Time Series Assignment 1

Iker Caballero Bragagnini

2023-04-07

PREGUNTAR LO DE QUE QUIERE DECIR ESTABILIDAD (SI ES ESTACIONARIEDAD O NO) PRE-GUNTAR SI DE LARGO ESTA BIEN PREGUNTAR CUAL ES LA DEFINICION DE FORECASTS COR-RECTOS ##

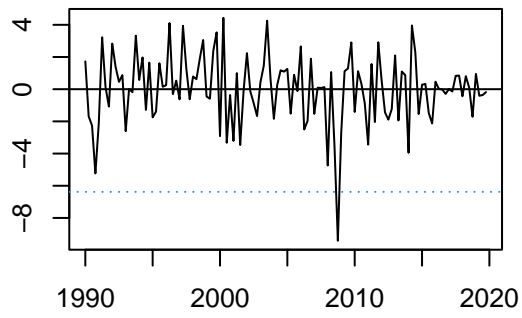## GDPUSAvar: Gross Domestic Product variation in USA

```
serie=ts((read.table("gdpUSAvar.dat",header=F)),start=1990,freq=4)
```

1) If we propose an ARMA(2,0) and an ARMA(1,1), the second is a better model Answer: *FALSE*
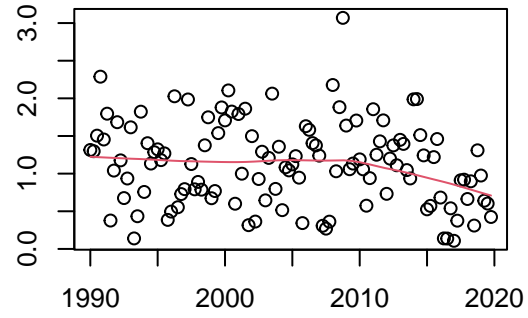
To assess whether a model is better than other, we need to carry on the validation of the two models proposed and the computation of model selection measures based on other criteria (explained after). This is because in case one of the models is not valid, then we can be sure that it will not be better than a valid model. We first start with the AR(2) model.

```
model_ar2=arima(serie,order=c(2,0,0))
validation(model_ar2,serie)
```
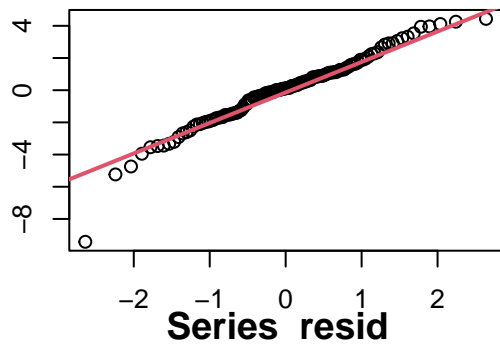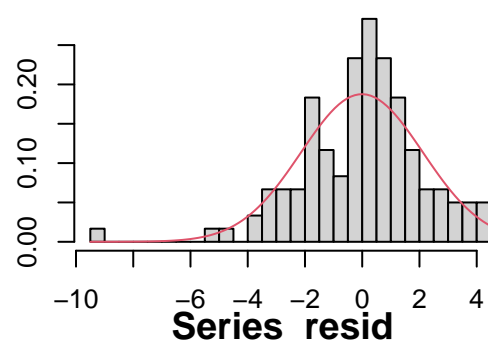
Residuals

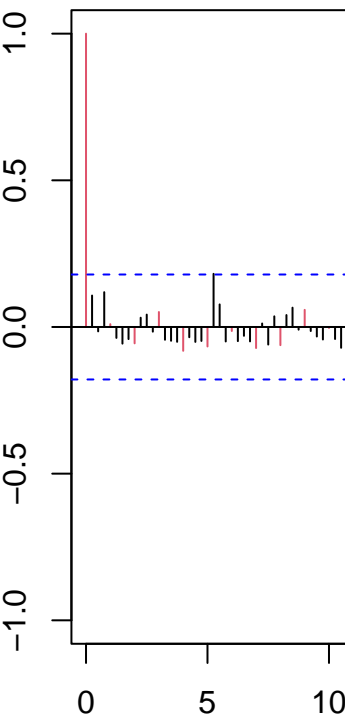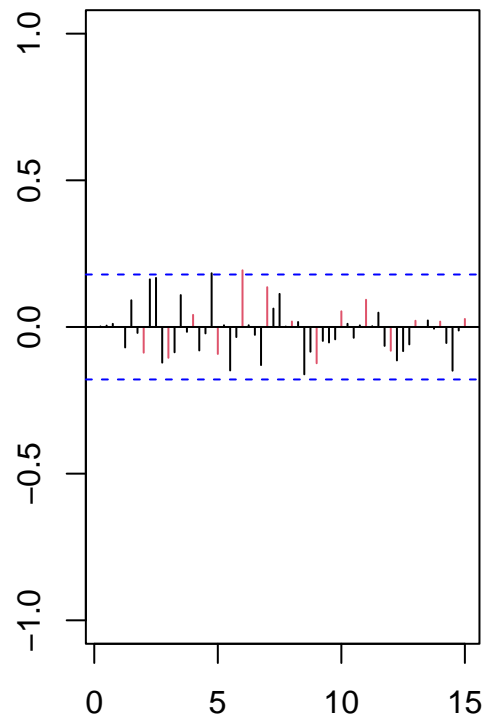# Square Root of Absolute residuals

Normal Q–Q Plot

Series resid

Histogram of resid

Series resid

Series resid

Standardized Residuals

p values for Ljung-Box statistic

```
## 
## -------------------------------------------------------------------------
## 
## Call:
## arima(x = serie, order = c(2, 0, 0))
## 
## Coefficients:
##           ar1     ar2  intercept
##        0.3048  0.1942     2.4980
## s.e.   0.0893  0.0892     0.3814
## 
## sigma^2 estimated as 4.484:  log likelihood = -260.41,  aic = 528.83
## 
## Modul of AR Characteristic polynomial Roots:  1.616438 3.185849
## 
## Modul of MA Characteristic polynomial Roots:
## 
## Psi-weights (MA(inf))
## 
## --------------------
##        psi 1         psi 2         psi 3         psi 4         psi 5         psi 6
## 3.047561e-01 2.870613e-01 1.466627e-01 1.004394e-01 5.908922e-02 3.751162e-02
##        psi 7         psi 8         psi 9        psi 10        psi 11        psi 12
## 2.290613e-02 1.426498e-02 8.795367e-03 5.450487e-03 3.368997e-03 2.085125e-03
##       psi 13        psi 14        psi 15        psi 16        psi 17        psi 18
## 1.289663e-03 7.979329e-04 4.936082e-04 3.053767e-04 1.889167e-04 1.168731e-04
```

```
##      psi 19       psi 20
## 7.230259e-05 4.472966e-05
##
## Pi-weights (AR(inf))
##
## --------------------
##      pi 1       pi 2       pi 3       pi 4       pi 5       pi 6       pi 7       pi 8
## 0.3047561 0.1941850 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
##      pi 9      pi 10      pi 11      pi 12      pi 13      pi 14      pi 15      pi 16
## 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
##     pi 17      pi 18      pi 19      pi 20
## 0.0000000 0.0000000 0.0000000 0.0000000
##
##  Shapiro-Wilk normality test
##
## data:  resid(model)
## W = 0.96332, p-value = 0.002358
```


Sample ACF
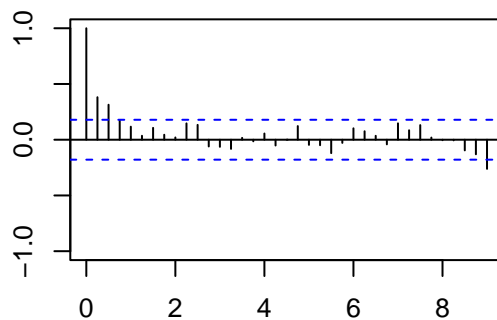
ACF Teoric
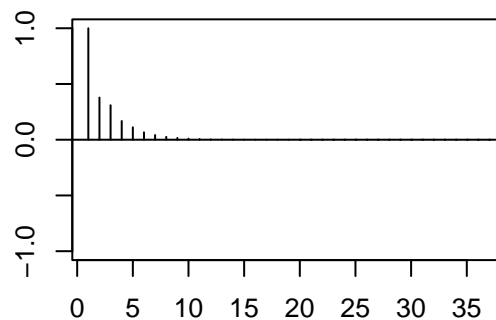
Sample PACF
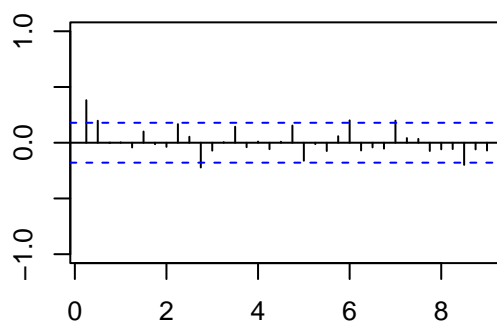
PACF Teoric
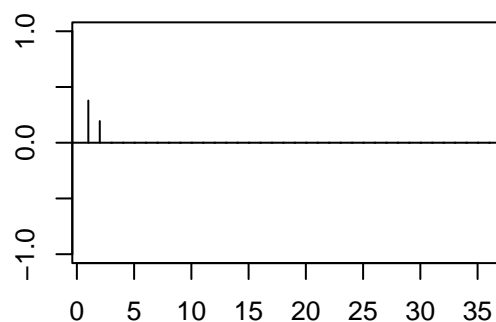
We can see that results show how the residuals behave like a white noise and that homoskedasticity (checked through squared residuals), independence of lags (checked through the Ljung-Box statistic) do hold, but normality (checked through normality tests and histograms) does not. Now, we can pass to the ARMA(1,1) model:

```
model_arma=arima(serie,order=c(1,0,1))
validation(model_arma,serie)
```

Residuals
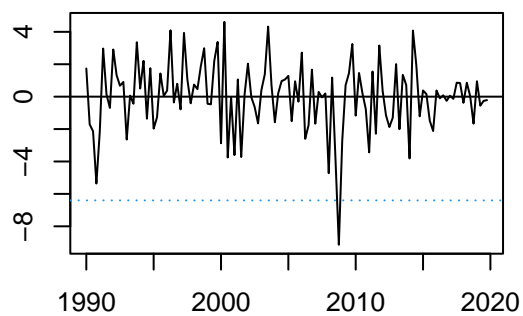
Square Root of Absolute residuals
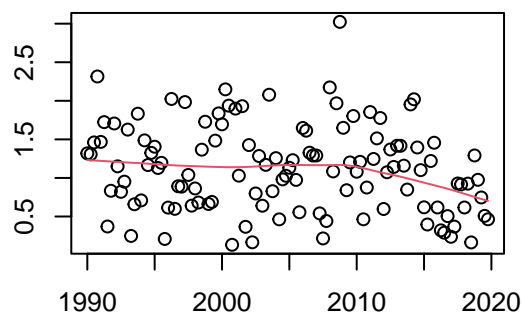
Normal Q–Q Plot

Histogram of resid

Series resid

Series resid

Series resid

5

Standardized Residuals

p values for Ljung–Box statistic
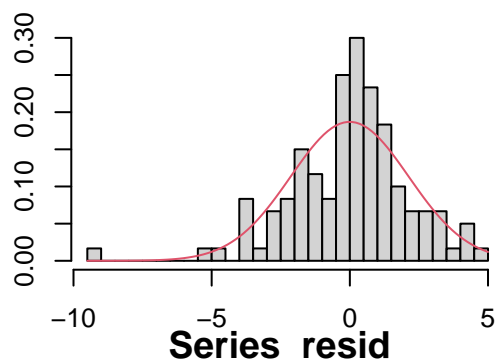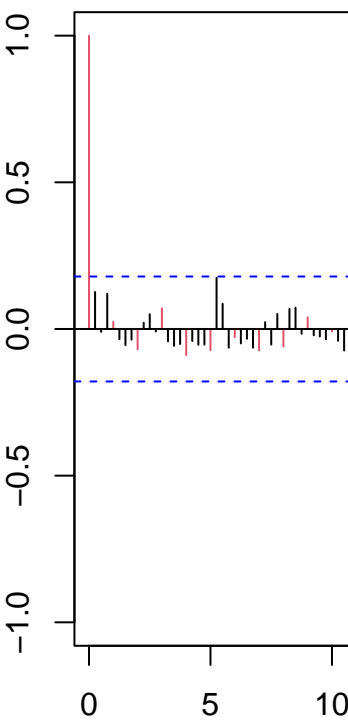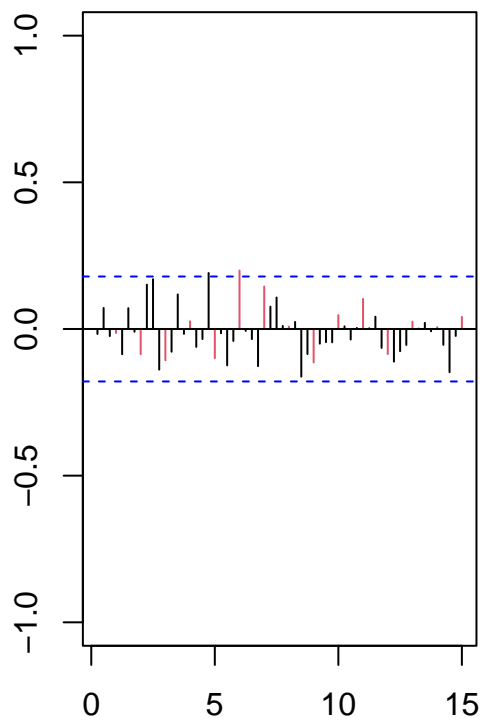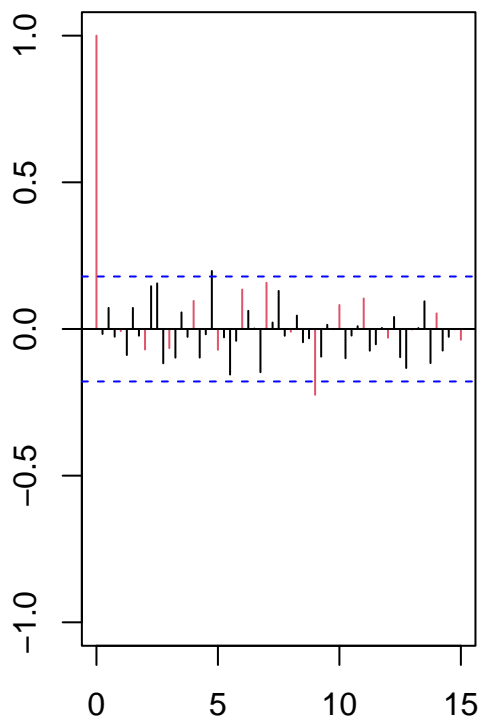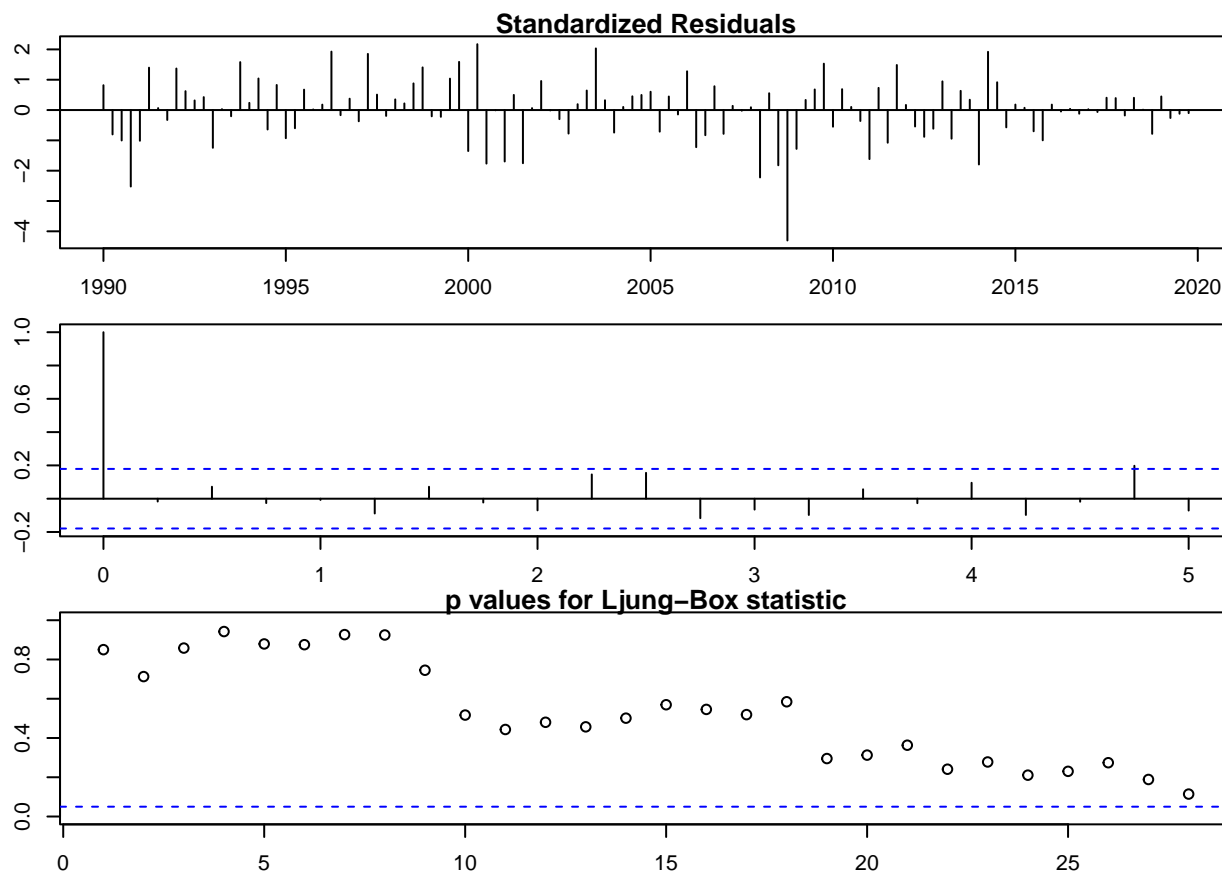
```
## 
## ---------------------------------------------------------------------
## 
## Call:
## arima(x = serie, order = c(1, 0, 1))
## 
## Coefficients:
##          ar1      ma1  intercept
##       0.7008  -0.3789     2.4917
## s.e.  0.1357   0.1747     0.3971
## 
## sigma^2 estimated as 4.514:  log likelihood = -260.82,  aic = 529.63
## 
## Modul of AR Characteristic polynomial Roots:  1.426985
## 
## Modul of MA Characteristic polynomial Roots:  2.63955
## 
## Psi-weights (MA(inf))
## 
## --------------------
##        psi 1         psi 2         psi 3         psi 4         psi 5         psi 6
## 0.3219256949 0.2255984736 0.1580944674 0.1107891389 0.0776386012 0.0544074306
##        psi 7         psi 8         psi 9        psi 10        psi 11        psi 12
## 0.0381275353 0.0267189414 0.0187240487 0.0131214031 0.0091951918 0.0064437889
##       psi 13        psi 14        psi 15        psi 16        psi 17        psi 18
## 0.0045156661 0.0031644799 0.0022175982 0.0015540442 0.0010890401 0.0007631755
```
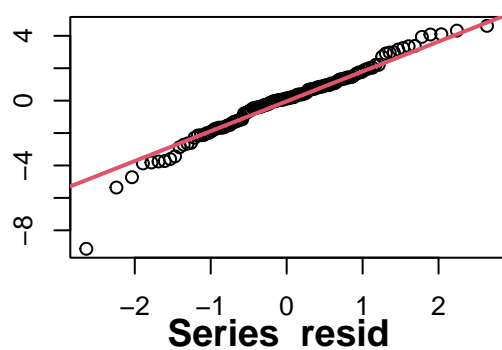
```
##        psi 19       psi 20
## 0.0005348166 0.0003747878
##
## Pi-weights (AR(inf))
##
## --------------------
##          pi 1         pi 2         pi 3         pi 4         pi 5         pi 6
## 3.219257e-01 1.219623e-01 4.620572e-02 1.750515e-02 6.631867e-03 2.512499e-03
##          pi 7         pi 8         pi 9        pi 10        pi 11        pi 12
## 9.518661e-04 3.606168e-04 1.366205e-04 5.175901e-05 1.960903e-05 7.428926e-06
##         pi 13        pi 14        pi 15        pi 16        pi 17        pi 18
## 2.814467e-06 1.066267e-06 4.039580e-07 1.530404e-07 5.797974e-08 2.196576e-08
##         pi 19        pi 20
## 8.321782e-09 3.152727e-09
##
##  Shapiro-Wilk normality test
##
## data:  resid(model)
## W = 0.96518, p-value = 0.003377
```


Sample ACF   **ACF Teoric**
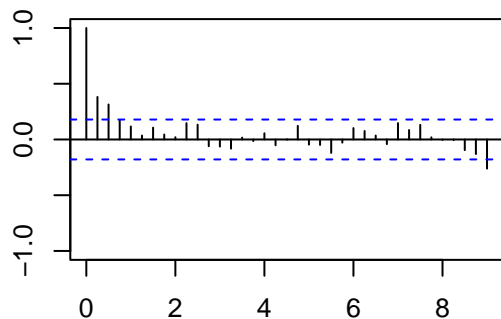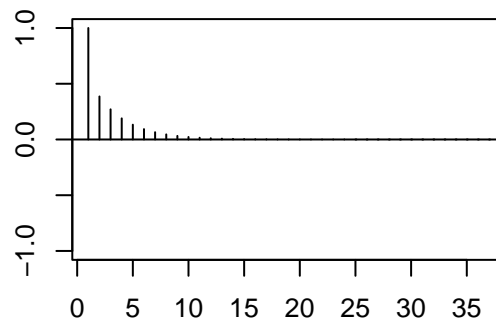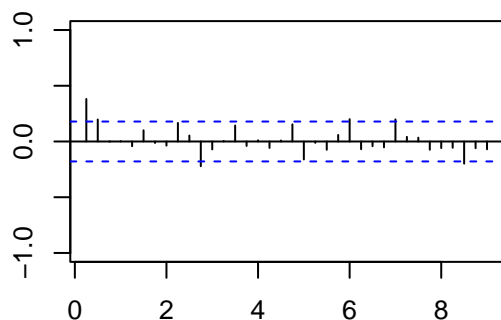
Sample PACF   **PACF Teoric**
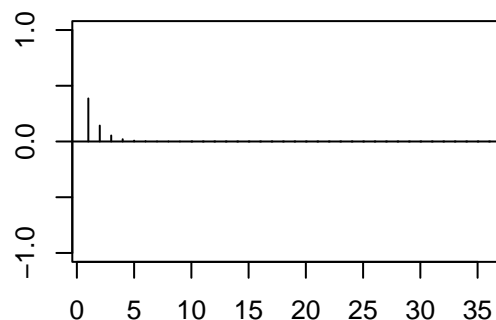
The results from the validation analysis are merely the same as with the previous model. All in all, both models seem to be valid, showing homoskedasticity (no volatiliy) and independence of lags, but having some normality problems. These latter problems might be due to a possible outlier, which can be detected at the left of the distribution, so we can think about a series that does not contain that observation. We can also check how all the coefficients are significant at the 5% significance level, which indicates no parameter should be eliminated.

```
abs(model_ar2$coef/sqrt(diag(model_ar2$var.coef)))
```

```
##       ar1       ar2 intercept
## 3.411880  2.176965  6.549268
```

```
abs(model_arma$coef/sqrt(diag(model_arma$var.coef)))
```

```
##       ar1       ma1 intercept
## 5.163750  2.168362  6.275043
```

We can now defend which model is best by using different measures related to goodness of fit and forecasting. To compute forecasting measures, we first need to compute forecasts and confidence intervals, so we compute the forecasts for the last 12 observations (we do not show them for the sake of brevity, but are available at the appendix).

```
obs=window(serie,start=ultim)
resul=data.frame(
  Sigma2Z=c(model_ar2$sigma2, model_arma$sigma2),
  AIC=c(AIC(model_ar2), AIC(model_arma)),
  BIC=c(BIC(model_ar2), BIC(model_arma)),
  RMSE=c(fRMSE(obs,prar2),fRMSE(obs,prarma)),
  MAE=c(fMAE(obs,prar2),fMAE(obs,prarma)),
  RMSPE=c(fRMSPE(obs,prar2),fRMSPE(obs,prarma)),
  MAPE=c(fMAPE(obs,prar2),fMAPE(obs,prarma)),
  CI_Length=c(fCI(tuar2,tlar2),fCI(tuarma,tlarma)))
row.names(resul)=c("AR(2)","ARMA(1,1)")
resul
```

```
##             Sigma2Z      AIC      BIC      RMSE       MAE     RMSPE      MAPE
## AR(2)      4.483507 528.8251 539.9751 0.6973069 0.5709096 0.4115847 0.2667970
## ARMA(1,1) 4.514371 529.6305 540.7804 0.7021945 0.5763017 0.4088805 0.2673888
##           CI_Length
## AR(2)      8.318700
## ARMA(1,1)  8.314336
```

Using these metrics, we can see how that the balance between goodness of fit and the complexity (given by the AIC and the BIC) seem to favour the AR(2) model, while some metrics for forecasting such as RMSPE and the mean length of the CI would indicate the ARMA(1,1) model is better, as the precision and the confidence of the forecasts would be slightly higher. However, looking at the measures jointly, we can see that the AR(2) seems to have a better fit and more or less the same forecasting precision and accuracy as the ARMA(1,1), so we can say that the ARMA(1,1) is not a better model than the AR(2).

We finally highlight that, even though we favour the AR(2) model, both are quite similar.

2) The fitted ARMA(1,1) model is stationary and invertible   Answer: $TRUE$

When assesing the stationarity and invertibility of the model, it is necessary to use the polinomical form of the model, which is the following:

$$(1 - 0.7008B)x_t = (1 - 0.3789B)w_t$$

To assess stationarity of the model, we can check the roots of the AR polynomial

$$1 - 0.7008B = 0 \rightarrow B = \frac{1}{0.7008}$$

$$\left| \frac{1}{0.7008} \right| > 1 \rightarrow Stationarity$$

Because of the absolute value of the root of the AR polynomial is greater than one, we can confirm that the ARMA(1,1) model is stationary.

Now, for the invertibility, we can check the roots of the MA polynomial.

$$1 - 0.3789B = 0 \rightarrow B = -\frac{1}{0.3789}$$

$$\left| -\frac{1}{0.3789} \right| > 1 \rightarrow Invertible$$

Because of the absolute value of the root of the MA polynomial is greater than one, we can confirm that the ARMA(1,1) model is stationary and invertible.

EURODOL: Monthly Average Dolar/Euro exchange

3) Although the mean is seemingly not constant, valid stationary-ARMA models can be fitted to the original series Answer: $TRUE$

```
serie=ts((read.table("eurodol.dat")),start=1995,freq=12)
```

In order to see whether this could be the case, we first do some exploratory analysis of the original series to see whether it needs modifications for stationarity and then we estimate different models to see if the possible models that are estimated can be valid and stationary.

Looking at the original series, one can observe that the variance seems to be constant and that there is no seasonal pattern to consider. However, the mean function seems to be more or less constant in the graph, but the ACF indicates that there might be non-stationarity (so the mean is not constant).

```
m=apply(matrix(serie,ncol=25),2,mean)
v=apply(matrix(serie,ncol=25),2,var)

par(mfrow=c(2,1))
plot(v~m,main="Mean-Variance plot")
plot(serie, ylab = "USD/EUR")
abline(h=mean(serie),col=2)
```

Mean−Variance plot
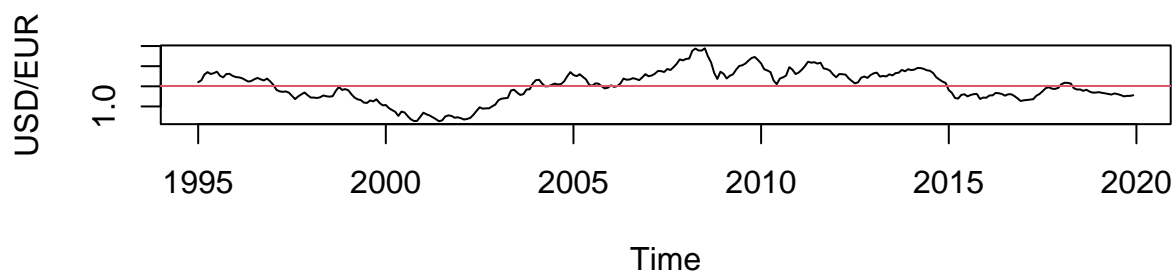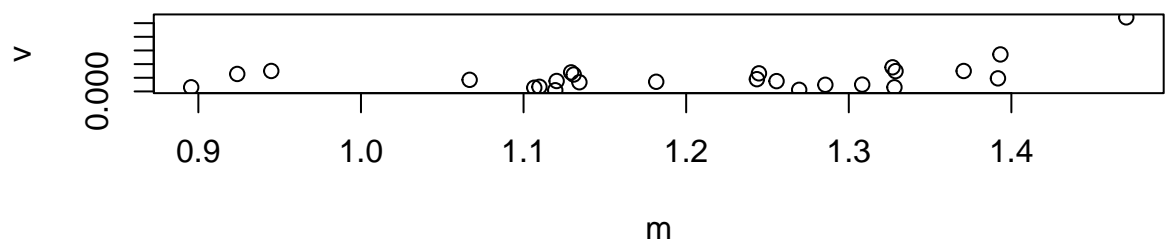

```
par(mfrow=c(1,1))
monthplot(serie)
```
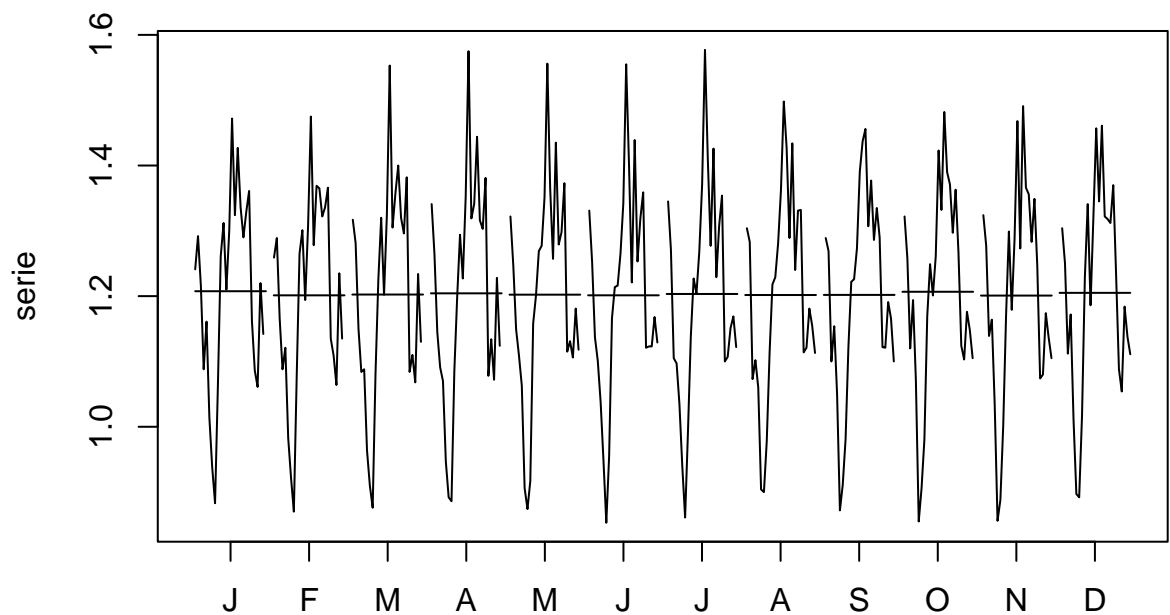


```
acf(serie,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,11)))
```

V1
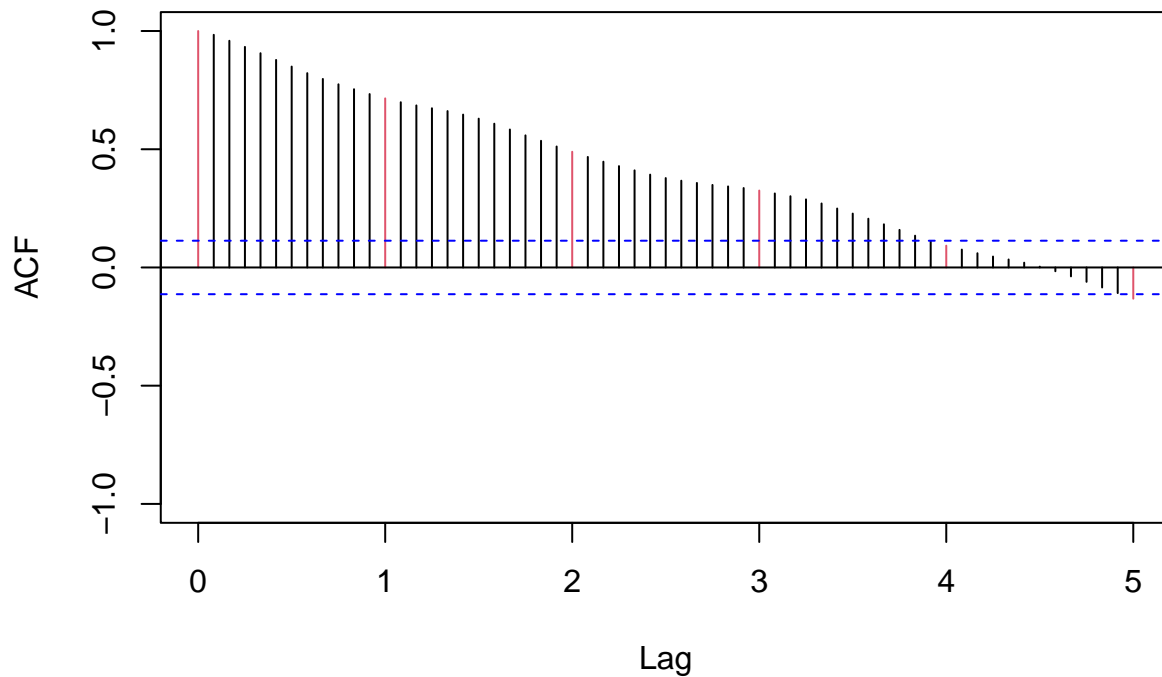
Hence, there are two cases: one in which the mean is assumed to be constant and stationary models can be used, and one in which the series is not stationary and one needs to use non-stationary models such as ARIMA models. In fact, the constant assumption seems not too adequate.

We just focus on the first case, so that we can analyze the validity of the models proposed (not considering ARIMA models nor ARMA models with differenciated series). From the ACF and the PACF, we can clearly identify an AR(2) model, which is stationary because all AR models are.

```
par(mfrow=c(1,2))
acf(serie,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,11)))
pacf(serie,ylim=c(-1,1),lag.max=60,col=c(rep(1,11),2))
```
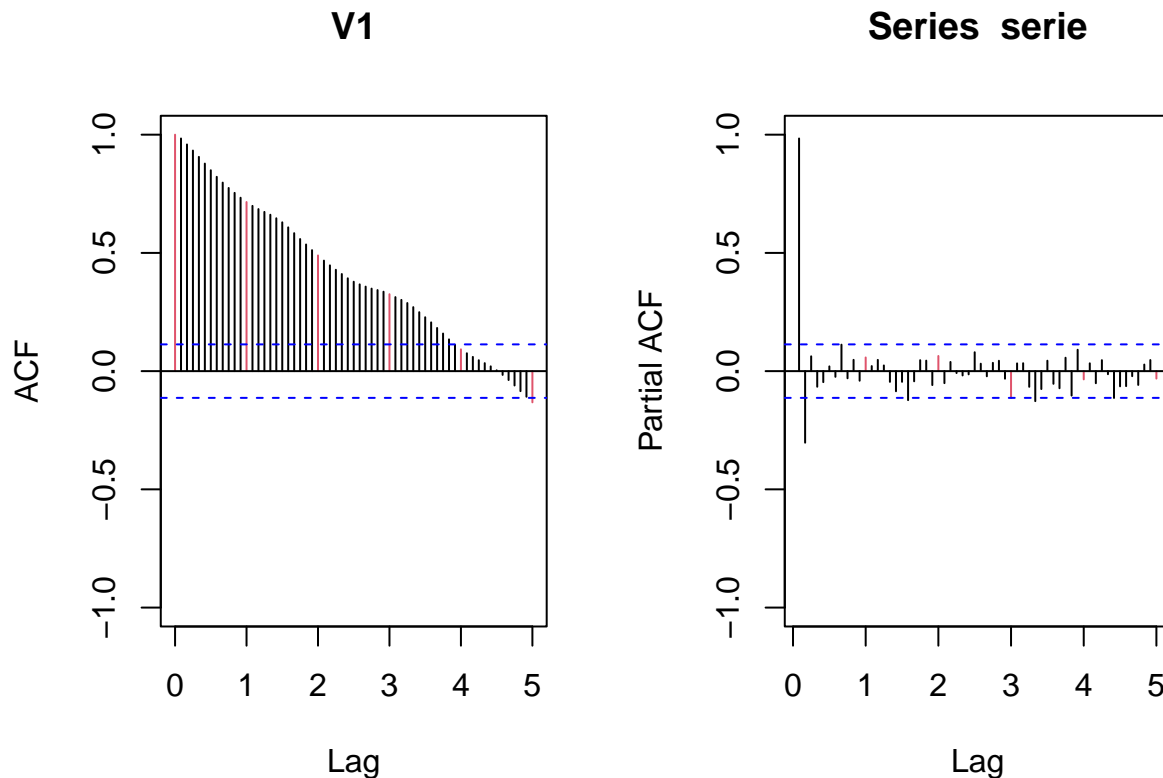
```
mod_ar2 <- arima(serie,order=c(2,0,0))
```

Finally, we check the validity of the model through the validation function. When checking the residuals, one can detect that the squared residuals could indicate the presence of volatility (it is a financial time series), but the general behaviour, the normality and independence of the residuals seem to indicate that the model could be validated. Moreover, comparing with the validation of other non-stationary models (not shown), one could see that the results are merely the same. Hence, we could validate this model, and we can say that we could obtain valid stationary-ARMA models assuming a constant mean, even though it is not (apparently).

PreuIndEner: Monthly Energy Price in Industry in Spain

```
serie=ts(read.table("preuIndEner.dat"),start=2000,freq=12)
```

4) This monthly series has a seasonal component and therefore, for the original series, we have to identify a regular and a seasonal part for the model Answer: *FALSE*

To check the presence of a seasonal component, we should use a monthplot. In this case, we can see that there is a slight increase in the median level starting in May and extending until October, where it begins to decrease.

```
serie=log(serie); monthplot(serie)
```
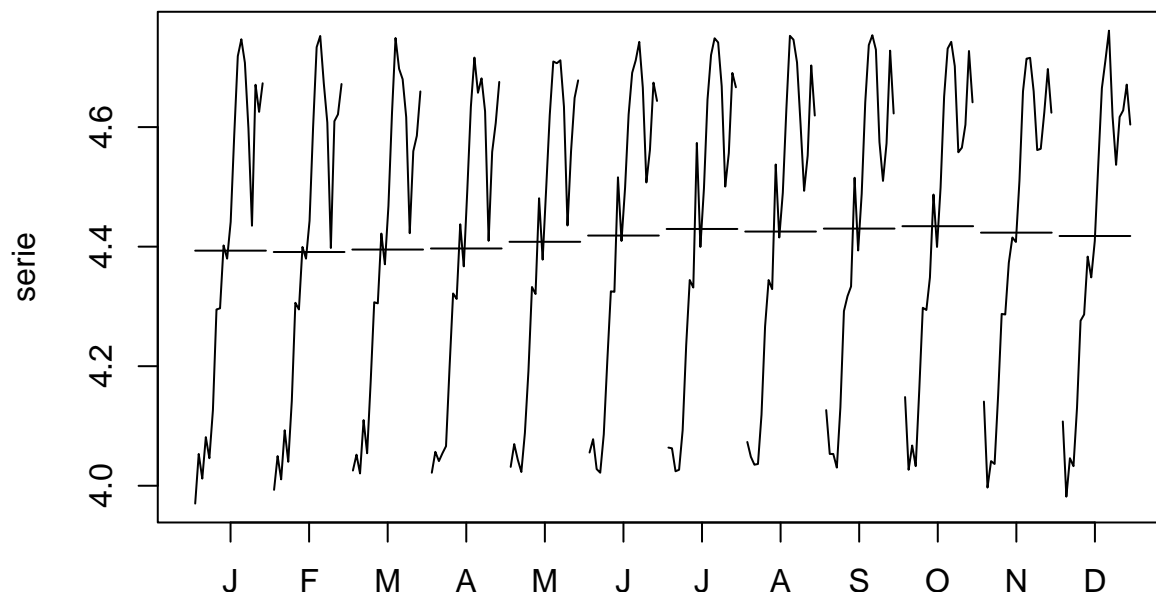
Nevertheless, the identification step allows as to see that there are no significant seasonal lags, so that we do not need to specify the seasonal part of the ARIMA model (even though it would be necessary if there were significant seasonal lags). We first detect that this series needs one differentiation to obtain stationarity (the ACF and the PACF to detect this are not shown), and then we can obtain the ACF and PACF of the stationary time series to check that this is in fact the situation: no seasonal lag except the very first one is significant.

```
d1serie=diff(serie);var(d1serie);var(diff(d1serie))
```

```
##              V1
## V1 0.0006293412
```

```
##              V1
## V1 0.0009185575
```

```
par(mfrow=c(1,2))
acf(d1serie,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,11)))
pacf(d1serie,ylim=c(-1,1),lag.max=60,col=c(rep(1,11),2))
```

V1	Series d1serie
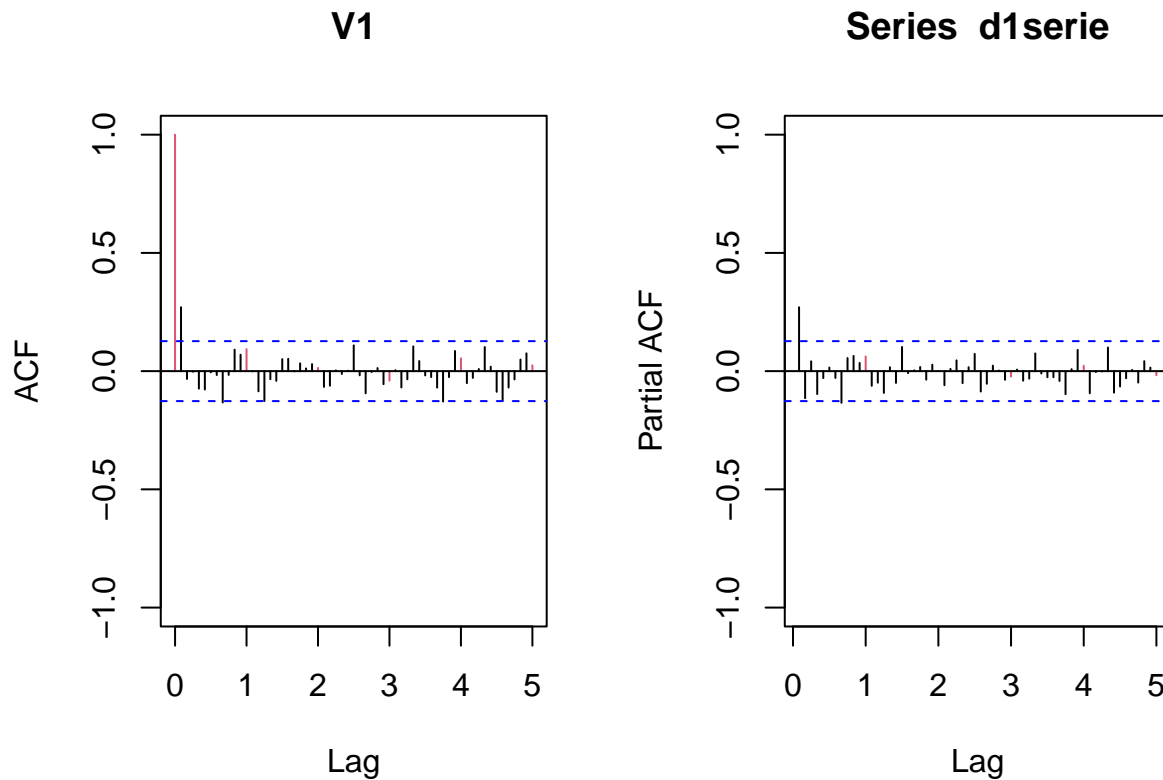
EURIBOR: Euro Interbank Offered Rate

```
serie=ts((read.table("euribor.dat")),start=1999,freq=12)
```

5) It seems that two regular differentiations are needed to transform the series into a stationary one   Answer: $TRUE$

To assess whether a second differentiation is needed, we transform using a logarithmic transformation (because variance is not constant) and, because there is no clear seasonal pattern, we can proceed checking the stationarity of the time series. One can observe that the series clearly needs a differentiation to obtain stationarity.

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

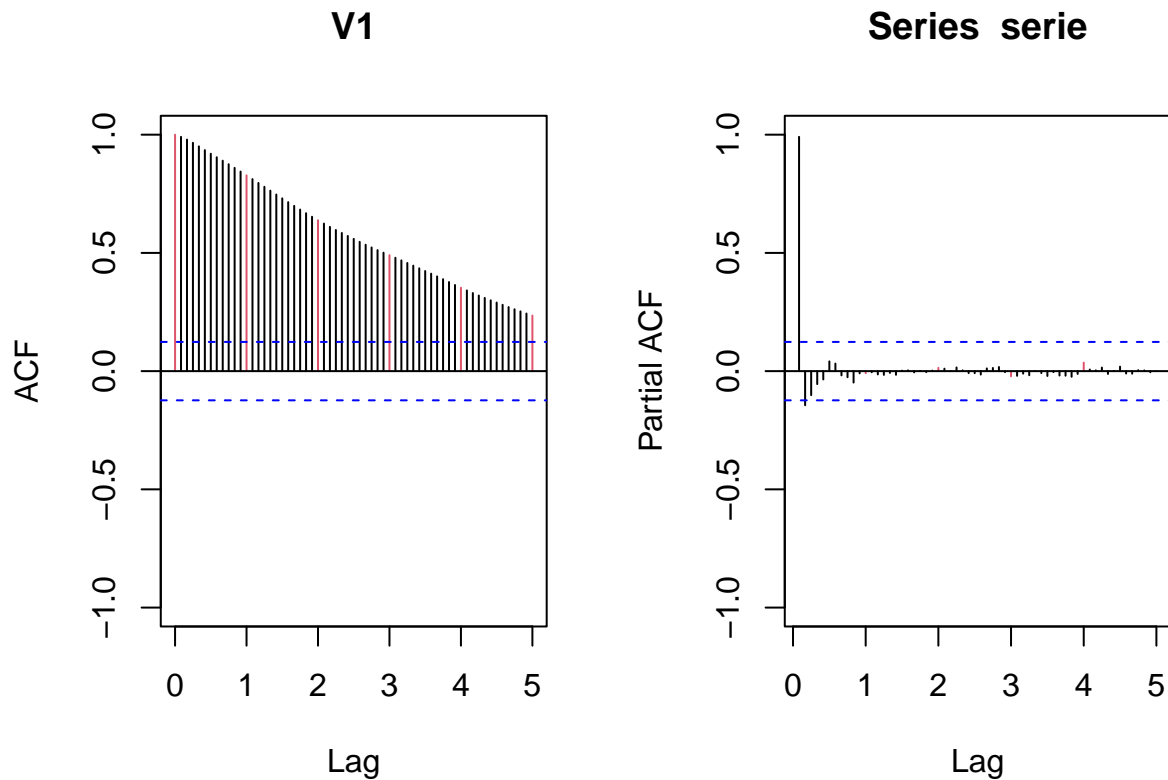**V1**   **Series serie**

Hence, to know how many differentiations should we use, we differentiate the series iteratively and check the variance at each step. If the variance increases, that means that the previous number of differences is enough. In this case, we can check that the variance decreases until the third difference, where it increases. Hence, we can see that a two differences is enough to transform the series to a stationary series.

```
var(serie)
```

```
##            V1
## V1 0.4601053
```

```
d1serie=diff(serie);var(d1serie)
```

```
##              V1
## V1 0.001564921
```

```
d1d1serie=diff(d1serie);var(d1d1serie)
```

```
##              V1
## V1 0.001006835
```

```
d1d1d1serie=diff(d1d1serie);var(d1d1d1serie)
```

```
##              V1
## V1 0.002246864
```

6) The once-differenced (d=1) log-transformed-series has a non-significant mean, and thus when fitting an AR(1) process, it is always better to drop this parameter (intercept) from the model   Answer: $TRUE$

To answer this question, we first need to estimate an AR(1) model and see whether the model is significant or not at certain significance levels. Estimating the model, we can see that the intercept is not significant at any common significance level, so this intercept could be eliminated from the model.

```
mod_ar1=arima(d1serie,order=c(1,0,0))
abs(mod_ar1$coef/sqrt(diag(mod_ar1$var.coef)))
```

```
##        ar1 intercept
## 14.698773  1.180459
```

Given that the expression for this model is $x_t = (1 - \phi)\mu + \phi x_{t-1} + w_t$ and the ϕ parameter would be significant in this case, the non-significance of this model comes from the mean, and we could obtain a model which does not include this parameter, such as an ARIMA(1,1,0) for the original time series. Hence, it is recommended not to include the intercept for an AR(1) model (it does not give any additional information as it is just a location shift)

7) When fitting an ARIMA(0, 2, 5) to the original series, the residuals are reasonably correct and thus we validate the model   Answer: $FALSE$

In order to determine whether the behaviour of residuals is correct, we can check them through the validation function.

```
mod_arim25=arima(serie,order=c(0,2,5))
validation(mod_arim25,d1d1serie)
```


16

Series resid   **Series resid**   **Series resid**

Standardized Residuals

p values for Ljung−Box statistic
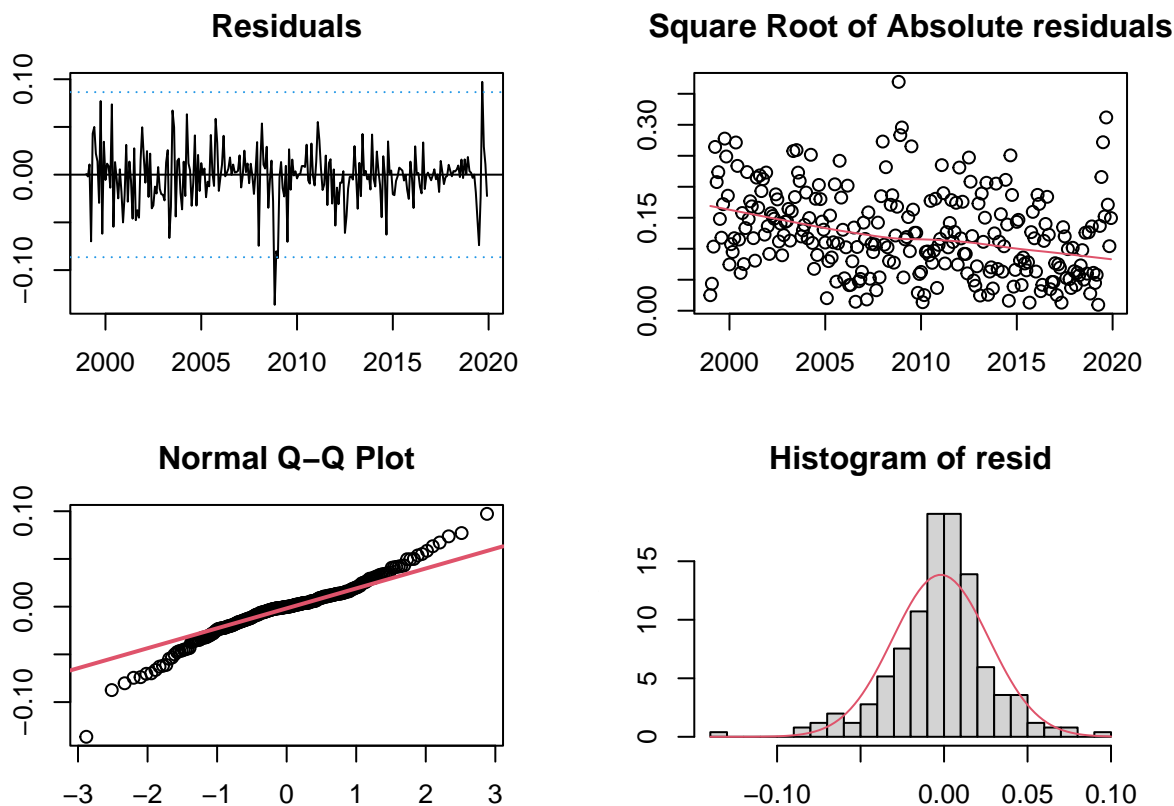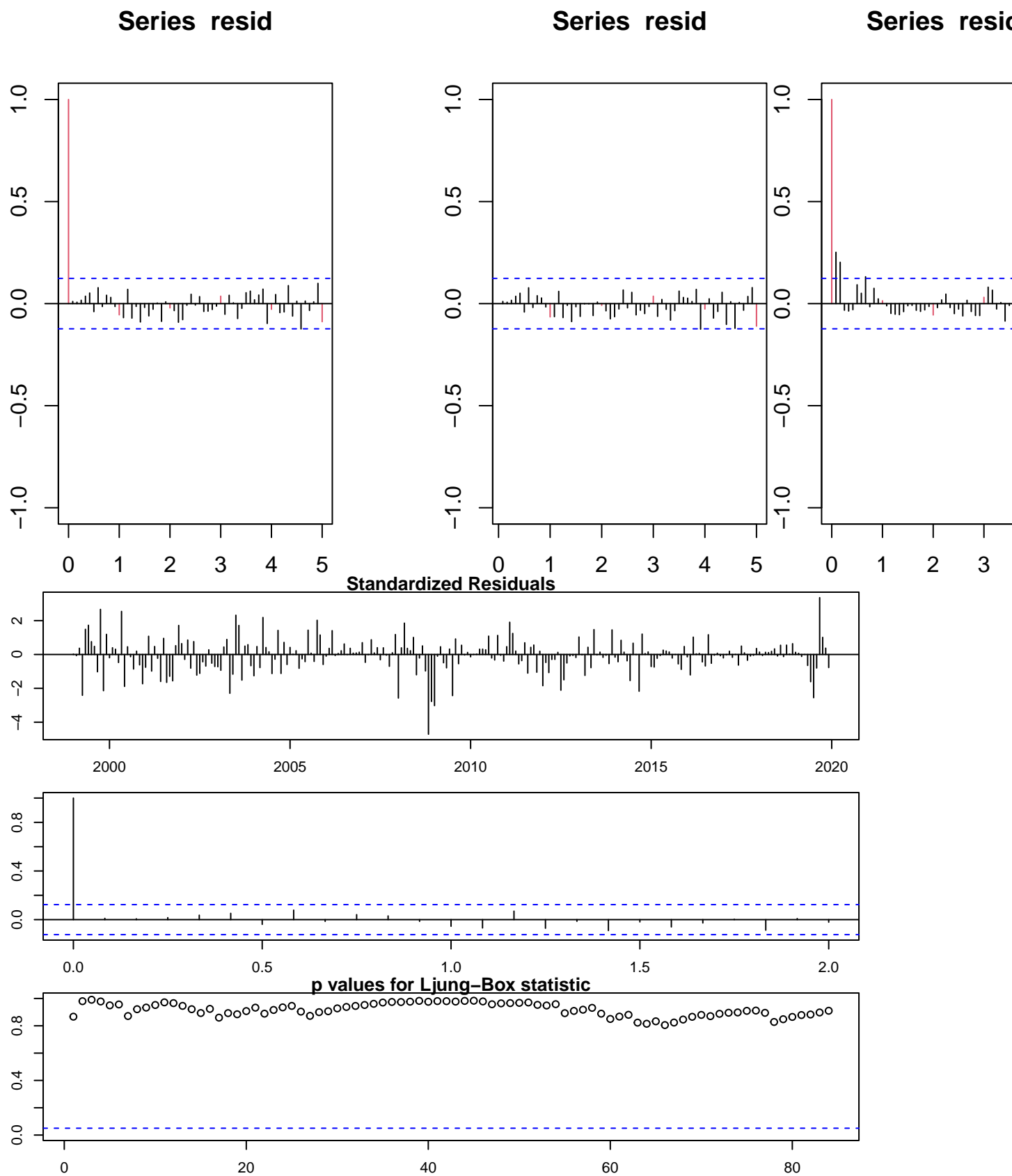
```
##
## -----------------------------------------------------------------------
##
## Call:
## arima(x = serie, order = c(0, 2, 5))
##
## Coefficients:
##           ma1      ma2      ma3      ma4      ma5
##       -0.2914  -0.2959  -0.1198  -0.1168  -0.1760
## s.e.   0.0750   0.0722   0.0778   0.0744   0.0664
##
## sigma^2 estimated as 0.0008378:  log likelihood = 528.72,  aic = -1045.45
##
## Modul of AR Characteristic polynomial Roots:
##
## Modul of MA Characteristic polynomial Roots:  1.000033 1.538402 1.538402 1.549302 1.549302
##
## Psi-weights (MA(inf))
##
## --------------------
##      psi 1       psi 2       psi 3       psi 4       psi 5       psi 6       psi 7
## -0.2914161 -0.2958725 -0.1197560 -0.1168451 -0.1760246  0.0000000  0.0000000
##      psi 8       psi 9      psi 10      psi 11      psi 12      psi 13      psi 14
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##     psi 15      psi 16      psi 17      psi 18      psi 19      psi 20
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##
## Pi-weights (AR(inf))
##
## --------------------
##       pi 1        pi 2        pi 3        pi 4        pi 5        pi 6        pi 7
## -0.2914161 -0.3807959 -0.3169481 -0.3567747 -0.4534238 -0.3714418 -0.3891889
##       pi 8        pi 9       pi 10       pi 11       pi 12       pi 13       pi 14
## -0.3750936 -0.3847224 -0.3929168 -0.3841084 -0.3865961 -0.3843402 -0.3860164
##      pi 15       pi 16       pi 17       pi 18       pi 19       pi 20
## -0.3865485 -0.3856695 -0.3859458 -0.3856288 -0.3858701 -0.3858706
##
##  Shapiro-Wilk normality test
##
## data:  resid(model)
## W = 0.9605, p-value = 2.121e-06
```
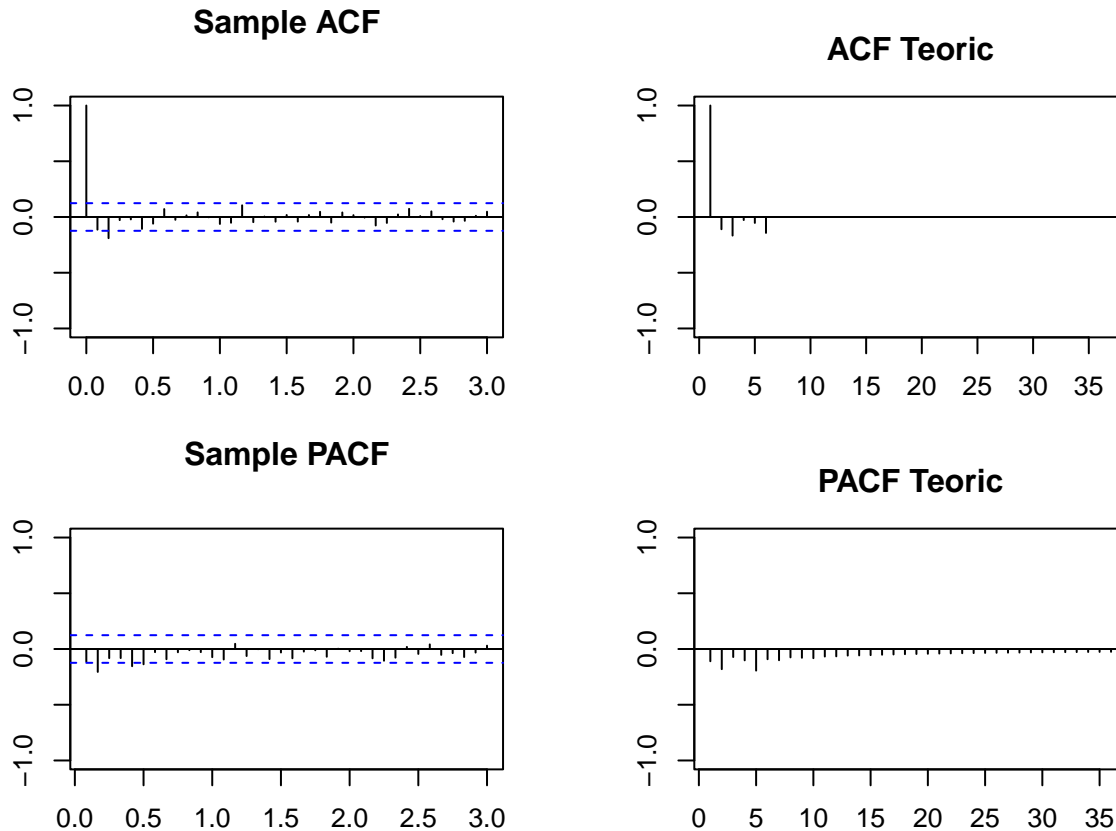
Sample ACF

ACF Teoric

Sample PACF

PACF Teoric

In the graphs, one can see that there seems to be some problems with the normality of the residuals, as observations at the extremes do not follow normal distribution tails, and the normality tests reject the normality hypothesis. When checking for homoskedasticity, we can see that there is no clear presence of volatility, even though there are some significant lags at the beginning, which may indicate the opposite. Finally, we can see that there is independence between lags. Consequently, we cannot say that the model is fully valid as the normality assumption does not hold and there could be volatility.

Atur: Unemployed registered people in Spain

```
serie=window(ts(read.table("atur.dat")/1000,start=1996,freq=12),start=1996)
```

8) For the log-transformed series, the $ARIMA(1,1,1)(0,1,1)_{12}$ is better than the $ARIMA(0,2,1)(0,1,1)_{12}$ model Answer: $FALSE$

In order to determine which model is better, we can first see the validity of both models and then check some measures to compare the models when it comes to goodness of fit and forecasting ability.

We first estimate both models using the logarithmic transformation, as it is needed due to a non-constant variance.

```
lnserie=log(serie)
mod1=arima(lnserie,order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12))
mod2=arima(lnserie,order=c(0,2,1),seasonal=list(order=c(0,1,1),period=12))
```

Now, we can check the validity of both models through the validation function. For the sake of brevity, we omit to show the graphs (as validation is not demanded), but we see that the behaviour of both models are

pretty similar: both models cannot be fully validated because the residuals do not behave as a white noise. Independence of some lags is not met, normality is rejected and volatility seems to be present.

Therefore we will need other criteria in order to choose a model, and we use goodness of fit and forecasting measures to assess both models.

```
resul=data.frame(
  Sigma2Z=c(mod1$sigma2, mod2$sigma2),
  AIC=c(AIC(mod1), AIC(mod2)),
  BIC=c(BIC(mod1), BIC(mod2)),
  RMSE=c(fRMSE(obs,pr),fRMSE(obs1,pr1)),
  MAE=c(fMAE(obs,pr),fMAE(obs1,pr1)),
  RMSPE=c(fRMSPE(obs,pr),fRMSPE(obs1,pr1)),
  MAPE=c(fMAPE(obs,pr),fMAPE(obs1,pr1)),
  CI_Length=c(fCI(tu,tl),fCI(tu1,tl1)))
row.names(resul)=c("Mod1","Mod2")
resul
```

```
##           Sigma2Z      AIC       BIC      RMSE       MAE      RMSPE       MAPE
## Mod1 9.056685e-05 -1766.337 -1751.870 86.69657 73.34856 0.02752382 0.02333661
## Mod2 9.586668e-05 -1739.830 -1728.991 83.55329 67.84670 0.02652930 0.02160159
##      CI_Length
## Mod1  486.4301
## Mod2  568.0208
```

The results show that the $ARIMA(1,1,1)(0,1,1)_{12}$ model seems to have a better goodness of fit than the $ARIMA(0,2,1)(0,1,1)_{12}$ model, but the latter has a better properties regarding forecasting, as it has more precision and accuracy and a smaller error. Because time series model are normally used for forecasting purposes, we prefer better forecasting features and hence we choose the $ARIMA(0,2,1)(0,1,1)_{12}$ model over the $ARIMA(1,1,1)(0,1,1)_{12}$ one.

9) The $ARIMA(1,1,1)(0,1,1)_{12}$ model for the series of logarithms is clearly stable   Answer: $FALSE$

In order to check the stability, we first need to express the model in its polinomial form and obtain the absolute value or modulus of the root for the autorregresive polinomial in order to check for the stability of the model.

$$(1 - 0.9381B)(1 - B)(1 - B^{12})x_t = (1 - 0.6542B - 0.5929B^{12})w_t$$
$$(1 - 0.9381B)(1 - B)(1 - B^{12}) = 0 \rightarrow unitary\ roots$$

Here, we can clearly see that the model has unitary roots, and hence, we can say that the model is not stable.

10) All the coefficients of the $ARIMA(8,1,0)(0,1,1)_{12}$ model are significant   Answer: $FALSE$

To check whether this is true or notm we can obtain the Z values for each coefficient in the following way, and see whether the values are lower than the common critical values for the common significance levels:

```
mod3=arima(lnserie,order=c(8,1,1),seasonal=list(order=c(0,1,1),period=12))
abs(mod3$coef/sqrt(diag(mod_ar1$var.coef)))
```

```
##        ar1         ar2         ar3         ar4         ar5         ar6         ar7
## 11.1998814 71.4585125   6.1949895 36.6989478   5.0874061 49.0174122   0.6525011
##        ar8         ma1         sma1
## 37.3219466 16.1182720 93.7731711
```

In this case, we can see how the seventh autorregreive coefficients is not significant at all common significance levels, and hence we can assure that not all coefficients of the model are significant.

EnerUSA: Total Primary Energy Consumption in USA

```
serie=ts(read.table("EnerUSA.dat", header=F)/1000,start=1995,freq=12)
```

11) The variance of the series increases with its level (mean), so we should take logarithms
Answer: *FALSE*

To assess the necessity of the logarithmic transformation, we need to plot the mean versus the variance and a boxplot. In this case, the first plot does not make clear whether variance increases with the mean, even though higher points are more common with higher values of the mean (but there are also similar variance values as with low levels). The second plot allows to have a better insight, and because the dispersion seems to be constant, we would not need a logarithmic transformation.

```
m=apply(matrix(serie[1:(25*12)],nr=12),2,mean)
v=apply(matrix(serie[1:(25*12)],nr=12),2,var)
plot(m,v,xlab="Annual Means",ylab="Annual Variances",main="serie")
```
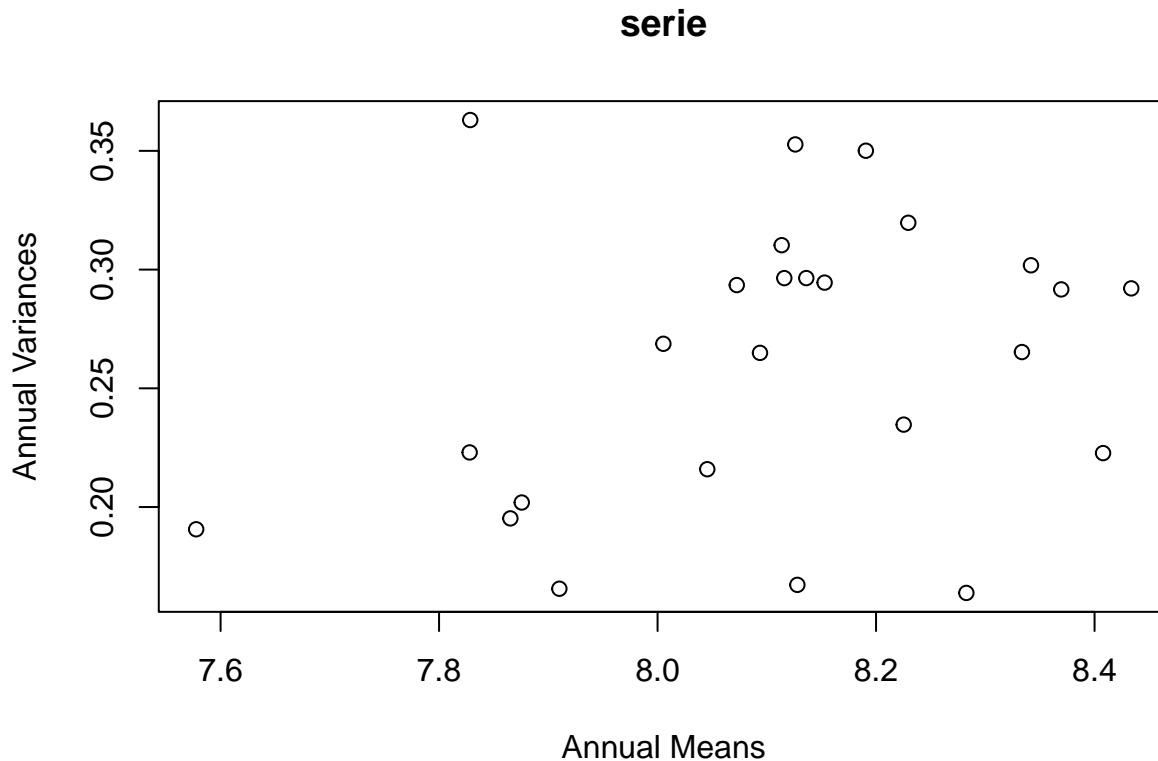

serie

```
boxplot(serie~floor(time(serie)))
```


floor(time(serie))
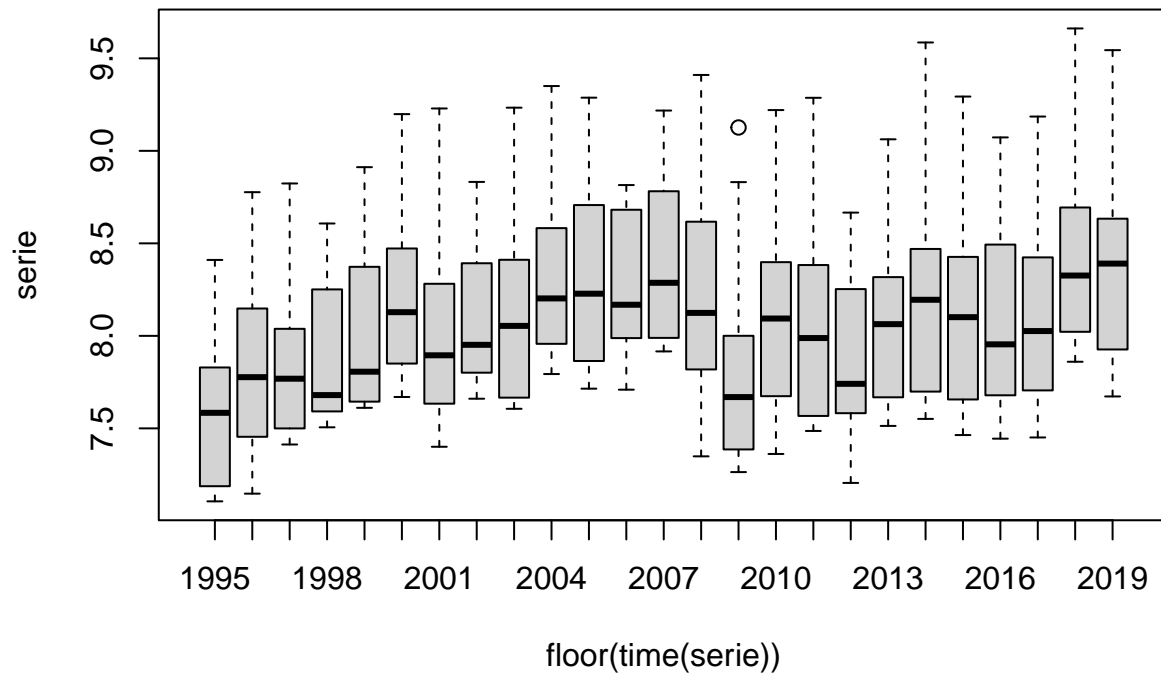
12) By looking at the ACF/PACF of the seasonally and regularly differenced series, an MA can be clearly identified for the seasonal part, and another MA for the regular part. Answer: *FALSE*

For this question, we need to use the ACF and the PACF graphs to assess whether MA models can be identified for both the seasonal and the regular part.

```
d12serie=diff(serie,12)
par(mfrow=c(1,2), mar=c(5,4,4,5))
acf(d12serie,ylim=c(-1,1),col=c(2,rep(1,11)),lwd=2,lag.max=72)
pacf(d12serie,ylim=c(-1,1),col=c(rep(1,11),2),lwd=2,lag.max=72)
```

V1   **Series d12serie**
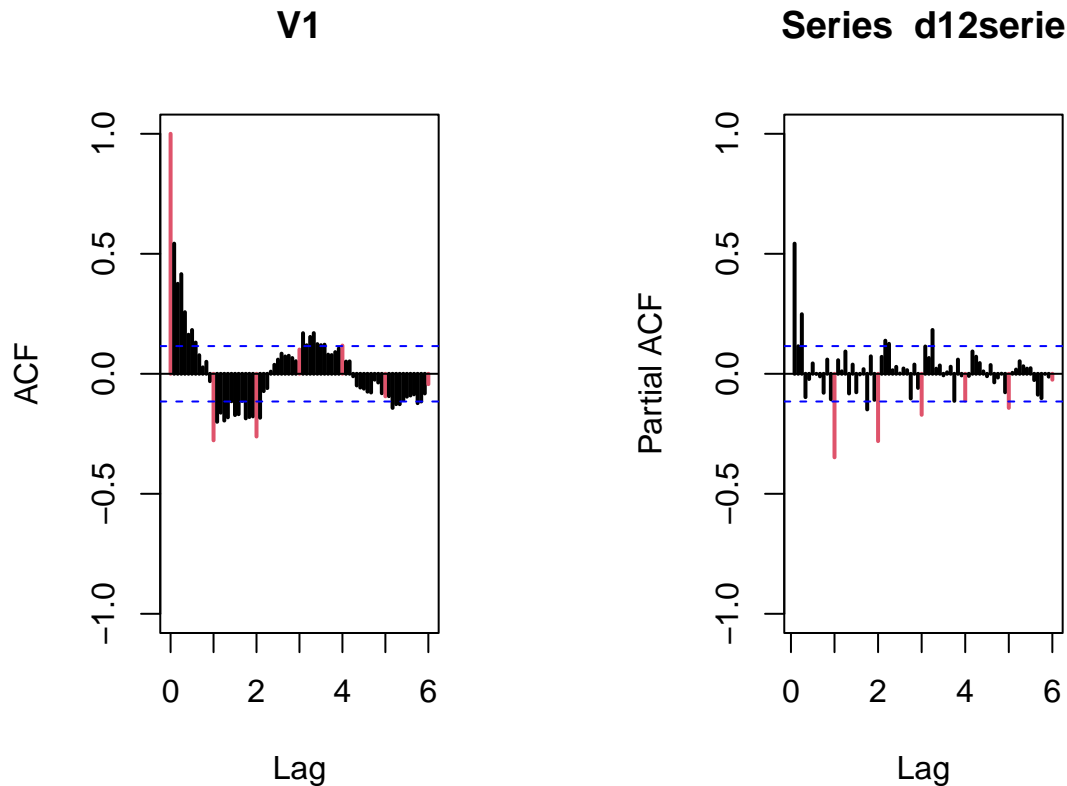
For the seasonal part, we can clearly identify and MA model because of the infinite decay of seasonal lags in the PACF, while the seasonal lags in the ACF seem to cut off at certain lags. In this part, we could use and MA(1) or MA(2). For the regular part, however, we would more likely identify an AR model, as there is a infinite decay of the regular lags in the ACF, while the regular lags in the PACF seem to cut off. For that part, we might identify an AR(2), for example. Consequently, we do not identify and MA for both the seasonal and the regular part.

13) The $ARIMA(0,1,2)(0,1,2)_{12}$ model yields better punctual predictions for the last 12 observations than the $ARIMA(3,0,0)(0,1,3)_{12}$ model with an intercept. Answer: $FALSE$

To obtain the answer about the quality of predictions of both models, we have to compute forecasting measures such as the RMSE, MAE, RMSPE, MAPE and the mean CI length, which allow to compare how the models forecast.

Hence, we compute, again, the table we have computed in previous exercises. Likewise, we do not show the computation of the necessary elements for predictions, but just the results obtained by doing so.

```
resul=data.frame(
  RMSE=c(fRMSE(obs,pr),fRMSE(obs1,pr1)),
  MAE=c(fMAE(obs,pr),fMAE(obs1,pr1)),
  RMSPE=c(fRMSPE(obs,pr),fRMSPE(obs1,pr1)),
  MAPE=c(fMAPE(obs,pr),fMAPE(obs1,pr1)),
  CI_Length=c(fCI(tu,tl),fCI(tu1,tl1)))
row.names(resul)=c("ARIMA(0,1, 2)(0,1,2)","ARIMA(3,0,0)(0,1,3)")
resul
```

```
##                          RMSE        MAE      RMSPE       MAPE CI_Length
## ARIMA(0,1, 2)(0,1,2) 0.1737406 0.15123593 0.02138503 0.01840705 0.8467449
## ARIMA(3,0,0)(0,1,3)  0.1264650 0.09868393 0.01537687 0.01188991 0.6336751
```

In this case, all measures favour the $ARIMA(3,0,0)(0,1,3)_{12}$ model when it comes to precision, accuracy, error and variability of punctual forecasts (measured by the mean length of the intervals). In this case, lower values indicate a better performance, so we can say that this model is better than the $ARIMA(0,1,2)(0,1,2)_{12}$ model.

APB: Mercantile Activity in the Port of Barcelona

```
serie=window(ts(read.table("apb.dat")[,1]/1000,start=1997,freq=12),start=2000)
```

14) The forecasts from the $ARIMA(0,1,1)(0,1,1)_{12}$ model for the last year are correct.   Answer: *FALSE*

In order to answer this question, we first need to estimate the model. We use a logarithmic transformation to estimate the model because of non-constant variance issues.

```
lnserie=log(serie)
mod1=arima(lnserie,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12))
```

Once estimated, we proceed to do the forecasts for the model. We represent these punctual forecasts and their confidence interval in the following plot.

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


Model ARIMA(0,1,1)(0,1,1)
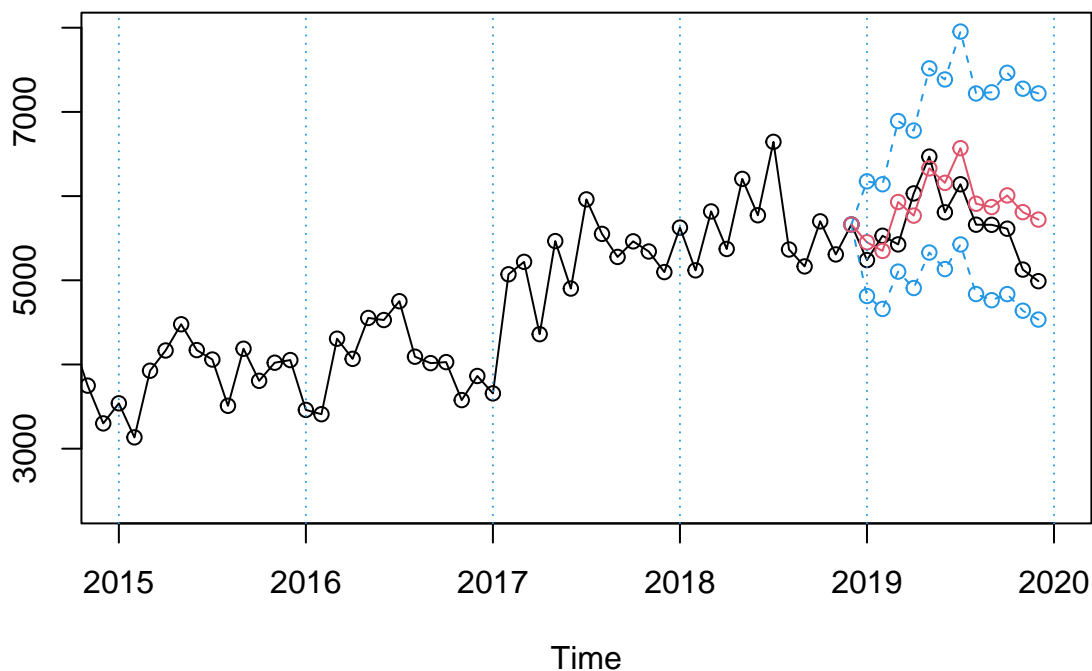
As we can see, the forecasts are not exactly the observations, but this is logical. In order to give a better sense of the accuracy and precision for the forecast, we present the RMSPE and the MAPE.

```
RMSPE=sqrt(sum(((obs-pr)/obs)^2)/12)
MAPE=sum(abs(obs-pr)/obs)/12
data.frame("RMSPE"= RMSPE,"MAPE"=MAPE)
```

```
##        RMSPE        MAPE
## 1 0.07620007 0.06603658
```

From the results we can clearly see that the model is quite precise and accurate, having very low values
for the different errors. However, the forecasts are not exactly the observations, so we would say that the
forecasts are not "correct" in that sense.

**15) The Ljung-Box statistics for the previous model have p-values higher than 0.05 for all the
lags, and that implies that the model is valid**   Answer: $FALSE$

Finally, to answer the question we would need to use the validation function used in previous exercises. In
order to use it, we do a seasonal and a regular difference to the logarithmic series, so that we work with a
stationary series.

```
d1d12lnserie=diff(diff(lnserie,12))
validation(mod1,d1d12lnserie)
```
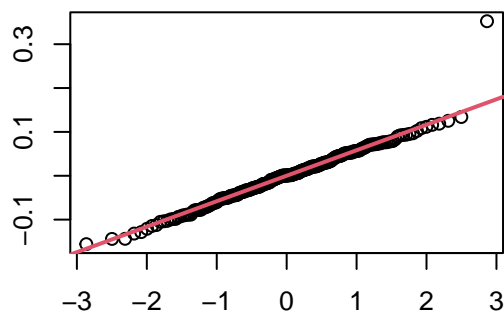
Series resid
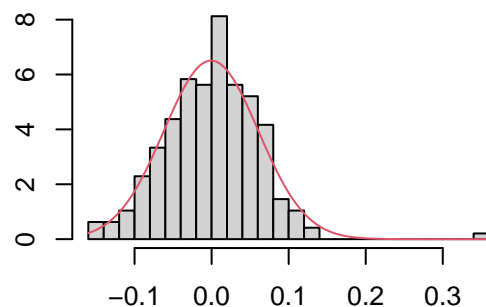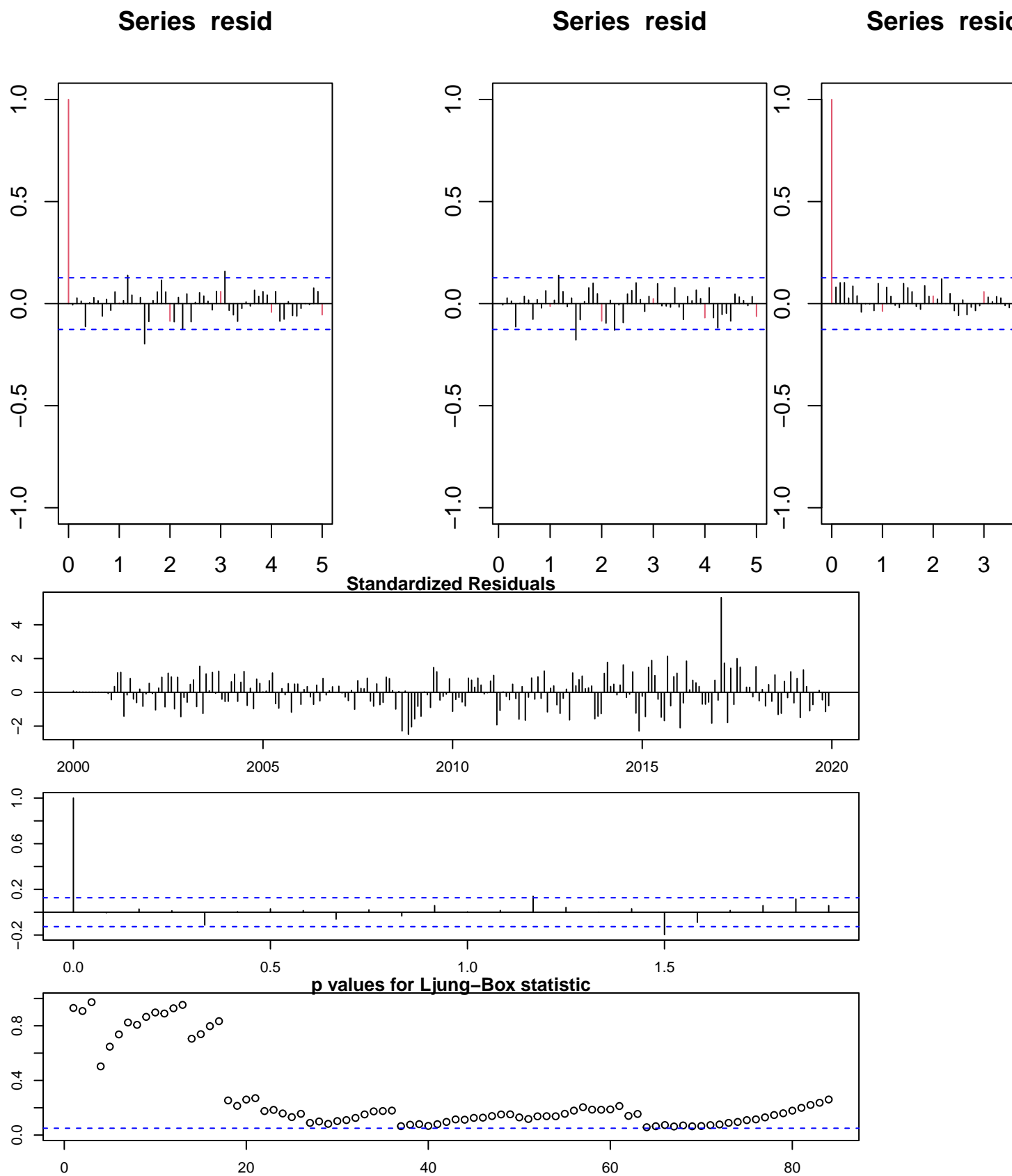
Series resid

Series resid

Standardized Residuals

p values for Ljung−Box statistic
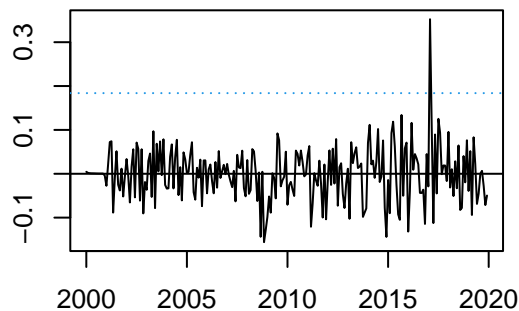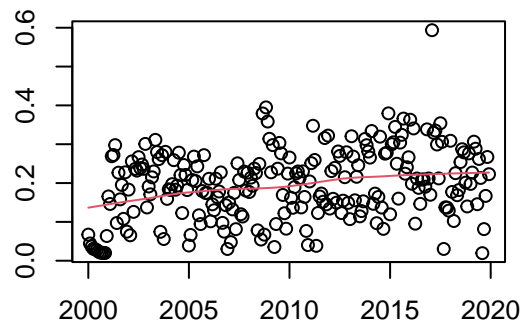
```
##
## --------------------------------------------------------------------------
##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))
##
## Coefficients:
##           ma1      sma1
##       -0.5299   -0.9278
## s.e.   0.0590    0.0762
##
## sigma^2 estimated as 0.00395:  log likelihood = 294.34,  aic = -582.68
##
## Modul of AR Characteristic polynomial Roots:
##
## Modul of MA Characteristic polynomial Roots:  1.006266 1.006266 1.006266 1.006266 1.006266 1.006266
##
## Psi-weights (MA(inf))
##
## --------------------
##      psi 1       psi 2       psi 3       psi 4       psi 5       psi 6       psi 7
## -0.5299476  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##      psi 8       psi 9      psi 10      psi 11      psi 12      psi 13      psi 14
##  0.0000000  0.0000000  0.0000000  0.0000000 -0.9277801  0.4916748  0.0000000
##     psi 15      psi 16      psi 17      psi 18      psi 19      psi 20
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##
## Pi-weights (AR(inf))
##
## --------------------
##           pi 1           pi 2           pi 3           pi 4           pi 5
## -0.5299475747 -0.2808444319 -0.1488328255 -0.0788735949 -0.0417988703
##           pi 6           pi 7           pi 8           pi 9          pi 10
## -0.0221512100 -0.0117389800 -0.0062210440 -0.0032968272 -0.0017471456
##          pi 11          pi 12          pi 13          pi 14          pi 15
## -0.0009258956 -0.9282707985 -0.4919348583 -0.2606996850 -0.1381571658
##          pi 16          pi 17          pi 18          pi 19          pi 20
## -0.0732160549 -0.0388006707 -0.0205623214 -0.0108969523 -0.0057748135
##
##  Shapiro-Wilk normality test
##
## data:  resid(model)
## W = 0.96325, p-value = 7.761e-06
```
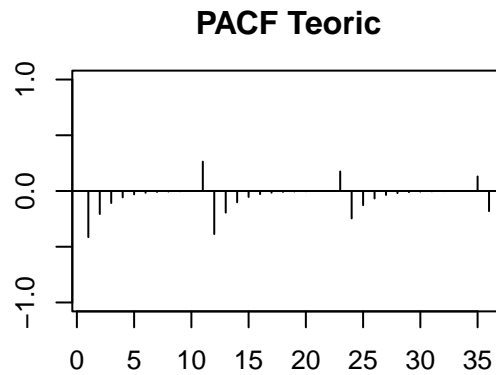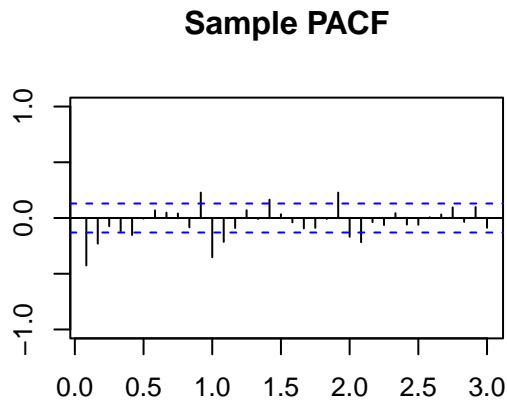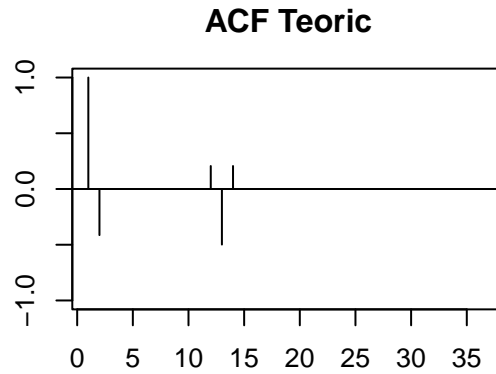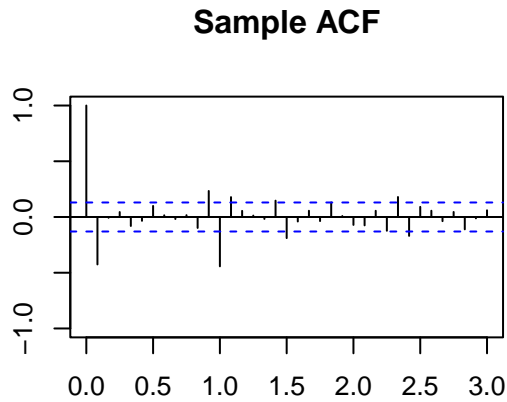
Sample ACF

ACF Teoric

Sample PACF

PACF Teoric

From the results, one can clearly see that the Ljung-Box statistics are all above 0.05. However, time series models as the ones we work with do not need to be validated if other assumptions made on the models do not hold, such as normality or homoskedasticity. In this case, for example, normality is mostly holding (because of a single outlier) and there is no sign of volatility, so we would validate this model. But, because of the Ljung-Box statistics above the significance level not being a sufficient condition to say that a model is valid, we would say this claim is false.