PROJECTE ANÀLISI DE DADES: Entrada Turística Espanya

David Anglada Rotger i Andreu Huguet Segarra

17/5/2019

Contents

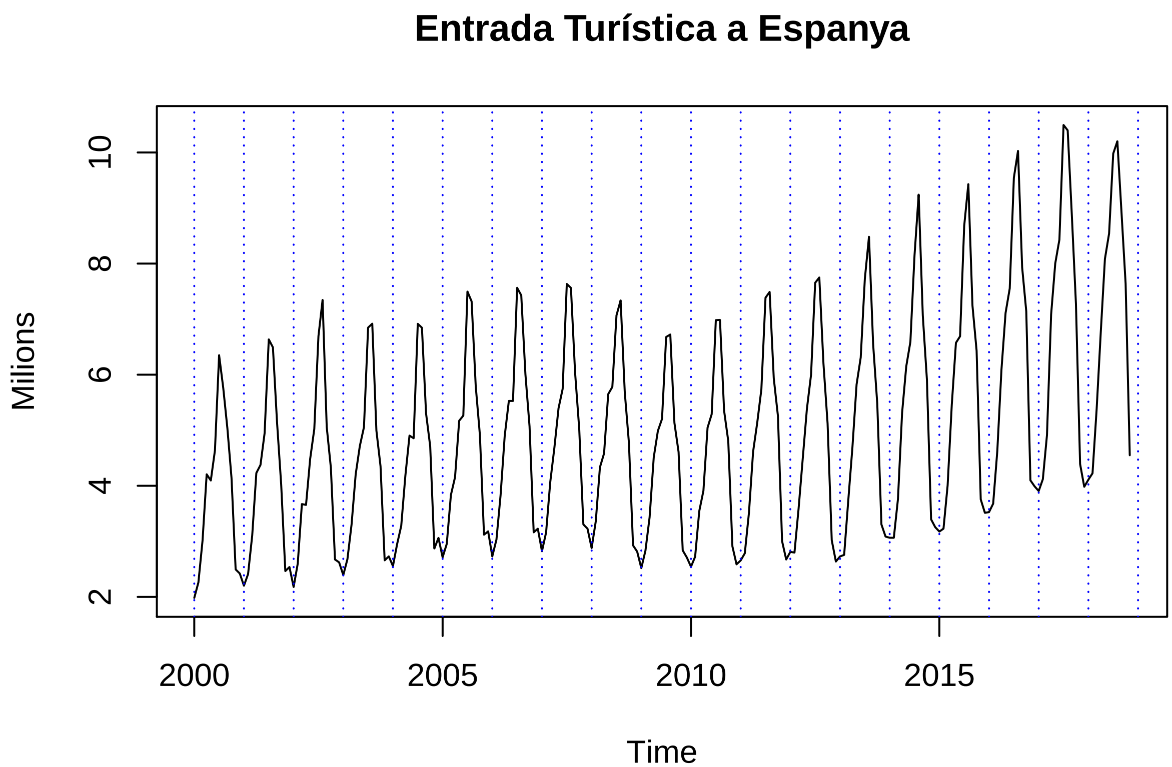
1 Introducció

2 Identificació

2.1 Representació gràfica de les dades

Un cop feta la representació de les dades, s'observa una clara tendència creixent. Tot i així, aquesta tendència no és constant, ja que és menys pronunciada entre els anys 2000 i 2010, fins i tot amb una petita baixada entre els anys 2007i 2010 i sembla que es pronuncia a partir de l'any 2011.

Pel que fa a la variància, s'observa que va augmentant a mesura que augmenta la mitjana dels valors de les dades, és a dir, a mesura que es pronuncia la tendència creixent. És a dir, en els anys 2000-2010, la variància és menor que en els anys 2011-2019, on el creixement augmenta.

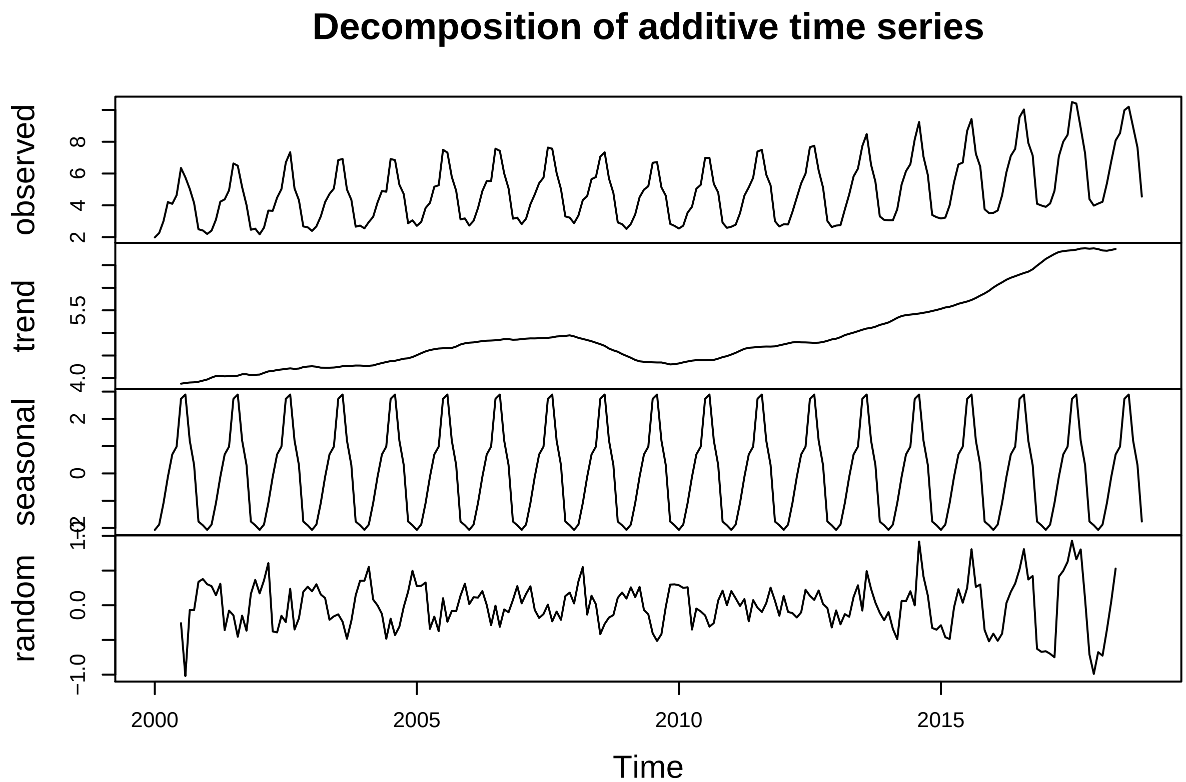


2.1.1 Descomposició en components bàsiques

Per poder analitzar millor les dades, es realitza la seva descomposició en les seves components bàsiques, és a dir, el model aditiu de la serie:

$$X_t = T_t + S_t + C_t + \omega_t$$

on: - T_t és la **tendència** de la sèrie a llarg termini. - S_t és el ***seasonal*** de la sèrie (patró repetit periòdicament amb període constant). - C_t és el **cicle** de la sèrie (patró repetit periòdicament amb període no constant). Aquesta part no surt representada en la descomposició. - ω_t és el soroll aleatòri.



S'observa, tal i com s'havia comentat anteriorment, la clara tendència creixent de la sèrie, amb un creixement menys pronunciat a l'inici, una petita baixada entre els anys 20017 i 2010 i una pujada més pronunciada més cap a l'actualitat. Pel que fa al patró estacional, observem que durant els mesos d'estiu, el número de turistes a Espanya augmenta molt considerablement. Aquest fet que no crida l'atenció, ja que és durant els mesos d'estiu quan més vacanses s'agafa la gent i més aprofiten per venir a les costes espanyoles. Durant els mesos de tardor-hivern, observem que el número de turistes cau en picat.

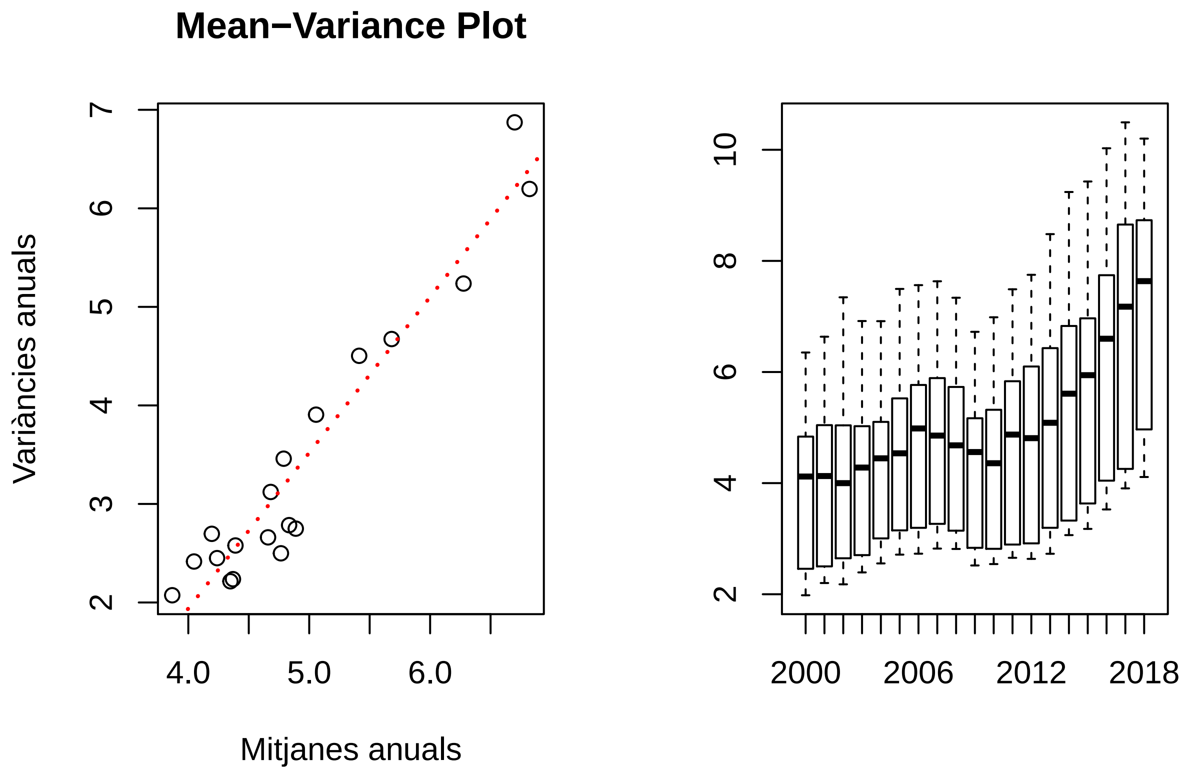
2.2 Transformació de les dades

A continuació s'analitzarà la necessitat de realitzar una sèrie de transformacions amb l'objectiu d'aconseguir estacionaritat en la nostra sèrie temporal.

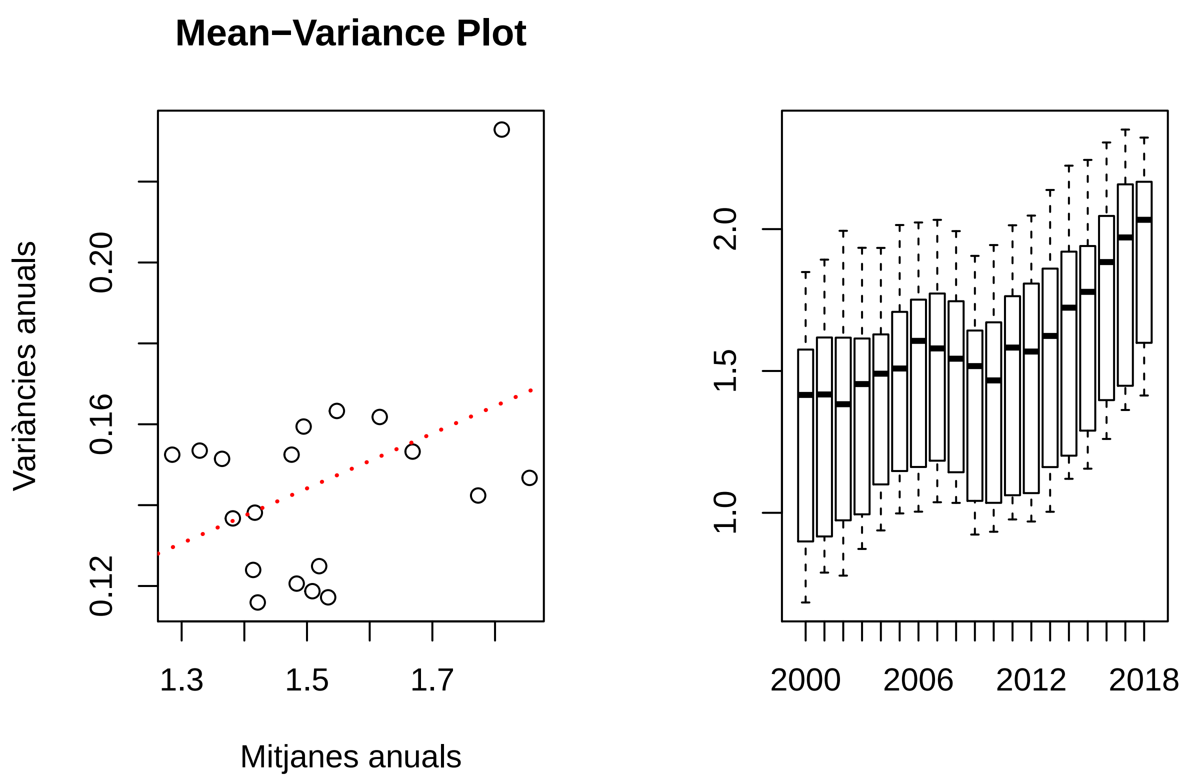
2.2.1 Variància constant

En primer lloc, s'estudiarà si es pot considerar que la variància de les dades sigui constant en el temps. Ja s'ha comentat que a simple vista semblava que no. Tot i així es comprova amb un plot de la variància front la mitjana i un *boxplot* de les dades cada 12 mesos (que és la freqüència de les nostres dades).

```
## Warning in matrix(serie, nr = 12): data length [227] is not a sub-multiple
## or multiple of the number of rows [12]

## Warning in matrix(serie, nr = 12): data length [227] is not a sub-multiple
## or multiple of the number of rows [12]
```



Tal i com s'havia observat a simple vista, la variància augmenta a mesura que agumenta la mitja. Per tant, no podem assumir variància constant. Amb el *boxplot* es confirma aquesta hipòtesis. Així doncs, es procedeix a realitzar una transformació logarítmica de la sèrie per homogeneïtzar la variància. Els resultats obtinguts són els següents:



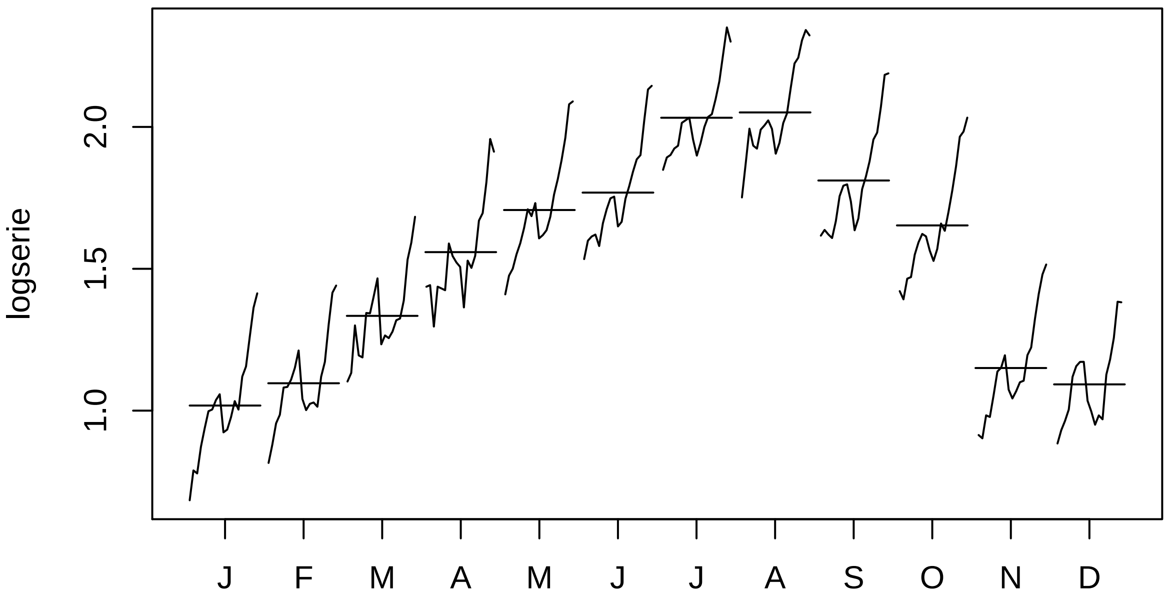
S’observa que la variància s’ha homogeneïtzat, és a dir, ja es pot considerar constant.

2.2.2 Patró estacional

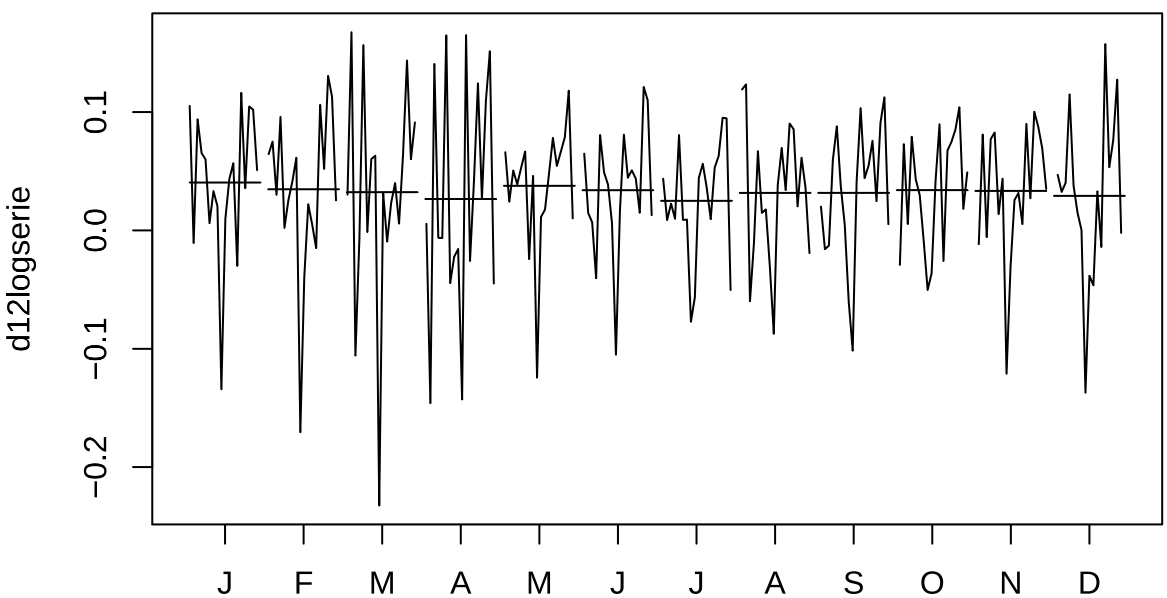
En segon lloc, s’estudiarà l’existència d’un patró estacional en les nostres dades. En cas que hi sigui present, es realitzarà una diferenciació d’ordre 12, és a dir,

$$W_t = X_t - X_{t-12} = (1 - B^{12})X_t$$

on B és el *backshift operator*, per eliminar aquest patró. Es realitza un *monthplot* per comprovar-ne l’existència.



Tal i com s’havia comenta, s’observa una clara pujada de la presència de turistes durant els mesos d’estiu i una baixada en picat en l’entrada de l’hivern/tardor. Així doncs, és necessària una diferenciació d’ordre 12 per eliminar aquest patró.



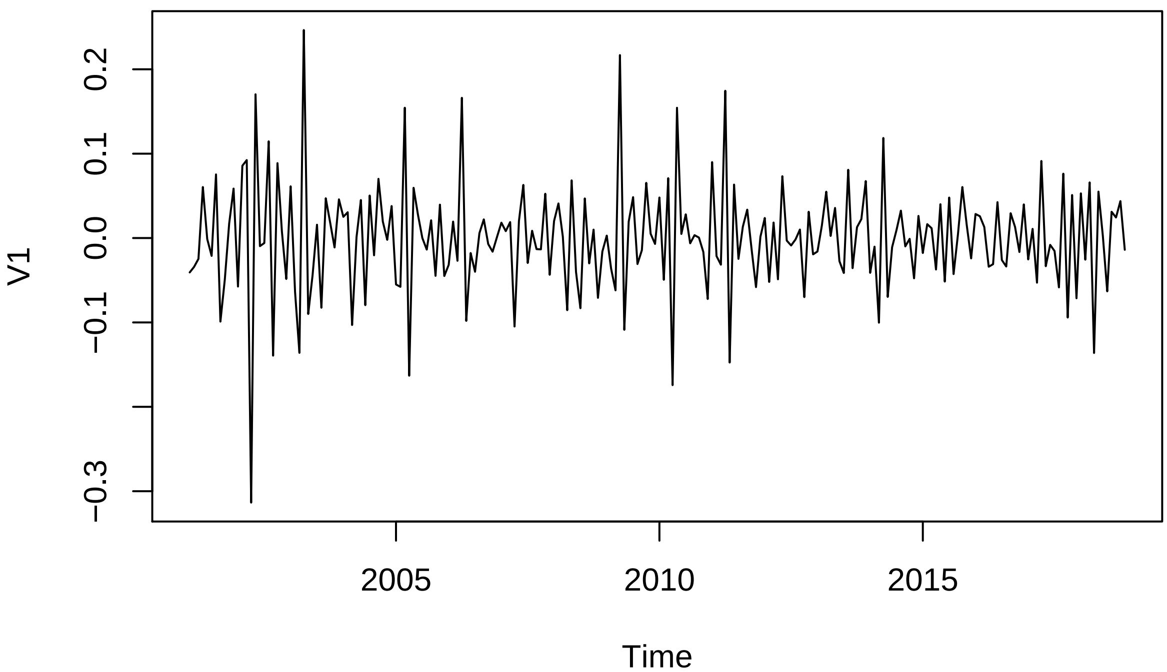
S'observa que amb una diferenciació d'ordre 12 s'ha eliminat el patró estacional. Ara bé, la mitjana de la sèrie encara no és constant.

2.2.3 Mitjana constant

Per últim, es vol aconseguir que la sèrie tingui mitjana constant igual (i si és possible igual a 0) per a poder considerar definitivament la sèrie com un procés estacionari. Per aconseguir-ho, es realitzaran diferenciacions regulars de la sèrie fins que s'obtingui el resultat desitjat

$$W_t = X_t - X_{t-1} = (1 - B)X_t$$

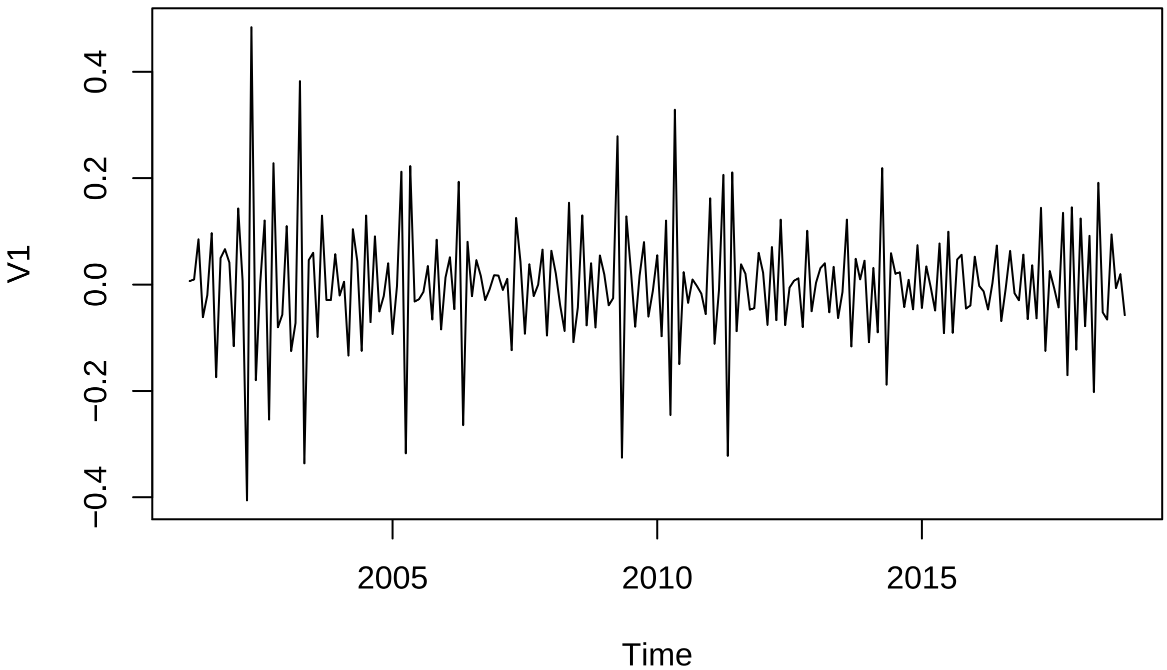
Es realitza la primera diferenciació. Els valors de mitjana i variància aconseguits són els seguents:




```
## [1] -0.0003272785
```

```
##             V1
## V1 0.004266965
```

Com es pot observar, la mitjana del procés diferenciat regularment un cop es pot considerar constant i nula. Ara bé, es mira de diferenciar un segon cop i s'observa que la variància augmenta i, per tant, es té *overdifferentiation*.



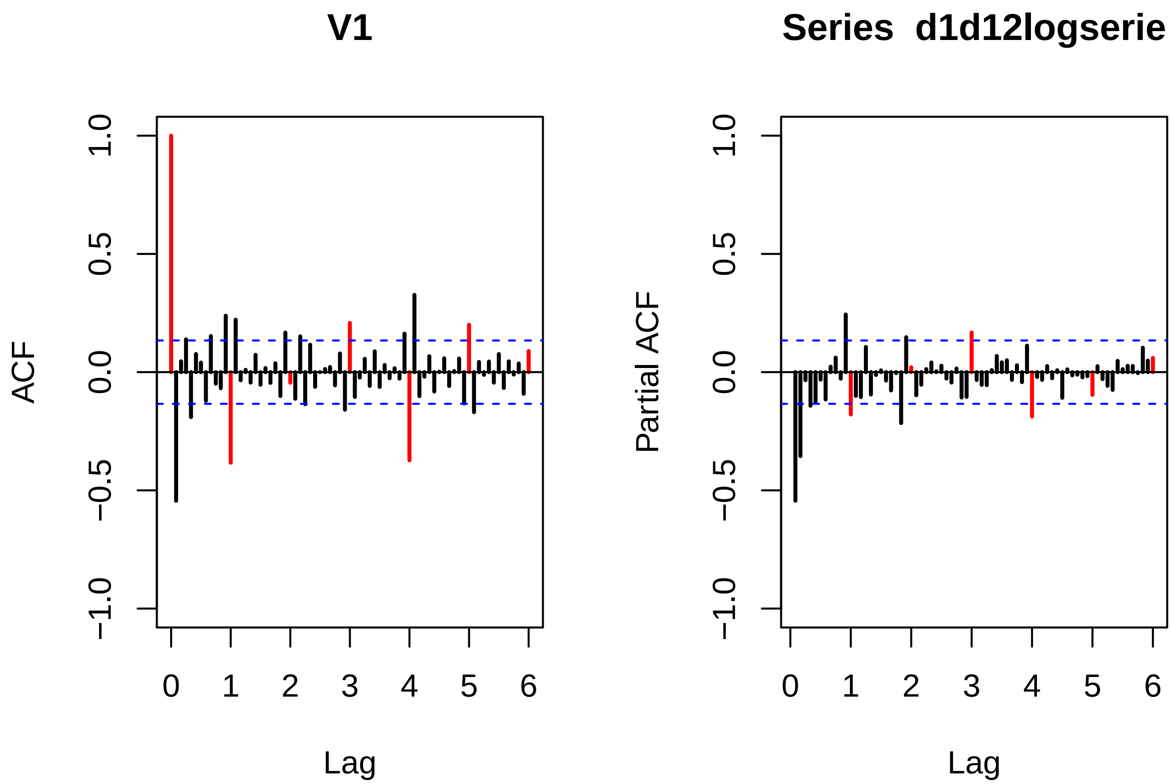

```
## [1] 0.0001260592
```

```
##            V1
## V1 0.0132293
```

En definitiva, la sèrie transformada pel logaritme, diferenciada un cop i amb una diferenciació d'ordre 12 per eliminar el patró estacional (`d1d12logserie`) és un procés estacionari de mitjana 0.

2.3 ACF/PACF de les dades i proposta de models

Tot seguit, es realitza un anàlisi de les funcions *AutoCorrelació* i de *Correlació Parcial* de la sèrie transformada, és a dir, de la sèrie estacionària.



2.3.1 Models proposats per la part regular (p,d,q)

En relació a la part regular de la sèrie, en la funció d'AutoCorrelació (ACF) s'observa que només sobresurt el primer valor. La resta de valors es poden considerar nuls, ja que o bé estan dintre de l'interval de confiança, o bé es poden assignar al cas d'aleatorietat del 5%. Per tant, per la part regular, es proposaria $q = 1$.

Pel que fa a la funció de Correlació Parcial (PACF) s'observa un decreixement exponencial dels primers valors. S'observen també valors fora de la banda de confiança, però poden ser assignats a la aleatorietat del cas 5%. Per tant, en aquest cas, es proposaria $p = 0$. En tot cas, si es volgués mirar d'incloure el primer valor que sobresurt més que la resta, es podria considerar també $p = 1$.

Donat que s'ha realitzat diferenciació 1 cop, es té que $d = 1$. Per tant, els models proposats per la part regular serien $MA(1)$ o, en tot cas, $ARMA(1,1)$ sobre la sèrie transformada regular.

2.4 Models proposats per la part estacional (P,D,Q)

En relació a la part estacional de la sèrie, en la funció d'AutoCorrelació (ACF) s'observa que el primer valor es força significatiu, però també ho són el tercer, el quart i el cinquè, sobretot el quart. Donat que volem intentar proposar un model simplificat, es proposa $Q = 0$.

Pel que fa a la funció de Correlació Parcial (PACF) s'observa que sobresurt el primer valor una mica i també sobresurten el tercer i el quart valor. Ara bé, no sobresurten de manera tant significativa com en el cas dels valors del ACF i, per tant, podem assignar-ho al cas d'aleatorietat del 5%. Per tant, en aquest cas, es proposaria $P = 1$.

Donat que s'ha realitzat una diferenciació d'ordre 12 per eliminar el patró estacional, es té que $D = 1$. Per tant, el model proposat per la part regular seria un $AR(1)$

2.5 Models proposats

En conclusió, es proposen per la sèrie diferenciada els models estacionals:

-$ARMA(0,1)(1,0)_s$ -$ARMA(1,1)(1,0)_s$

I per la sèrie original, tenint en compte les diferenciacions, es proposen:

-$SARIMA(0,1,1)(1,1,0)_s$ -$SARIMA(1,1,1)(1,1,0)_s$

3 Estimació dels models

A continuació, s'estimen els coeficients dels dos models proposats i es mira que tots siguin significatius. Per mirar-ho, es realitza el test següent (suposant que estem davant d'un model MA:

$$H_0 : \theta_i = 0$$

$$H_1 : \theta_i \neq 0$$

amb l'estadístic

$$\hat{t} = \frac{\hat{\theta}_i}{\text{se}(\hat{\theta}_i)}$$

~

$$t - student_{T-k}$$

, on k és el nombre total de paràmetres i T és el període. Ara bé, a la pràctica es diu que un coeficient és significant si $|\hat{t}| > 2$.

En primer lloc, s'estimen els coeficients dels models proposats, amb intercept i sense.

```
##
## Call:
## arima(x = d1d12logserie, order = c(0, 0, 1), seasonal = list(order = c(1, 0,
##     0), period = 12))
##
## Coefficients:
##           ma1     sar1  intercept
##       -0.6947  -0.3559     -1e-04
## s.e.   0.0463   0.0648      8e-04
##
## sigma^2 estimated as 0.002268:  log likelihood = 346.73,  aic = -685.46
```

```
##
## Call:
## arima(x = logserie, order = pdq.1, seasonal = list(order = PDQ.1, period = 12))
##
## Coefficients:
##           ma1     sar1
##       -0.6946  -0.3560
## s.e.   0.0463   0.0648
##
## sigma^2 estimated as 0.002268:  log likelihood = 346.72,  aic = -687.44
```

```
##
## Call:
## arima(x = d1d12logserie, order = c(1, 0, 1), seasonal = list(order = c(1, 0,
##     0), period = 12))
```

```
##
## Coefficients:
##           ar1     ma1     sar1  intercept
##       -0.1469  -0.615  -0.3517     -1e-04
## s.e.   0.0989   0.081   0.0650      8e-04
##
## sigma^2 estimated as 0.002245:  log likelihood = 347.81,  aic = -685.63
```

```
##
## Call:
## arima(x = logserie, order = pdq.2, seasonal = list(order = PDQ.2, period = 12))
##
## Coefficients:
##           ar1      ma1     sar1
##       -0.1469  -0.6149  -0.3518
## s.e.   0.0989   0.0811   0.0650
##
## sigma^2 estimated as 0.002245:  log likelihood = 347.8,  aic = -687.61
```

S'observa que, en ambdós casos, l'intercept no és significatiu i, per tant, es descarten aquests dos models. En el cas del model $SARIMA(0,1,1)(1,1,0)_s$, els dos coeficients són significatius. Ara bé, en l'altre model proposat, el model $SARIMA(1,1,1)(1,1,0)_s$, s'observa que el coeficient `ar1` no és significatiu. Tot i així, el primer model proposat ($SARIMA(0,1,1)(1,1,0)_s$), té un pitjor AIC que l'altre model i té una *loglikelihood* més baixa. Ara bé, donat que eliminant el coeficient no significatiu del segon model, queda el primer model, *a-priori* s'escolliria el primer model. Falta, òbviament, la seva validació.

```
##       ma1      sar1 intercept
##      TRUE      TRUE     FALSE
```

```
##  ma1 sar1
## TRUE TRUE
```

```
##       ar1       ma1      sar1 intercept
##     FALSE      TRUE      TRUE     FALSE
```

```
##   ar1   ma1  sar1
## FALSE  TRUE  TRUE
```

4 Validació dels Models

Tot seguit, es realitzarà la validació dels dos models proposat. En el procés de validació es realitzarà un anàlisi dels residus (Z_t) dels models, es comprovarà que aquests siguin estacionaris i invertibles, es verificarà la seva estabilitat i s'avaluarà la seva capacitat de previsió.

4.1 Estudi dels residus dels models

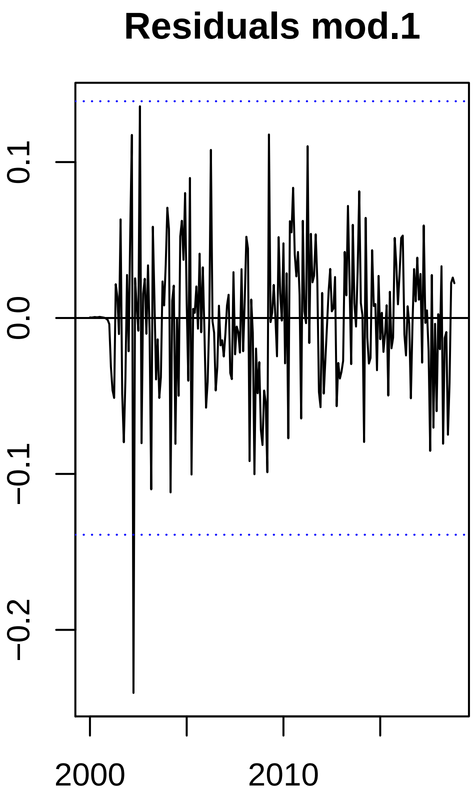
Així doncs, en primer lloc, s'estudiaràn els residus del model i es comprovaran els següents aspectes:

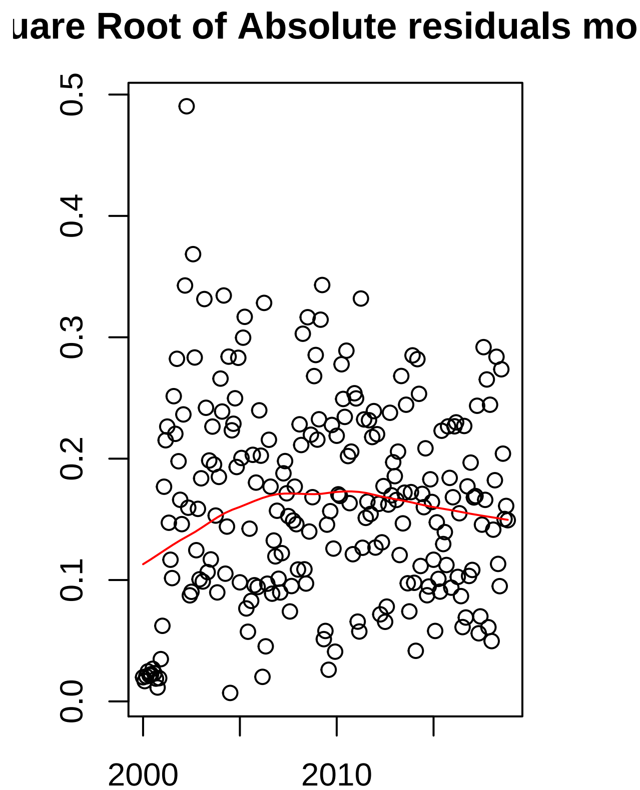
- Homogeneïtat de la variància residual (σ_Z^2 constant).
- Normalitat ($Z_T \sim$ Normal).
- Independència ($\rho(k) = 0 \; \forall k > 0$).

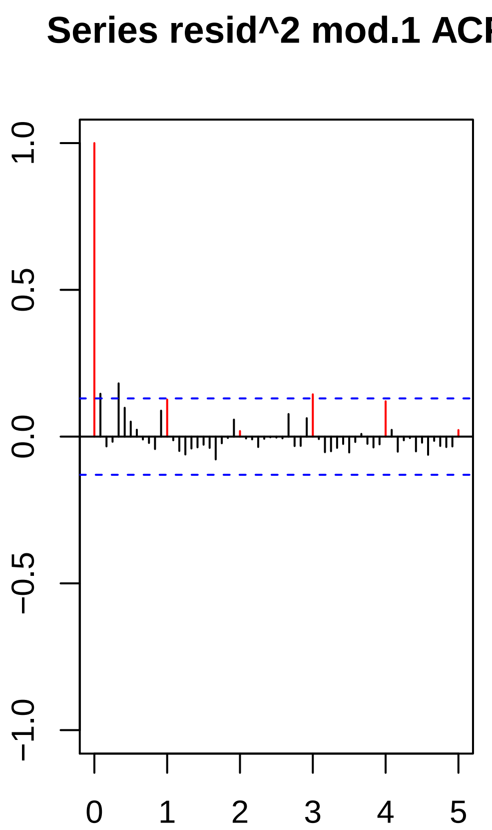
4.1.1 Homogeneïtat de la variància

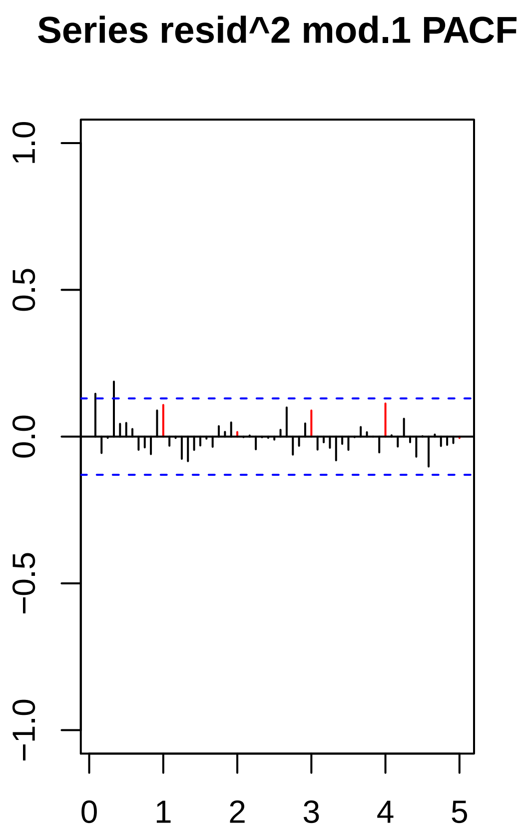
Per comprovar l'homogeneïtat de la variància dels residus, s'analitzen el plot dels mateixos residus, el plot de l'arrel quadrada del seu valor absolut i les funcions ACF i PACF del seu quadrat.

En el cas del primer model (`mod.1`) no s'observa cap tipus de patró (ni creixent ni decreixent) en el plot dels residus o en el plot de l'arrel quadrada del seu valor absolut. A més, en l'ACF i el PACF del quadrat dels residus tots els valors estan dintre de la banda de confiança i, per tant, els podem considerar nuls.

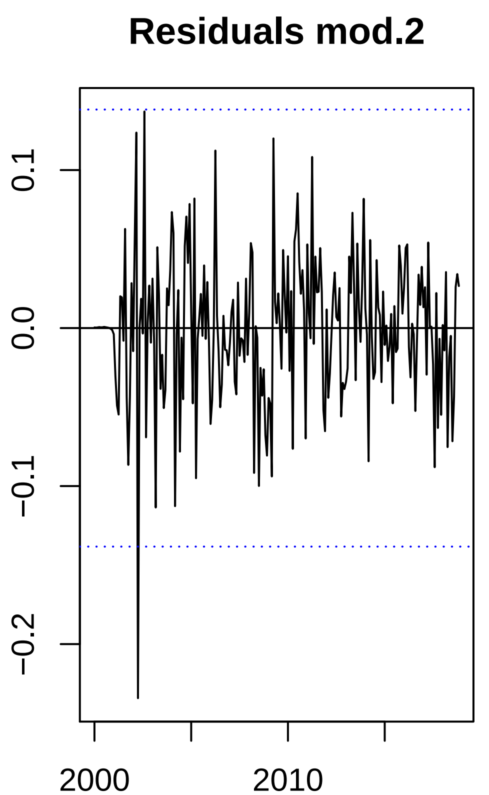


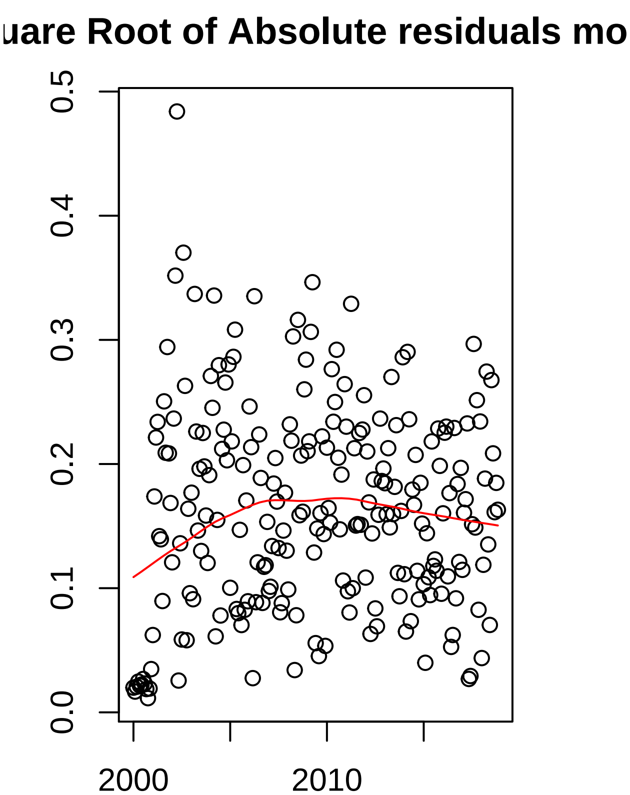


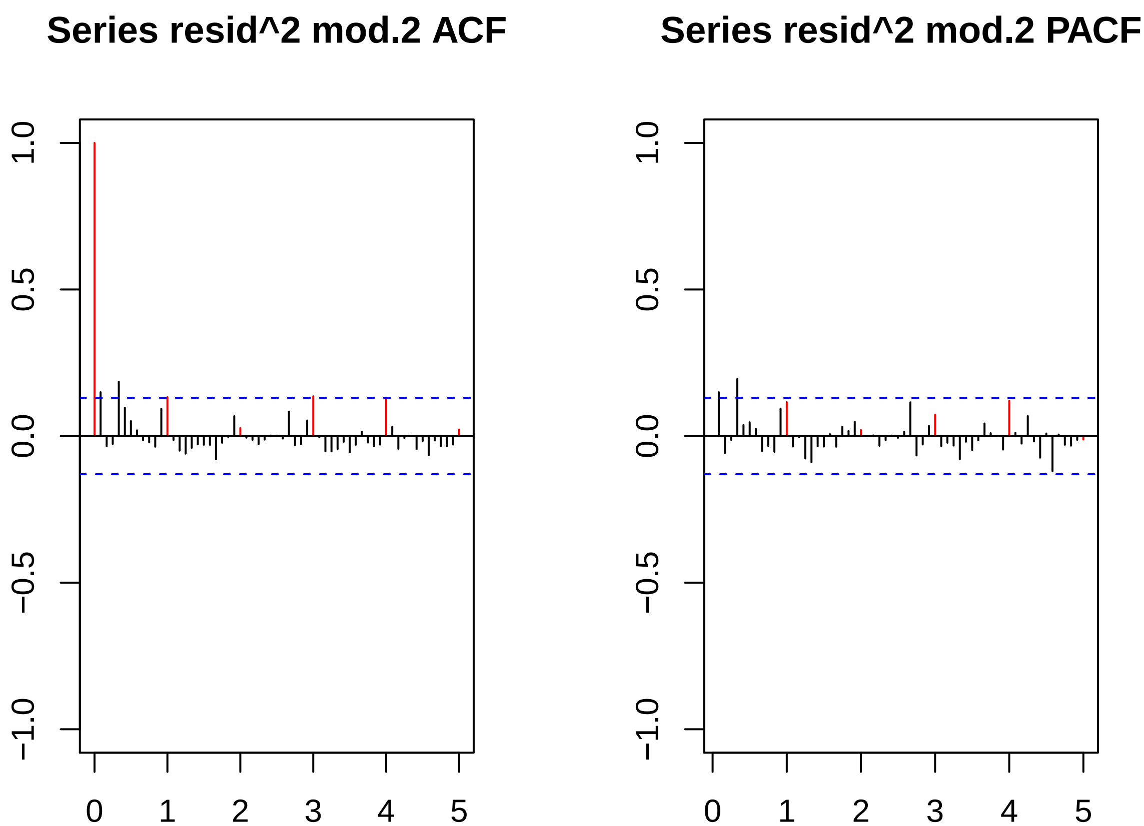




En el cas del segon model (`mod.1`), es poden extreure les mateixes conclusions que en el primer model i, per tant, també es pot assumir homogeneïtat de variància residual.



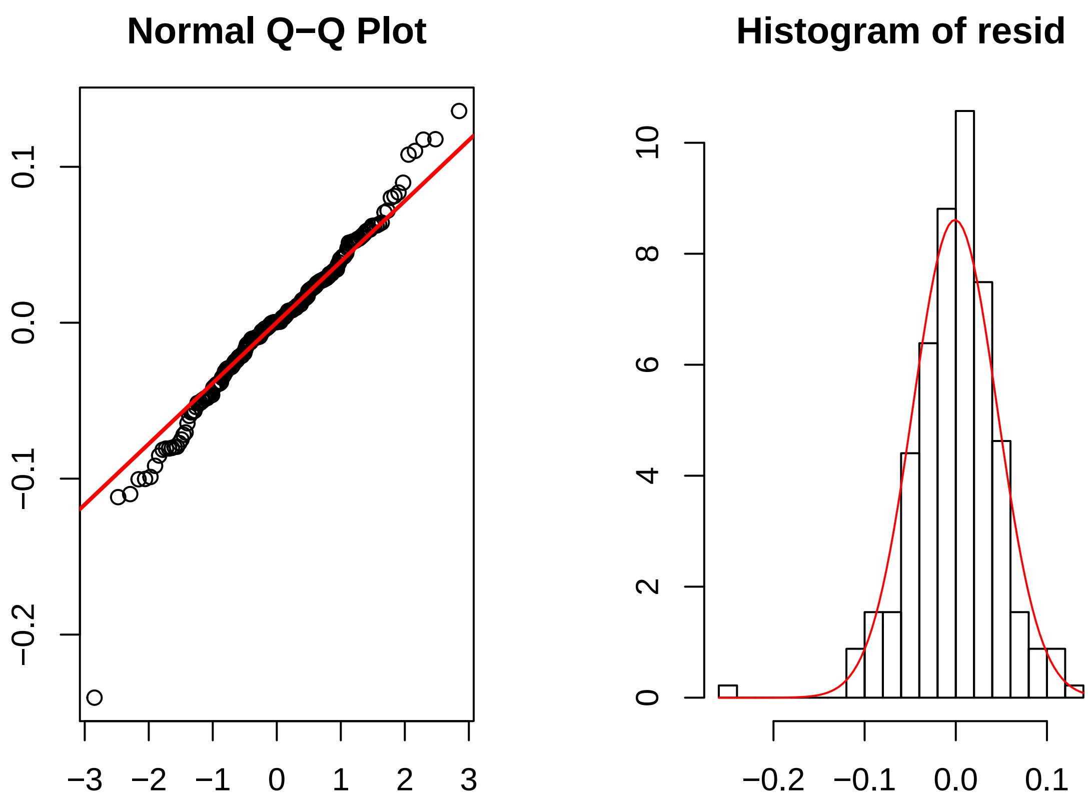




4.1.2 Normalitat

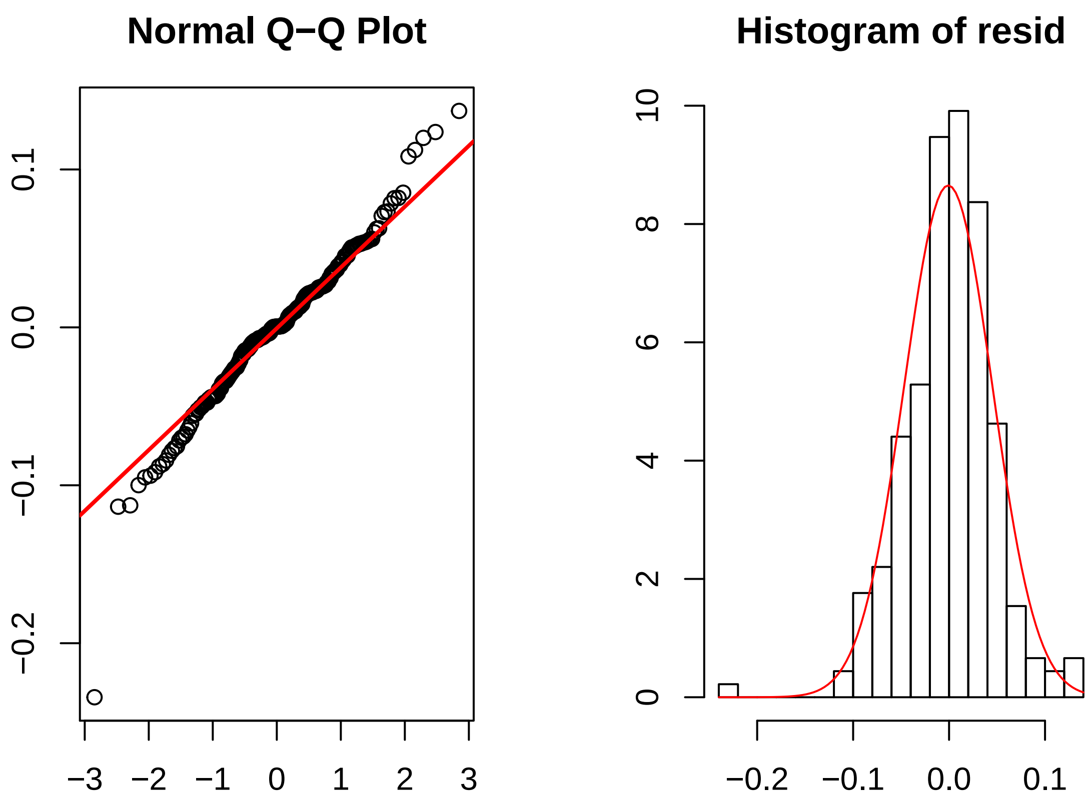
Per comprovar la normalitat dels residus dels models proposats s'estudiarà el Q-Q plot, l'histograma dels residus amb la normal que s'hauria de seguir sobreposada i es realitzarà el test de Sharipo-Wilks.

En el cas del model `mod.1`, s'observa en el Q-Q plot que els quartils es situen sobre la línia dels quartils teòrics i que l'histograma s'ajusta a la distribució normal a la que s'hauria d'ajustar. A més, el *p-value* del test de Sharipo-Wilks és 4.879×10^{-05}, menor que 0.05 i, per tant, es pot assumir la hipòtesi de normalitat en els residus.




```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model)
## W = 0.96771, p-value = 4.879e-05
```

En el cas del model `mod.2`, les conclusions que s'extreuen són les mateixes. En aquest cas, el *p-value* és de 7.675×10^{-05}. Per tant, també assumim normalitat en aquest cas.

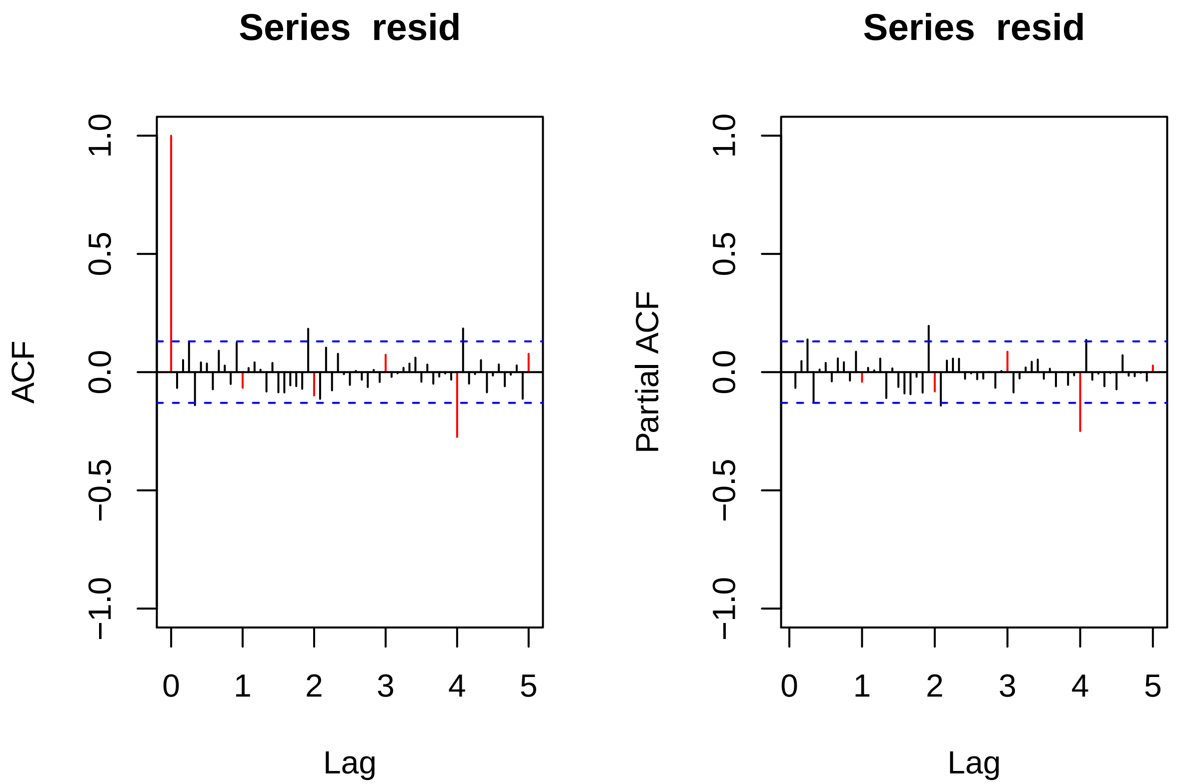


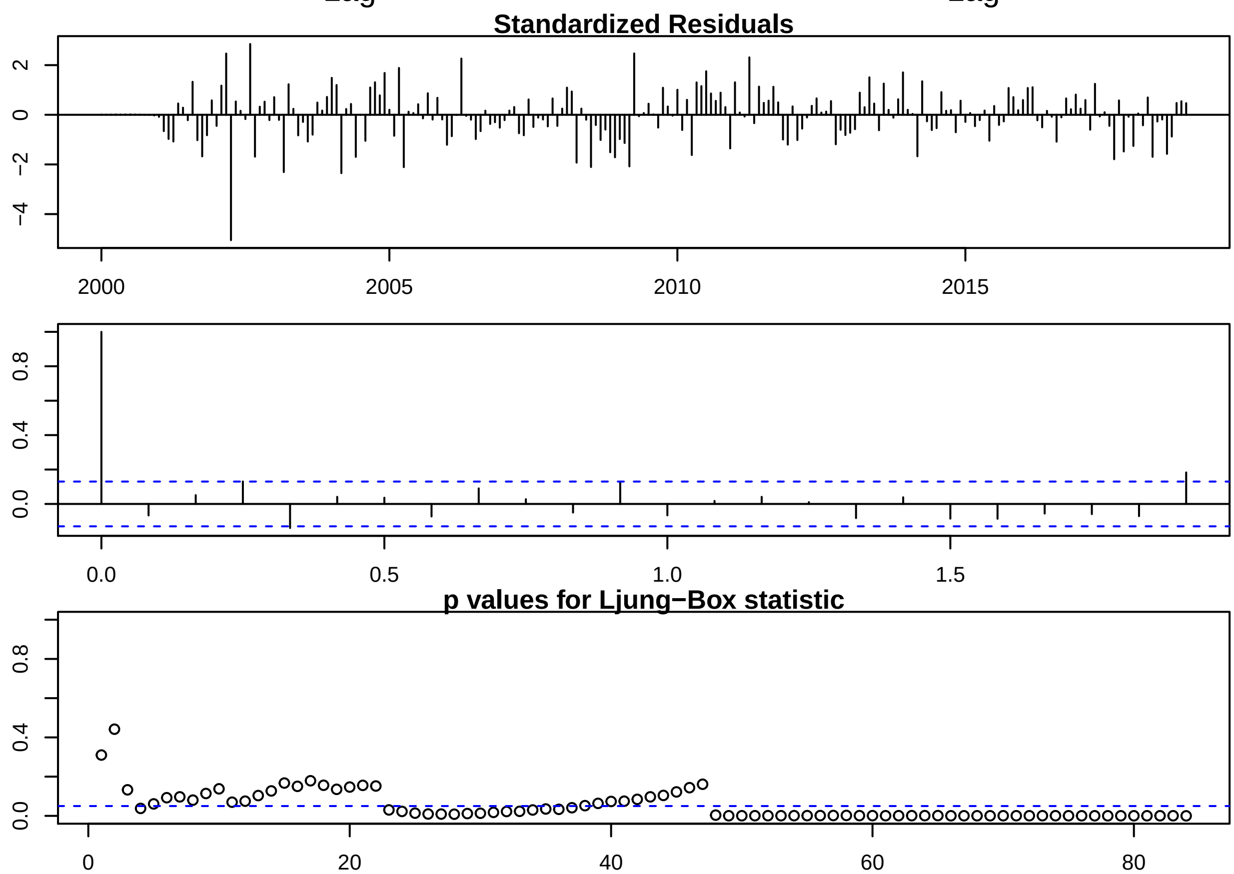

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model)
## W = 0.96922, p-value = 7.675e-05
```

4.1.3 Independència

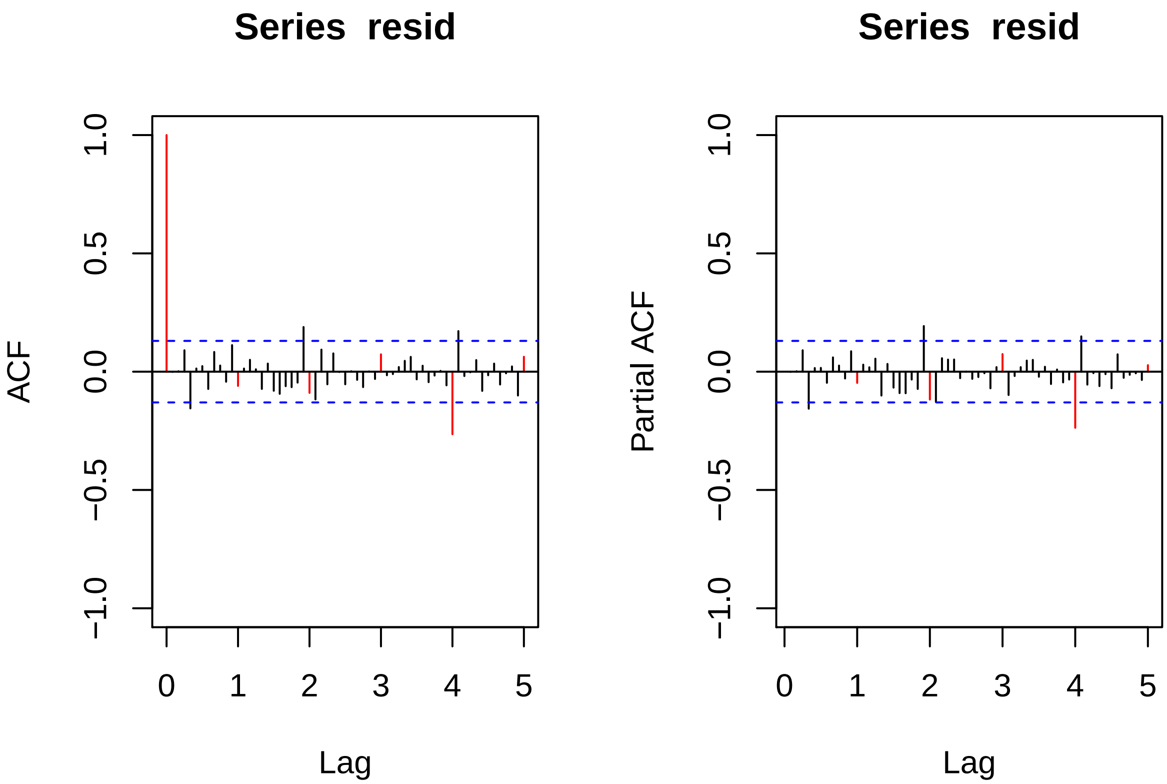
Per comprovar la independència residual, és a dir, que $\rho(k) = 0 \ \forall k > 0$ s'estudiarà el ACF i el PACF dels residus i es realitzarà el test de Ljung-Box.

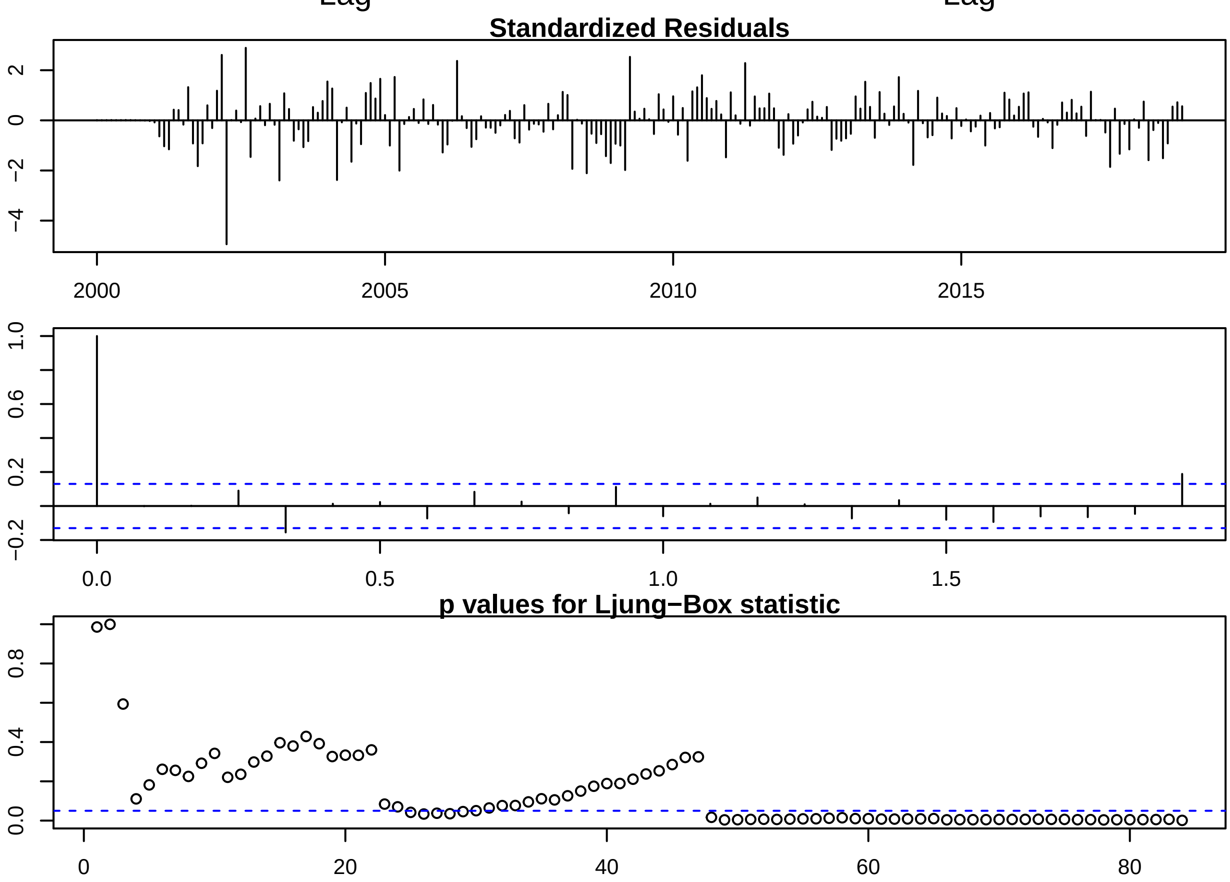
Pel que fa al primer model, en primer lloc observem que les funcions ACF i PACF prenen valors pràcticament iguals, cosa que ja fa intuïr que es complirà la independència. Els residus estandaritzats prenen valors dintre de la franja de (-2,2), la gran majoria, que és el comportament esperat. A més, els *p-values* del test Ljung-Box, que gairebé tots menors que 0.05, confirmen que es pot assumir la independència dels residus.





En el segon model, l'anàlisi es pràcticament el mateix, tret que, en aquest cas, els *p-values* els costa més assolir un valor per sota de 0.05. Tot i així, també podem assumir la independència (tot i que no de manera tant clara com en el cas anterior).





4.1.4 Estacionaritat i invertibilitat dels models

Per analitzar l'estacionaritat i la invertibilitat dels models proposats, s'expresaran els models com a models $AR(\infty)$ i $MA(\infty)$:

$$
\begin{aligned}
&(1-\phi_1 B-\cdots-\phi_p B^p)X_t=(1+\theta_1 B+\cdots+\theta_q B^q)Z_t\\
AR(\infty):\quad &\frac{1-\phi_1 B-\cdots-\phi_p B^p}{1+\theta_1 B+\cdots+\theta_q B^q}X_t=(1-\pi_1 B-\pi_2 B-\cdots)X_t=Z_t\\
MA(\infty):\quad &\frac{1+\theta_1 B+\cdots+\theta_q B^q}{1-\phi_1 B-\cdots-\phi_p B^p}Z_t=(1+\psi_1 B+\psi_2 B+\cdots)Z_t=X_t
\end{aligned}
$$

A partir d'aquí el models seran *invertibles* si el mòdul de totes les arrels de $\theta_q(B)=1+\theta_1 B+\cdots+\theta_q B^q$ és major que 1, és a dir, si $\sum_{i\geq 0}\pi_i^2<\infty$. Per per altra banda, seran *estacionaris* si el mòdul de totes les arrels de $\phi_q(B)=1-\phi_1 B-\cdots-\phi_q B^q$ és major que 1, és a dir, si $\sum_{i\geq 0}\psi_i^2<\infty$.

En el cas del primer model, s'observa que es compleixen totes les condicions i, per tant, el `mod.1` és estacionari i invertible.

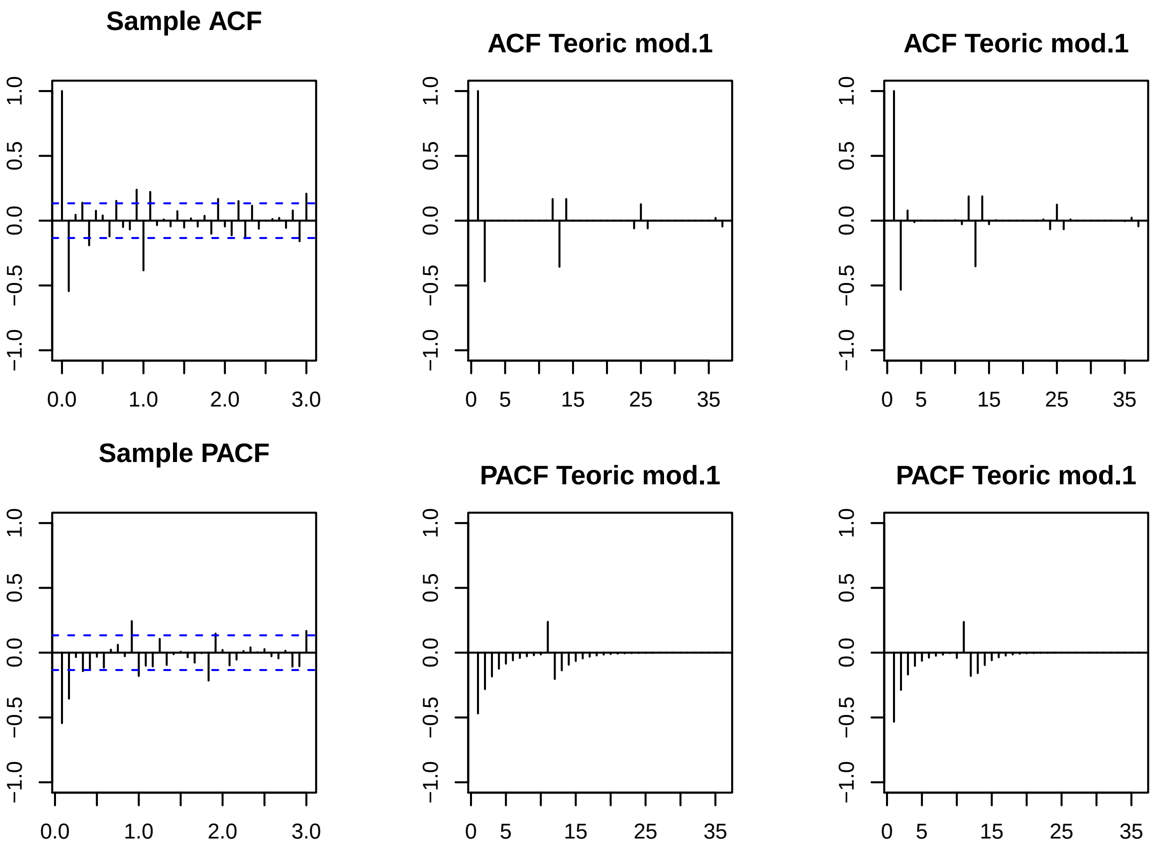
```
##
## Modul of AR Characteristic polynomial Roots:  1.089872 1.089872 1.089872 1.089872 1.089872 1.089872

##
## Modul of MA Characteristic polynomial Roots:  1.439675

##
## Psi-weights (MA(inf))

##
## --------------------

##      psi 1      psi 2      psi 3      psi 4      psi 5      psi 6
## -0.6946012  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##      psi 7      psi 8      psi 9     psi 10     psi 11     psi 12
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000 -0.3560354
##     psi 13     psi 14     psi 15     psi 16     psi 17     psi 18
##  0.2473027  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##     psi 19     psi 20
##  0.0000000  0.0000000

##
## Pi-weights (AR(inf))

##
## --------------------

##        pi 1        pi 2        pi 3        pi 4        pi 5        pi 6
## -0.69460120 -0.48247083 -0.33512482 -0.23277810 -0.16168795 -0.11230864
##        pi 7        pi 8        pi 9       pi 10       pi 11       pi 12
## -0.07800972 -0.05418564 -0.03763741 -0.02614299 -0.01815895 -0.36864868
##       pi 13       pi 14       pi 15       pi 16       pi 17       pi 18
## -0.25606382 -0.17786224 -0.12354332 -0.08581334 -0.05960605 -0.04140243
##       pi 19       pi 20
## -0.02875818 -0.01997547

##
## Modul of AR Characteristic polynomial Roots:  1.090959 1.090959 1.090959 1.090959 1.090959 1.090959

##
## Modul of MA Characteristic polynomial Roots:  1.626315
```

```
##
## Psi-weights (MA(inf))
##
## --------------------
##          psi 1          psi 2          psi 3          psi 4          psi 5
## -7.618201e-01  1.119367e-01 -1.644722e-02  2.416642e-03 -3.550850e-04
##          psi 6          psi 7          psi 8          psi 9         psi 10
##  5.217378e-05 -7.666061e-06  1.126399e-06 -1.655054e-07  2.431824e-08
##         psi 11         psi 12         psi 13         psi 14         psi 15
## -3.573157e-09 -3.518036e-01  2.680111e-01 -3.937973e-02  5.786190e-03
##         psi 16         psi 17         psi 18         psi 19         psi 20
## -8.501835e-04  1.249202e-04 -1.835492e-05  2.696948e-06 -3.962712e-07
##
## Pi-weights (AR(inf))
##
## --------------------
##          pi 1          pi 2          pi 3          pi 4          pi 5
## -0.761820141 -0.468433247 -0.288033481 -0.177108023 -0.108901408
##          pi 6          pi 7          pi 8          pi 9         pi 10
## -0.066962053 -0.041174091 -0.025317410 -0.015567345 -0.009572157
##         pi 11         pi 12         pi 13         pi 14         pi 15
## -0.005885794 -0.355422705 -0.270236410 -0.166164836 -0.102172586
##         pi 16         pi 17         pi 18         pi 19         pi 20
## -0.062824588 -0.038630018 -0.023753093 -0.014605466 -0.008980710
```

En el cas del segon model, també es compleix tot i, per tant, també és invertible i estacionari.

4.1.5 Comparació entre els ACF/PACF mostrals i els ACF/PACF teòrics

Per últim, comparem els valors del ACF i el PACF de les dades amb els valors teòric. S'observa que, en el cas del model `mod.1`, els valors teòrics s'aproximen gairebé perfectament als valors mostrals. En el cas del segon model també es podria dir el mateix, tot i que el quart valor de l'ACF teòric és negatiu i el mostral és positiu. Per tant, ambdós models s'aproximen als valors de ACF/PACF de les mostres, potser una mica millor el `mod.1`.



4.2 Estabilitat dels Models

Per comprovar l'estabilitat dels models proposats, calculem els models de la serie ocultant les 12 últimes observacions, és a dir, l'últim període d'observacions. Així doncs, s'observa que el valor dels coeficients varia molt poc, de l'ordre de menys de 0.07 en gairebé tots els casos. Per tant, podem confirmar que els models són estables.

```
## ########## Model ARIMA(0,1,1)(1,1,0)12 amb i sense les 12 últimes observacions ##########

##
## Call:
## arima(x = lnserie1, order = pdq.1, seasonal = list(order = PDQ.1, period = 12))
##
## Coefficients:
##           ma1     sar1
##       -0.6946  -0.3560
## s.e.   0.0463   0.0648
##
## sigma^2 estimated as 0.002268:  log likelihood = 346.72,  aic = -687.44

##
## Call:
## arima(x = lnserie2, order = pdq.1, seasonal = list(order = PDQ.1, period = 12))
##
## Coefficients:
##           ma1     sar1
##       -0.7019  -0.3478
```

```
## s.e.   0.0483   0.0664
##
## sigma^2 estimated as 0.002306:  log likelihood = 327.18,  aic = -648.36
```

```
## ########## Model ARIMA(1,1,1)(1,1,0)12 amb i sense les 12 últimes observacions ##########
```

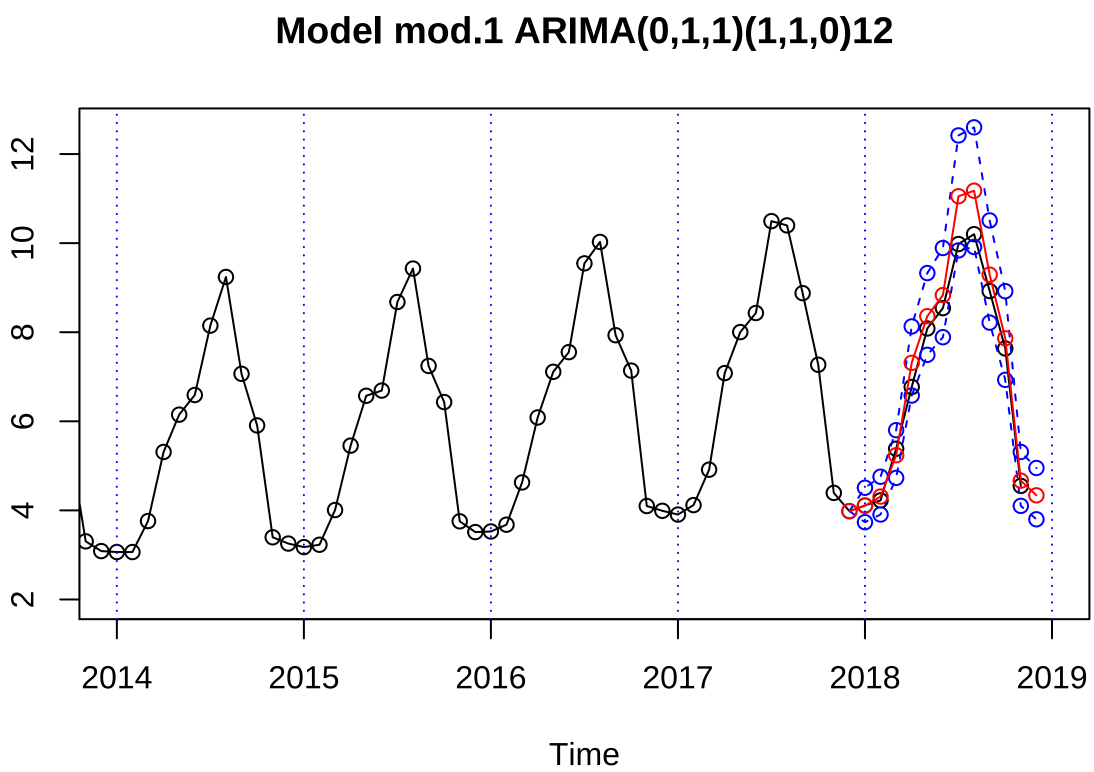
```
##
## Call:
## arima(x = lnserie1, order = pdq.2, seasonal = list(order = PDQ.2, period = 12))
##
## Coefficients:
##          ar1      ma1     sar1
##      -0.1469  -0.6149  -0.3518
## s.e.   0.0989   0.0811   0.0650
##
## sigma^2 estimated as 0.002245:  log likelihood = 347.8,  aic = -687.61
```

```
##
## Call:
## arima(x = lnserie2, order = pdq.2, seasonal = list(order = PDQ.2, period = 12))
##
## Coefficients:
##          ar1      ma1     sar1
##      -0.1471  -0.6204  -0.3441
## s.e.   0.1023   0.0848   0.0666
##
## sigma^2 estimated as 0.002283:  log likelihood = 328.19,  aic = -648.39
```

4.3 Capacitat de predicció

A continuació s'avaluarà la capacitat de predicció dels dos models proposats fent-los predir el valor de les 12 útlimes observacions utilitzant la resta d'observacions conegudes.

```
#mod.1

pred=predict(mod12,n.ahead=12)
pr<-ts(c(tail(lnserie2,1),pred$pred),start=ultim,freq=12)

se<-ts(c(0,pred$se),start=ultim,freq=12)

#Intervals
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)


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 mod.1 ARIMA(0,1,1)(1,1,0)12")
abline(v=(ultim[1]-3):(ultim[1]+2),lty=3,col=4)
```




```
(previs=window(cbind(tl,pr,tu,serie,error=round(serie-pr,3)),start=ultim))
```

```
##                tl        pr        tu     serie  error
## Dec 2017 3.982530  3.982530  3.982530  3.982530  0.000
## Jan 2018 3.734627  4.103191  4.508130  4.110137  0.007
## Feb 2018 3.907489  4.310720  4.755563  4.224826 -0.086
## Mar 2018 4.729512  5.238112  5.801406  5.383687  0.146
## Apr 2018 6.578010  7.313009  8.130133  6.770845 -0.542
## May 2018 7.492564  8.360233  9.328383  8.084173 -0.276
## Jun 2018 7.888484  8.833173  9.890992  8.541181 -0.292
## Jul 2018 9.836784 11.052612 12.418716  9.979779 -1.073
## Aug 2018 9.913943 11.176461 12.599758 10.201456 -0.975
## Sep 2018 8.215819  9.292140 10.509466  8.924326 -0.368
## Oct 2018 6.930365  7.863066  8.921291  7.635569 -0.227
## Nov 2018 4.101182  4.667478  5.311970  4.549899 -0.118
## Dec 2018 3.800263  4.338037  4.951911        NA     NA
```

```
obs=window(serie,start=ultim)
cat("\n")
```

```
cat("###### Errors de predicció del model mod.1 ######\n")
```

```
## ###### Errors de predicció del model mod.1 ######
```

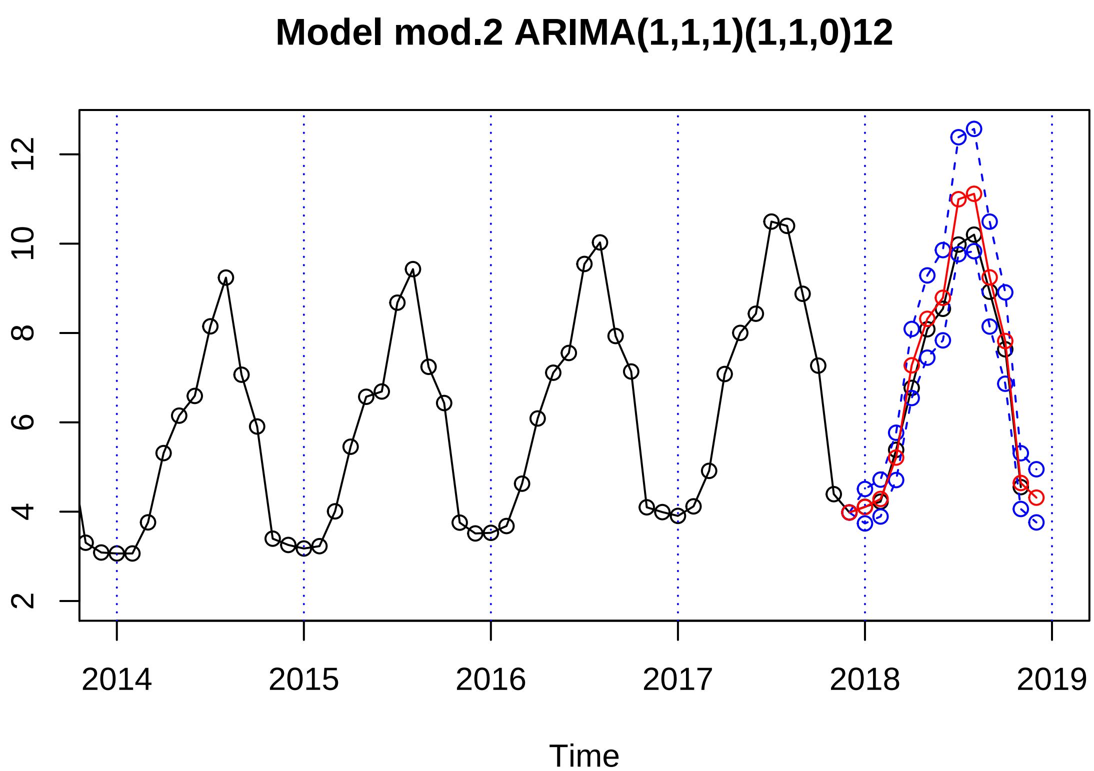
```
(mod.EQM1=sqrt(sum(((obs-pr)/obs)^2)/12))
```

```
## [1] 0.05310294
```

```
(mod.EAM1=sum(abs(obs-pr)/obs)/12)
```

```
## [1] 0.04144964
```

```r
#mod.2

pred=predict(mod22,n.ahead=12)
pr<-ts(c(tail(lnserie2,1),pred$pred),start=ultim,freq=12)

se<-ts(c(0,pred$se),start=ultim,freq=12)

#Intervals
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)


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 mod.2 ARIMA(1,1,1)(1,1,0)12")
abline(v=(ultim[1]-3):(ultim[1]+2),lty=3,col=4)
```




```r
(previs=window(cbind(tl,pr,tu,serie,error=round(serie-pr,3)),start=ultim))
```

```
##                tl        pr        tu     serie  error
## Dec 2017 3.982530  3.982530  3.982530  3.982530  0.000
## Jan 2018 3.737531  4.104492  4.507483  4.110137  0.006
## Feb 2018 3.892287  4.285127  4.717617  4.224826 -0.060
## Mar 2018 4.708133  5.210834  5.767209  5.383687  0.173
## Apr 2018 6.544441  7.276383  8.090188  6.770845 -0.506
## May 2018 7.447661  8.317507  9.288947  8.084173 -0.233
## Jun 2018 7.835241  8.787742  9.856034  8.541181 -0.247
## Jul 2018 9.763159 10.995148 12.382598  9.979779 -1.015
## Aug 2018 9.831280 11.115953 12.568498 10.201456 -0.914
```

```
## Sep 2018 8.144641  9.244436 10.492740  8.924326 -0.320
## Oct 2018 6.863957  7.819978  8.909156  7.635569 -0.184
## Nov 2018 4.060406  4.642777  5.308674  4.549899 -0.093
## Dec 2018 3.759502  4.313942  4.950149        NA     NA
```

```r
obs=window(serie,start=ultim)
cat("\n")
```

```r
cat("###### Errors de predicció del model mod.2 ######\n")
```

```
## ###### Errors de predicció del model mod.2 ######
```

```r
(mod.EQM1=sqrt(sum(((obs-pr)/obs)^2)/12))
```

```
## [1] 0.04928805
```

```r
(mod.EAM1=sum(abs(obs-pr)/obs)/12)
```

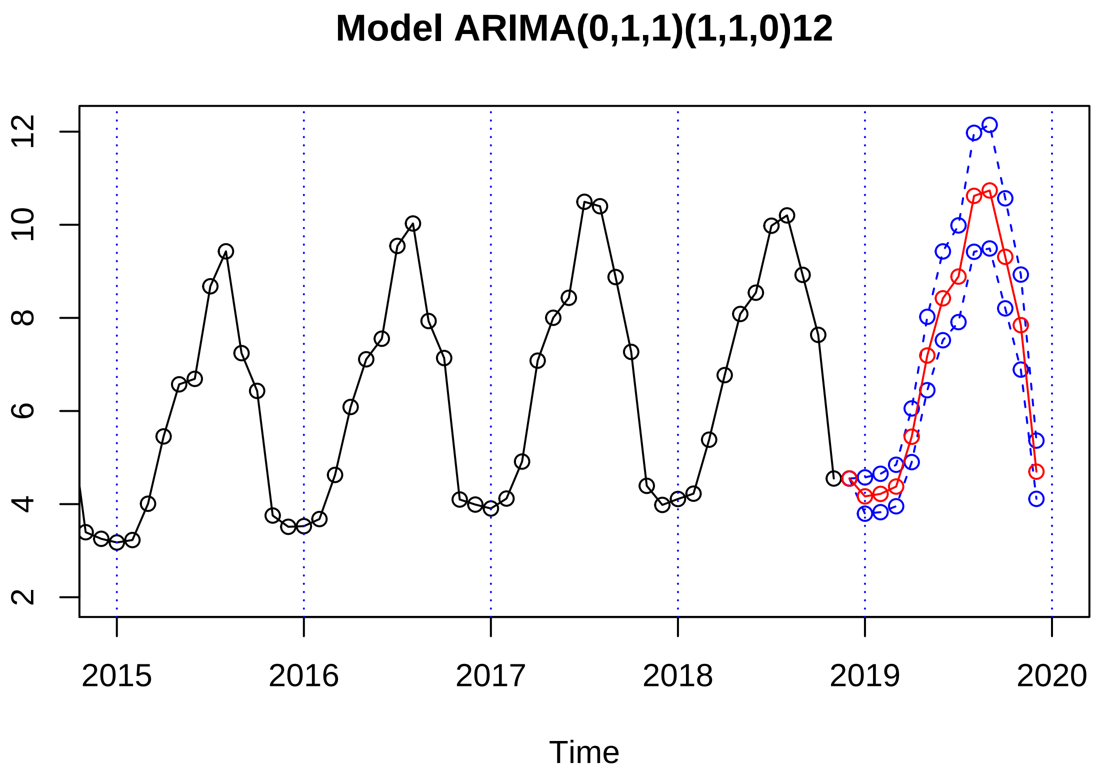
```
## [1] 0.03766399
```

Com es pot veure, les prediccions de les últimes 12 observacions són semblants i prou bones, ja que en amdós casos s'apropen força a la realitat. A més, el valor real de les observacions que dins l'interval de confiança dels valors predits. Per tant, es pot concloure que els models tenen bona capacitat de predicció. A més, els errors de predicció (l'**Error Quadràtic Mitjà** i l'**Error Absolut Mitjà**) dels dos models són semblants i molt petits (> 0.06 en ambdós casos).

4.4 Elecció de model

En definitiva, donat que els dos models han passat la prova de validació i que els dos presenten un comportament similar en la predicció de les últimes 12 observacions, s'escull el primer model, el model $ARIMA(0,1,1)(1,1,0)_{12}$. El principal motiu és que el coeficient que diferencia els dos models surt no significatiu i, per tant, ens quedem amb el model més senzill que, tal i com hem vist, té un bon comportament predictiu i és totalment vàlid.

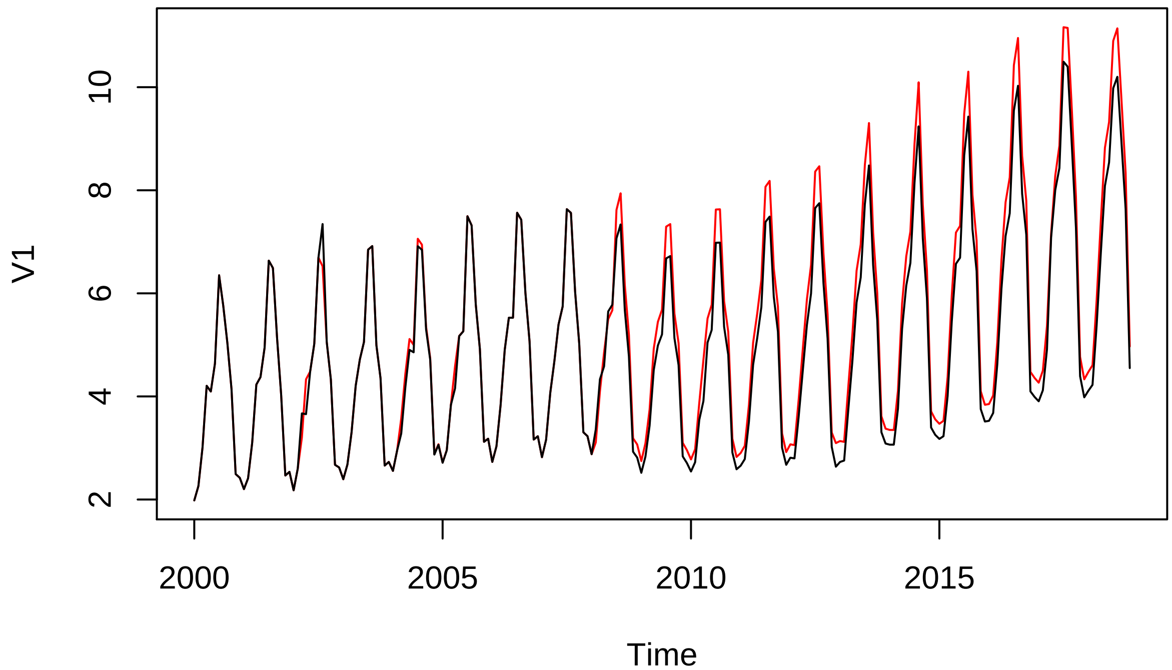
5 Predicció a llarg termini

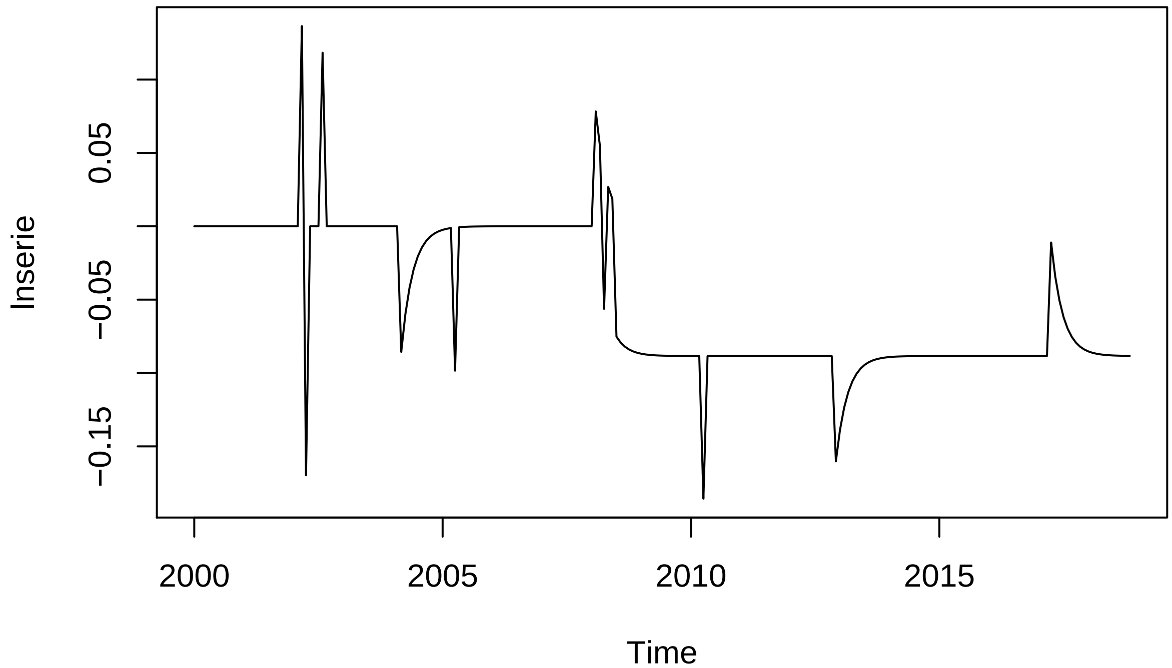


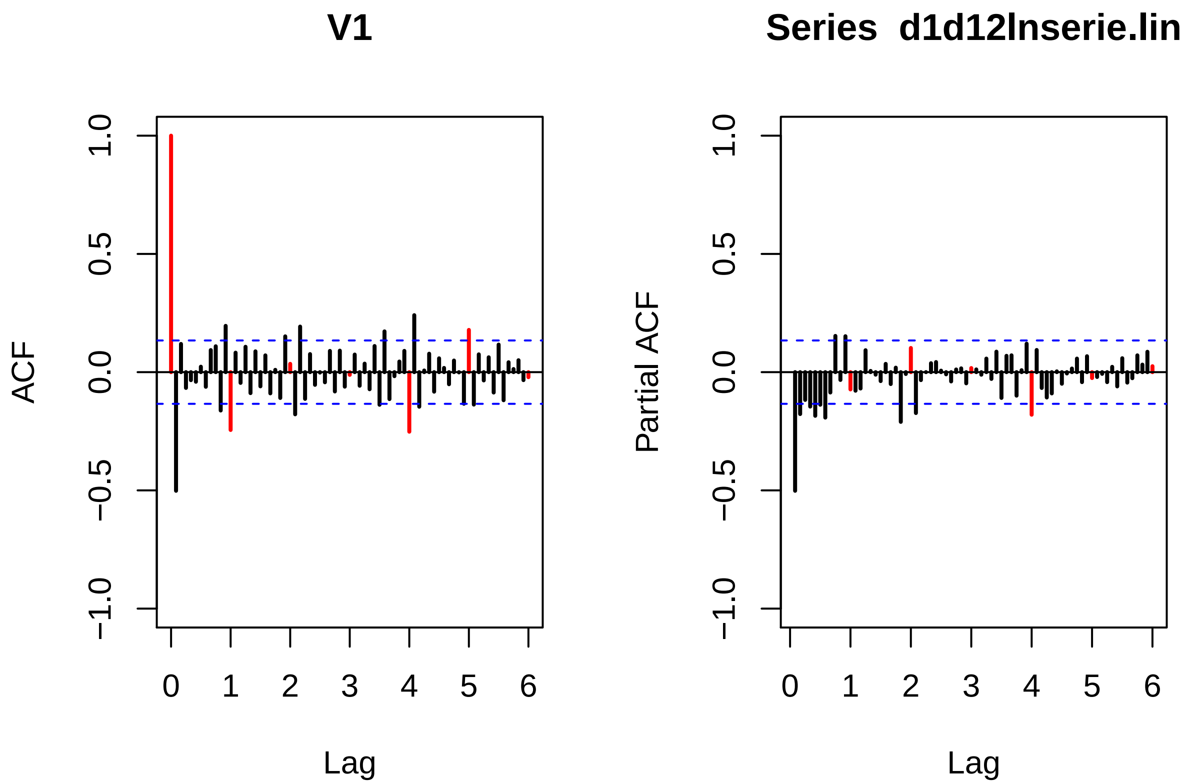

```
##                 tl1       pr1       tu1
## Dec 2018 4.549899  4.549899  4.549899
## Jan 2019 3.794995  4.166268  4.573864
## Feb 2019 3.827086  4.219417  4.651967
## Mar 2019 3.953745  4.376872  4.845281
## Apr 2019 4.902111  5.448048  6.054784
## May 2019 6.447088  7.192220  8.023472
## Jun 2019 7.520711  8.420638  9.428251
## Jul 2019 7.909996  8.887904  9.986711
## Aug 2019 9.420050 10.621064 11.975203
## Sep 2019 9.491309 10.737202 12.146640
## Oct 2019 8.204643  9.311865 10.568507
## Nov 2019 6.889348  7.843893  8.930695
## Dec 2019 4.112863  4.697227  5.364619
```

6 Tractament de *outliers*

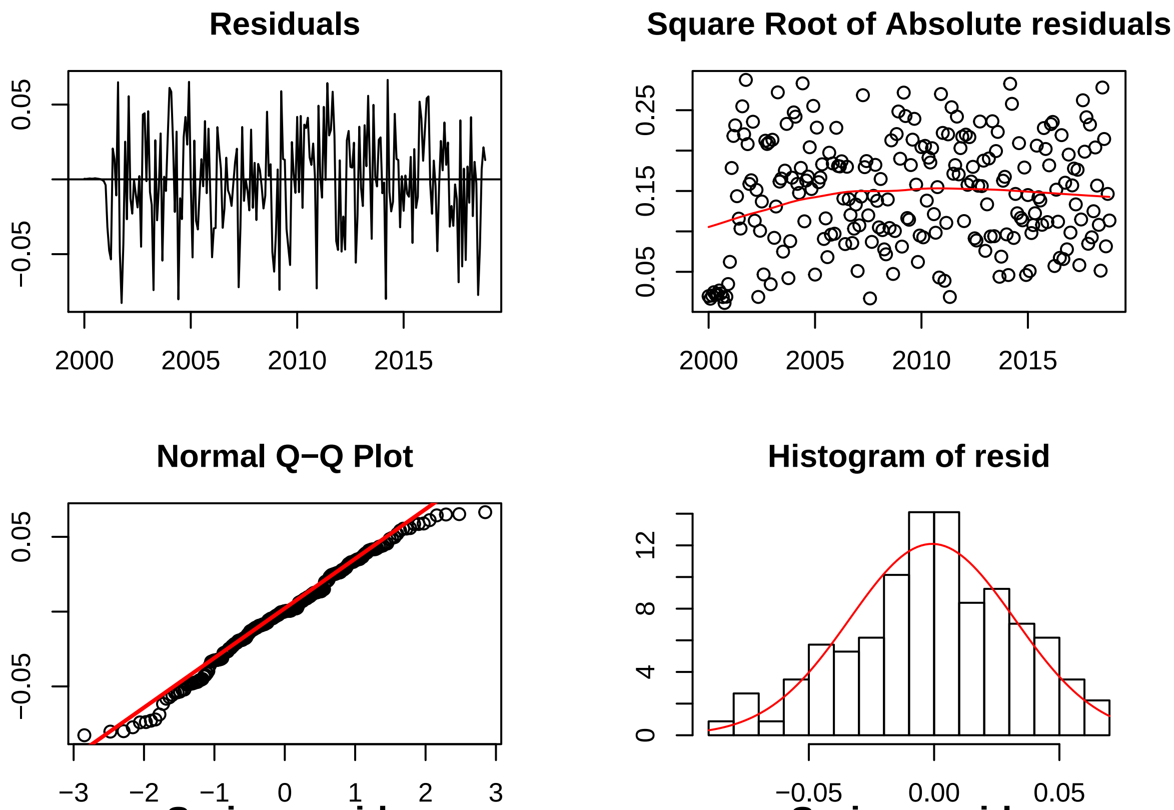
```
## [1] 0.001091865
```

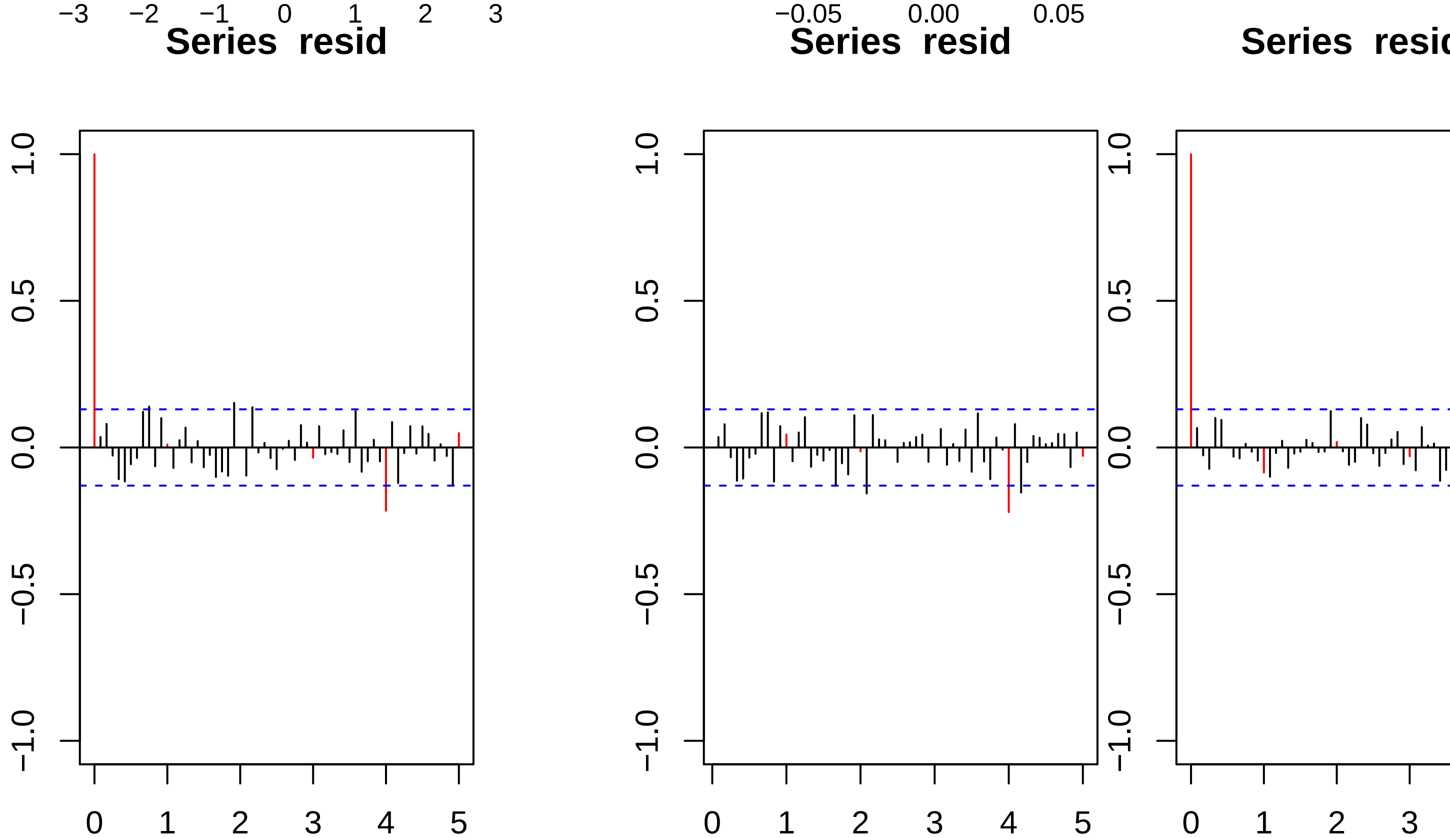

V1
2
4
6
8
10
2000
2005
2010
2015
Time

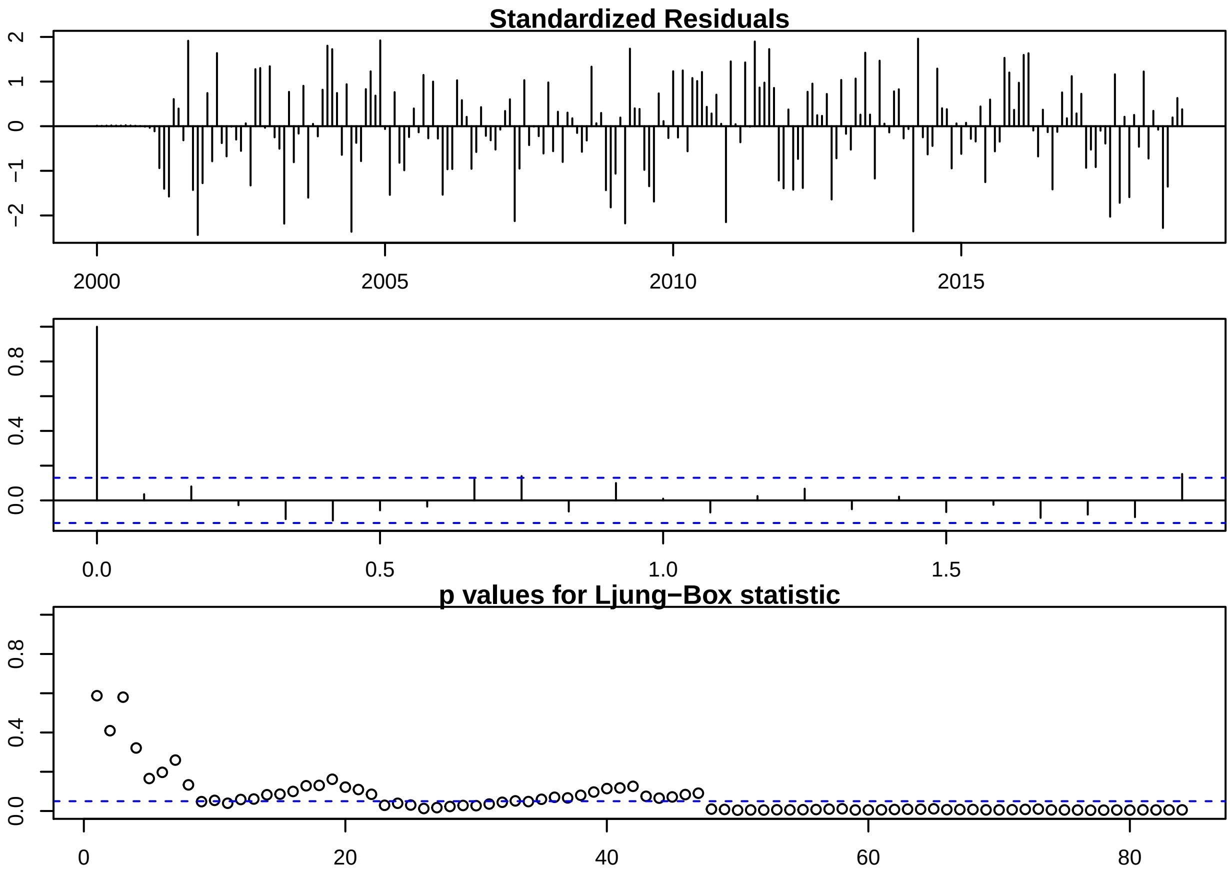

lnserie
−0.15
−0.05
0.05
2000
2005
2010
2015
Time




```
##
## Call:
## arima(x = lnserie.lin, order = pdq.1, seasonal = list(order = PDQ.1, period = 12))
##
## Coefficients:
##           ma1     sar1
##       -0.7267  -0.2585
## s.e.   0.0541   0.0696
##
## sigma^2 estimated as 0.00115:  log likelihood = 419.75,  aic = -833.5
```

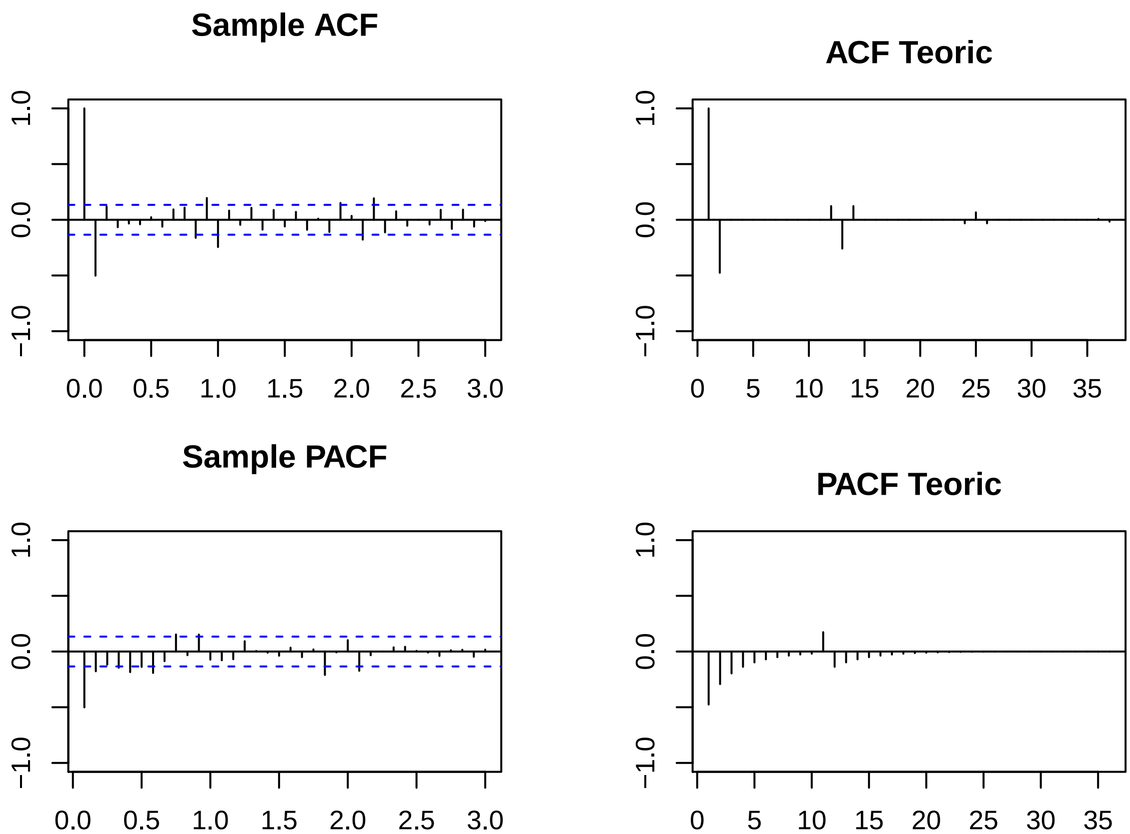

Residuals
Square Root of Absolute residuals
Normal Q−Q Plot
Histogram of resid


Series resid
Series resid




```
##
## ---------------------------------------------------------------------
##
## Call:
## arima(x = lnserie.lin, order = pdq.1, seasonal = list(order = PDQ.1, period = 12))
##
## Coefficients:
##           ma1     sar1
##       -0.7267  -0.2585
## s.e.   0.0541   0.0696
##
## sigma^2 estimated as 0.00115:  log likelihood = 419.75,  aic = -833.5
##
## Modul of AR Characteristic polynomial Roots:  1.119328 1.119328 1.119328 1.119328 1.119328 1.119328
##
## Modul of MA Characteristic polynomial Roots:  1.375992
##
## Psi-weights (MA(inf))
##
## --------------------
##      psi 1      psi 2      psi 3      psi 4      psi 5      psi 6
## -0.7267484  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##      psi 7      psi 8      psi 9     psi 10     psi 11     psi 12
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000 -0.2585311
##     psi 13     psi 14     psi 15     psi 16     psi 17     psi 18
##  0.1878871  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
```

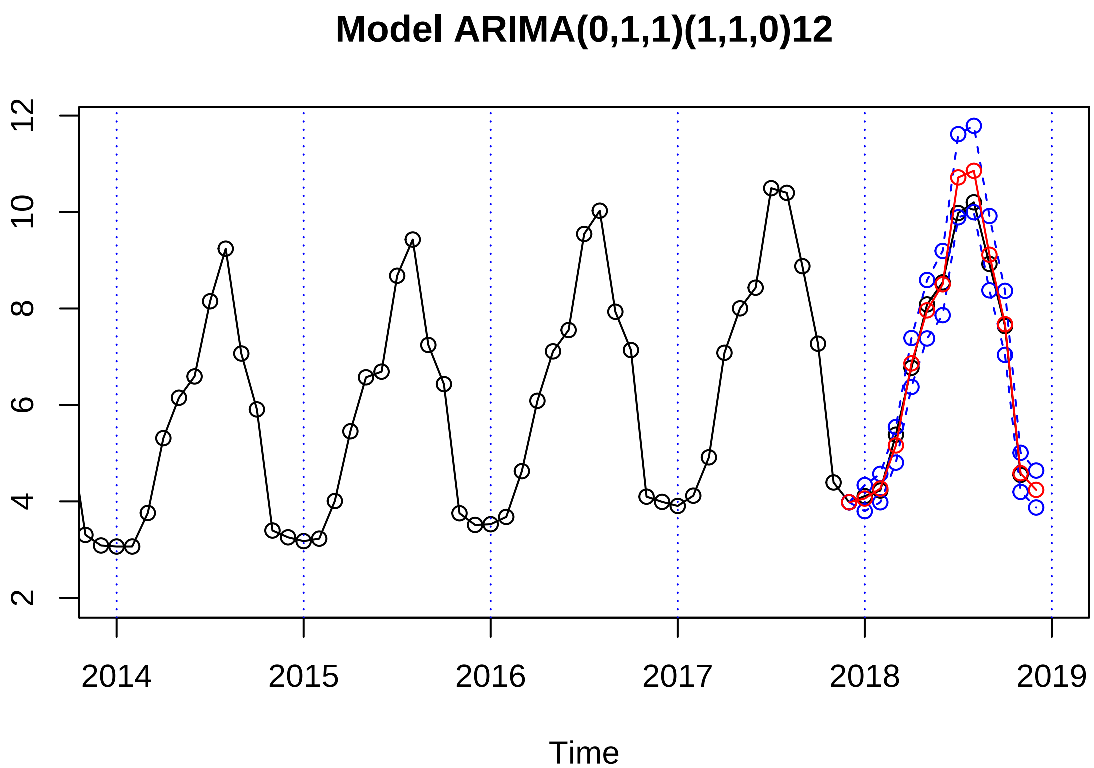
```
##     psi 19     psi 20
##  0.0000000  0.0000000
##
## Pi-weights (AR(inf))
##
## --------------------
##        pi 1        pi 2        pi 3        pi 4        pi 5        pi 6
## -0.72674837 -0.52816320 -0.38384175 -0.27895636 -0.20273108 -0.14733449
##        pi 7        pi 8        pi 9       pi 10       pi 11       pi 12
## -0.10707510 -0.07781665 -0.05655313 -0.04109989 -0.02986928 -0.28023858
##       pi 13       pi 14       pi 15       pi 16       pi 17       pi 18
## -0.20366293 -0.14801171 -0.10756727 -0.07817434 -0.05681307 -0.04128881
##       pi 19       pi 20
## -0.03000657 -0.02180723
##
##  Shapiro-Wilk normality test
##
## data:  resid(model)
## W = 0.9868, p-value = 0.03398
```




```
## Warning in window.default(x, ...): 'end' value not changed

##
## Call:
## arima(x = lnserie1.lin, order = c(0, 1, 1), seasonal = list(order = c(2, 1,
##     0), period = 12))
##
## Coefficients:
```

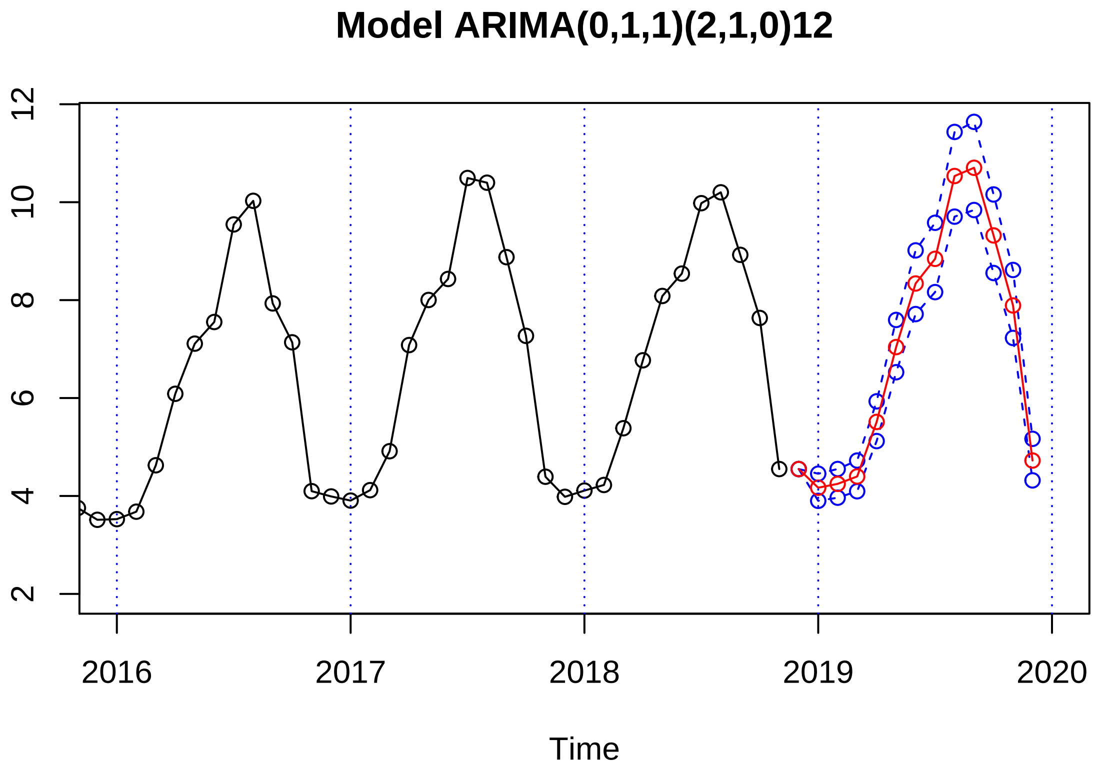
```
##           ma1     sar1    sar2
##       -0.7266  -0.2575  0.0046
## s.e.   0.0542   0.0715  0.0735
##
## sigma^2 estimated as 0.00115:  log likelihood = 419.75,  aic = -831.51
```

```
##
## Call:
## arima(x = lnserie2.lin, order = c(0, 1, 1), seasonal = list(order = c(2, 1,
##     0), period = 12))
##
## Coefficients:
##           ma1     sar1     sar2
##       -0.7245  -0.2656  -0.0158
## s.e.   0.0578   0.0738   0.0756
##
## sigma^2 estimated as 0.001155:  log likelihood = 397.68,  aic = -787.37
```




```
##                 tl        pr        tu     serie  error
## Dec 2017 3.982530  3.982530  3.982530  3.982530  0.000
## Jan 2018 3.797510  4.059052  4.338607  4.110137  0.051
## Feb 2018 3.981625  4.266419  4.571583  4.224826 -0.042
## Mar 2018 4.804731  5.160742  5.543133  5.383687  0.223
## Apr 2018 6.373840  6.861996  7.387540  6.770845 -0.091
## May 2018 7.378594  7.961567  8.590600  8.084173  0.123
## Jun 2018 7.861871  8.501558  9.193292  8.541181  0.040
## Jul 2018 9.890334 10.717794 11.614483  9.979779 -0.738
## Aug 2018 9.996226 10.854971 11.787489 10.201456 -0.654
## Sep 2018 8.376610  9.114589  9.917583  8.924326 -0.190
## Oct 2018 7.038509  7.673706  8.366227  7.635569 -0.038
## Nov 2018 4.197010  4.584596  5.007975  4.549899 -0.035
```

```
## Dec 2018 3.873660  4.239371  4.639608        NA      NA
## [1] 0.03228679
## [1] 0.02240873
```




```
##                tl2       pr2       tu2
## Dec 2018 4.549899  4.549899  4.549899
## Jan 2019 3.900656  4.168694  4.455150
## Feb 2019 3.964901  4.247701  4.550672
## Mar 2019 4.095893  4.398384  4.723215
## Apr 2019 5.121097  5.511856  5.932431
## May 2019 6.527095  7.040689  7.594697
## Jun 2019 7.714422  8.339344  9.014888
## Jul 2019 8.163448  8.843216  9.579588
## Aug 2019 9.704386 10.533911 11.434344
## Sep 2019 9.839793 10.702151 11.640086
## Oct 2019 8.553485  9.321199 10.157818
## Nov 2019 7.227038  7.890670  8.615242
## Dec 2019 4.318102  4.723392  5.166721
```

```
##          previs1.tl1 previs1.pr1 previs1.tu1 previs2.tl2 previs2.pr2
## Dec 2018    4.549899    4.549899    4.549899    4.549899    4.549899
## Jan 2019    3.794995    4.166268    4.573864    3.900656    4.168694
## Feb 2019    3.827086    4.219417    4.651967    3.964901    4.247701
## Mar 2019    3.953745    4.376872    4.845281    4.095893    4.398384
## Apr 2019    4.902111    5.448048    6.054784    5.121097    5.511856
## May 2019    6.447088    7.192220    8.023472    6.527095    7.040689
## Jun 2019    7.520711    8.420638    9.428251    7.714422    8.339344
## Jul 2019    7.909996    8.887904    9.986711    8.163448    8.843216
## Aug 2019    9.420050   10.621064   11.975203    9.704386   10.533911
## Sep 2019    9.491309   10.737202   12.146640    9.839793   10.702151
```

```
## Oct 2019    8.204643    9.311865   10.568507    8.553485    9.321199
## Nov 2019    6.889348    7.843893    8.930695    7.227038    7.890670
## Dec 2019    4.112863    4.697227    5.364619    4.318102    4.723392
##          previs2.tu2
## Dec 2018    4.549899
## Jan 2019    4.455150
## Feb 2019    4.550672
## Mar 2019    4.723215
## Apr 2019    5.932431
## May 2019    7.594697
## Jun 2019    9.014888
## Jul 2019    9.579588
## Aug 2019   11.434344
## Sep 2019   11.640086
## Oct 2019   10.157818
## Nov 2019    8.615242
## Dec 2019    5.166721
```