica Anàlisi de Series Temporals

Garcia Sató, Pol
Nualart Sanz, Arnau

02 gener, 2023



UNIVERSITAT DE
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

El objetivo de este informe es ver el comportamiento del número de pasajeros en el puerto de Barcelona durante los años 2012 - 2019 y hacer una predicción para el período 2020 - 2022 si no se tuvieran en cuenta los efectos causados por la pandemia del SARS-CoV-2. Se observa que los datos presentan tendencia creciente y estacionalidad multiplicativa. Es decir, con el paso de los años la demanda fue aumentando. Además, hay un claro patrón estacional que muestra que en los meses de verano el número es superior a los meses de otoño e invierno. En cuanto a la predicción, la demanda sigue aumentando de manera multiplicativa y siguiendo el mismo patrón estacional.

Índex

Introducción	1
Descripción del entorno y el problema - motivación del trabajo	1
Definición precisa de la variable	1
Descripción de la fuente de información	1
Resumen de la estructura del trabajo	2
Aplicación empírica	3
Análisis descriptivo de la serie	3
Identificación del modelo	3
Transformación de la serie temporal en una serie estacionaria	3
Identificación de modelos pausibles	7
Forma compacta y usando el operador de retardo B de los modelos identificados	8
Estimación de los modelos	8
Modelo 1A, MA(1)SMA(1) ₁₂	8
Modelo 1B, usando la serie no-estacionaria $\ln(X_t)$ (lnserie)	9
Modelo 2A, AR(2)SMA(1) ₁₂	9
Modelo 2B, usando la serie no-estacionaria $\ln(X_t)$ (lnserie)	9
Modelo 3A, MA(1)SAR(1) ₁₂	10
Modelo 3B, usando la serie no-estacionaria $\ln(X_t)$ (lnserie)	10
Modelo 4A, AR(2)SAR(1) ₁₂	10
Modelo 4B, usando la serie no-estacionaria $\ln(X_t)$ (lnserie)	11
Tabla resumen	11
Validación de los modelos ajustados	13
Validación modelo 1A	13
Validación modelo 2A	15
Predicción de modelos ARIMA ajustados y validados	18
Modelo 1A	18
Verificación estabilidad del modelo	18
Predicción out-of-sample y cálculo RMSPE/MAPE	18
Modelo 2A	19
Verificación estabilidad del modelo	19
Predicción out-of-sample y cálculo RMSPE/MAPE	20
Selección del mejor modelo	21
Realización predicción a largo plazo	21

Conclusiones	23
Anexo	24

Introducción

Descripción del entorno y el problema - motivación del trabajo

Desde el estallido de la pandemia del SARS-CoV-2 las preferencias de la población respecto a sus destinos vacacionales parecen haber cambiado. Actualmente, los viajeros están más interesados en lugares naturales, al aire libre, debido a que en la época del confinamiento mucha parte de la población se sintió agobiada dentro de sus hogares. Adicionalmente, los españoles se han centrado en un turismo nacional debido a la situación de crisis sanitaria que estamos viviendo. Así pues, el número de pasajeros en el Puerto de Barcelona puede haber disminuido drásticamente ya que los cruceros son espacios cerrados y suelen ser vacaciones internacionales.

El objetivo principal de este informe es ver cuál era el comportamiento del número de pasajeros en el Puerto de Barcelona, ya sean de cruceros como de ferris, en los años anteriores a la pandemia. De esta manera se podrá ver de manera correcta la tendencia creciente que se tenía antes de la crisis sin tener en cuenta la posible caída debida a la pandemia.

Para la elección de los años analizados, se ha evitado estudiar aquellos momentos temporales en los que había un factor condicionante. Por tanto, los datos tomados dejan fuera la crisis económica del 2008 y la pandemia del 2020. También tendremos en cuenta que a principios del 2012 hubo el desastre de la naviera italiana Costa Concordia con 32 muertos, lo cual hizo que hubiera un descenso significativo en el número de pasajeros en los cruceros del Mediterráneo. Teniendo esto en cuenta, solo se han seleccionado los datos mensuales pertenecientes al espacio temporal comprendido entre 2012 y 2019, ambos períodos incluidos.

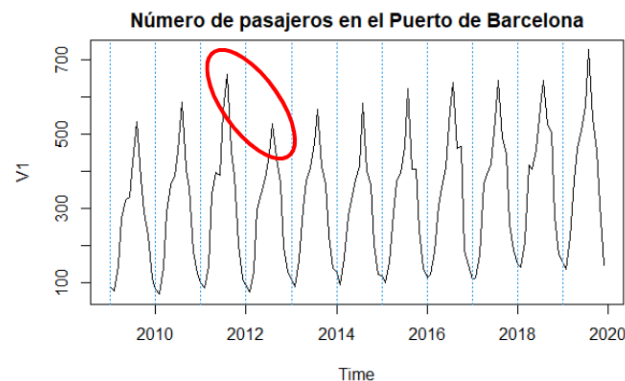


Figura 1. src*='#center'

Definición precisa de la variable

La variable a estudiar es el número de pasajeros que llegan al Puerto de Barcelona durante el periodo ya comentado anteriormente. Cabe destacar que esta variable recoge tanto el número de llegadas internas (nacionales) como externas (internacionales).

Descripción de la fuente de información

Los datos han sido extraídos de los casos propuestos del repositorio que precisamos en el CV de la UB. Proviene de la web del Ministerio del Fomento (<http://www.fomento.gob.es>) en el apartado Ministerio de Fomento / Marítimo / Información Estadística y corresponden al número de pasajeros que llegan al Puerto de Barcelona.

Resumen de la estructura del trabajo

El estudio sigue el siguiente esquema:

- *Aplicación empírica*

En este apartado se observan cuáles son las características principales de la serie (media, varianza, tendencia, estacionalidad, residuos, . . .). Es decir, se analiza descriptivamente la serie para así tener un conocimiento previo de los datos.

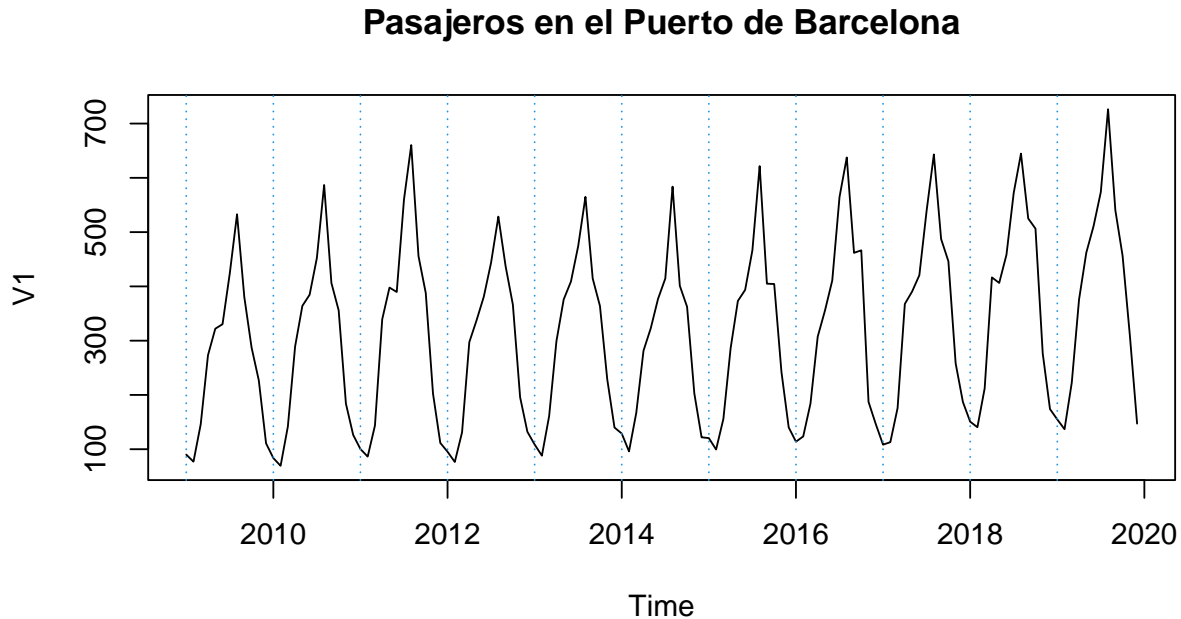
- *Metodología de alisado*

A partir del análisis realizado en el apartado anterior se decide qué tipo de serie se ajusta mejor a nuestros datos. A partir de aquí, se hace una predicción para observar cuál sería el número de viajeros en alojamientos rurales en los próximos años si no existiera ningún factor condicionante que actuará como brecha, en el caso expuesto, si no existiera la pandemia del Covid-19.

Aplicación empírica

Análisis descriptivo de la serie

El análisis descriptivo de los datos seleccionados comenzará con una herramienta visual, un gráfico que permitirá observar la distribución de éstos a lo largo de los 8 años estudiados. El gráfico temporal tiene la siguiente forma:



Como ya se explicó con detalle en la práctica 1, la serie estudiada presenta una tendencia creciente y una estacionalidad multiplicativa de orden $s = 12$.

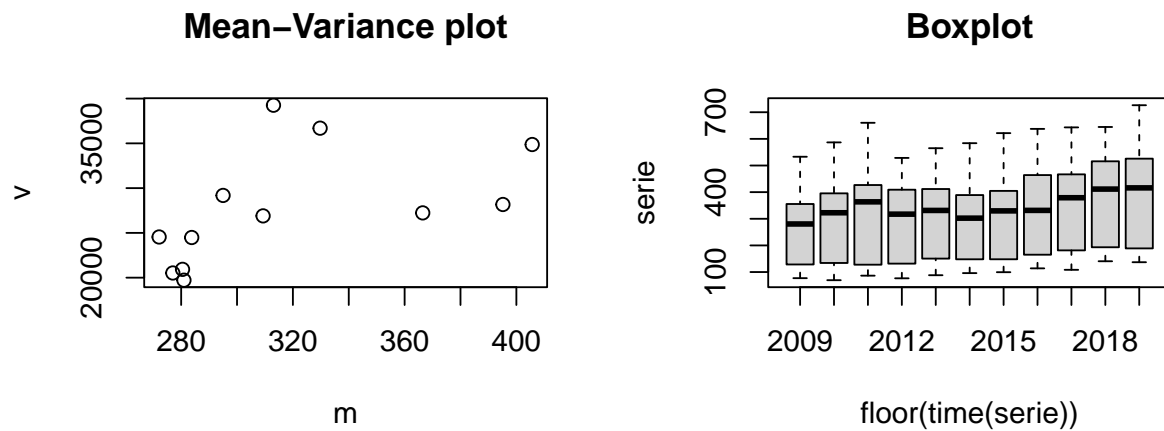
Identificación del modelo

Transformación de la serie temporal en una serie estacionaria

Para que la serie estudiada sea estacionaria estudiaremos que la varianza sea constante, que no exista componente estacional y que presente una media constante.

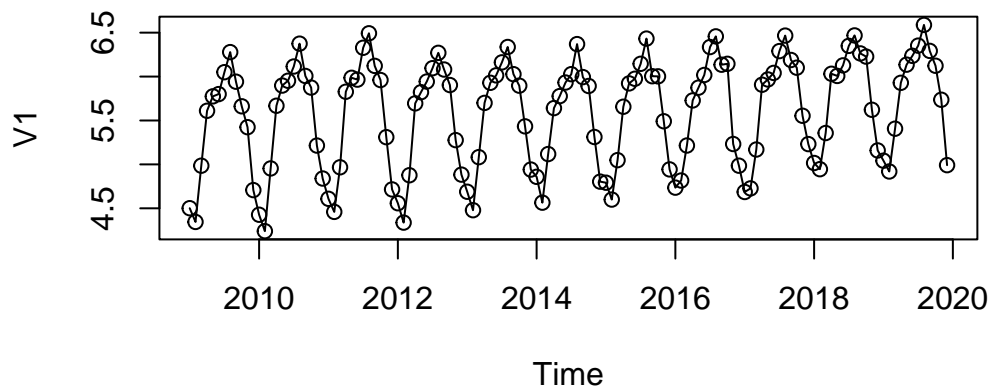
Varianza constante

Se dispone de dos soportes gráficos que permiten ver si la varianza de la serie es constante o no: *Gráfico medias vs varianzas* y *Boxplot por periodos*.



En el primer gráfico podemos observar que, a medida que aumenta la media, también lo hace la varianza. Lo mismo podemos decir en el segundo gráfico boxplot, donde el rango intercuartílico aumenta al mismo paso que aumenta el nivel de la serie.

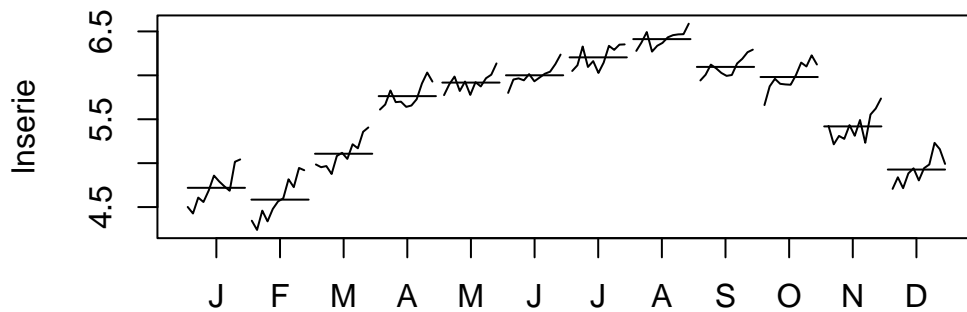
Entonces podemos concluir que la varianza de la serie no es constante. Para que lo sea, usaremos una transformación logarítmica graficando de nuevo los datos.



Ahora sí podríamos decir que la varianza de la serie se ha estabilizado al aplicar la transformación logarítmica.

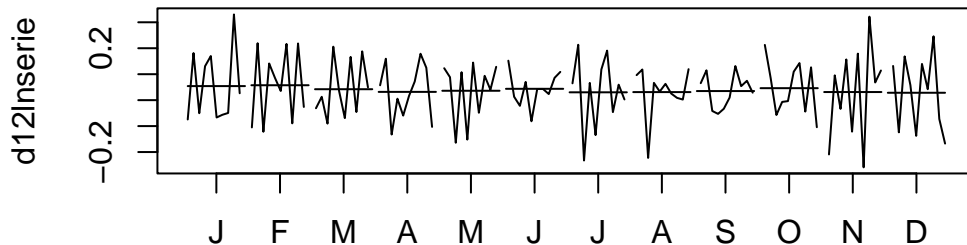
Patrón estacional

Para saber si existe patrón estacional usaremos la función `monthplot`, que calcula la media en cada orden estacional. En nuestro caso, mensualmente.



Vemos claramente que hay irregularidades en la media para cada mes, ya que no hay equilibrio en sus líneas horizontales. Por lo que podemos decir que la serie presenta un patrón estacional de orden $s=12$, como se ha dicho anteriormente. Para poder eliminarlo, se realiza una diferenciación estacional ($D = 1$), es decir, se restan las observaciones del patrón estacional anterior. Por tanto, se eliminarán $S = 12$ observaciones, las 12 primeras.

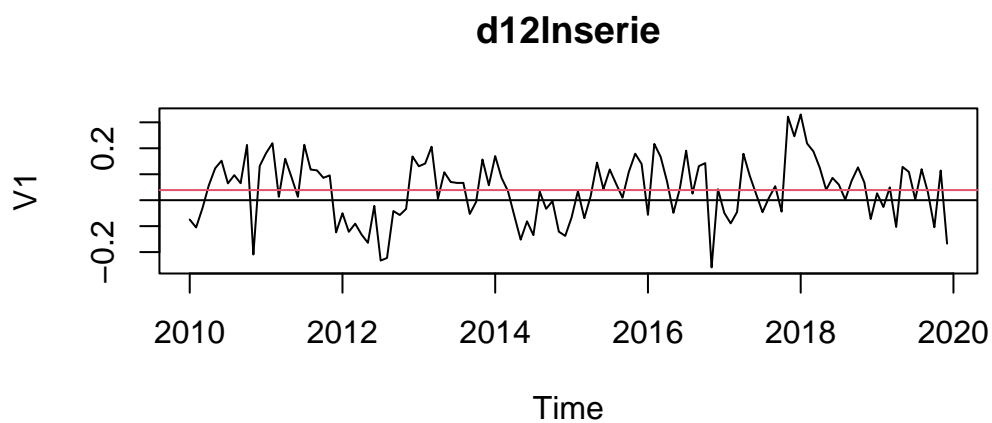
Volvemos a utilizar la función `monthplot` sobre la serie diferenciada estacionalmente para comprobar que se haya eliminado el patrón estacional.



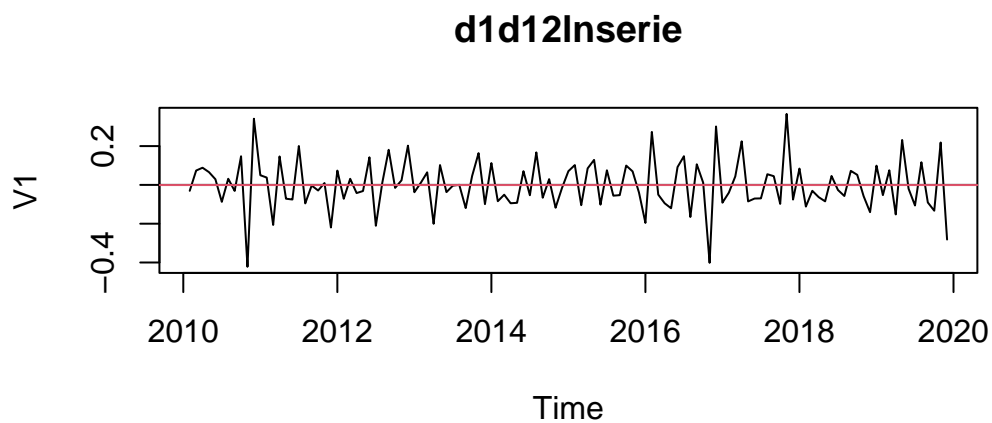
Vemos que, efectivamente, se han estabilizado las medias de cada mes, por lo que podemos afirmar que hemos eliminado el patrón estacional.

Media constante

Se obtiene el gráfico temporal de la serie en escala logarítmica y con una diferenciación estacional.



Podemos ver que la media de la serie parece constante pero no llega al valor nulo deseado. Por ello, haremos una diferenciación regular ($d=1$) y volveremos a graficar la serie para verificar que es 0.



Ahora podemos ver como la media sí tiene valor nulo pero para comprobar que la serie no presenta sobrediferenciación compararemos las varianzas de las tres transformaciones.

```
##          V1
## V1 0.378022

##          V1
## V1 0.01363138

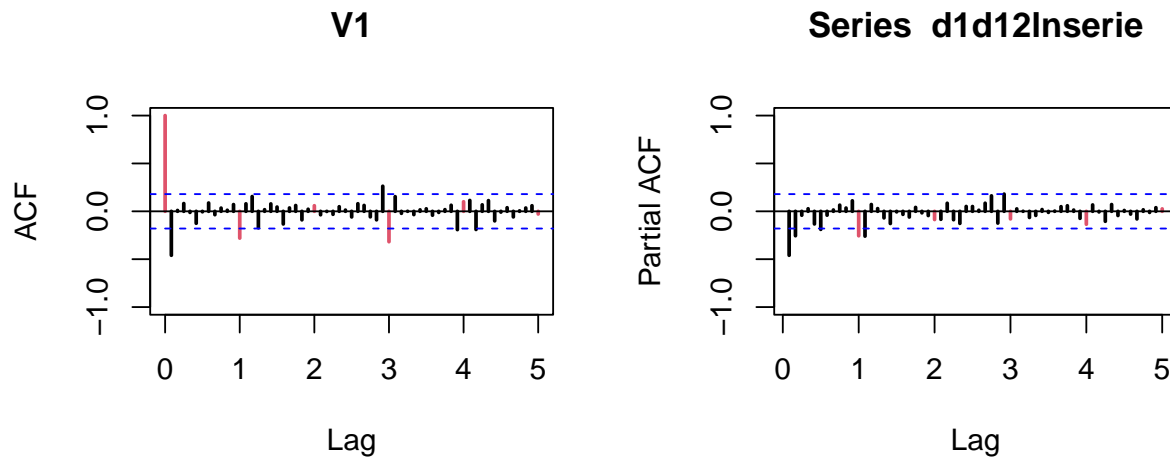
##          V1
## V1 0.01699842
```

Modelo	lnserie	d12lnserie	d1d12lnserie
Varianza	0.378022	0.0136314	0.0169984

Se observa que la segunda serie presenta una menor varianza pero no se aleja significativamente de la tercera y, como deseábamos tener un valor nulo de la media, nos quedamos con esta.

Identificación de modelos pausibles

Para identificar algunos modelos graficamos el ACF y el PACF de la serie estacionaria. En el siguiente gráfico se muestra en color rojo los “s” rezagos estacionales de la función (P)ACF. En negro, todos los rezagos regulares.



Podemos confirmar que se ha llegado a la serie estacionaria ya que los valores de ACF disminuyen hacia el 0.

Observando las gráficas se concluye que:

- *Opciones para la parte regular:* MA(1), AR(2)
- *Opciones para la parte estacional:* SMA(1), SAR(1) con $S=12$

Todas ellas con las diferenciaciones aplicadas, una regular ($d = 1$) y otra estacional ($D = 1$ con $S = 12$).

Así pues, se obtienen 4 posibles combinaciones, 4 modelos propuestos.

Forma compacta y usando el operador de retardo B de los modelos identificados

$$W_t = (1 - B)^d(1 - B^S)^D \ln(W_t)$$

Modelo 1: $MA(1)SMA(1)_{12}$ para W_t

Forma compacta de modelo ARIMA para $\ln(X_t)$

$$(1 - B)(1 - B^{12})\ln(X_t) = \theta_1(B)\Theta_1(B^{12})Z_t, d = 1, y D = 1 con S = 12$$

Sustituyendo cada polinomio característico se obtiene:

$$(1 - B)(1 - B^{12})\ln(X_t) = (1 + \theta_1 B)(1 + \Theta_1 B^{12})Z_t$$

Modelo 2: $AR(2)SMA(1)_{12}$ para W_t

Forma compacta de modelo ARIMA para $\ln(X_t)$

$$\phi_2(B)(1 - B)(1 - B^{12})\ln(X_t) = \Theta_1(B^{12})Z_t, d = 1, y D = 1 con S = 12$$

Sustituyendo cada polinomio característico se obtiene:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^{12})\ln(X_t) = (1 + \Theta_1 B^{12})Z_t$$

Modelo 3: $MA(1)SAR(1)_{12}$ para W_t

Forma compacta de modelo ARIMA para $\ln(X_t)$

$$(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})\ln(X_t) = \theta_1(B)Z_t, d = 1, y D = 1 con S = 12$$

Sustituyendo cada polinomio característico se obtiene:

$$(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})\ln(X_t) = (1 + \theta_1 B)Z_t$$

Modelo 4: $AR(2)SAR(1)_{12}$ para W_t

Forma compacta de modelo ARIMA para $\ln(X_t)$

$$\phi_2(B)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})\ln(X_t) = Z_t, d = 1, y D = 1 con S = 12$$

Sustituyendo cada polinomio característico se obtiene:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})\ln(X_t) = Z_t$$

Estimación de los modelos

Una vez se han expuesto los modelos posibles, se procede a realizar su estimación a partir de la función **arima**. Cabe destacar que dicha estimación se hará para la serie estacionaria y para la serie no estacionaria. Así pues, en total se verán seis modelos. Además, se calcularán los **T-ratios** para cada uno de los coeficientes. Éstos, deberán tomar un valor superior a 2 para que sean necesarios en la explicación de los datos. Por contra, si son inferiores a 2, no serán necesarios. Por esta misma razón, en los modelos de la serie estacionaria, se quiere que el valor del **intercept** salga no significativo, ya que se ha realizado una diferenciación regular extra para conseguir este mismo propósito, que la media, el **intercept**, sea igual a 0.

Modelo 1A, $MA(1)SMA(1)_{12}$

Se impone a la serie estacional **d1d12lnserie** el modelo identificado: $ARIMA(0, 0, 1)(0, 0, 1)_{12}$.

Se calculan los T-ratios para cada coeficiente para ver cuáles son estadísticamente significativos:

```
## Modelo 1a
## T-ratios: -7.14 -6.58 -0.03
```

El T-ratio del `intercept` es menor a 2, así que se plantea el modelo sin este coeficiente.

Modelo 1B, usando la serie no-estacionaria $\ln(X_t)$ (`lnserie`)

Se elimina el coeficiente `intercept`, estimando el modelo con la serie original en logaritmos (sin aplicar ninguna diferenciación).

Se calculan los T-ratios para cada coeficiente para ver cuáles son estadísticamente significativos:

```
## Modelo 1b
## T-ratios: -7.14 -6.59
```

Todos los coeficientes son estadísticamente significativos, no es necesario eliminar nada más. Se calcula el AIC de los modelos con y sin `intercept` para saber cuál de ellos tiene un valor menor de esta medida.

Modelo	AIC
Con <code>intercept</code>	-203.7712101
Sin <code>intercept</code>	-205.7718431

Es mejor el modelo con `intercept`, aunque la diferencia entre ambos modelos es muy pequeña.

Modelo 2A, $AR(2)SMA(1)_{12}$

Se impone a la serie estacional `d1d12lnserie` el modelo identificado: $ARIMA(2, 0, 0)(0, 0, 1)_{12}$

Se calculan los T-ratios para cada coeficiente para ver cuáles son estadísticamente significativos:

```
## Modelo 2a
## T-ratios: -6.64 -2.35 -5.95 -0.04
```

El T-ratio del `intercept` es menor a 2, así que se plantea el modelo sin este coeficiente.

Modelo 2B, usando la serie no-estacionaria $\ln(X_t)$ (`lnserie`)

Se elimina el coeficiente `intercept`, estimando el modelo con la serie original en logaritmos (sin ninguna diferenciación).

Se calculan los T-ratios para cada coeficiente para ver cuáles son estadísticamente significativos:

```
## Modelo 2b
## T-ratios: -6.64 -2.35 -5.95
```

Todos los coeficientes son estadísticamente significativos, no es necesario eliminar nada más. Se calcula el AIC de los modelos con y sin `intercept` para saber cuál de ellos tiene un valor menor de esta medida.

Modelo	AIC
Con <code>intercept</code>	-199.1634482
Sin <code>intercept</code>	-201.1625283

Es mejor el modelo con `intercept`, aunque la diferencia entre ambos modelos es muy pequeña.

Modelo 3A, MA(1)SAR(1)₁₂

Se impone a la serie estacional `d1d12lnserie` el modelo identificado: ARIMA(0, 0, 1)(1, 0, 0)₁₂

Se calculan los T-ratios para cada coeficiente para ver cuáles son estadísticamente significativos:

```
## Modelo 3a
## T-ratios: -7.26 -4.51 -0.15
```

El T-ratio del `intercept` es menor a 2, así que se plantea el modelo sin este coeficiente.

Modelo 3B, usando la serie no-estacionaria $\ln(X_t)$ (`lnserie`)

Se calculan los T-ratios para cada coeficiente para ver cuáles son estadísticamente significativos:

```
## Modelo 3b
## T-ratios: -7.27 -4.5
```

No es necesario eliminar nada más. Se calcula el AIC de los modelos con y sin `intercept`, para saber cuál de ellos tiene un menor valor de esta medida.

Modelo	AIC
Con <code>intercept</code>	-195.9228906
Sin <code>intercept</code>	-197.9027586

En este caso, es mejor el modelo sin `intercept`, aunque la diferencia entre ambos modelos es muy pequeña.

Modelo 4A, AR(2)SAR(1)₁₂

Se impone a la serie estacional `d1d12lnserie` el modelo identificado: ARIMA(2, 0, 0)(1, 0, 0)₁₂

Se calculan los T-ratios para cada coeficiente para ver cuáles son estadísticamente significativos:

```
## Modelo 4a
## T-ratios: -7.04 -2.61 -4.47 -0.09
```

El T-ratio del `intercept` es menor a 2, así que se plantea el modelo sin este coeficiente.

Modelo 4B, usando la serie no-estacionaria $\ln(X_t)$ (Inserie)

Se calculan los T-ratios para cada coeficiente para ver cuáles son estadísticamente significativos:

```
## Modelo 4b
## T-ratios: -7.04 -2.61 -4.47
```

No es necesario eliminar nada más. Se calcula el AIC de los modelos con y sin `intercept`, para saber cuál de ellos tiene un menor valor de esta medida.

Modelo	AIC
Con <code>intercept</code>	-192.8037495
Sin <code>intercept</code>	-194.7956946

En este caso, es mejor el modelo sin `intercept`, aunque la diferencia entre ambos modelos es muy pequeña.

Tabla resumen

Realizamos una tabla resumen para los cuatro modelos definitivos con el objetivo de poder ver de forma más clara la comparación entre ellos.

```
##
## Results
## =====
##                               Dependent variable:
##                               -----
##                               did12lnserie
##                               1A      2A      3A      4A
##                               (1)     (2)     (3)     (4)
## -----
## ma1          t = -7.136          t = -7.262
##              -0.673             -0.679
##
## ar1          t = -6.641          t = -7.038
##              -0.614             -0.642
##
## ar2          t = -2.347          t = -2.611
##              -0.220             -0.239
##
## sma1         t = -6.585 t = -5.949
##              -0.620     -0.593
##
## sar1          t = -4.506 t = -4.468
##              -0.407     -0.404
##
## intercept    t = -0.027 t = -0.044 t = -0.146 t = -0.093
##              -0.00004  -0.0001  -0.0003  -0.0003
## -----
## Observations      119      119      119      119
```

```
## Log Likelihood      105.886      104.582      101.961      101.402
## sigma2              0.009       0.010       0.010       0.010
## Akaike Inf. Crit.  -203.771    -199.163    -195.923    -192.804
## =====
## Note:                t = T-statistic value = coeff/SE(coeff)
```

En la tabla anterior se observan las estimaciones y los T-ratios de los coeficientes de los 4 modelos estimados. Antes de proceder a la explicación, se recuerda a que hace referencia cada uno de ellos:

- Modelo 1: $MA(1)SMA(1)_{12}$
- Modelo 2: $AR(2)SMA(1)_{12}$
- Modelo 3: $MA(1)SAR(1)_{12}$
- Modelo 4: $AR(2)SAR(1)_{12}$
- Modelos A: estimación con la serie estacionaria
- Modelos B: estimación con la serie NO estacionaria

Como se deseaba, todos los T-ratios son mayores de 2, excepto los que hacen referencia al **intercept** de la serie estacionaria. Se concluye que todos los coeficientes son necesarios para explicar los datos.

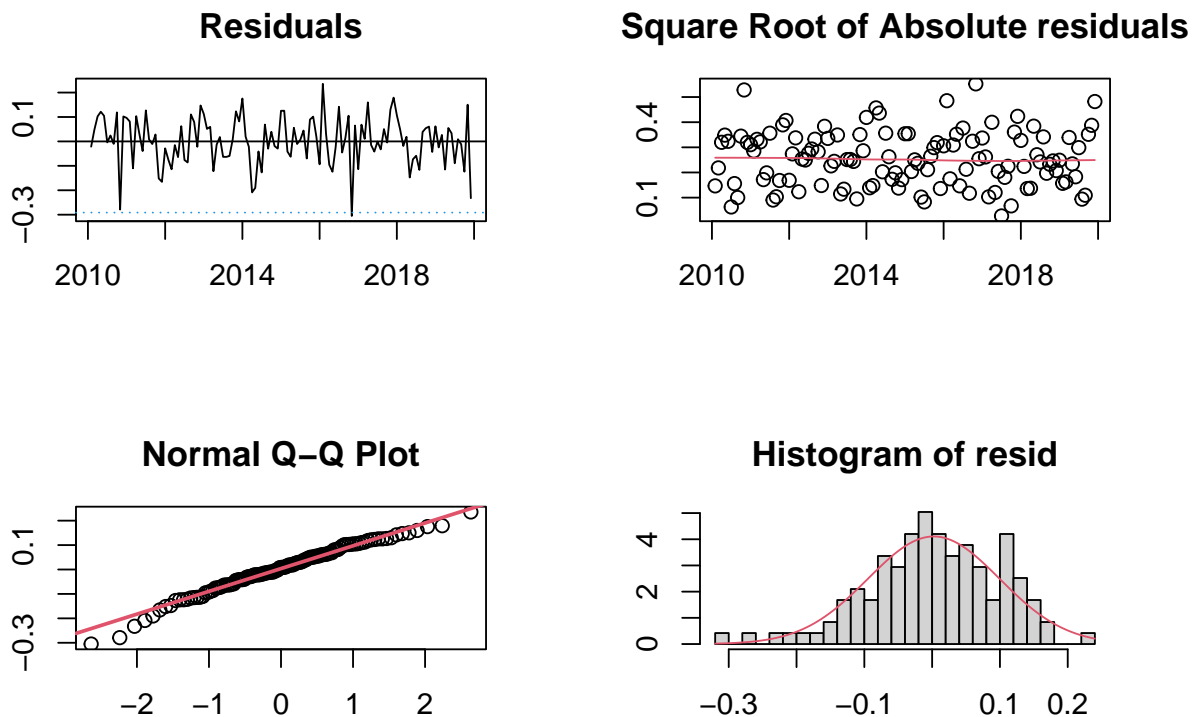
Observando el criterio del AIC, se puede decir que los mejores modelos son aquellos que tienen un valor menor en esta medida de calidad relativa. Así pues, a partir de ahora, solo se seguirán estudiando los modelos 1A y 2A ya que son estos los que tienen un valor más negativo.

Validación de los modelos ajustados

Para validar los modelos propuestos, seguiremos los siguientes pasos:

-> Verificaremos que los residuos tengan varianza constante y se distribuyan de manera normal e independiente. -> Revisaremos el ACF (autocorrelación) y PACF (autocorrelación parcial) de los residuos para asegurarnos de que se distribuyen como un “ruido blanco”, es decir, sin patrones o dependencias. -> Aseguraremos que el modelo sea causal e invertible.

Validación modelo 1A



Residuals: Los residuos parecen estar distribuidos de manera aleatoria y la mayoría se encuentra dentro de los límites de confianza establecidos. Además, los residuos se centran en torno a cero, lo que indica que el modelo está haciendo predicciones precisas.

Square Root of Absolute residuals: La línea roja en el gráfico de los residuos es constante, lo que sugiere que la varianza también es constante.

Normal QQ-Plot: Hay algunos valores atípicos en las colas.

Histogram of resid: Parece que hay algunos valores atípicos en los datos y la distribución presenta kurtosis, lo que significa que no se ajusta del todo a una distribución normal. La presencia de outliers puede estar contribuyendo a la kurtosis, y la distribución tiende a ser asimétrica y a inclinarse hacia la derecha.

Conclusión: Parece que la varianza es constante y hay presencia de atípicos en los residuos.

Realizamos el test de Breusch-Pagan y contrastamos las siguientes hipótesis:

- H_0 : Los residuos son homocedásticos

- H_1 : Los residuos no son homocedasticos

Obtenemos un p-valor superior a 0.05, así que afirmamos que los residuos son homocedasticos.

Realizamos los tests de Shapiro-Wilk, Anderson-Darling y Jarque Bera y contrastamos las siguientes hipótesis:

- H_0 : Los residuos son normales
- H_1 : Los residuos no son normales

Después de realizar los tests de normalidad, obtuvimos p-valores superiores a 0.05, lo que nos permite afirmar que los residuos se distribuyen de manera normal. Es importante tener en cuenta que estos tests son muy sensibles a la presencia de valores atípicos, por lo que es posible que los resultados no sean del todo precisos. Por esta razón, también es útil examinar el gráfico QQ-plot, donde se puede ver que los residuos centrales sí cumplen con la normalidad, pero los que pertenecen a las colas no. Esto sugiere que es necesario prestar especial atención a los valores atípicos y tratar de identificarlos y eliminarlos para mejorar la precisión del modelo.

Contrastamos hipótesis:

Durbin-Watson test

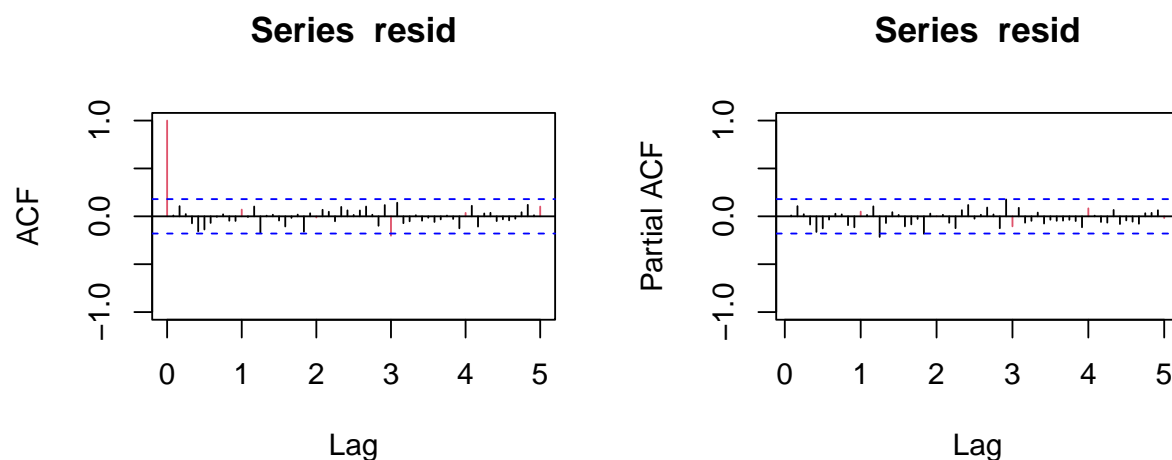
-> H_0 : No hay autocorrelación en los residuos en el retardo k -> H_1 : Hay autocorrelación en los residuos en el retardo k

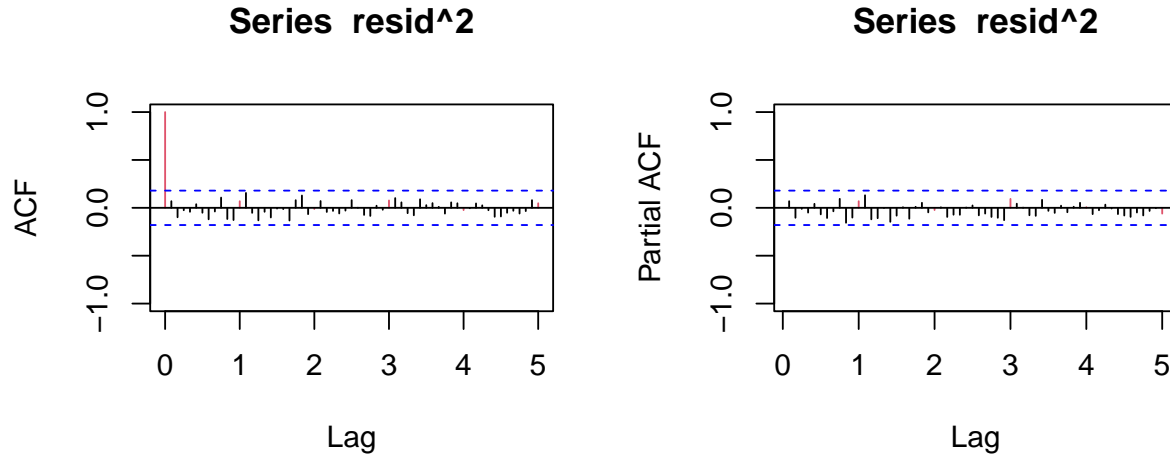
Se obtiene un p-valor superior a 0.05, así que afirmamos que no hay autocorrelación en los residuos.

Ljung-Box test

-> H_0 : No hay autocorrelación hasta el retardo k -> H_1 : Hay autocorrelación hasta el retardo k

Después de realizar los tests de autocorrelación, obtuvimos un p-valor superior a 0.05 en el retardo (lag) 36. Esto significa que, según el programa utilizado, al menos hasta el retardo 48 no hay una estructura de autocorrelación evidente en los datos. Esto es una buena señal, ya que la autocorrelación puede afectar la precisión del modelo y es importante asegurarse de que los datos no tengan patrones de dependencia

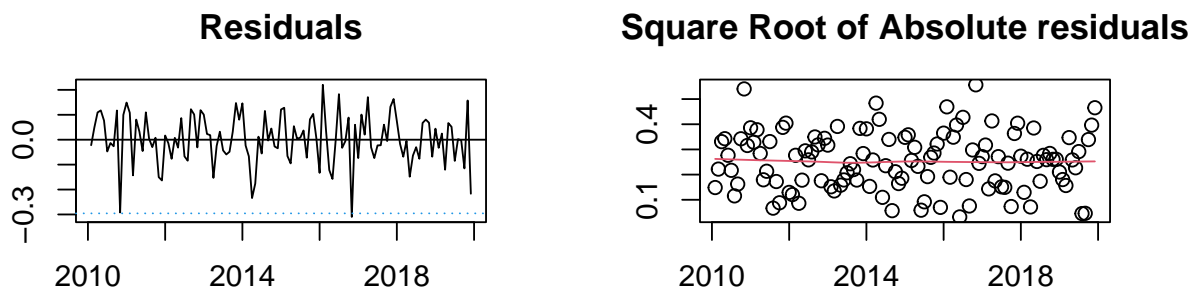


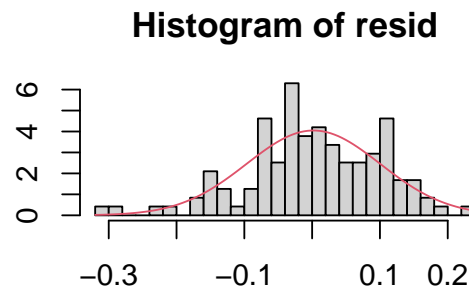
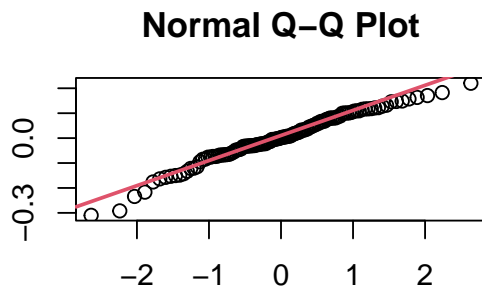


Los gráficos ACF (autocorrelación) y PACF (autocorrelación parcial) de los residuos nos permiten visualizar la dependencia entre los datos y nos dan una idea de si hay autocorrelación presente. Al examinar estos gráficos, observamos que todos los retardos se encuentran dentro de las bandas de confianza, lo que nos permite reafirmar el resultado obtenido en el test de Durbin-Watson: no hay autocorrelación en los residuos.

Además, al examinar los gráficos ACF y PACF de los residuos al cuadrado, también observamos que no hay autocorrelación presente. Esto nos permite afirmar que no hay presencia de outliers ni hay volatilidad en los datos, lo que es una buena señal para la validez del modelo.

Validación modelo 2A





Residuals: Los residuos parecen estar distribuidos de manera aleatoria y la mayoría se encuentra dentro de los límites de confianza establecidos. Además, los residuos se centran en torno a cero, lo que indica que el modelo está haciendo predicciones precisas

Square Root of Absolute residuals: La línea roja en el gráfico de los residuos es ligeramente constante, lo que sugiere que la varianza también es constante

Normal QQ-Plot: Hay algunos valores atípicos en las colas.

Histogram of resid: Parece que hay algunos valores atípicos presentes en los datos, especialmente en la cola izquierda. Además, la parte central de los datos no se ajusta bien a una distribución normal, ya que los datos están desplazados hacia la izquierda.

Conclusión: Parece que la varianza es constante y hay presencia de atípicos en los residuos.

A continuación se realiza el test de Breusch-Pagan y se contrastan las siguientes hipótesis:

- H_0 : Los residuos son homocedásticos
- H_1 : Los residuos no son homocedásticos

Se obtiene un p-valor superior a 0.05, así que se afirma que los residuos son homocedásticos.

A continuación se realizan los tests de Shapiro-Wilk, Anderson-Darlin y Jarque Bera y se contrastan las siguientes hipótesis:

- H_0 : Los residuos son normales
- H_1 : Los residuos no son normales

Después de realizar los tests de normalidad, obtuvimos p-valores superiores a 0.05, lo que nos permite afirmar que los residuos se distribuyen de manera normal. Es importante tener en cuenta que estos tests son muy sensibles a la presencia de valores atípicos, por lo que es posible que los resultados no sean del todo precisos.

Por esta razón, también es útil examinar el gráfico QQ-plot, donde se puede ver que los residuos centrales sí cumplen con la normalidad, pero los que pertenecen a las colas no. Esto sugiere que es necesario prestar especial atención a los valores atípicos y tratar de identificarlos y eliminarlos para mejorar la precisión del modelo.

Contrastamos hipótesis:

Durbin-Watson test

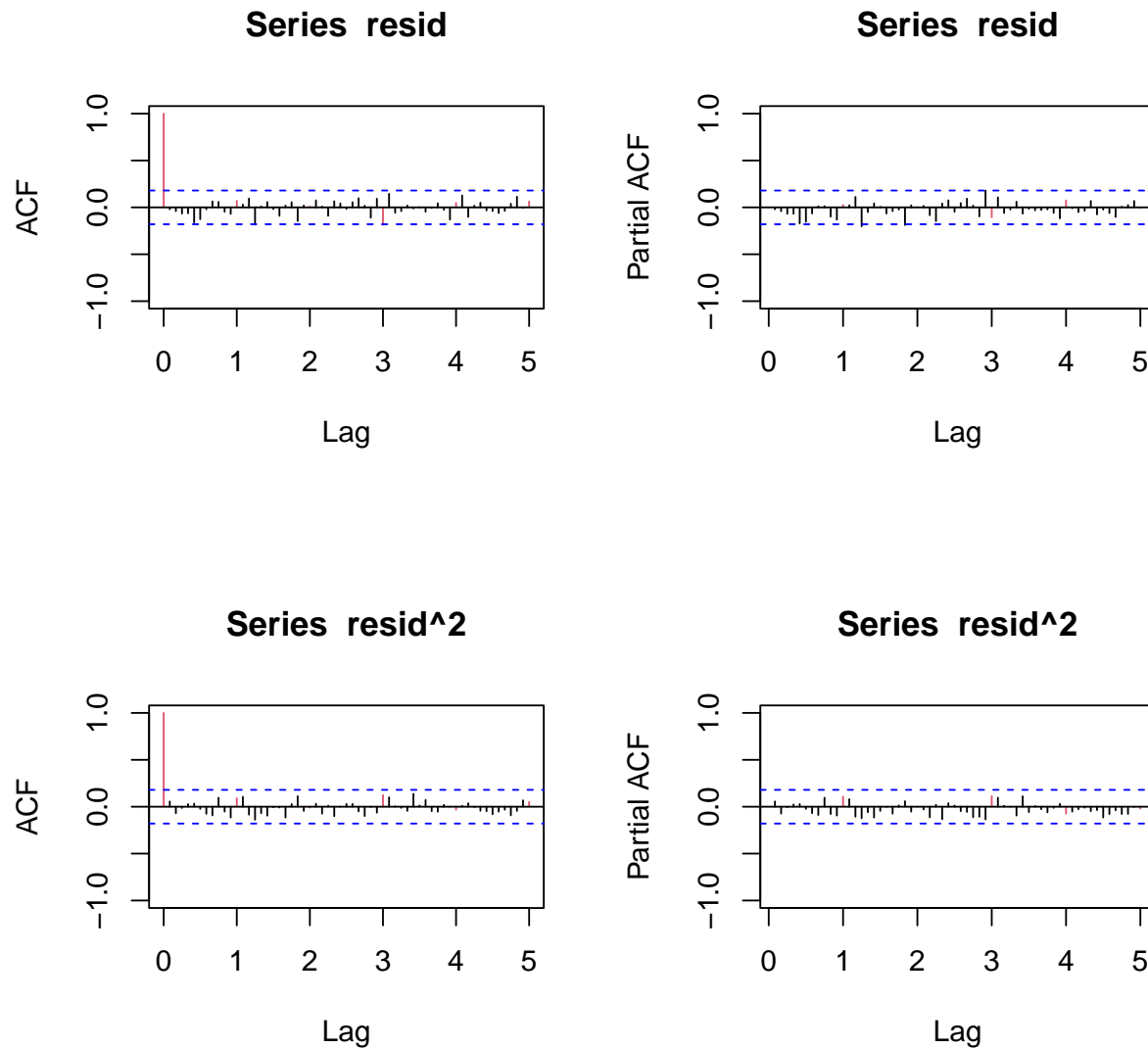
-> H_0 : No hay autocorrelación en los residuos en el retardo k -> H_1 : Hay autocorrelación en los residuos en el retardo k

Se obtiene un p-valor superior a 0.05, así que se afirma que no hay autocorrelación en los residuos.

Ljung-Box test

-> H_0 : No hay autocorrelación hasta el retardo k -> H_1 : Hay autocorrelación hasta el retardo k

Se obtiene un p-valor superior a 0.05 en el retardo (lag) 36. Tal y como está programada la función significa que como mínimo hasta el retardo 48 no hay una estructura de autocorrelación.



Los gráficos ACF (autocorrelación) y PACF (autocorrelación parcial) de los residuos nos permiten visualizar la dependencia entre los datos y nos dan una idea de si hay autocorrelación presente. Al examinar estos gráficos, observamos que casi todos los retardos se encuentran dentro de las bandas de confianza, lo que nos permite reafirmar el resultado obtenido en el test de Durbin-Watson: no hay autocorrelación en los residuos.

Además, al examinar los gráficos ACF y PACF de los residuos al cuadrado, también observamos que no hay autocorrelación presente. Esto nos permite afirmar que no hay presencia de outliers ni hay volatilidad en los datos, lo que es una buena señal para la validez del modelo.

Predicción de modelos ARIMA ajustados y validados

Modelo 1A

Verificación estabilidad del modelo

Para evaluar la estabilidad del modelo, se ha excluido el último año de datos y se ha creado un nuevo modelo. Si el modelo es estable, esperaríamos que los coeficientes estimados obtenidos con la serie completa y con la serie incompleta sean similares. Por “similares” nos referimos a que tienen la misma magnitud, el mismo signo y la misma significancia.

```
##
## Call:
## arima(x = lnserie1, order = c(0, 0, 1), seasonal = list(order = c(0, 0, 1),
##   period = 12))
##
## Coefficients:
##          ma1      sma1  intercept
##          0.7434  0.7568      5.5935
## s.e.  0.0501  0.0643      0.0669
##
## sigma^2 estimated as 0.06835:  log likelihood = -15.69,  aic = 39.38

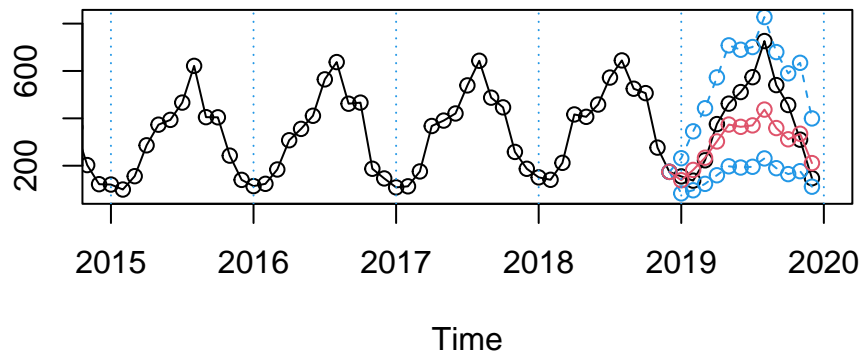
##
## Call:
## arima(x = lnserie2, order = c(0, 0, 1), seasonal = list(order = c(0, 0, 1),
##   period = 12))
##
## Coefficients:
##          ma1      sma1  intercept
##          0.7256  0.7989      5.5776
## s.e.  0.0548  0.0833      0.0699
##
## sigma^2 estimated as 0.06655:  log likelihood = -14.1,  aic = 36.19
```

El modelo es estable pues se cumplen las 3 condiciones establecidas.

Predicción out-of-sample y cálculo RMSPE/MAPE

A continuación se grafica la serie completa (**sin dejar el último año fuera**) y se hace una predicción del último año para ver si el modelo estima correctamente los datos.

Model ARIMA(0,0,1)(0,0,1)_{12}



El modelo propuesto estima de manera regular los datos (estimación en color rojo). Aunque, los datos reales se encuentran dentro de las bandas de confianza (estimaciones en color azul).

Cálculo de medidas RMSPE/MAPE y mean Length

Medida	Valor
EQM	0.284744
EAM	0.2565645

Modelo 2A

Verificación estabilidad del modelo

Para comprobar si el modelo es estable se usa el mismo criterio explicado en el modelo 1A.

```
##
## Call:
## arima(x = lnserie1, order = c(2, 0, 0), seasonal = list(order = c(0, 0, 1),
##   period = 12))
##
## Coefficients:
##      ar1      ar2      sma1  intercept
##    1.2323 -0.5443  0.5923    5.5814
## s.e.  0.0778  0.0782  0.0603    0.0890
##
## sigma^2 estimated as 0.04299:  log likelihood = 16.92,  aic = -23.84

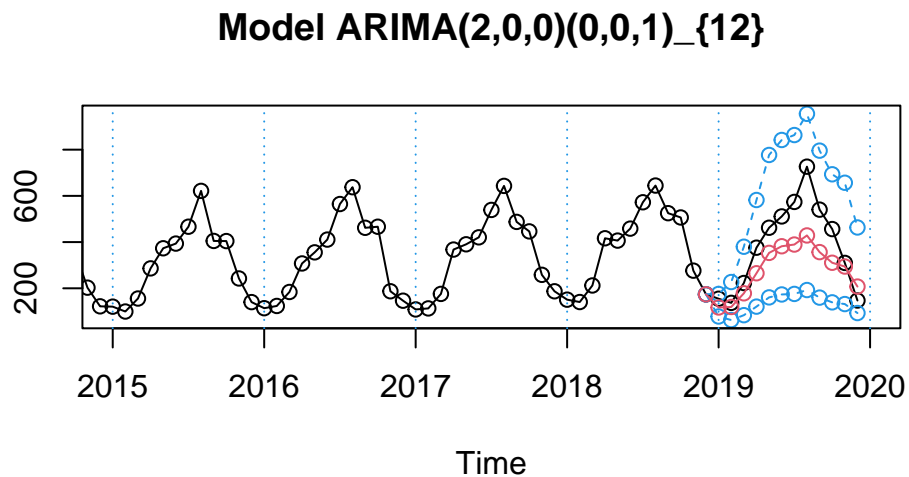
##
## Call:
## arima(x = lnserie2, order = c(2, 0, 0), seasonal = list(order = c(0, 0, 1),
##   period = 12))
##
```

```
## Coefficients:
##          ar1      ar2      sma1  intercept
##          1.2176 -0.5335  0.6152     5.5682
## s.e.  0.0809   0.0813  0.0660     0.0926
##
## sigma^2 estimated as 0.04256:  log likelihood = 15.42,  aic = -20.85
```

El modelo es estable pues se cumplen las 3 condiciones establecidas.

Predicción out-of-sample y cálculo RMSPE/MAPE

A continuación, se ha graficado la serie completa de datos (sin excluir el último año) y se ha realizado una predicción para el último año para evaluar si el modelo estima correctamente los datos. Esto nos permite determinar si el modelo es adecuado para hacer predicciones y si es capaz de captar los patrones presentes en los datos.



Igual que en el modelo 1A, la estimación no es muy cercana a los valores reales pero las bandas de confianza ajustan bien los datos reales.

Cálculo de medidas RMSPE/MAPE y mean Length

Medida	Valor
EQM	0.2861041
EAM	0.2673542

Selección del mejor modelo

Después de validar los modelos y analizar sus capacidades predictivas, es necesario elegir un solo modelo para llevar a cabo el objetivo principal del estudio, que es la realización de predicciones a largo plazo.

Para tomar esta decisión, es importante comparar las medidas de adecuación a los datos (AIC) y sus capacidades de predicción, dando más peso a estas últimas.

Es necesario crear una tabla que incluya los valores más relevantes para poder comparar los dos modelos propuestos y seleccionar el mejor. Esto nos permitirá utilizar el modelo seleccionado para hacer predicciones precisas y confiables a largo plazo.

```
##
## =====
##              mod1A      mod2A
## -----
## Log Likelihood 105.886  16.919
## AIC            -203.771 -23.838
## RMSPE          0.285   0.286
## MAPE           0.257   0.267
## Mean Length    408.507  486.750
## -----
```

En cuanto a la medida de adecuación a los datos, se observa que el modelo 1A es preferible al 2A, ya que éste tiene un AIC menor. Para la capacidad predictiva realmente no importa cual escogamos ya que el error es de 28% en el caso del EQM y un 25% en EAM. Sin embargo, respecto el EAM, el mod2A lo tiene un 1% superior por lo tanto es mejor el modelo 1A. Cabe destacar que ambos modelos tienen un error en la capacidad predictiva superior al 25%, por lo que se concluye que los dos son bastante regulares. Además, existe otra medida comparable entre modelos: la amplitud de la predicción. Este valor, interesa que sea pequeño, pues querrá decir que el intervalo de confianza de los valores predichos estará más acotado, será más estrecho. Observando la tabla anterior, se puede decir que el modelo 1A tiene una amplitud de predicción bastante menor que la del modelo 2A.

Así pues, después de toda esta explicación, se concluye que el mejor modelo para realizar la predicción a largo plazo es el modelo 1A, ya que es el que tiene menor AIC y menor amplitud de predicción.

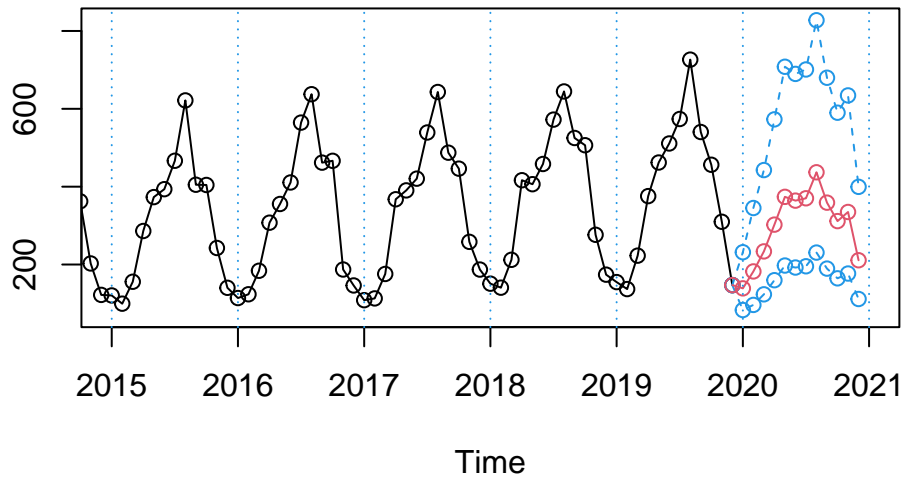
Antes de pasar a realizar la predicción, es necesario recordar la expresión del modelo elegido. Esta era la siguiente: $MA(1)SMA(1)_{12}$.

Realización predicción a largo plazo

Una vez que se ha seleccionado el mejor modelo, es hora de hacer la predicción. Para llevarla a cabo, se utilizan los valores de la serie completa y se estima cuántos pasajeros habría habido en el puerto de Barcelona en el año 2020 si no se hubiera producido la pandemia del Covid-19.

Esto nos permite obtener una idea de cómo habría evolucionado la cantidad de pasajeros en el puerto sin la influencia de la pandemia y nos permite comparar el resultado obtenido con la realidad para evaluar la precisión del modelo.

Model ARIMA(0,0,1)(0,0,1)_12



Se observa que la predicción en un tiempo extrapolable al de los datos sigue el mismo patrón que lo visto hasta el momento pero con una disminución en la tendencia lineal. Parece ser que se sigue el patrón de los años anteriores pero en una escala mucho menor.

Por tanto, gracias a la predicción observada, se puede decir que ya se ha cumplido el objetivo principal del estudio.

Conclusiones

Después de realizar las evaluaciones correspondientes y utilizar los métodos ya conocidos y mencionados con anterioridad, se concluye que la serie temporal correspondiente al número de pasajeros en el puerto de Barcelona se puede explicar como un modelo $MA(1)SMA(1)_{12}$.

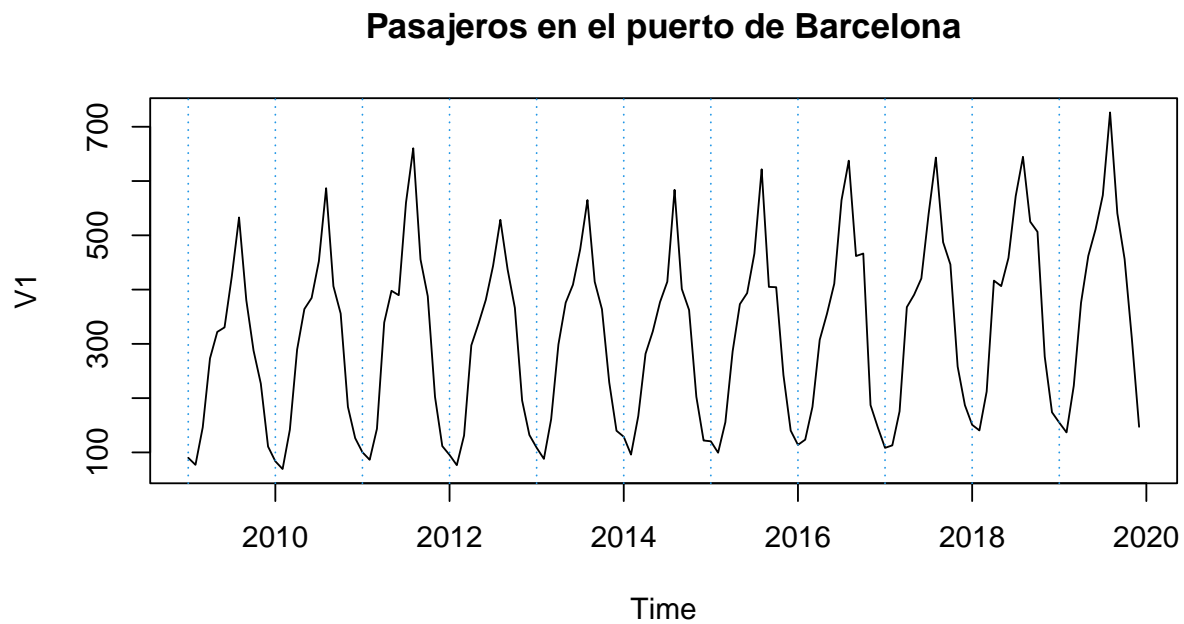
Para llegar a esta conclusión ha sido necesario transformar la serie original con el fin de convertirla en una serie estacionaria. Para ello, se ha debido de aplicar una transformación logarítmica y hacer una diferenciación regular y otra estacional. Una vez conseguida la serie en el formato deseado, se han identificado 4 posibles modelos que podrían encajar con los datos estudiados, aunque dos de ellos han sido descartados por tener una medida de adecuación a los datos bastante inferiores en comparación al resto de opciones. Así pues, a los dos modelos resultantes, se les ha hecho una validación, se ha verificado su estabilidad y se ha observado su capacidad predictiva. A partir de los resultados obtenidos, se ha seleccionado un solo modelo, el mejor de ellos, el cual ha resultado ser un $MA(1)SMA(1)_{12}$. Este modelo tenía un AIC igual a -203.7712, un error de predicción de 28.5% y 25.7% en los casos de EQM y EAM, respectivamente y una amplitud de predicción de 408.5.

Finalmente y a partir del modelo seleccionado, se ha llevado a cabo el objetivo principal del estudio: hacer una predicción para el pasajeros en el puerto de Barcelona que hubiera habido en el año 2020 si no hubiera existido la pandemia del Covid-19. En este supuesto, se hubiera esperado la misma tendencia lineal creciente y el mismo patrón mensual que lo visto hasta el momento.

Anexo

#Aplicación empírica

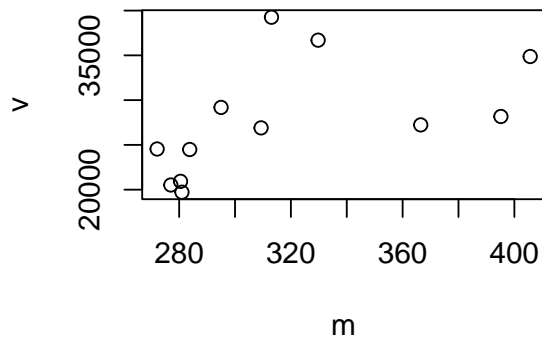
```
serie=window(ts(read.table("PasajeBCN.dat"),start=1997,freq=12),start=2009)
plot(serie, main = "Pasajeros en el puerto de Barcelona")
abline(v = 2009:2019, col = 4, lty = 3)
```



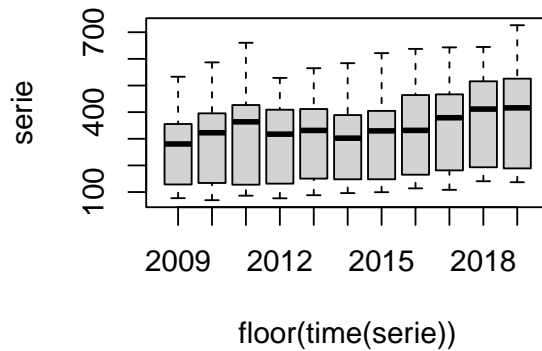
##Identificación del modelo

```
par(mfrow=c(1,2))
m = apply(matrix(serie,ncol=12),2,mean)
v = apply(matrix(serie,ncol=12),2,var)
plot(v~m,main="Mean-Variance plot")
boxplot(serie~floor(time(serie)), main = "Boxplot")
```

Mean-Variance plot



Boxplot



```
lnserie=log(serie)
plot(lnserie,type="o")
```

```
monthplot(lnserie)
```

```
d12lnserie<-diff(lnserie,lag=12)
```

```
monthplot(d12lnserie)
```

```
plot(d12lnserie,main="d12lnserie")
abline(h=0)
abline(h=mean(d12lnserie), col=2)
```

```
d1d12lnserie <- diff(d12lnserie)
```

```
plot(d1d12lnserie,main="d1d12lnserie")
abline(h=0)
abline(h=mean(d1d12lnserie), col=2)
```

```
v1 <- var(lnserie)
v2 <- var(d12lnserie)
v3 <- var(d1d12lnserie)
```

```
par(mfrow=c(1,2))
acf(d1d12lnserie,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,11)),lwd=2)
pacf(d1d12lnserie,ylim=c(-1,1),lag.max=60,col=c(rep(1,11), 2),lwd=2)
```

##Estimación de los modelos

```
mod1A = arima(d1d12lnserie, order = c(0,0,1), seasonal = list(order = c(0,0,1), period=12))
```

```
mod1B = arima(lnserie, order = c(0,1,1), seasonal = list(order = c(0,1,1), period=12))
```

```
mod2A = arima(d1d12lnserie, order = c(2,0,0), seasonal = list(order = c(0,0,1), period=12))
```

```
mod2B = arima(lnserie, order = c(2,1,0), seasonal = list(order = c(0,1,1), period=12))
```

```
mod3A = arima(d1d12lnserie, order = c(0,0,1), seasonal = list(order = c(1,0,0), period=12))
```

```
mod3B = arima(lnserie, order = c(0,1,1), seasonal = list(order = c(1,1,0), period=12))
```

```
mod4A = arima(d1d12lnserie, order = c(2,0,0), seasonal = list(order = c(1,0,0), period=12))
```

```
mod4B = arima(lnserie, order = c(2,1,0), seasonal = list(order = c(1,1,0), period=12))
```

```
#install.packages("stargazer")
```

```
library("stargazer")
```

```
stargazer(mod1A, mod2A, mod3A, mod4A, title="Results", type="text", notes.append = FALSE, report = "vtc",
notes = c("t = T-statistic value = coeff/SE(coeff)", digits = 3, column.labels = c("1A", "2A", "3A", "4A"))
```

```
##
## Results
## =====
##                               Dependent variable:
##                               -----
##                               d1d12lnserie
##                               1A      2A      3A      4A
##                               (1)    (2)    (3)    (4)
## -----
## ma1          t = -7.136          t = -7.262
##              -0.673              -0.679
##
## ar1          t = -6.641          t = -7.038
##              -0.614              -0.642
##
## ar2          t = -2.347          t = -2.611
##              -0.220              -0.239
##
## sma1         t = -6.585 t = -5.949
##              -0.620   -0.593
##
## sar1         t = -4.506 t = -4.468
##              -0.407   -0.404
##
## intercept    t = -0.027 t = -0.044 t = -0.146 t = -0.093
##              -0.00004 -0.0001  -0.0003  -0.0003
## -----
## Observations      119      119      119      119
## Log Likelihood    105.886  104.582  101.961  101.402
## sigma2            0.009    0.010    0.010    0.010
```

```
## Akaike Inf. Crit.  -203.771  -199.163  -195.923  -192.804
## =====
## Note:              t = T-statistic value = coeff/SE(coeff)
```

#Validación de los modelos ajustados

```
validation_grafic <- function(model, dades){

s = frequency(get(model$series))
resid = model$residuals
par(mfrow = c(2, 2), mar = c(3, 3, 3, 3))
#Residuals plot
par(mfrow=c(1,2))
plot(resid, main = "Residuals")
abline(h = 0)
abline(h = c(-3*sd(resid, na.rm = T), 3*sd(resid, na.rm = T)), lty = 3, col = 4)
#Square Root of absolute values of residuals (Homocedasticity)
scatter.smooth(sqrt(abs(resid)), main = "Square Root of Absolute residuals",
lpars = list(col = 2))
#Normal plot of residuals
par(mfrow=c(1,2))
qqnorm(resid)
qqline(resid, col = 2, lwd = 2)
#Histogram of residuals with normal curve
hist(resid, breaks = 20, freq = F)
curve(dnorm(x, mean = mean(resid, na.rm = T), sd = sd(resid, na.rm = T)), col = 2, add = T)
}
```

```
validation_test_homoce <- function(model, dades){
suppressMessages(require(lmtest, quietly = TRUE, warn.conflicts = FALSE))
##Breusch-Pagan test (vs. order)
obs = get(model$series)
print(bptest(resid(model)~I(1:length(resid(model)))))
##Breusch-Pagan test (vs. predictions)
obs = get(model$series)
print(bptest(resid(model)~I(obs-resid(model))))
}
```

```
validation_test_normal <- function(model, dades){

##Shapiro-Wilks Normality test
print(shapiro.test(resid(model)))
suppressMessages(require(nortest, quietly = TRUE, warn.conflicts = FALSE))
##Anderson-Darling test
print(ad.test(resid(model)))
suppressMessages(require(tseries, quietly = TRUE, warn.conflicts = FALSE))
##Jarque-Bera test
print(jarque.bera.test(na.omit(c(resid(model)))))
}
```

```
validation_test_indepen <- function(model, dades){
s = frequency(get(model$series))
resid = model$residuals
```

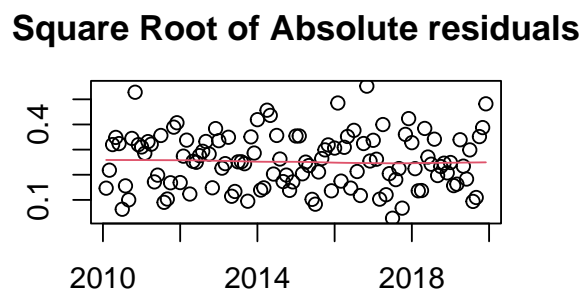
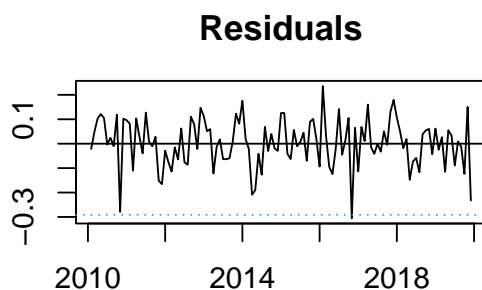
```
##Durbin-Watson test
print(dwtest(resid(model)~I(1:length(resid(model)))))
##Ljung-Box test
cat("\nLjung-Box test\n")
print(t(apply(matrix(c(1:4, (1:4)*s)), 1, function(e1) {
te = Box.test(resid(model), type = "Ljung-Box", lag = e1)
c(lag = (te$parameter), statistic = te$statistic[[1]], p.value = te$p.value)})))
}
```

```
validation_acf_pacf <- function(model, dates){
  s = frequency(get(model$series))
  resid = model$residuals
  #ACF & PACF of residuals
  par(mfrow=c(1,2))
  acf(resid,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,s-1)),lwd=1)
  pacf(resid,ylim=c(-1,1),lag.max=60,col=c(rep(1,s-1),2),lwd=1)
  par(mfrow=c(1,1))

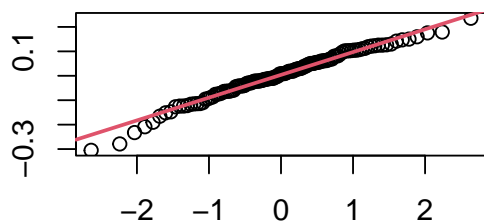
  #ACF & PACF of square residuals
  par(mfrow=c(1,2))
  acf(resid^2,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,s-1)),lwd=1)
  pacf(resid^2,ylim=c(-1,1),lag.max=60,col=c(rep(1,s-1),2),lwd=1)
  par(mfrow=c(1,1))
}
```

##Validación modelo 1A

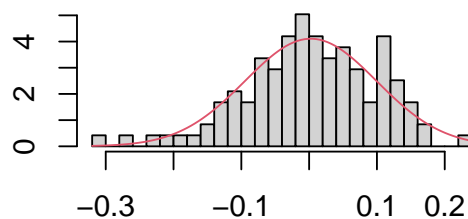
```
model = mod1A
validation_grafic(model)
```



Normal Q-Q Plot



Histogram of resid



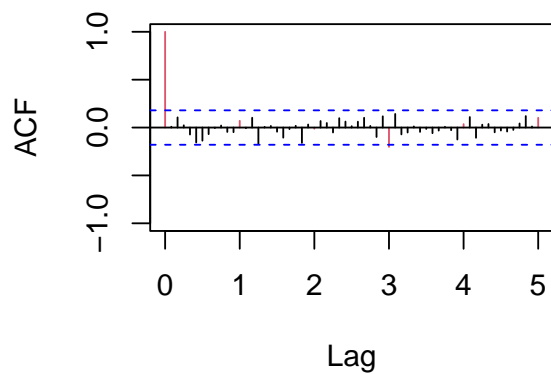
```
validation_test_homoce(model)
```

```
validation_test_normal(model)
```

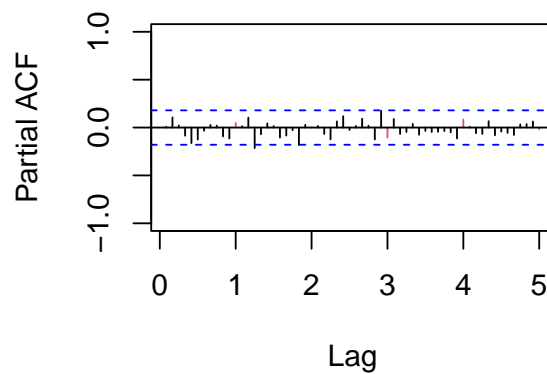
```
validation_test_indepen(model)
```

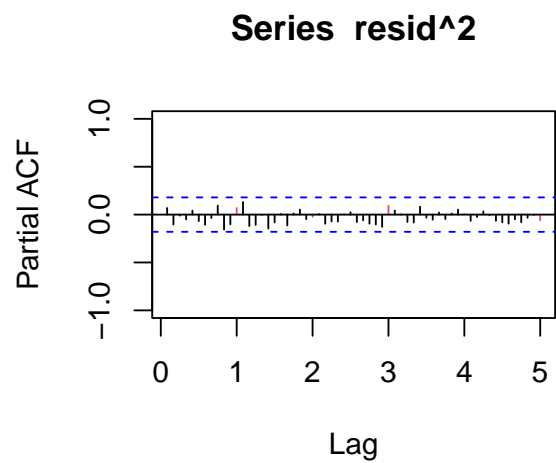
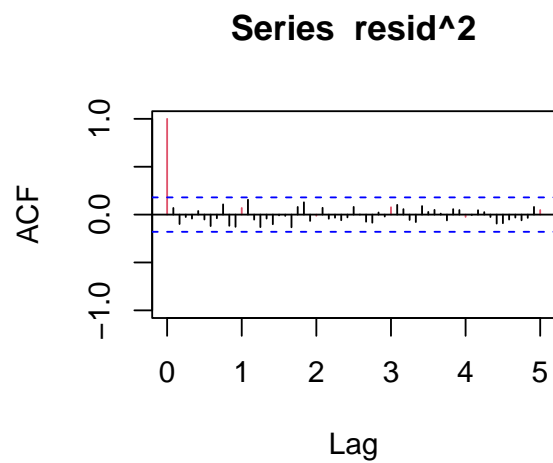
```
validation_acf_pacf(model)
```

Series resid



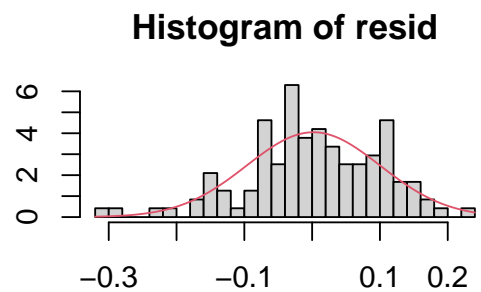
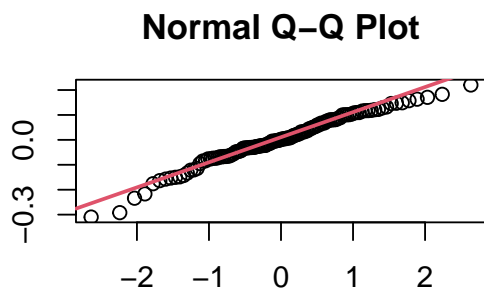
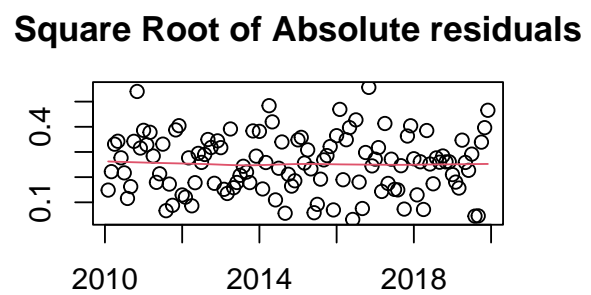
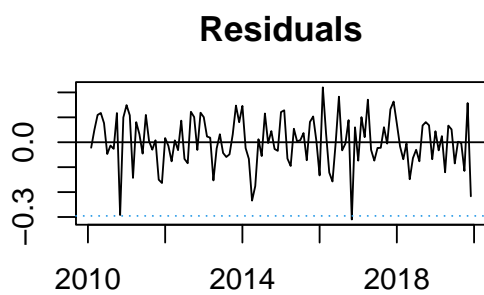
Series resid





##Validación modelo 2A

```
model = mod2A
validation_grafic(model)
```

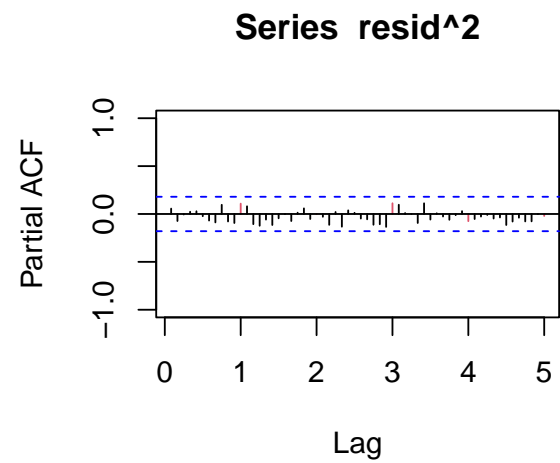
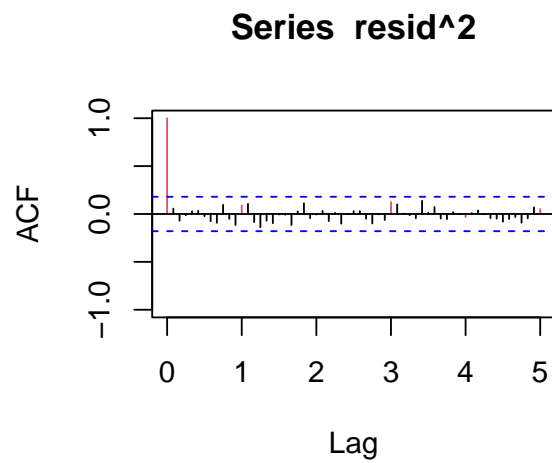
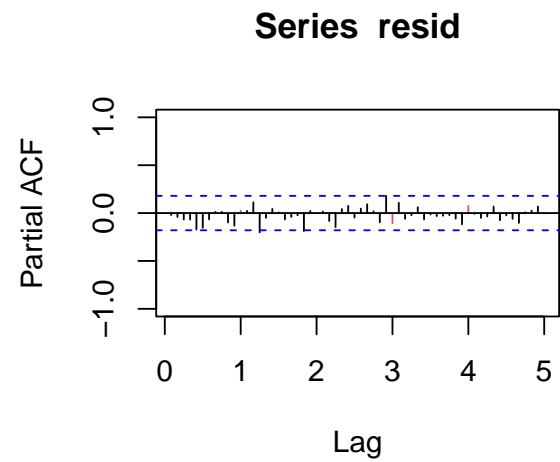
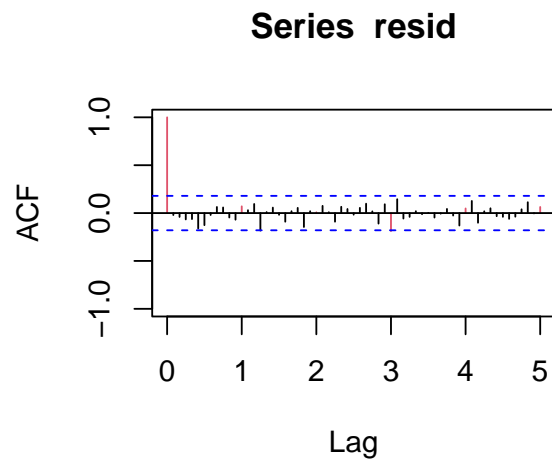


```
validation_test_homoce(model)
```

```
validation_test_normal(model)
```

```
validation_test_indepen(model)
```

```
validation_acf_pacf(model)
```



#Predicción de modelos ARIMA ajustados y validados

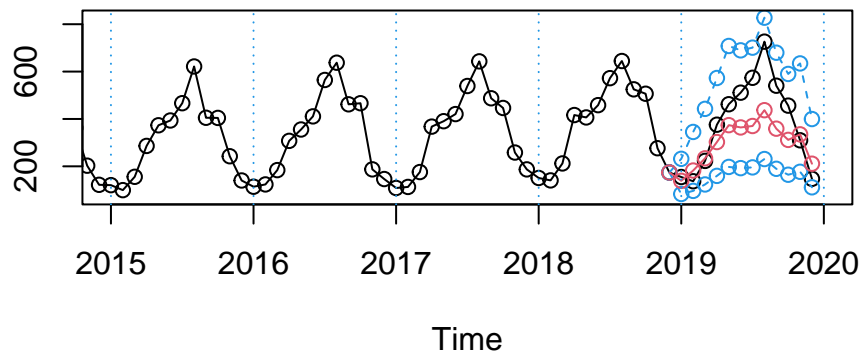
```
ultim = c(2018,12)                                #Dic 2018
serie1 = window(serie, end = ultim + c(1,0))        #complete series: 2009-2019
lnserie1 = log(serie1)                              #log transformed
serie2 = window(serie, end = ultim)                 #series without last year observations: 2009-2018
lnserie2 = log(serie2)                              #log transformed
```

##Modelo 1A

```
# Fit the model to the complete series: lnserie1
(mod1A2 = arima(lnserie1, order = c(0,0,1), seasonal = list(order = c(0,0,1), period=12)))
#Fit the model to the subset series (without 2019 data): lnserie2
(mod1A22 = arima(lnserie2,order = c(0,0,1), seasonal = list(order = c(0,0,1), period=12)))

pred = predict(mod1A2, n.ahead=12) #outputs point predictions and correspo
pr <- ts(c(tail(lnserie2,1),pred$pred),start = ultim, freq=12) #point predictions
se <- ts(c(0,pred$se), start = ultim, freq=12) #Standard errors for point predictions
#Prediction Intervals (back transformed to original scale using exp-function)
tl <- ts(exp(pr-1.96*se), start = ultim, freq = 12)
tu <- ts(exp(pr+1.96*se), start = ultim, freq = 12)
pr <- ts(exp(pr), start = ultim, freq = 12) #predictions in original scale
#Plot of the original airbcn series (thousands) and out-of-sample predictions: only time window 2015-20
ts.plot(serie,tl,tu,pr,lty=c(1,2,2,1),col=c(1,4,4,2),xlim=ultim[1]+c(-3,+2),type="o",main="Model ARIMA(
abline(v=(ultim[1]-3):(ultim[1]+2),lty=3,col=4)
```

Model ARIMA(0,0,1)(0,0,1)_{12}



```
obs=window(serie,start=ultim)
mod.EQM1=sqrt(sum(((obs-pr)/obs)^2)/12) # Error = obs - pred
mod.EAM1=sum(abs(obs-pr)/obs)/12
mod.ML1=sum(tu-tl)/12
```

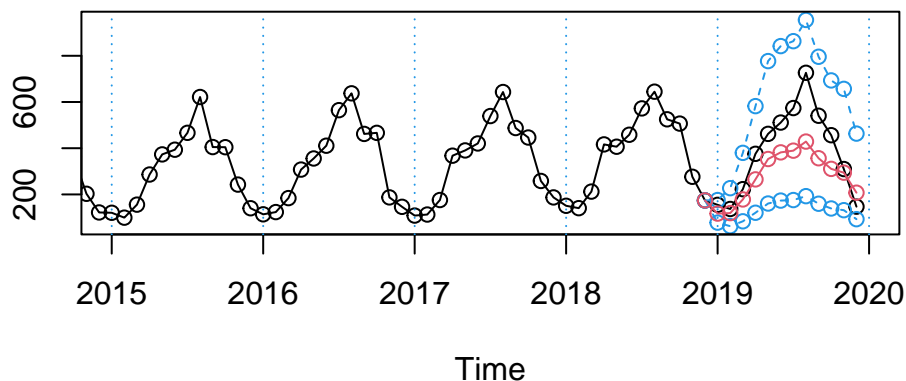
##Modelo 1A

```
##
## Call:
## arima(x = lnserie1, order = c(2, 0, 0), seasonal = list(order = c(0, 0, 1),
## period = 12))
##
## Coefficients:
##          ar1          ar2          sma1  intercept
##          1.2323   -0.5443   0.5923         5.5814
## s.e.    0.0778    0.0782   0.0603         0.0890
##
## sigma^2 estimated as 0.04299: log likelihood = 16.92, aic = -23.84
```

```
##
## Call:
## arima(x = lnserie2, order = c(2, 0, 0), seasonal = list(order = c(0, 0, 1),
##     period = 12))
##
## Coefficients:
##          ar1      ar2      sma1  intercept
##       1.2176 -0.5335  0.6152    5.5682
## s.e.  0.0809  0.0813  0.0660    0.0926
##
## sigma^2 estimated as 0.04256:  log likelihood = 15.42,  aic = -20.85
```

```
pred = predict(mod2A, n.ahead=12) #outputs point predictions and correspon
pr <- ts(c(tail(lnserie2,1),pred$pred),start = ultim, freq=12) #point predictions
se <- ts(c(0,pred$se), start = ultim, freq=12) #Standard errors for point predictions
#Prediction Intervals (back transformed to original scale using exp-function)
tl <- ts(exp(pr-1.96*se), start = ultim, freq = 12)
tu <- ts(exp(pr+1.96*se), start = ultim, freq = 12)
pr <- ts(exp(pr), start = ultim, freq = 12) #predictions in original scale
#Plot of the original airbcn series (thousands) and out-of-sample predictions: only time window 2015-20
ts.plot(serie,tl,tu,pr,lty=c(1,2,2,1),col=c(1,4,4,2),xlim=ultim[1]+c(-3,+2),type="o",main="Model ARIMA(2,0,0)(0,0,1)_12")
abline(v=(ultim[1]-3):(ultim[1]+2),lty=3,col=4)
```

Model ARIMA(2,0,0)(0,0,1)_12



```
obs=window(serie,start=ultim)
mod.EQM2=sqrt(sum(((obs-pr)/obs)^2)/12) # Error = obs - pred
mod.EAM2=sum(abs(obs-pr)/obs)/12
mod.ML2=sum(tu-tl)/12
```

#Selección del mejor modelo

```
selection=function(model){
  s=frequency(get(model$series))
  resid=model$residuals
```

```

par(mfrow=c(2,2),mar=c(3,3,3,3))
resumen<-data.frame(Pruebas=1:5)
colnames(resumen)<-paste0("mod1B")
rownames(resumen)<-c("Log Likelihood","AIC","RMSPE", "MAPE","Mean Length")

resumen[1,1]=model$loglik
resumen[2,1]=model$aic
resumen[3,1]=NA
resumen[4,1]=NA
resumen[5,1]=NA
return(resumen)
}

```

```

model = mod1A
resumen1 <- selection(model)
colnames(resumen1) <- c("mod1A")
resumen1[3,1] = mod.EQM1
resumen1[4,1] = mod.EAM1
resumen1[5,1] = mod.ML1
model = mod2A
resumen2 <- selection(model)
colnames(resumen2) <- c("mod2A")
resumen2[3,1] = mod.EQM2
resumen2[4,1] = mod.EAM2
resumen2[5,1] = mod.ML2
tablef<-cbind.data.frame(resumen1,resumen2)
stargazer(tablef, summary=FALSE, type="text")

```

```

##
## =====
##              mod1A      mod2A
## -----
## Log Likelihood 105.886  16.919
## AIC            -203.771 -23.838
## RMSPE          0.285   0.286
## MAPE           0.257   0.267
## Mean Length    408.507  486.750
## -----

```

Previsions a llarg termini amb el model complet

```

pred=predict(mod1A2,n.ahead=12)
pr<-ts(c(tail(lnserie,1),pred$pred),start=ultim+c(1,0),freq=12) #starts Dec 2019!
se<-ts(c(0,pred$se),start=ultim+c(1,0),freq=12)
#Intervals
tl1<-ts(exp(pr-1.96*se),start=ultim+c(1,0),freq=12)
tu1<-ts(exp(pr+1.96*se),start=ultim+c(1,0),freq=12)
pr1<-ts(exp(pr),start=ultim+c(1,0),freq=12)
ts.plot(serie,tl1,tu1,pr1,lty=c(1,2,2,1),col=c(1,4,4,2),xlim=c(ultim[1]-3,ultim[1]+3),type="o",main="Model")
abline(v=(ultim[1]-3):(ultim[1]+3),lty=3,col=4)

```

Model ARIMA(0,0,1)(0,0,1)_12

