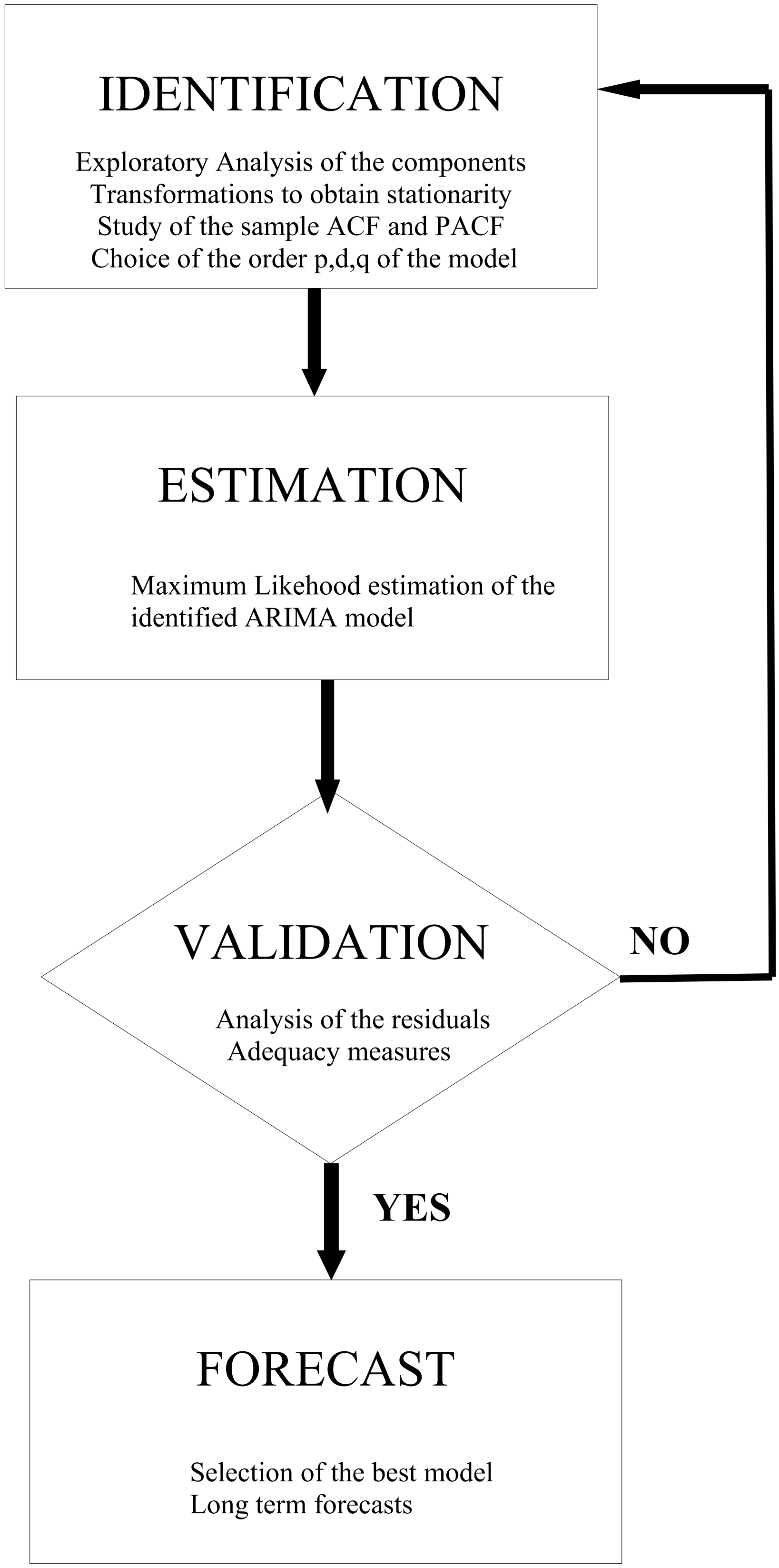
ARIMA Models for Time Series with R. Practical Cases

Contents

Box-Jenkins Methodology for Time Series



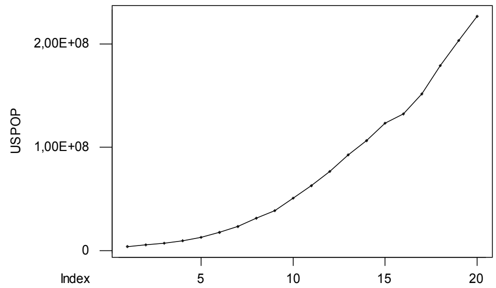
Practices: SESSION 1

1) Preliminary Exploratory Analysis

1) Recommended series to work with (all of them can be found in MINITAB \..\Guio_Lab\L01\SERIES.MTW).:

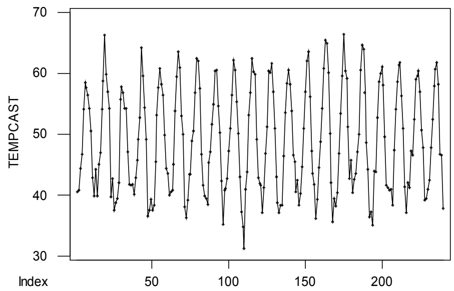
USPOP

US population evolution (in millions) with respect to the decennial census made within the period 1790-1970



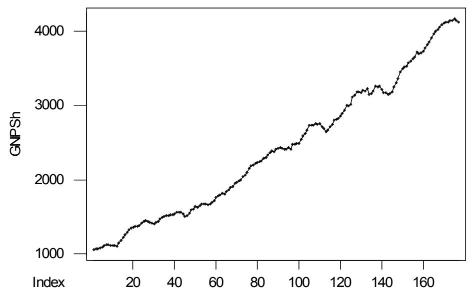
NOTTEM

Monthly average air temperatures in Nottingham Castle in Fahrenheit degrees from January of 1920 until 1939.

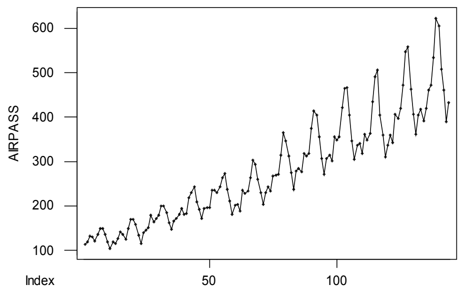


GNPSH

177 quarterly observations (seasonal adjusted data) of the USA Gross Domestic Product from January 1947 to December 1991. (Shumway R.H. & Stoffer D. S., 1999) Notice the Economic crisis in 1974-75 and 1979-80

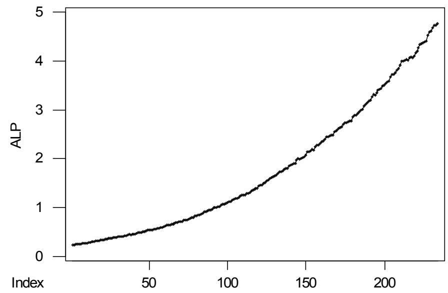


AIRPASSENGERS
144 monthly total amount (in thousands) of passengers in the international airlines of the US between 1949 and 1960 (Box and Jenkins, 1994)



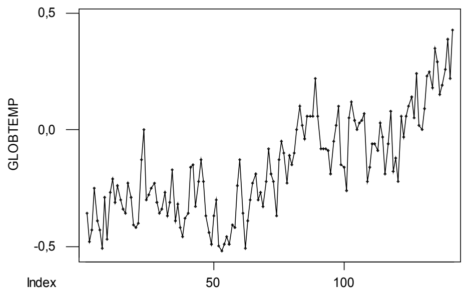
ALP

234 monthly observations of the monetary aggregate of Spain from January of 1972 to June of 1991.



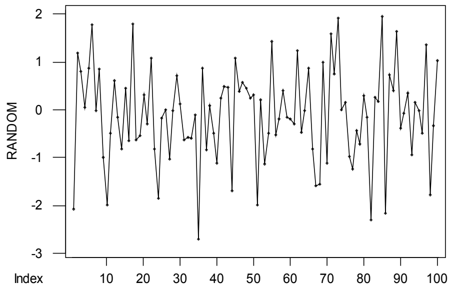
GLOBTEMP

Annual World temperatures (differences from an average rate) between 1855 and 1997 (Shumway, 2000)



RANDOM

Simulated data with a Normal(0,1) distribution.



2) Methodology:

a) Preliminary Exploratory Analysis. Series components Determination.

As a result of the basic Exploratory data analysis obtained with MINITAB/R, which of the proposed series presents each of the following characteristics?

...**Trend**?
Which kind of trend it looks like, stochastic or deterministic? If it is deterministic, lineal, squared, ...?

...**Seasonality**?
What is the period?

...**Heterocedasticity** (non-constant variance)?

From the answers given below, which of the series could be considered already stationary? Why?

b) Transformations to have stationary data.

A Box-Cox transformation must be applied to the series with non-constant variance, (most of the times, the transformation is the logarithm which corresponds to λ=0)

$$W_t = \frac{X_t^{\lambda} - 1}{\lambda}$$

Fit the corresponding model for the series with deterministic trend and study the random noises. If they do not behave as expected, the trend may be stochastic.

$$X_t = \alpha_0 + \alpha_1 t + W_t$$

For series with stochastic (and deterministic) trend, differentiation (1-B) remove this component. The process can be repeated as many times as necessary until a constant mean is obtained (the trend is then completely removed). Overdifferencing can be detected by the increase of the variance in the last differenced series.

$$W_t = (1 - B)X_t$$

For seasonal series, the differencing must be of the same order of the period of the seasonal component (for example 12 for an annual series, 4 for quarterly data, ...)

$$W_t = (1 - B^s)X_t$$

It might be necessary to work on the increment per unit on the economical series, this transformation corresponds to the presence of a lineal trend and non-constant variance. The transformation in this case results as a differentiation of the logarithm series.

$$W_t = (1 - B)\log X_t \cong \frac{X_t - X_{t-1}}{X_{t-1}}$$

In stationary series, the autocorrelation decreases quickly, that means that any observation is strongly correlated with the previous observations close to it but not with the large lags. This can be observed analysing the **Sample Autocorrelation Function** (sample ACF).

Taking in account the previous indications, determine which transformations are necessary to achieve a stationary series from the series proposed before:

SERIES	COMPONENTS	TRANSFORMATIONS	% Variance Reduction
NOTTEM			
GNPSH			
AIRPASS			
ALP			
GLOBTEMP			
RANDOM			

2) Study of the Autocorrelation Function

The autocorrelation function must be studied in order to identify the model..
In R, it can be calculated as follows:

1	Construction of the variable x_{t-1}	`lag<-1`
		`x0<-x[1:(length(x)-lag)]`
		`x1<-x[(1+lag):length(x)]`

2	Calculation of $\sum_{t=2}^{n}(x_t - \bar{x})(x_{t-1} - \bar{x})$	`a<-sum((x0-mean(x))*(x1-mean(x)))`
3	Calculation of $\sum_{t=1}^{n}(x_t - \bar{x})^2$	`b<-sum((x-mean(x))^2)`
4	Calculation of $\hat{\rho}(1)$	`r1<-a/b`
5	Repeat the previous steps for $\hat{\rho}(2), \hat{\rho}(3),...$	`lag<-2` `...`

For the proposed series, once the stationary adjusted series are obtained, create the plot of the ACF and describe its behaviour (amount of significant lags, decreasing shape, ...)

SERIES	ACF	DESCRIPTION
NOTTEM		
GNPSH		
AIRPASS		
ALP		
GLOBTEMP		
RANDOM		

3) Exercise 1.1: Exploratory Analysis and transformation. Proposed Series.

Summary of R instructions:

`plot(serie)`	Plot of the series
`plot(decompose(serie))`	Plot of the decomposition of the basic model

`ng=length(serie)%/%12*12`
`m=apply(matrix(serie[1:ng],nrow=12),2,mean)`
`s=apply(matrix(serie[1:ng],nrow=12),2,sd)`
`plot(m,s,xlab="means",ylab="StandardDeviations")`
`abline(lm(s~m),col=2,lty=3,lwd=3)`
`summary(lm(s~m))`
	Plot of the means and variances

`library(MASS)` `boxcox(serie~1)`	Box-Cox Transformation
`lnserie=log(serie)` `plot(lnserie)`	$W_t=log(X_t)$
`d1serie=diff(serie)` `plot(d1serie)`	$W_t=(1-B)X_t$
`d12serie=diff(lnserie,lag=12)` `plot(d12serie)`	$W_t=(1-B^{12})(X_t)$
`d1lnserie=diff(log(serie))` `plot(d1lnserie)`	$W_t=(1-B)log(X_t)$
`var(serie)`	Variance of the series

`acf(serie,ylim=c(-1,1))`
`win.graph()`
`pacf(serie,ylim=c(-1,1))`
	Plot of the ACF and the PACF of the series

Indicate which components of the basic model can be found in the proposed series. Determine which transformations are necessary to obtain a stationary series from each of the series proposed:

SERIES	BASIC MODEL COMPONENT	TRANSFORMATIONS	% Variance Reduction
PIBsp			
IPCsp			
AirBCN			
Tuberc			

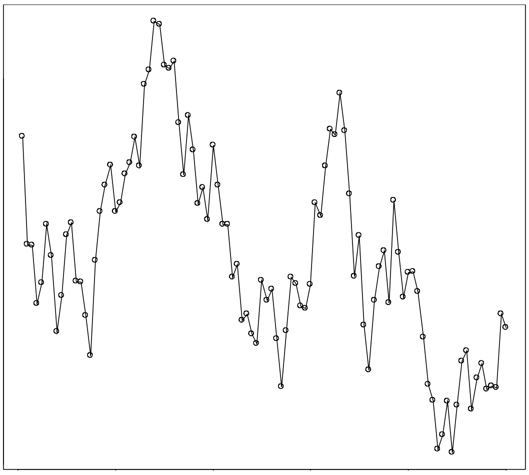
4) Simulation of the series. Comparison between the theoretical and the sample ACF and PACF.

Simulation of a series with R and comparison between the sample ACF and PACF of the series and the ACF and PACF of the theoretical model.

Example 1: AR(1) Model

$$x_t = 0.9x_{t-1} + z_t \qquad (1-0.9B)x_t = z_t$$

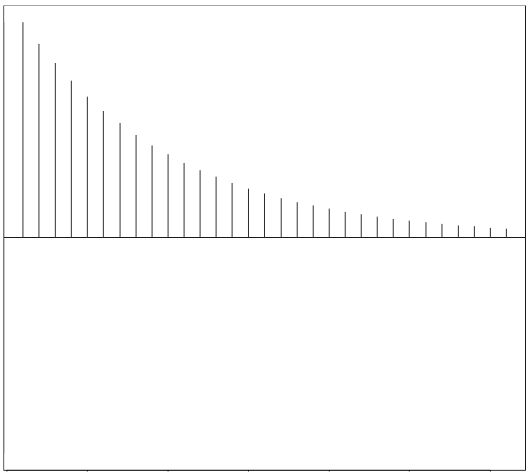
```
#Model Parameters Xt=0.9Xt-1+Zt
model.ar<-c(0.9)
#Generate of a  100 observation series and save it as a "ts" file
ser<-ts(arima.sim(list(ar=model.ar),100))
plot(ser)
```

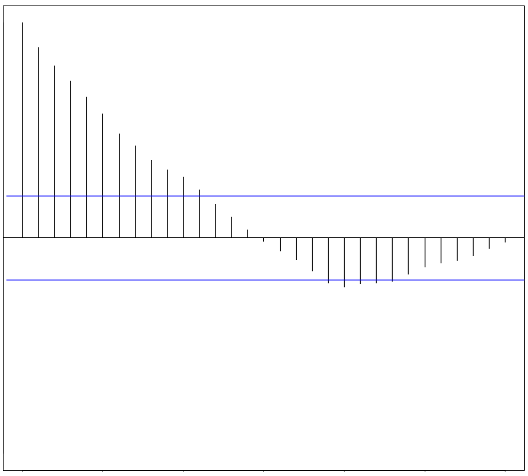


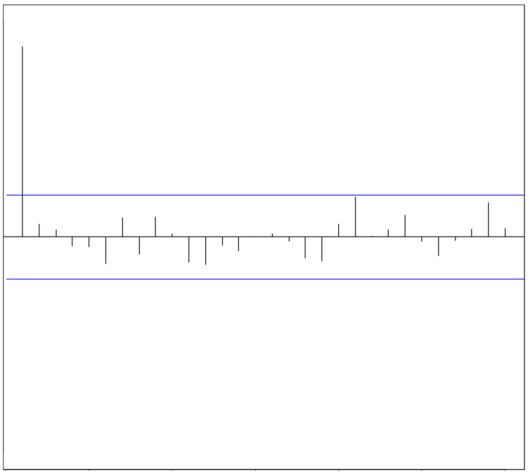
```
#Comparison between the first 10 autocorrelation of the model and the sample ones
data.frame(model=ARMAacf(ar=model.ar,lag.max=10),mostra=acf(ser,lag=10)$acf)

       model    mostra
0  1.0000000 1.0000000
1  0.9000000 0.8872802
2  0.8100000 0.8006852
3  0.7290000 0.7290888
4  0.6561000 0.6541672
5  0.5904900 0.5779644
6  0.5314410 0.4859813
7  0.4782969 0.4295134
8  0.4304672 0.3604706
9  0.3874205 0.3184554
10 0.3486784 0.2840197

#Calculation of the ACF values until lag 30 for the theoretical model
model.acf<-ARMAacf(ar=model.ar,lag.max=30)

#Plot of the value between vertical lines (type="h") and limitation of the vertical
#axis to values between -1 and 1. In addition we plot a horizontal line that will do as
#x-axis
plot(model.acf,type="h",ylim=c(-1,1))
abline(h=0)

#Plot of the sample ACF obtained from the series generated before
acf(ser,ylim=c(-1,1),lag.max=30)
```

```
#Same process for the PACF
model.pacf<-ARMAacf(ar=model.ar,lag.max=30,pacf=T)
plot(model.pacf,type="h",ylim=c(-1,1))
abline(h=0)

#Plot of the sample PACF obtained from the series generated before
pacf(ser,ylim=c(-1,1),lag.max=30)
```

Model **Series**



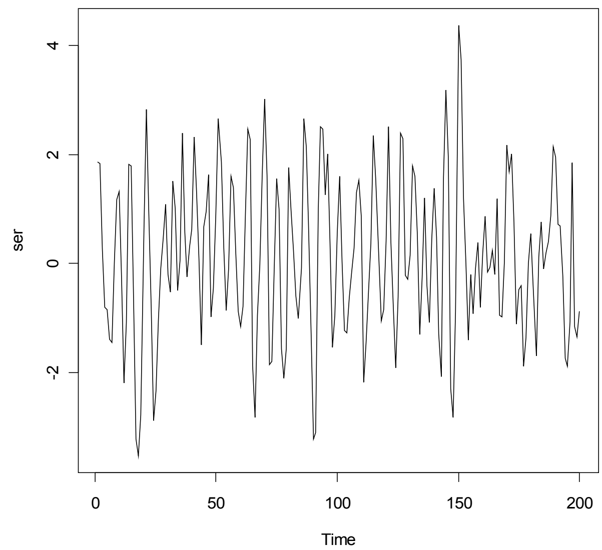




Example 2: AR(2) Model

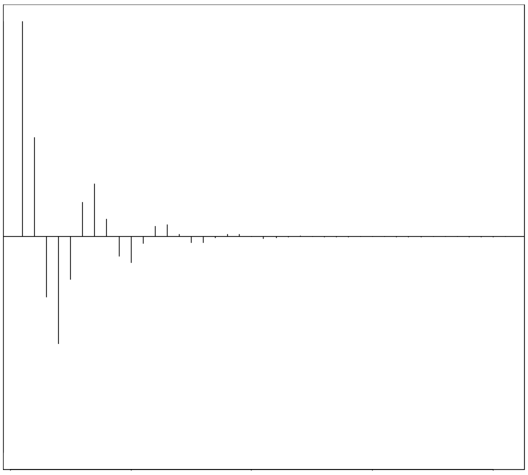
$$x_t = 0.75x_{t-1} - 0.625x_{t-2} + z_t \qquad (1 - 0.75B + 0.625B^2)x_t = z_t$$

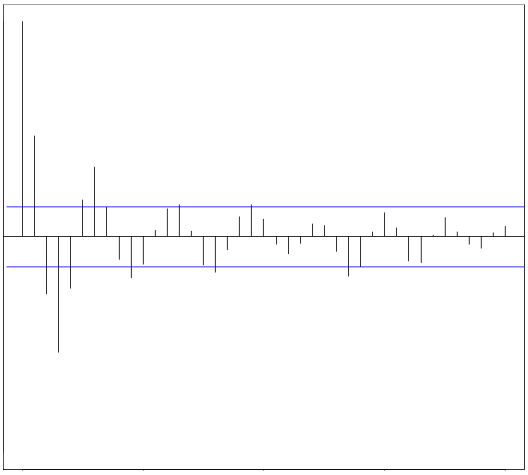
```
# AR(2)
model.ar<-c(0.75,-.625)
ser<-ts(arima.sim(list(ar=model.ar),200))
plot(ser)

# Comparison between the ACF's
model.acf<-ARMAacf(ar=model.ar,lag.max=40)
plot(model.acf,type="h",ylim=c(-1,1))
abline(h=0)
#Open a new Window to view the plot
win.graph()
acf(ser,lag.max=40,ylim=c(-1,1))
# Comparison between the PACF's
model.pacf<-ARMAacf(ar=model.ar,lag.max=40,pacf=T)
win.graph()
plot(model.pacf,type="h",ylim=c(-1,1))
abline(h=0)
win.graph()
pacf(ser,lag.max=40,ylim=c(-1,1))

#Calculation of the characteristic polynomial roots
polyroot(c(1,-0.75,.625))
[1] 0.6+1.113553i 0.6-1.113553i

#Calculation of the characteristic polynomial roots module
Mod(polyroot(c(1,-0.75,.625)))
[1] 1.264911 1.264911

#Both roots are outside the unit circle, so the AR model is stationary
```

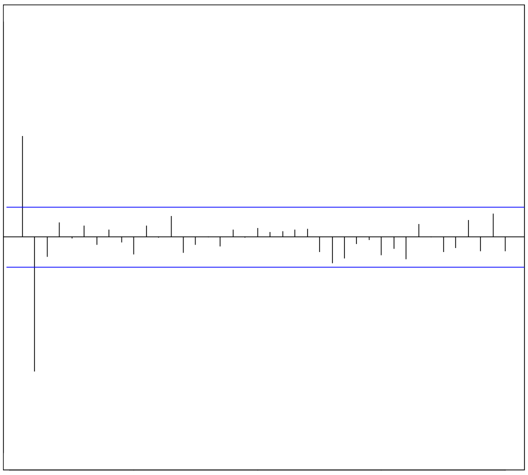


Model

Series







5) Exercise 1.2: Simulation of ARMA models

The objective of this exercise is to see how the Autocorrelation (ACF) and Partial Autocorrelation (PACF) functions of a stationary time series behave when the sample size increases for different models. During the whole exercise it is assumed that $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ with $\sigma^2 = 1$.

1. Different theoretical models are proposed . For each model, fill the gaps in the table below with a brief description of the results of the following steps:
   a) Representation of the theoretical ACF and PACF. Description of both. For increasing sample size (n = 25, 100, 500) generate series of the model with the same seed for each n.
   b) For each of the generated series, plot the theoretical and sample ACF and PACF simultaneously. Write down the value of n from which the theoretical and the sample ACF are similar.
   c) Calculate the roots of the characteristic polynomial.

	Model: MA(1) $X_t = Z_t - 0.75Z_{t-1}$	**Model: MA(1)** $X_t = Z_t + 0.75Z_{t-1}$
Description		

	Model: MA(1) $X_t = Z_t - 0.05Z_{t-1}$	**Model: MA(2)** $X_t = Z_t + 0.4Z_{t-1} - 0.32Z_{t-2}$
Description		

	Model: MA(2) $X_t = Z_t + 0.3Z_{t-1} - 0.4Z_{t-2}$	**Model: MA(2)** $X_t = Z_t + 0.4Z_{t-1} - 0.05Z_{t-2}$
Description		

	Model: AR(1) $X_t = 0.75X_{t-1} + Z_t$	**Model: AR(1)** $X_t = -0.75X_{t-1} + Z_t$
Description		

	Model: AR(1) $X_t = 0.999X_{t-1} + Z_t$	**Model: AR(2)** $X_t = 0.7X_{t-1} - 0.1X_{t-2} + Z_t$
Description		

	Model: AR(2) $X_t = -0.4X_{t-1} + 0.45X_{t-2} + Z_t$	**Model: AR(2)** $X_t = 0.75X_{t-1} - 0.5625X_{t-2} + Z_t$
Description		

	Model: ARMA(1,1) $X_t = 0.8X_{t-1} + Z_t + 0.8Z_{t-1}$	**Model: ARMA(1,1)** $X_t = -0.8X_{t-1} + Z_t + 0.8Z_{t-1}$
Description		

2. Overall conclusions:
    a) Indicate the models for which the sample function are similar to the theoretical's for small values of *n.*
    b) Find the relation between this selection and the situation of the characteristic polynomial roots relative to the unite circle.
    c) Does the size of the sample affect in the same way to identification of the AR and MA part? Why?
    d) Full study of the AR(2) case (Chapter 3).

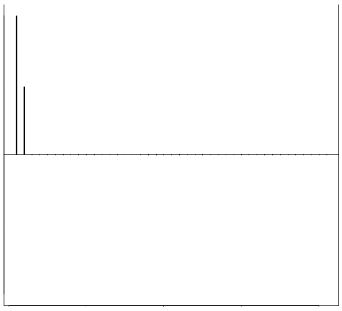
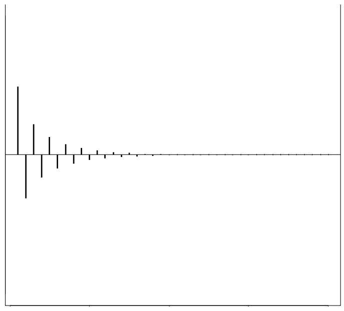
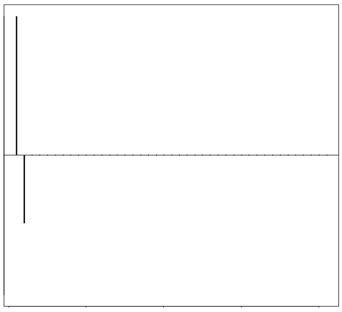
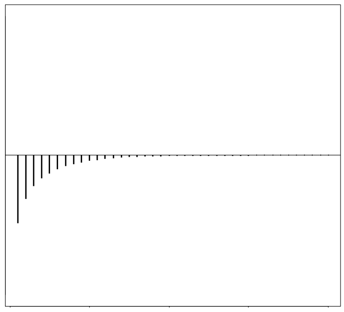
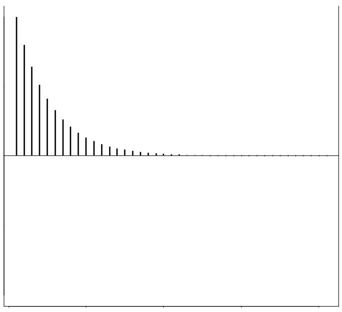
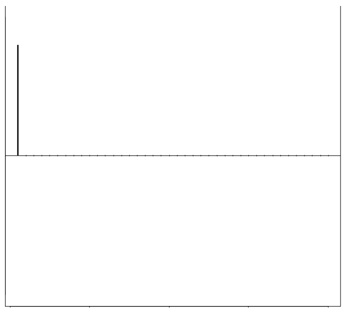
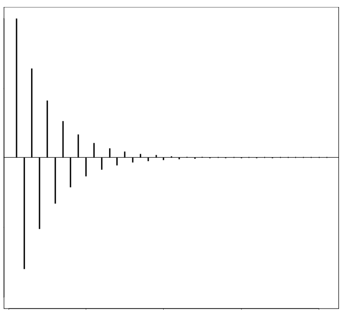
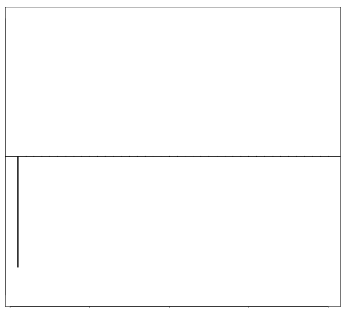
SUMMARY TABLE

(Let $\mu = 0$, otherwise $X_t = Y_t - \mu$)

	AR*(p)*	MA*(q)*	ARMA*(p,q)*
ACF	**Infinitely many values, linear combination of attenuated exponential and/or sinusoid functions**	**Zero after the first (finite) q lags**	**Infinitely many values, linear combination of attenuated exponential and/or sinusoid functions after the first *q-p* values**
PACF	**Zero after the first (finite) p lags**	**Infinitely many values, linear combination of attenuated exponential and/or sinusoid functions**	**Infinitely many values, linear combination of attenuated exponential and/or sinusoid functions after the first *p-q* values**
Stationarity Condition (Causality)	**Roots of $\phi_p(B) = 0$ outside the unit circle**	**Always stationary**	**Roots of $\phi_p(B) = 0$ outside the unit circle**
Invertibility Condition	**The model itself it is expressed as a function of previous random noises**	**Roots of $\theta_q(B) = 0$ outside the unit circle**	**Roots of $\theta_q(B) = 0$ outside the unit circle**

AR Expression	$\phi_p(B)X_t = Z_t$	$\pi(B)X_t = \frac{1}{\theta_q\left(B\right)}X_t = Z_t$	$\pi(B)X_t = \frac{\phi_q\left(B\right)}{\theta_q\left(B\right)}X_t = Z_t$
MA Expression	$X_t = \frac{1}{\theta_q(B)}Z_t = \psi(B)Z_t$	$X_t = \theta_q(B)Z_t$	$X_t = \frac{\theta_q(B)}{\phi_q(B)}Z_t = \psi(B)Z_t$
Weights $\pi(B)X_t = Z_t$	**Infinitely many**	**Infinitely many**	**Infinitely many**
Weights $X_t = \psi(B)Z_t$	**Infinitely many**	**Infinitely many**	**Infinitely many**

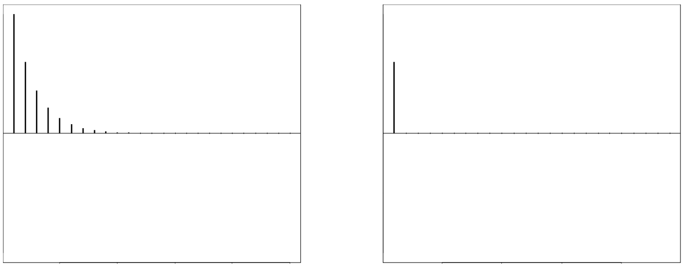
Basic Models MA(1) and AR(1)

Model	Param.	ACF	PACF
MA(1)	θ>0		
	θ<0		
AR(1)	ϕ>0		
	ϕ<0		

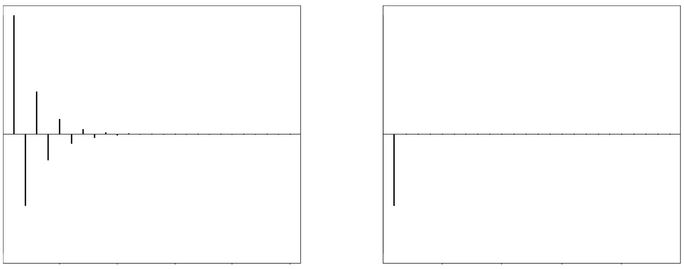
Model AR(1)

```
model.acf<-ARMAacf(ar=-0.6,ma=NULL,lag.max=24)
plot(model.acf,type="h",ylim=c(-1,1),lwd=2,main=model)
abline(h=0)
model.pacf<-ARMAacf(ar=model.ar,ma=model.ma,lag.max=24,pacf=T)
plot(model.pacf,type="h",ylim=c(-1,1),lwd=2)
abline(h=0)
```

$\phi = 0.6$



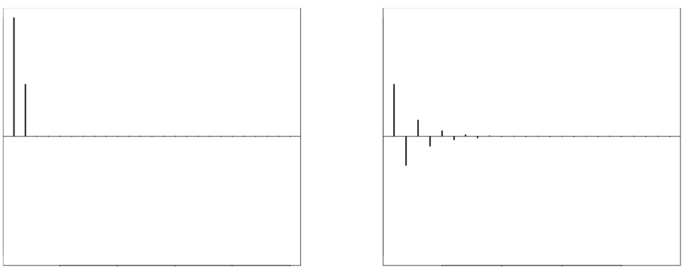
$\phi = -0.6$



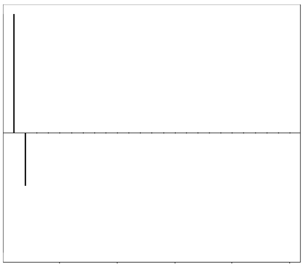
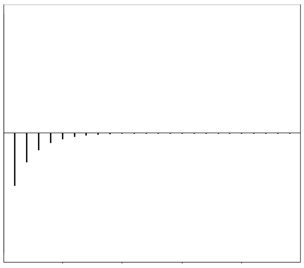
MA(1) Model

```
model.acf<-ARMAacf(ar=NULL,ma=0.6,lag.max=24)
plot(model.acf,type="h",ylim=c(-1,1),lwd=2,main=model)
abline(h=0)
model.pacf<-ARMAacf(ar=model.ar,ma=model.ma,lag.max=24,pacf=T)
plot(model.pacf,type="h",ylim=c(-1,1),lwd=2)
abline(h=0)
```

$\theta = 0.6$



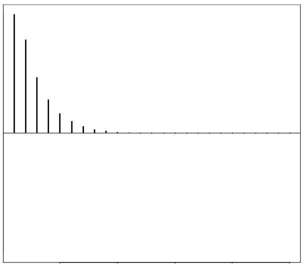
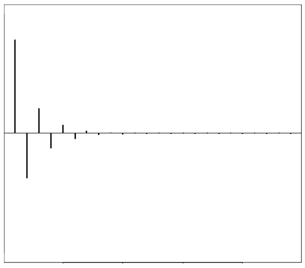
θ = −0.6

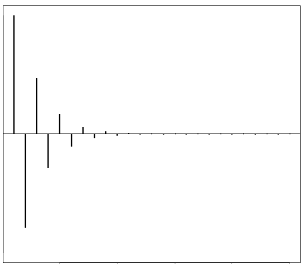
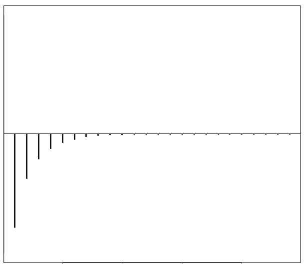
Model ARMA(1,1)

```
model.acf<-ARMAacf(ar=0.6,ma=0.6,lag.max=24)
plot(model.acf,type="h",ylim=c(-1,1),lwd=2,main=model)
abline(h=0)
model.pacf<-ARMAacf(ar=model.ar,ma=model.ma,lag.max=24,pacf=T)
plot(model.pacf,type="h",ylim=c(-1,1),lwd=2)
abline(h=0)
```

ϕ = 0.6 θ = 0.6

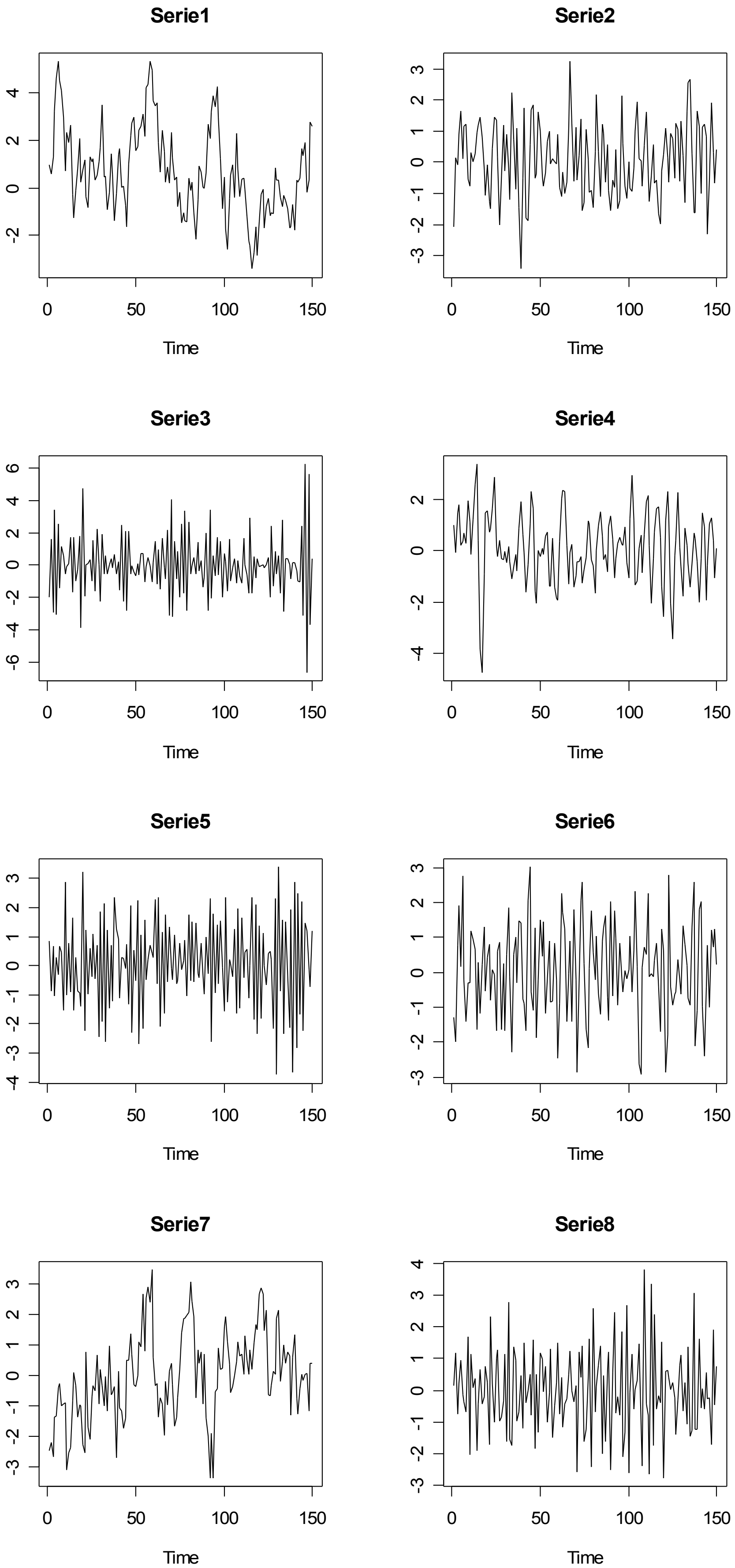
ϕ = −0.6 θ = −0.6

6) Exercise 1.3: Identification of the ARMA Models

Series from SERIE1.dat to serie8.dat have been generated by simple ARMA models (not necessarily the same). For each model calculate the characteristic polynomial roots and decide whether they are or not stationary and/or invertible. Identify which series correspond to each model and describe their ACF and PACF.

Model	Roots	Stationary	Invertible	Series	ACF and PACF
AR(1) $X_t=0.8X_{t-1}+Z_t$					
AR(1) $X_t=-0.78X_{t-1}+Z_t$					
AR(2) $X_t=0.75X_{t-1}-0.55X_{t-1}+Z_t$					
MA(1) $X_t=Z_t-0.85Z_{t-1}$					
MA(2) $X_t=Z_t+0.6Z_{t-1}-0.4Z_{t-2}$					
MA(2) $X_t=Z_t-0.5Z_{t-1}-0.9Z_{t-2}$					
ARMA(1,1) $X_t=-0.8X_{t-1}+Z_t-0.45Z_{t-1}$					
ARMA(1,1) $X_t=0.85X_{t-1}+Z_t-0.38Z_{t-1}$					

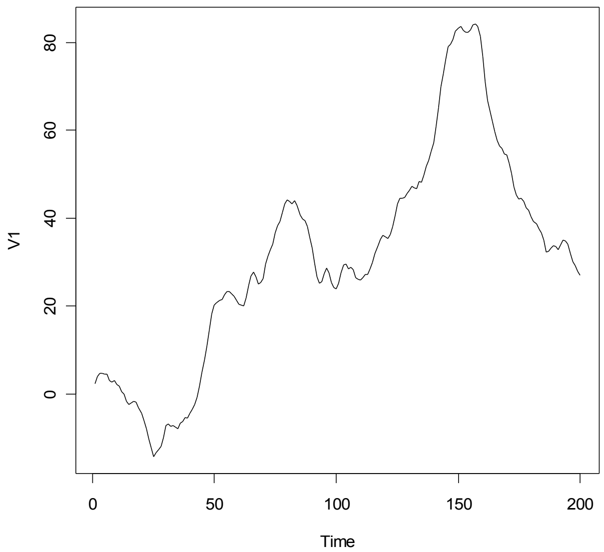

Serie1
Serie2
Serie3
Serie4
Serie5
Serie6
Serie7
Serie8
0 50 100 150
Time

7) Identification of models for simulated series. Cases E9 (Brockwell and Davis).

Data can be found in Brockwell P.J., Davis R.A. (1991) *Time Series: Theory and Methods. Cap. 9*. Series have been simulated with the following models:: E911, E921, E923, E951

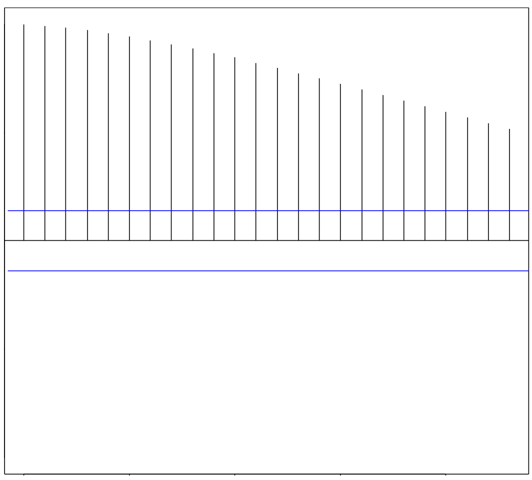
Case E911

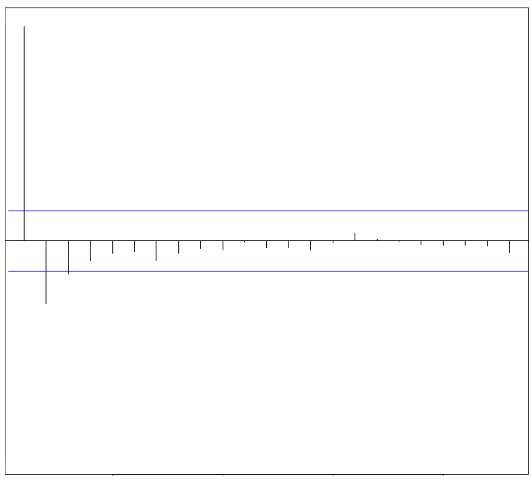
```
e911<-ts(read.table("C:\\e911.dat"))
plot(e911)
```



Analysis of the autocorrelation structure:

```
acf(e911,ylim=c(-1,1))
pacf(e911,ylim=c(-1,1))
```





Interpretation: Decreasing ACF and two significant lags in the PACF: Estimate a AR(2).

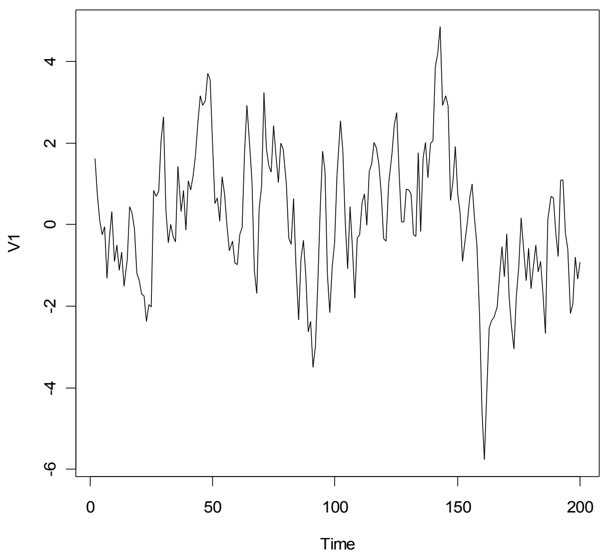
```
arima(e911,order=c(2,0,0),include.mean=T)

Call:
arima(x = e911, order = c(2, 0, 0), include.mean = T)

Coefficients:
         ar1      ar2  intercept
      1.8029  -0.8071    24.3666
s.e.  0.0407   0.0407    14.0351

sigma^2 estimated as 0.9659:  log likelihood = -284.06,  aic = 576.11
```

Estimated model:

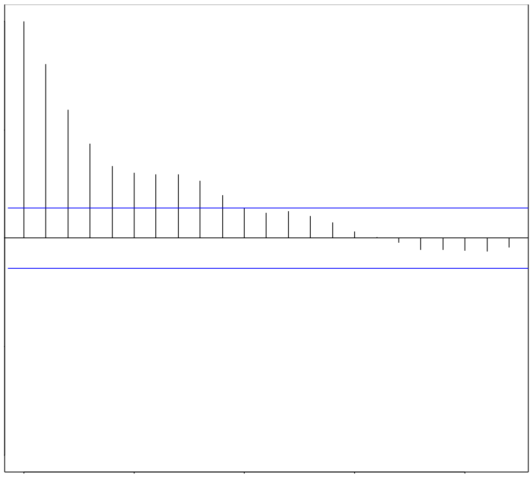
$$(1 - 1.8029B + 0.8071B^2)(X_t - \mu) = Z_t$$

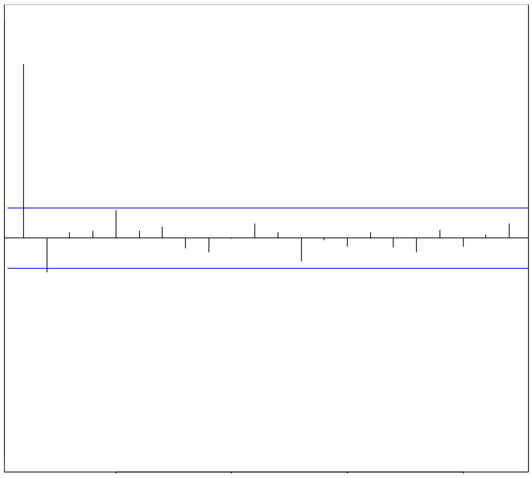
One of the roots of the characteristic polynomial is on the unit circle (the module of one of the roots is near 1). The variance decreases when differencing.

```
Mod(polyroot(c(1,-1.8029,0.8071)))
[1] 1.024912 1.208888

d1e911<-diff(e911)
var(e911)
          V1
V1 612.7864
var(d1e911)
          V1
V1 2.832163
```



```
acf(d1e911,ylim=c(-1,1))
pacf(d1e911,ylim=c(-1,1))
```



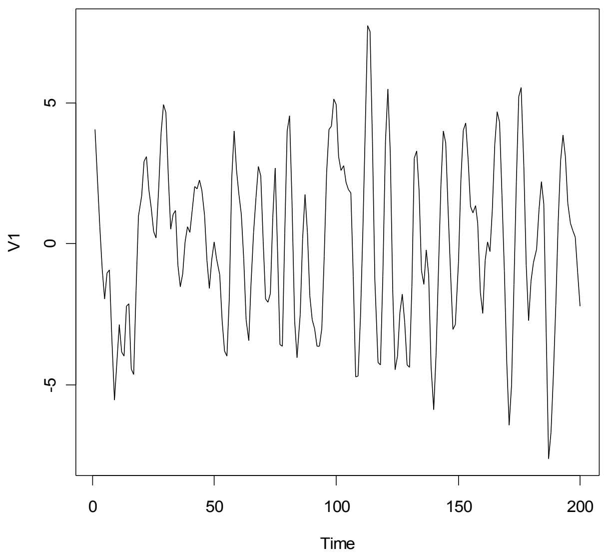


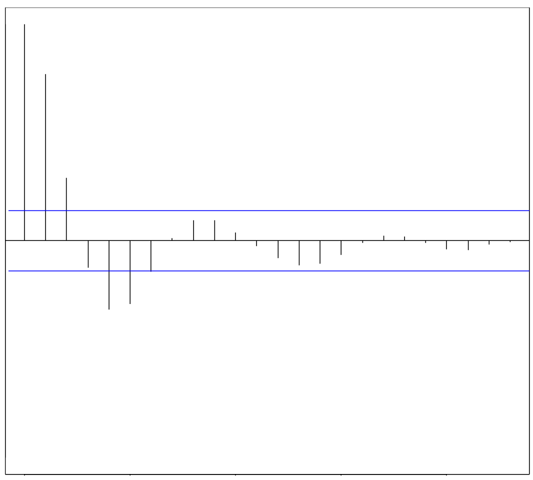
Model ARIMA(1,1,0).

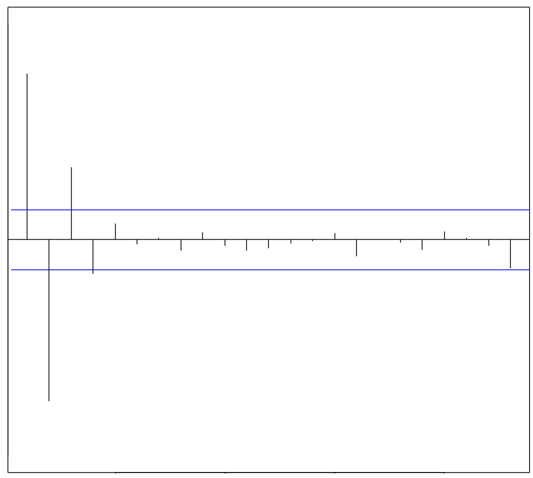
```
arima(d1e911,order=c(1,0,0),include.mean=F)
Call:
arima(x = d1e911, order = c(1, 0, 0), include.mean = F)

Coefficients:
         ar1
      0.8078
s.e.  0.0412

sigma^2 estimated as 0.9775:  log likelihood = -280.63,  aic = 565.27
```

$$(1 - 0.8078B)(1 - B)X_t = Z_t$$

Case E923





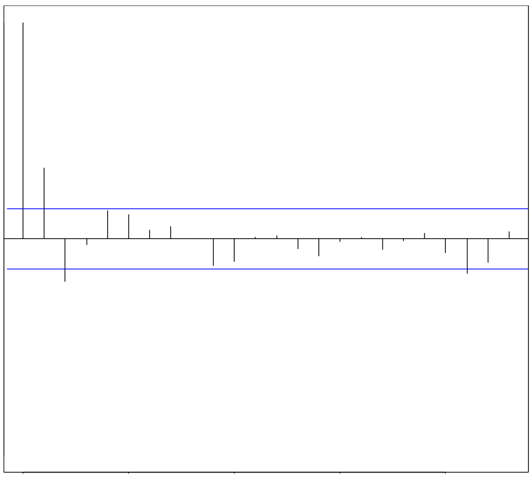


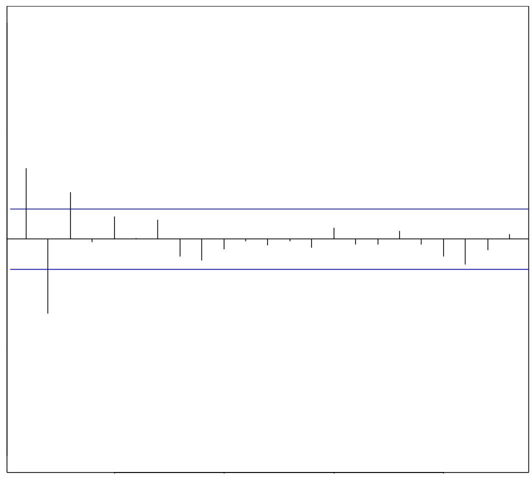
Fit an AR(5). Too many significant parameters.

```
arima(e923,order=c(5,0,0),include.mean=F)
Call:
arima(x = e923, order = c(5, 0, 0), include.mean = F)

Coefficients:
         ar1      ar2     ar3      ar4     ar5
      1.8917  -1.9328  1.2490  -0.6277  0.1866
s.e.  0.0695   0.1442  0.1781   0.1438  0.0691

sigma^2 estimated as 0.9922:  log likelihood = -284.95,  aic = 581.91
```

Fit with an AR(2) and check the residuals: They do not behave as white noise.

```
arima(e923,order=c(2,0,0),include.mean=F)
Call:
arima(x = e923, order = c(2, 0, 0), include.mean = F)

Coefficients:
         ar1      ar2
      1.3794  -0.7728
s.e.  0.0444   0.0442

sigma^2 estimated as 1.380:  log likelihood = -317.35,  aic = 640.71

resid<- arima(e923,order=c(2,0,0),include.mean=F)$residuals
acf(resid,ylim=c(-1,1))
pacf(resid,ylim=c(-1,1))
```





Add MA(2) and check the signification of the coefficients

```
arima(e923,order=c(2,0,2),include.mean=F)
Call:
arima(x = e923, order = c(2, 0, 2), include.mean = F)

Coefficients:
         ar1      ar2     ma1     ma2
      1.1179  -0.5803  0.7983  0.1029
s.e.  0.1026   0.0839  0.1246  0.1201

sigma^2 estimated as 0.9822:  log likelihood = -283.98,  aic = 577.96
```

Remove the non-significant coefficient. Finally ARIMA(2,0,1)

```
arima(e923,order=c(2,0,1),include.mean=F)
Call:
arima(x = e923, order = c(2, 0, 1), include.mean = F)

Coefficients:
         ar1      ar2     ma1
      1.1846  -0.6244  0.7031
s.e.  0.0594   0.0588  0.0526

sigma^2 estimated as 0.9858:  log likelihood = -284.34,  aic = 576.68
```

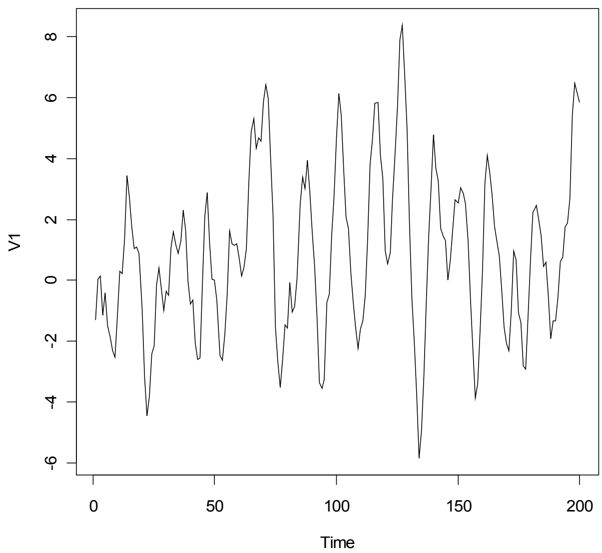
$$(1 - 1.1846B + 0.6244B^2)X_t = (1 + 0.7031B)Z_t$$

Test the residuals and the characteristic polynomial.

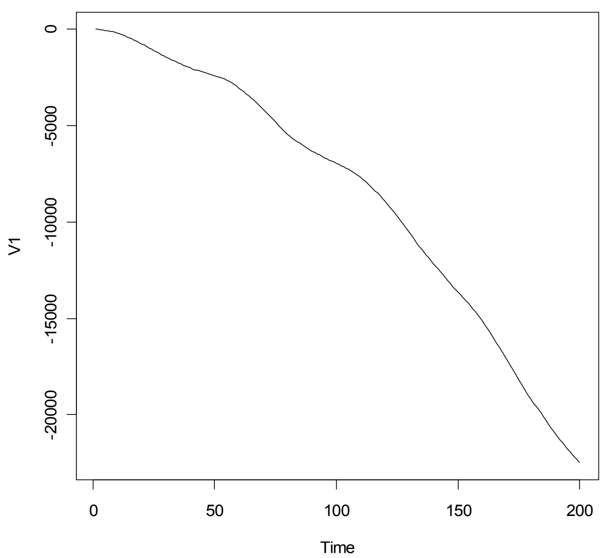
8) Exercise 1.4: Cases E9

Identify, estimate and verify the models for the series:

Case E921



Case E951

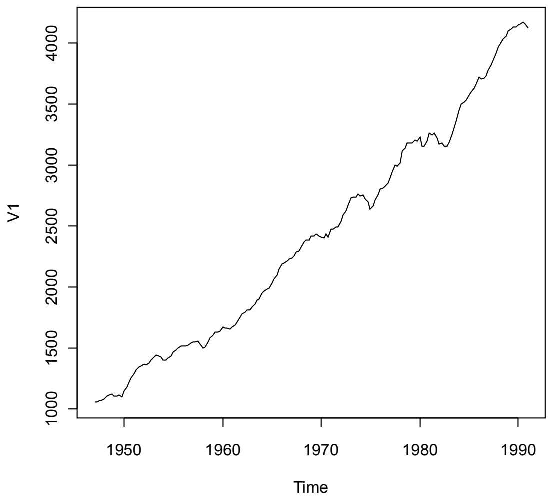


Practices: Session 2

9) Practical cases: Models Identification and estimation for *gnpsh*

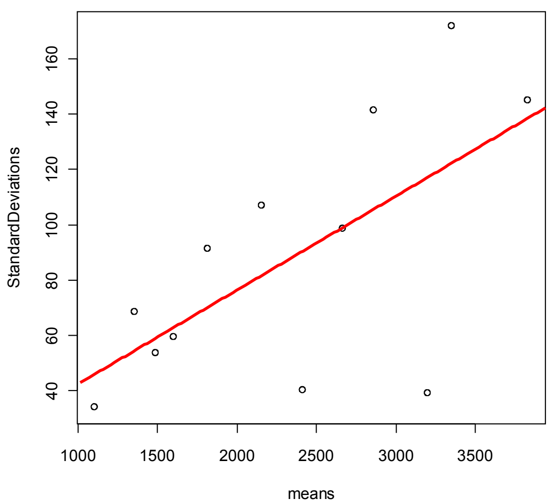
Plot the series. It is quarterly data (s=4) but it has been seasonal adjusted and so, decomposition is unnecessary. It cannot be considered stationarity because the mean is not constant, it presents a clear (approximately linear) increasing trend. As the series is an economical index, the transformation $(1-B)log(X_t)$ might be useful. It is equivalent to calculate the increments per unit between consecutive quarters. It will be analyse the possible presence of heteroscedasticity (non constant variance) using a plot of the mean-bias.

```
> plot(gnpsh)
```



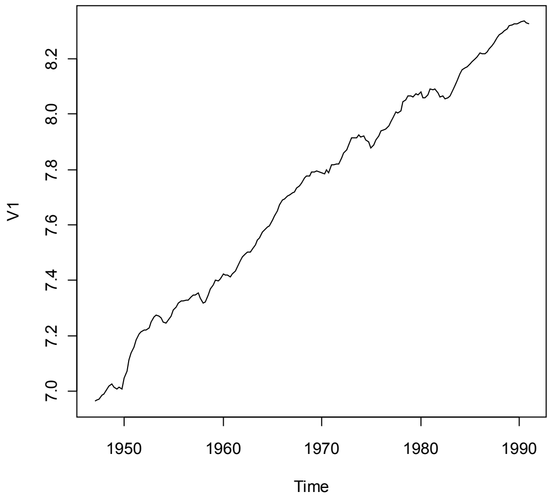
After a moving average of order 12[1], it can be appreciated that the standard deviation increases when the mean raises, which is a heteroscedasticity symptom. Hence the next step is to apply the logarithm transformation (particular case of the Box-Cox transformation for λ=0)

```
> ng=length(serie)%/%12*12
> m=apply(matrix(serie[1:ng],nrow=12),2,mean)
> s=apply(matrix(serie[1:ng],nrow=12),2,sd)
> plot(m,s,xlab="means",ylab="StandardDeviations")
> abline(lm(s~m),col=2,lty=3,lwd=3)
```



[1]The mean value of 12 consecutive observations

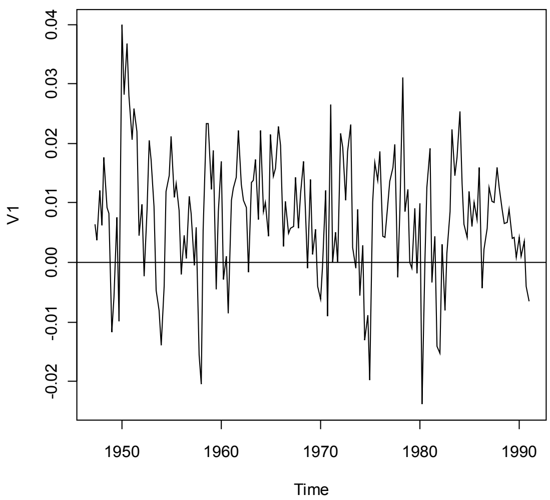
```
> summary(lm(s~m))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.03841   25.33753  -0.199  0.84571
m            0.03369    0.01026   3.283  0.00654 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 32.24 on 12 degrees of freedom
Multiple R-squared: 0.4732,     Adjusted R-squared: 0.4293
F-statistic: 10.78 on 1 and 12 DF,  p-value: 0.006542

> lngnpsh=log(gnpsh)
> plot(lngnpsh)
```




```
> var(lngnpsh)
V1 0.1595917
```

The logarithm transformation changes the scale and makes the trend more linear. A difference of order 1 will be applied to remove the trend.

```
> d1lngnpsh=diff(lngnpsh)
> plot(d1lngnpsh)
```




```
> mean(d1lngnpsh)
[1] 0.007741268
```

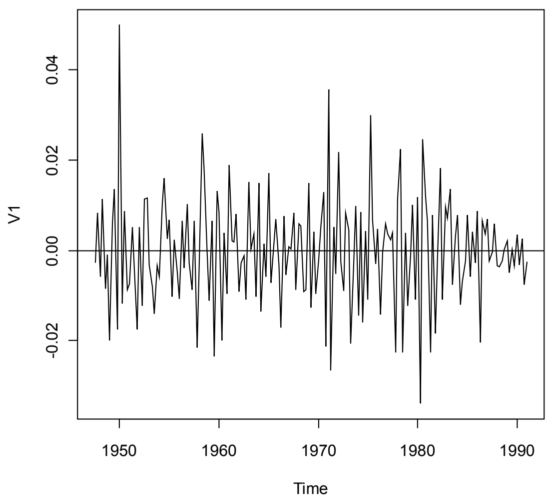
```
> var(d1lngnpsh)
             V1
V1 0.0001150796
```

After the differentiation, the trend is gone. The mean is different from zero because the difference of a series with a linear trend has the slope of the previous series as the mean.
The variance has reduced compared to the logarithms series, the reduction is of the 99.92%:

$$\frac{0.1595917 - 0.0001150796}{0.1595917} = 0.999279$$

apparently the mean and the variance are constant now and so the series can be considered stationary. Another difference would move the mean to zero. We will check if it also reduces the variance.

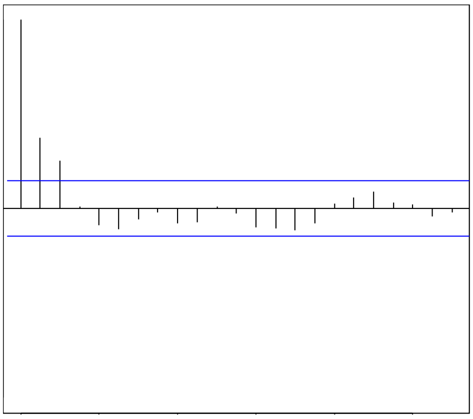
```
> d1d1lngnpsh=diff(d1lngnpsh)
> plot(d1d1lngnpsh)
```

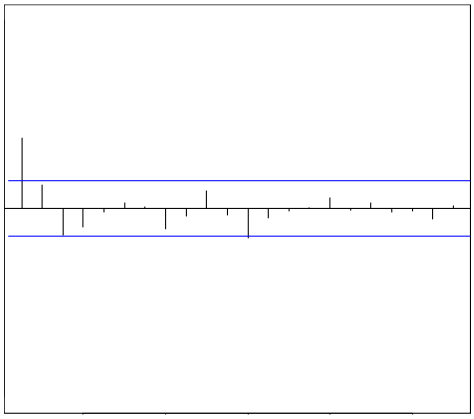



```
> var(d1d1lngnpsh)
             V1
V1 0.0001430604
```

The variance is slightly higher in this case . This indicates overdifferencing on the series. Thus the transformation of the original series that gives a stationary series may be the difference of the logarithm (this corresponds to work with the relative variation of the original rate).

```
> acf(lngnpsh,ylim=c(-1,1))
> win.graph()
> pacf(lngnpsh,ylim=c(-1,1))
```



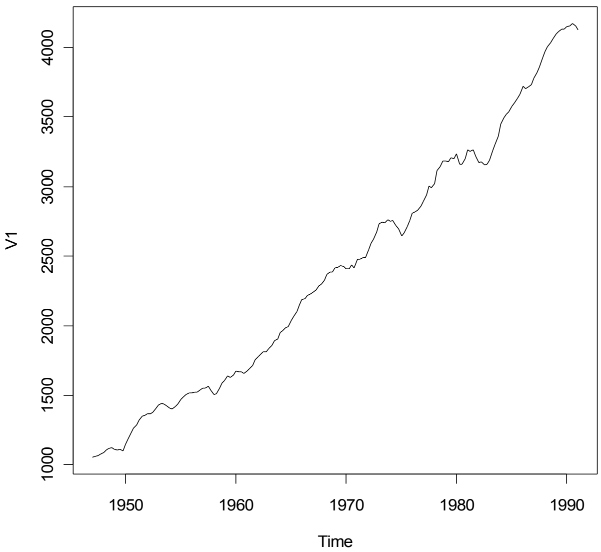


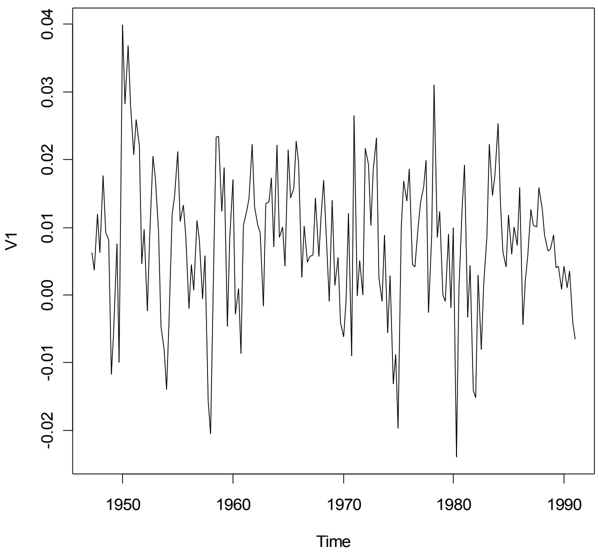
Both the ACF and the PACF decrease fast, so it can be deduced that the autocorrelation structure between the observations does not depend on the origin but only on the lag. So we must conclude that the adjusted series is second order stationary.
Now it is possible to proceed with the next step of the identification: to identify a model for this series.

Conclusion:

```
gnpsh<-ts(read.table("c:\\gnpsh.dat"),start=1947,frequency=4)
plot(lngnpsh)                              (Plot of the series)
lngnpsh<-log(gnpsh)                        (Logarithm Transformation)
d1lngnpsh<-diff(lngnpsh,lag=1)             (Differentiation of order1)
plot(d1lngnpsh)                            (Plot of the differenced series)
```

ORIGINAL SERIES (X_t) **TRANSFORMED SERIES (W_t)**



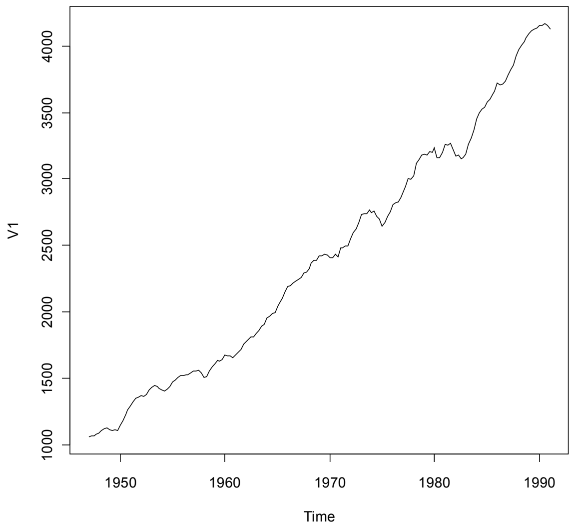


The original series corresponds to quarterly economical data and the logarithm transformation has been applied to stabilize the variance. The series presents trend but not seasonality (it was seasonal adjusted before). The series can be considered stationary after one only difference of order 1 and the variances increases after one more difference (overdifferencing) hence the series with one only difference will be the one we have to fit.

$$(1 - B)\log X_t$$

This is the series of the increments per unit which is very interesting from an econometric point of view.

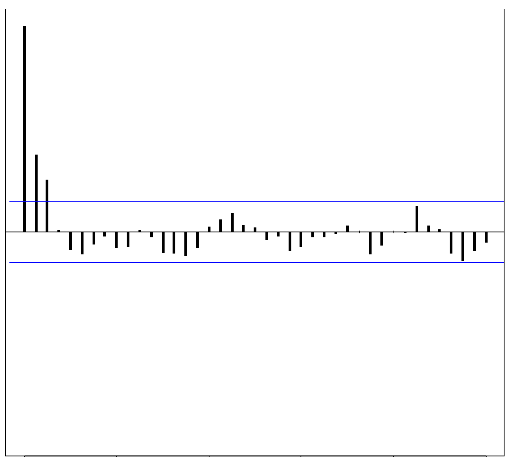
Gnpsh

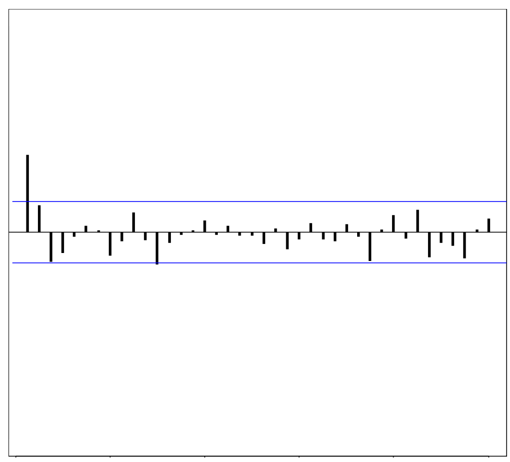


Transformations:

- Heteroscedasticity ➔ *log(X_t)*
- No seasonality, already seasonally adjusted data.
- Linear trend (non constant mean) ➔ (1-B) *log(X_t)*
- New differentiation increases the variance ➔ W_t=(1-B) *log(X_t)*

ACF and PACF de W_t:





Models identification:

✓ The last lag different from zero in the ACF is the second one and the PACF has a decreasing pattern ➔ MA(2)

✓ The last lag different from zero in the PACF is the first or the third and the ACF is decreasing ➔ AR(1) or AR(3)

✓ It could be consider that there are decreasing pattern in both plots . The simplest ARMA model has only two parameters➔ ARMA(1,1)

Possible models:

Wt ~ ARMA(0,2) ➔ p=0, q=2
Wt ~ ARMA(1,0) ➔ p=1, q=0
Wt ~ ARMA(3,0) ➔ p=3, q=0
Wt ~ ARMA(1,1) ➔ p=1, q=1

Estimation of the model for lnGnpsh

<u>Model 1</u>: *ARIMA(0,1,2)* with a constant

```
lngnp.arima <-arima(d1lngnpsh,order=c(0,0,2), include.mean = TRUE)
Call:
arima(x = d1lngnpsh, order = c(0, 0, 2), include.mean = TRUE)

Coefficients:
         ma1     ma2     intercept
       0.3121  0.2714      0.0077
 s.e.  0.0736  0.0678      0.0012

sigma^2 estimated as 9.504e-05:  log likelihood = 565.14,  aic = -1122.29
```

$$(1-B)(\log X_t - 0.0077) = (1 + 0.3121B + 0{,}2714B^2)Z_t \qquad Z_t \sim N(0, \sigma^2 = 9.504 \cdot 10^{-5})$$

<u>Model 2</u>: *ARIMA(3,1,0)* with a constant

```
lngnp.arima2 <-arima(d1lngnpsh,order=c(3,0,0), include.mean = TRUE)
Call:
arima(x = d1lngnpsh, order = c(3, 0, 0), include.mean = T)

Coefficients:
         ar1     ar2      ar3  intercept
      0.3480  0.1793  -0.1423     0.0077
s.e.  0.0745  0.0778   0.0745     0.0012

sigma^2 estimated as 9.427e-05:  log likelihood = 565.84,  aic = -1121.69
```

$$(1 - 0.3480B + 0.1793B^2 - 0.1423B^3)(1-B)(\log X_t - 0.0077) = Z_t \qquad Z_t \sim N(0, \sigma^2 = 9.427 \cdot 10^{-5})$$

Discuss the signification of the coefficients. What is the meaning of the parameter intercept?

Discuss the validity of the model using the results obtained and reproduce the validation of the model ARIMA(3,1,0). Which is the conclusion?

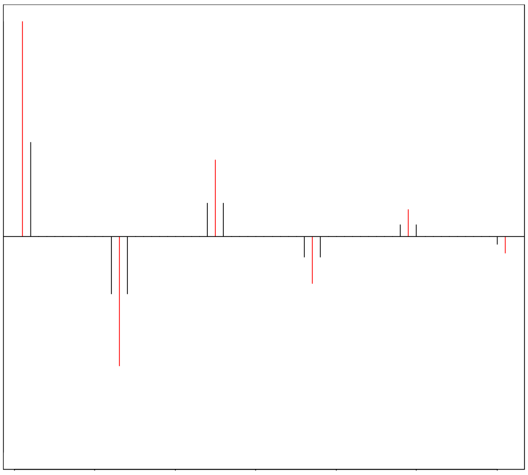
10) Seasonal Models. Representation of the theoretical ACF and PACF

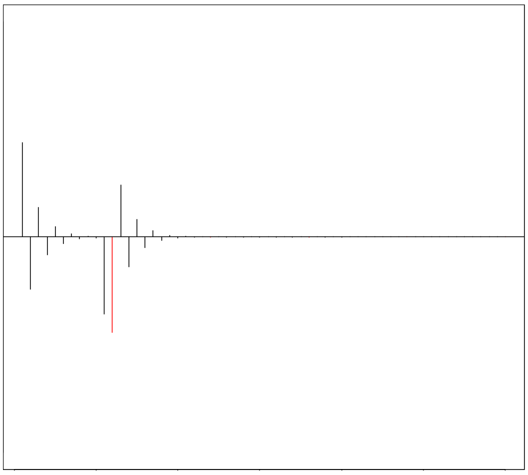
Express all the models in the ARIMA format giving the equation for the characteristic polynomial and specifying the parameters *(p,d,q)(P,D,Q)_s*

- ***MA(1)xAR(1)₁₂*** $\theta = 0.6, \Phi = -0.6$

ARIMA(0,0,1)(1,0,0)₁₂ $\qquad (1+0.6B^{12})X_t = (1+0.6B)Z_t$

```
#Parameters of the model
model.ar<-c(rep(0,11),-0.6)
model.ma<-c(0.6)

#The values of the ACF for lags until 30 are obtained
model.acf<-ARMAacf(ar=model.ar,ma=model.ma,lag.max=60)

#Plot of the value between vertical lines (type="h") and limitation of the vertical
#axis to values between -1 and 1. In addition an horizontal line has been plotted do
#as x-axis
plot(model.acf,type="h",ylim=c(-1,1),col=c(2,rep(1,11)))
abline(h=0)

#Repetition for the PACF
model.pacf<-ARMAacf(ar=model.ar,ma=model.ma,lag.max=60,pacf=T)
plot(model.pacf,type="h",ylim=c(-1,1),col=c(rep("black",11),"red"))
abline(h=0)
```





- ***MA(1)₁₂*** $\Theta = 0.6$
- ***MA(1)xMA(1)₁₂*** $\theta = 0.6, \Theta = 0.6$
- ***AR(1)₁₂*** $\Phi = 0.6$
- ***AR(1)xMA(1)₁₂*** $\phi = 0.6, \Theta = 0.6$
- ***AR(1)xAR(1)₁₂*** $\phi = 0.6, \Phi = 0.6$

11) Seasonal Models. Identification

ARMA(p,q)(P,Q)$_s$ Models

$$\phi_p(B)\Phi_P(B^s)X_t = \theta(B)_q\Theta_Q(B^s)Z_t$$

$$(1-\phi_1 B\ldots-\phi_p B^p)(1-\Phi_s B^s\ldots-\Phi_{Ps}B^{Ps})X_t = (1+\theta_1 B\ldots+\theta_q B^q)(1+\Theta_s B^s\ldots+\Theta_{Qs}B^{Qs})Z_t$$

Characteristic polynomial of the **regular part:**

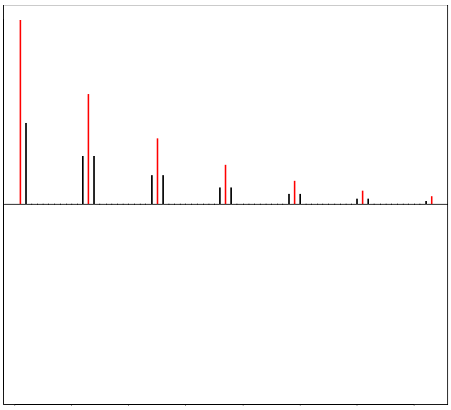
- AR (Autoregressive) Part (*p* degrees): $\phi_p(B) = (1-\phi_1 B\ldots-\phi_p B^p)$
- MA (Moving Average) Part (*q* degrees): $\theta_q(B) = (1+\theta_1 B\ldots+\theta_q B^q)$

Characteristic polynomial of the **seasonal part**

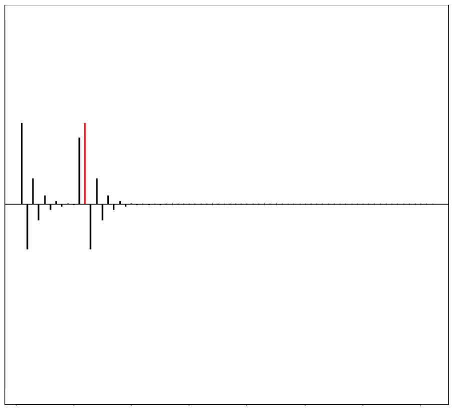
- AR (Autoregressive) Part (P degrees): $\Phi_P(B^s) = (1-\Phi_s B^s\ldots-\Phi_{Ps}B^{Ps})$
- MA (Moving Average) Part (Q degrees): $\Theta_Q(B^s) = (1+\Theta_s B^s\ldots+\Theta_{Qs}B^{Qs})$

When treating seasonality (for example, if it is monthly data and s=12), it is helpful to represent the lags that are multiple of *s* with a different colour in order to make the identification of the multiplicative model easier:

ACF



PACF

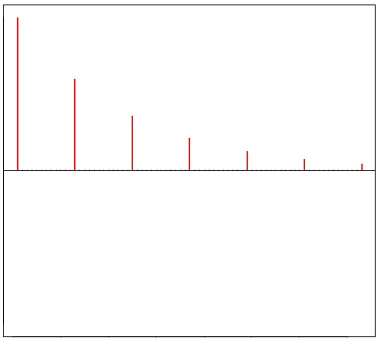


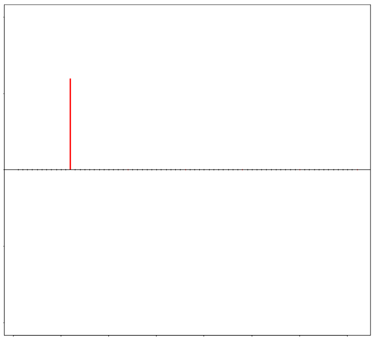
We must determine the corresponding *ARMA(p,q)(P,Q)$_s$ model*. The value of *s* is known because it depends on the seasonality period (monthly data ➔ s=12, quarterly data ➔ s=4, etc...)

Each part (regular and seasonal) will be treated separately using the same ARMA models identification criteria (identification table).

Seasonal Part: In this part we only consider the lags that are a multiple of the seasonality *s (*for example, if $s=12$ the only lags considered are 12, 24, 36, 48 and so on)

ACF **PACF**



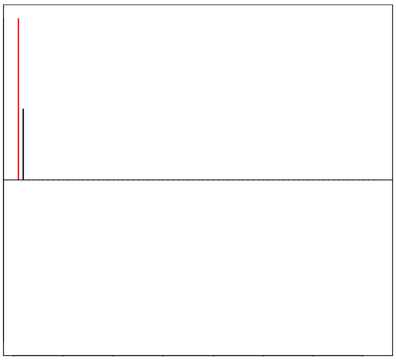


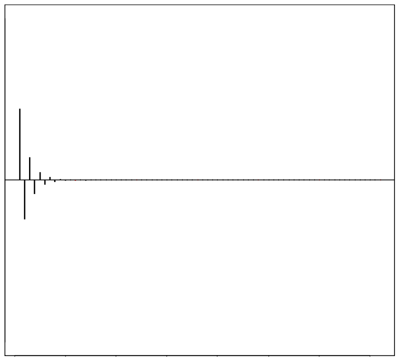
There is a clear (exponentially) decreasing pattern in the ACF and there is one only lag (the lag 12 to be more precise, corresponding to the first seasonal lag) different from zero in the PACF. This behaviour suggest an $AR(1)_{12}$ model for the seasonal part. Moreover, this exponentially decreasing shape and the sign of the 12th lag of the PACF confirm that the value of the parameter (Φ_{12}) is <u>positive</u> (see basic models table).

- Seasonal AR (Autoregressive) Part (P=1): $\Phi_P(B^{12}) = (1 - \Phi_{12} B^{12})$
- Seasonal MA (Moving Average) Part (Q=0): $\Theta_Q(B^{12}) = 1$

Regular part: Only first lags before the seasonality will be considered and the rest will be desestimated.

ACF **PACF**





There is one only lag different from zero in the ACF and alternatively exponentially decreasing behaviour on the lags of the PACF. This is the behaviour expected for a MA(1) model with <u>positive</u> θ.

- Regular AR (Autoregressive) Part (p=0): $\phi_p(B) = 1$
- Regular MA (Moving Average) Part (q=1): $\theta_q(B) = (1 + \theta B)$

If the model can only have parameters 0.6 or -0.6, the ACF and PACF correspond to the following model:

$ARMA(0,1)(1,0)_{12}$ (i.e. $MA(1)xAR(1)_{12}$)
$$(1 - 0.6B^{12})X_t = (1 + 0.6B)Z_t$$

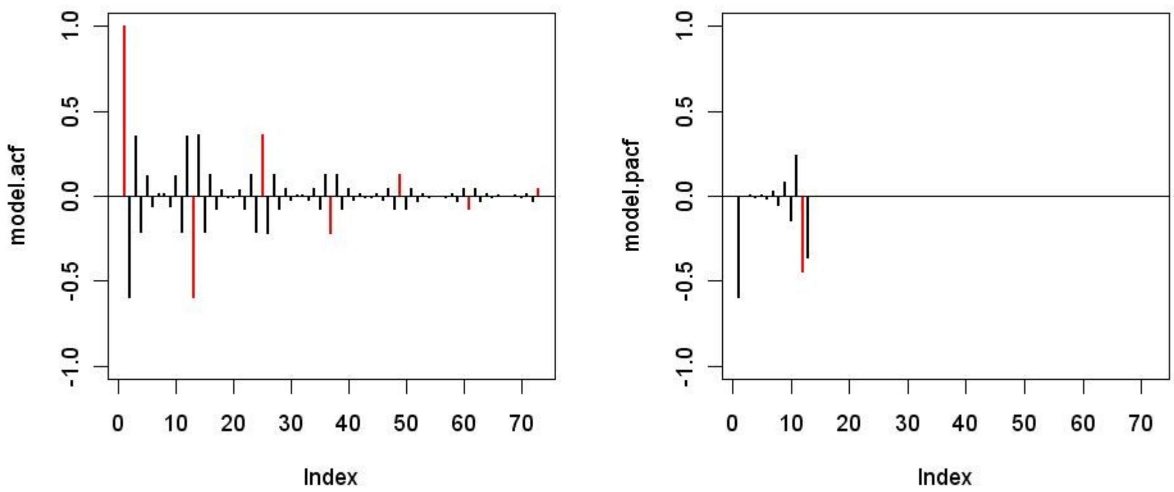
12) Exercise 2.2: Identification of Seasonal Models.

The following ACF and PACF plots correspond to *ARMA(p,q)(P,Q)₁₂* models, in the following way:

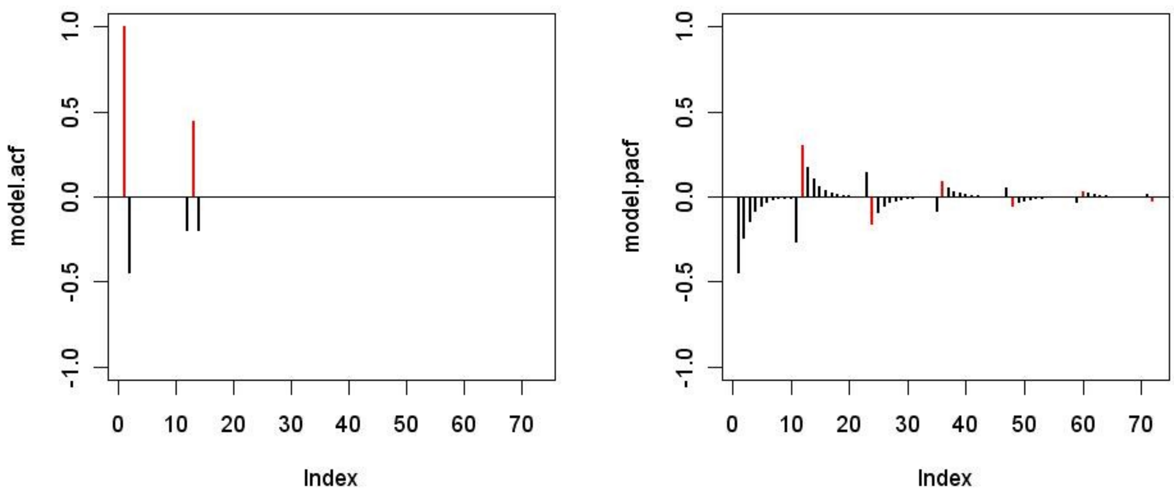
$$(1-\phi B)(1-\Phi B^{12})X_t = (1+\theta B)(1+\Theta B^{12})Z_t$$
$$\phi, \Phi, \theta, \Theta \in \{-0.6, 0, 0.6\}$$

Deduce which model generated each case and justify your answer.

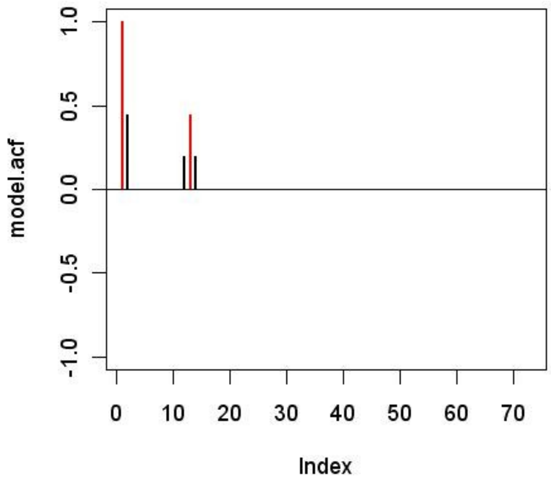
Model 1

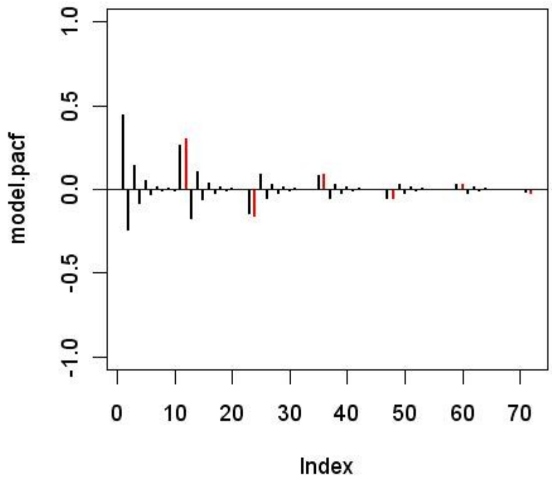


Model 2

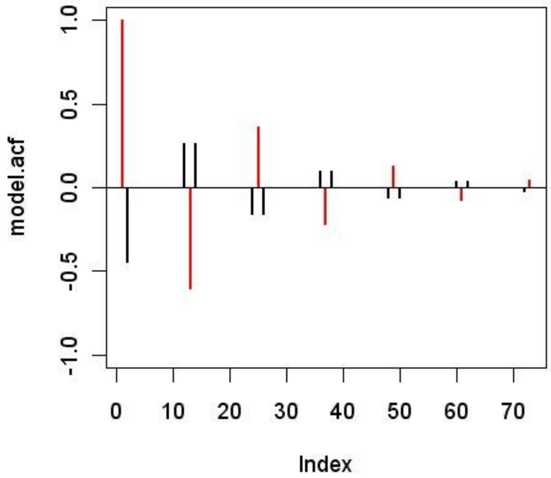


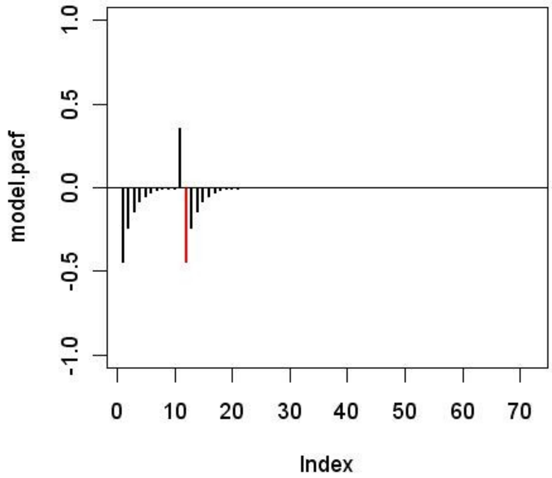
Model 3



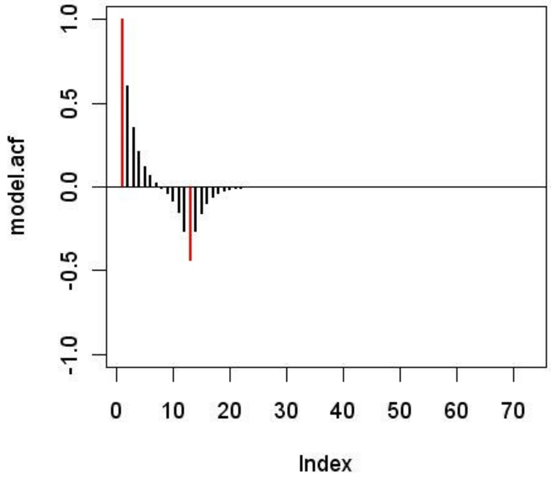


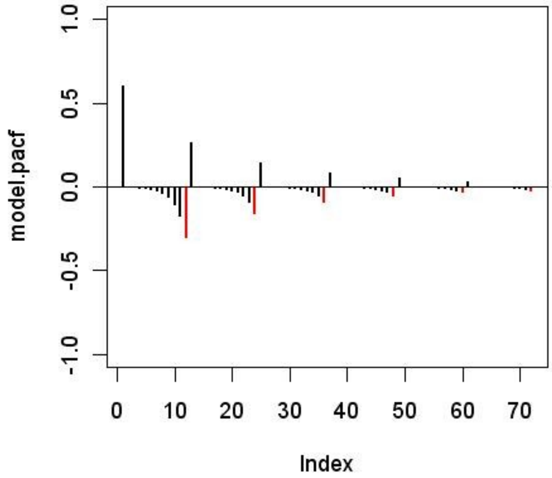
Model4



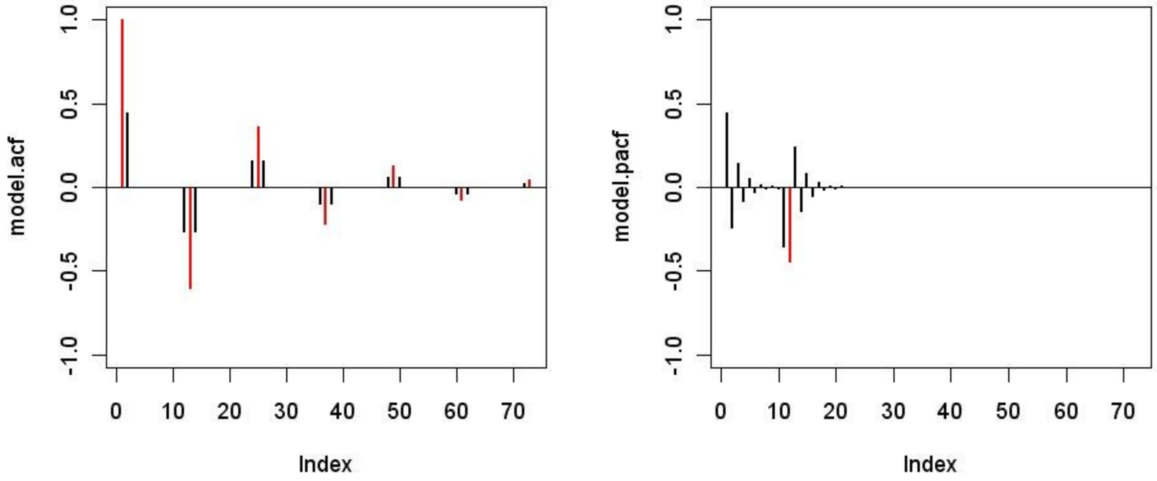


Model 5

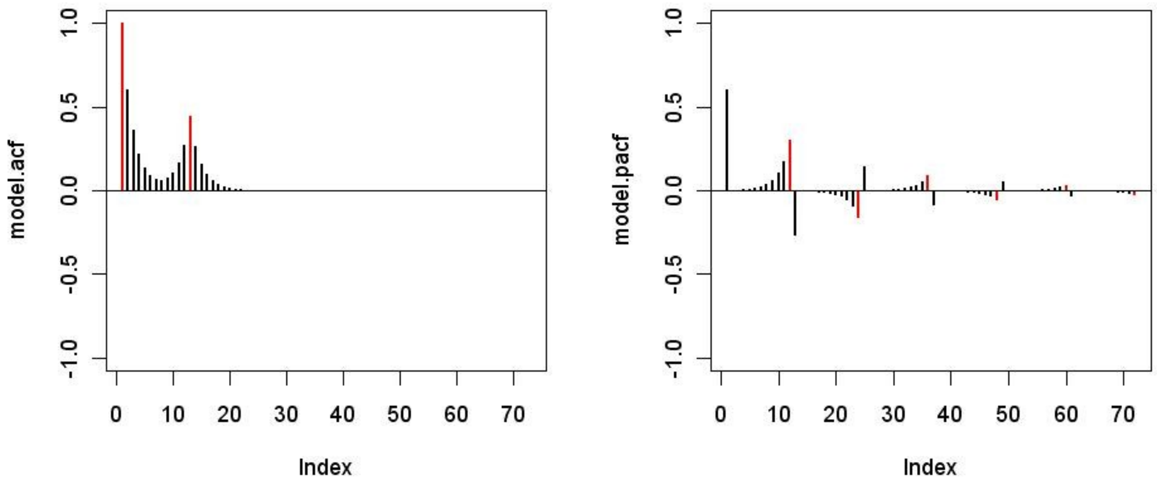




Model 6



Model 7



Solution 2.2: Identification of seasonal models

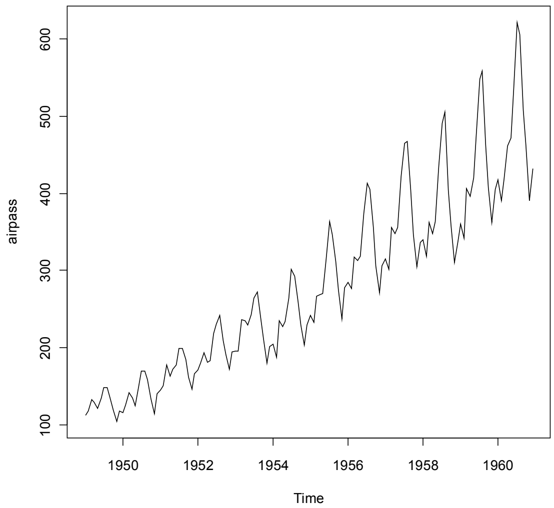
Model	Regular Part		Seasonal Part		Model
Model 1	AR(1)	$\phi<0$	AR(1)	$\phi<0$	$(1+0.6B)(1+0.6B^{12})X_t=Z_t$
Model 2	MA(1)	$\theta<0$	MA(1)	$\theta>0$	$X_t=(1-0.6B)(1+0.6B^{12})Z_t$
Model 3	MA(1)	$\theta>0$	MA(1)	$\theta>0$	$X_t=(1+0.6B)(1+0.6B^{12})Z_t$
Model 4	MA(1)	$\theta<0$	AR(1)	$\phi<0$	$(1+0.6B^{12})X_t=(1-0.6B)\ Z_t$
Model 5	AR(1)	$\phi>0$	MA(1)	$\theta<0$	$(1-0.6B)X_t=(1-0.6B^{12})Z_t$
Model 6	MA(1)	$\theta>0$	AR(1)	$\phi<0$	$(1+0.6B^{12})X_t=(1+0.6B)Z_t$
Model 7	AR(1)	$\phi>0$	MA(1)	$\theta>0$	$(1-0.6B)X_t=(1+0.6B^{12})Z_t$

Model	R Code
Model 1	`model.ar = c(-0.6,rep(0,10),-0.6,-0.36)` `model.ma = NULL`
Model 2	`model.ar = NULL` `model.ma = c(-0.6,rep(0,10),0.6,-0.36)`
Model 3	`model.ar = NULL` `model.ma = c(0.6,rep(0,10),0.6,0.36)`
Model 4	`model.ar = c(rep(0,11),-0.6)` `model.ma = -0.6`
Model 5	`model.ar = 0.6` `model.ma = c(rep(0,11),-0.6)`
Model 6	`model.ar = c(rep(0,11),-0.6)` `model.ma = 0.6`
Model 7	`model.ar = 0.6` `model.ma = c(rep(0,11),0.6)`

```
#Plots of the ACF and ACF of each model (72 lags and degree of seasonality s=12)

model.acf<-ARMAacf(ar=model.ar,ma=model.ma,lag.max=72)
plot(model.acf,type="h",ylim=c(-1,1),col=c(2,rep(1,11)),lwd=2)
abline(h=0)

win.graph()

model.pacf<-ARMAacf(ar=model.ar,ma=model.ma,lag.max=72,pacf=T)
plot(model.pacf,type="h",ylim=c(-1,1),col=c(rep(1,11),2),lwd=2)
abline(h=0)
```

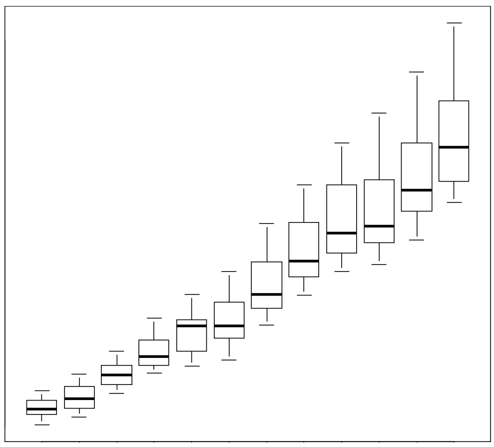
13) Practical cases: Models Identification and Estimation for *AirPassengers*

```
> airpass=AirPassengers
> plot(airpass)
```



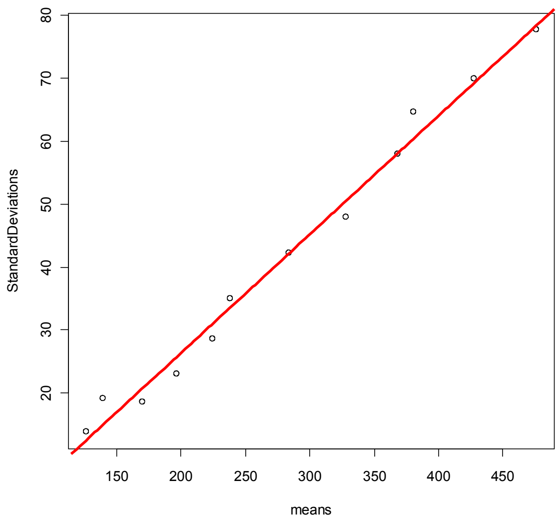
The plot of the series shows a clear increase of the variance with the level of the series. It can be checked with some other graphic representations.

```
> plot(c(serie)~gl(length(serie)/12,12))
```



Doing a Box-plot with each year, it is easy to appreciate how the interquartile rang raises. It is represented by the boxes on the Box-plot. So it is advisable to take logarithms on the series to remove the heteroscedasticity. The mean-variance plot also confirms this hypothesis.

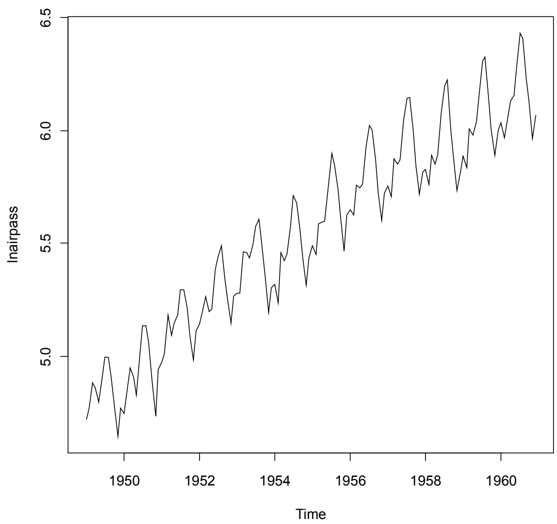
```
> ng=length(airpass)%/%12*12
> m=apply(matrix(airpass[1:ng],nrow=12),2,mean)
> s=apply(matrix(airpass[1:ng],nrow=12),2,sd)
> plot(m,s,xlab="means",ylab="StandardDeviations")
> abline(lm(s~m),col=2,lty=3,lwd=3)
```

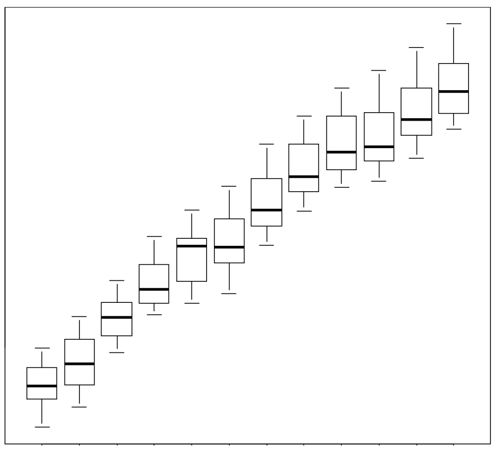



```
> summary(lm(s~m))
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.403254   1.982789  -5.751 0.000185 ***
m             0.188613   0.006577  28.676 6.19e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
Residual standard error: 2.528 on 10 degrees of freedom
Multiple R-squared: 0.988,    Adjusted R-squared: 0.9868
F-statistic: 822.3 on 1 and 10 DF,  p-value: 6.192e-11
```

After applying the logarithm transformation, the variance is more homogeneous.

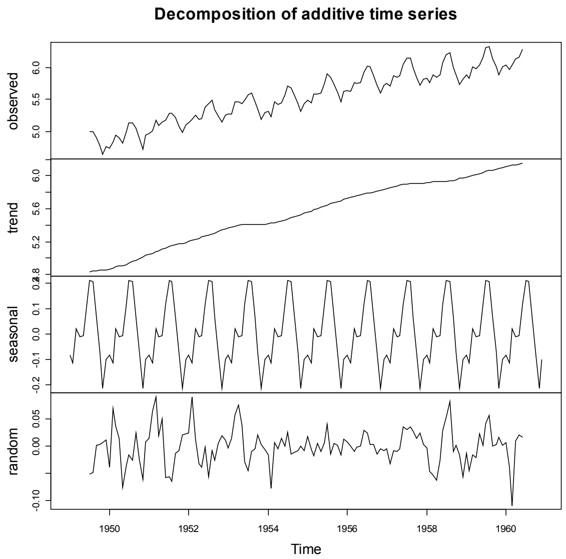
```
> lnairpass=log(airpass)
> plot(lnairpass)
> plot(c(lnairpass)~gl(length(lnairpass)/12,12))
> var(lnairpass)
[1] 0.1948838
```

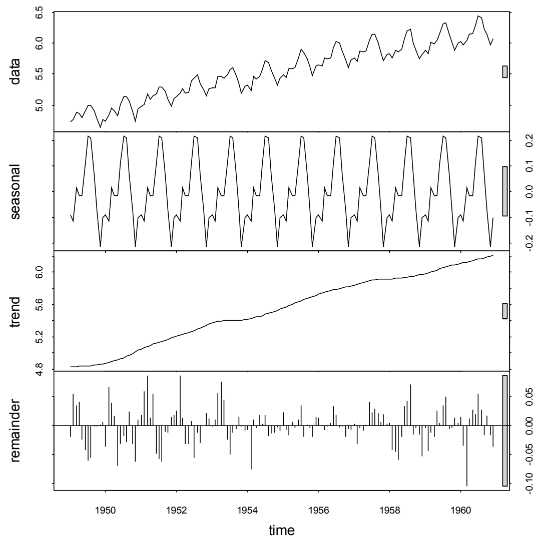




Now we are going to analyse the seasonality of the series which is clearly shown in the first plot. There are two instructions in R to decompose the series. In the first case the series X_t is separated in the trend component T_t (calculated with centred moving average of order 12), a seasonal component S_t (the rate for each month is estimated with the values of the detrended series X_t-T_t) and random noises (X_t-T_t-S_t).
The other option is equivalent if we indicate in the windows for the seasonal calculus that it is periodic (that means that the seasonal component is constant). This function allow to decompose the series with a variable seasonal part.

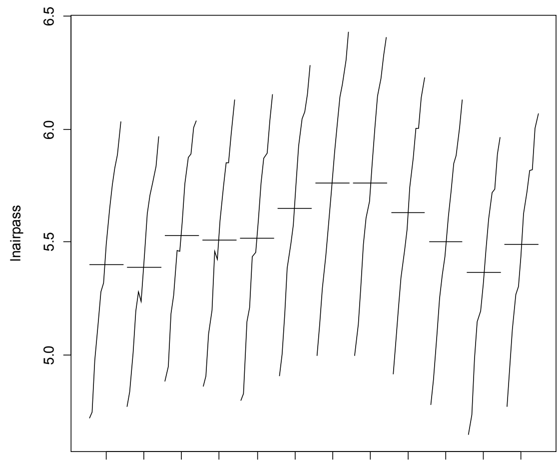
```
> plot(decompose(lnairpass))
> plot(stl(lnairpass,s.window="periodic"))
```





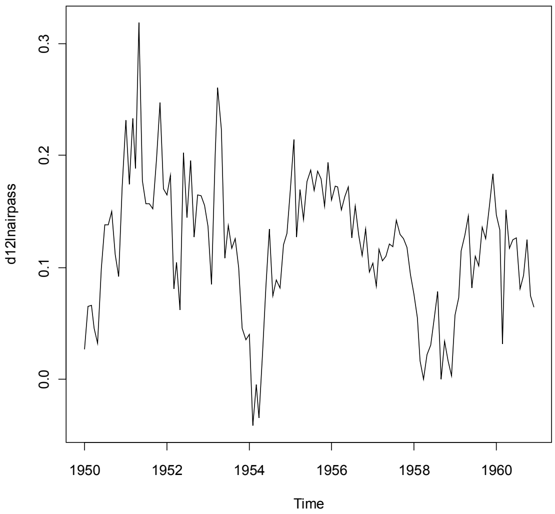
It is also possible to plot evolution for each month separately indicating the average value that also confirms that there is a seasonal pattern in the series.

```
> monthplot(lnairpass)
```



Thus it is necessary to difference the series with lag 12 to remove the seasonal component. This transformation implies a reduction of the variance.

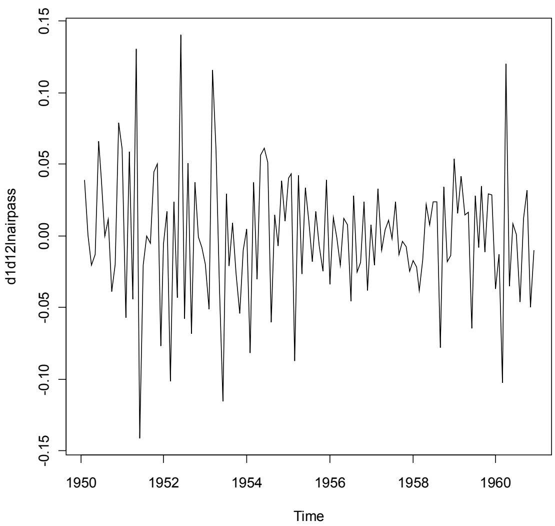
```
> d12lnairpass=diff(lnairpass,lag=12)
> plot(d12lnairpass)
```




```
> var(d12lnairpass)
[1] 0.003800061
```

The mean of the differenced series does not look constant, it is apparently a strong autoregressive series),so it is advisable to difference with order 1. The variance decreases again with this transformation so the differentiation is appropriate.

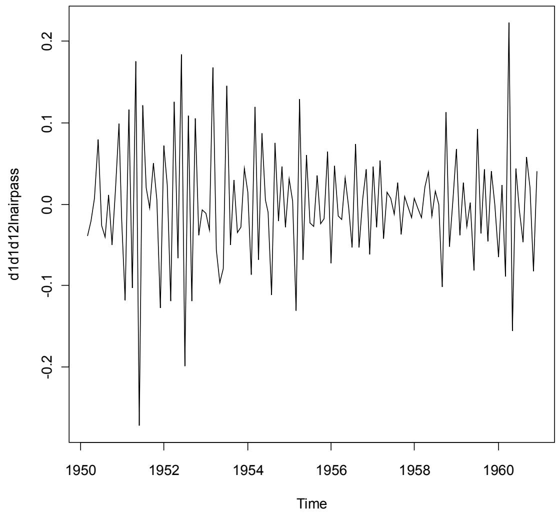
```
> d1d12lnairpass=diff(d12lnairpass)
> plot(d1d12lnairpass)
```




```
> var(d1d12lnairpass)
[1] 0.002102066
```

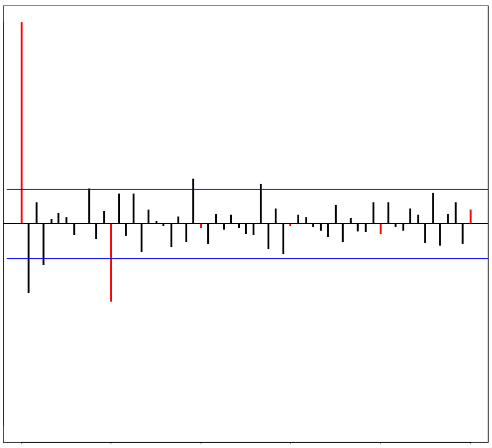
Even though the series obtained after these two transformations (seasonal and regular) on the logarithm of the original series is apparently stationary, this must be checked by applying another regular differentiation and calculating how it affects to the variance. In this case, after one more differentiation, we obtain a new stationary series with larger variances.

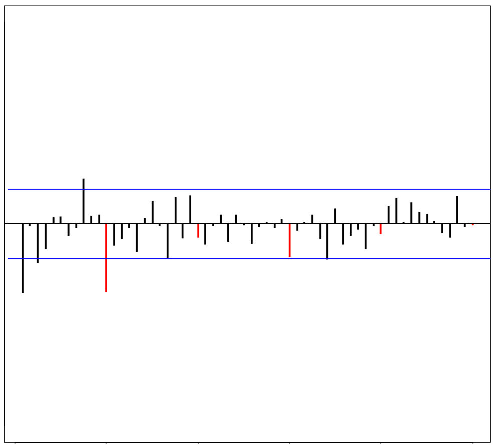
```
> d1d1d12lnairpass=diff(d1d12lnairpass)
> plot(d1d1d12lnairpass)
```




```
> var(d1d1d12lnairpass)
[1] 0.005669296
```

```
>acf(d1d12lnairpass,ylim=c(-1,1),lag.max=60, col=c(2,rep(1,11)),lwd=2)
>win.graph()
>pacf(d1d12lnairpass,ylim=c(1,1),lag.max=60, col=c(rep(1,11),2),lwd=2)
```

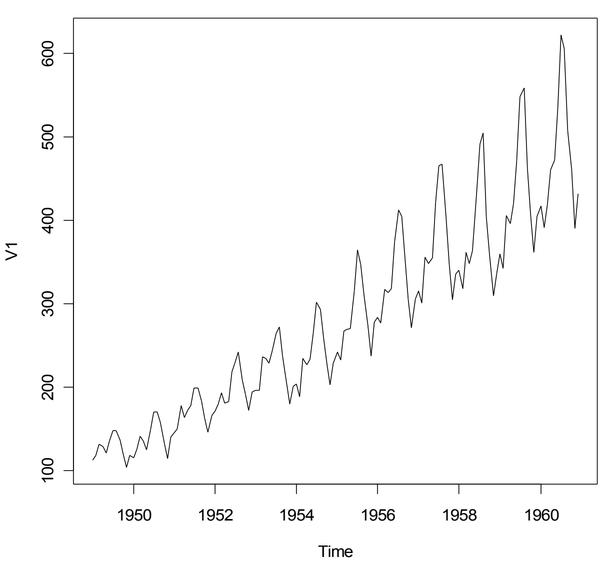




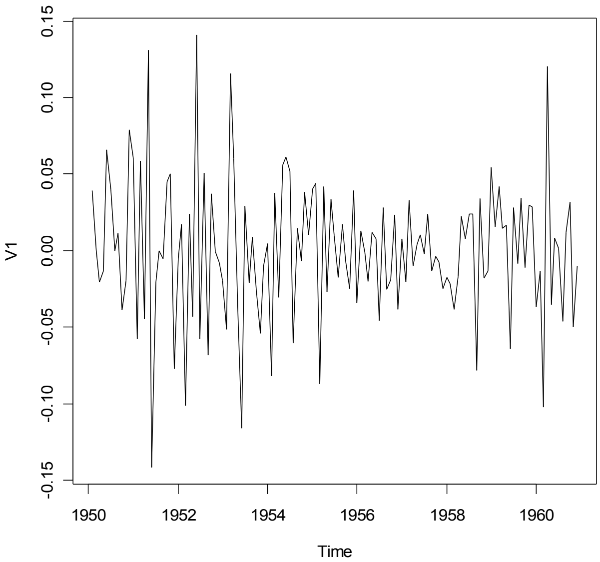
Conclusion

```
lnairpass<-log(AirPassengers)                    (logarithm transformation)
d12lnairpass<-diff(lnairpass,lag=12)             (differentiation with lag 12)
d1d12lnairpass<-diff(d12lnairpass)               (differentiation with lag 1)
```

SÈRIE ORIGINAL (X_t)



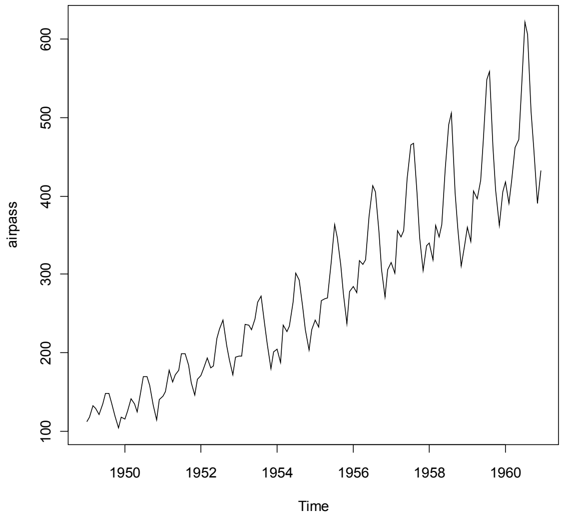
SÈRIE TRANSFORMADA (W_t)



The original series corresponds to monthly data and a logarithm transformation has been applied to stabilize the variance. The series presents trend and seasonality (s=12). We applied a difference with lag 12 but it was not enough because the resulting series did not have yet a constant mean so it was not stationary. With one more differentiation of order 1 it can be considered stationary so we will apply the identification phase to this last adjusted series.

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

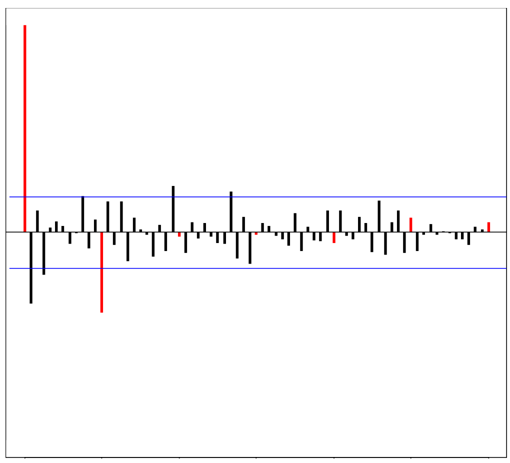
Airpass



Transformations:

- Heteroscedasticity ➔ $log(X_t)$
- Stationarity ➔ $(1-B^{12})\ log(X_t)$
- Non-constant mean ➔ $(1-B)(1-B^{12})\ log(X_t)$
- New differentiation increases the variance ➔ $W_t=(1-B)(1-B^{12})\ log(X_t)$

ACF and PACF de W_t:



Identification of the Models:

- Seasonal Part:
  One only significant lag multiple of 12 in the ACF and exponential decrease in the lags 12, 24, 36, 48, ... of the PACF ➔ $MA(1)_{12}$
- Regular Part:
  ✓ First lag (or until the third) of the ACF different from zero and the PACF has an exponentially decreasing pattern in the first lags ➔ MA(1) or MA(3)
  ✓ First lag (or until the third) of the PACF different from zero and the ACF has an alternatively exponentially decreasing pattern in the first lags ➔ AR(1) or AR(3)

Possible Models:

Wt ~ $ARMA(0,1)(0,1)_{12}$ ➔ p=0, q=1, P=0, Q=1
Wt ~ $ARMA(1,0)(0,1)_{12}$ ➔ p=1, q=0, P=0, Q=1
Wt ~ $ARMA(0,3)(0,1)_{12}$ ➔ p=0, q=3, P=0, Q=1
Wt ~ $ARMA(3,0)(0,1)_{12}$ ➔ p=3, q=0, P=0, Q=1

Estimation of a Model for lnAirpass

<u>Model 1</u>: *ARIMA(3,1,0)(0,1,1)₁₂*

```
airpass.arima1<-arima(lnairpass,order=c(3,1,0),
seasonal=list(order=c(0,1,1),period=12))

Call:
arima(x = lnairpass, order = c(3, 1, 0), seasonal = list(order = c(0, 1, 1),
    period = 12))

Coefficients:
          ar1      ar2      ar3     sma1
      -0.3698  -0.1026  -0.1071  -0.5492
s.e.   0.0872   0.0922   0.0873   0.0765

sigma^2 estimated as 0.001348:  log likelihood = 244.76,  aic = -479.52
```

$$(1+0.37B+0.103B^2+0.107B^3)(1-B)(1-B^{12})\log X_t=(1-0.549B^{12})Z_t \quad Z_t\sim N(0,\sigma^2=0.001348)$$

<u>Model 2</u>: *ARIMA(1,1,0)(0,1,1)₁₂*

```
airpass.arima2<-arima(lnairpass,order=c(1,1,0),
seasonal=list(order=c(0,1,1),period=12))

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

Coefficients:
          ar1     sma1
      -0.3395  -0.5619
s.e.   0.0822   0.0748

sigma^2 estimated as 0.001367:  log likelihood = 243.74,  aic = -481.49
```

$$(1+0.3395B)(1-B)(1-B^{12})\log X_t=(1-0.5619B^{12})Z_t \qquad Z_t\sim N(0,\sigma^2=0.001367)$$

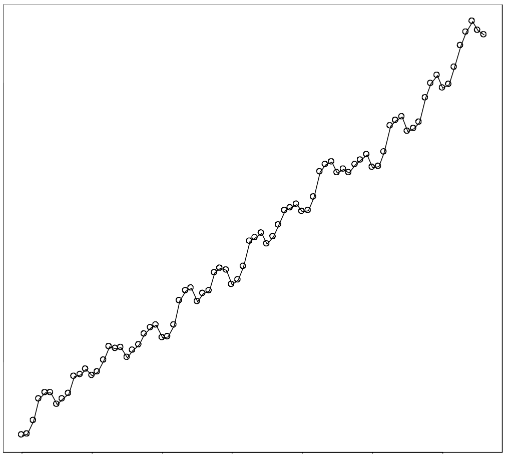
<u>Model 3</u>: *ARIMA(0,1,1)(0,1,1)₁₂*

```
airpass.arima3<-arima(lnairpass,order=c(0,1,1),
seasonal=list(order=c(0,1,1),period=12))

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

Coefficients:
          ma1     sma1
      -0.4018  -0.5569
s.e.   0.0896   0.0731

sigma^2 estimated as 0.001348:  log likelihood = 244.7,  aic = -483.4
```

$$(1-B)(1-B^{12})\log X_t=(1-0.4018B)(1-0.5569B^{12})Z_t \qquad Z_t\sim N(0,\sigma^2=0.001348)$$

14) Exercise 2.3: Identification of the proposed series

IPCsp



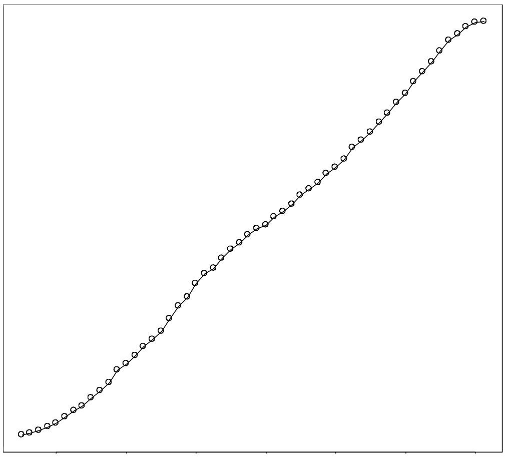
Transformations:

ACF and PACF of W_t:

Identification of the Model:

Possible models:

PIBsp



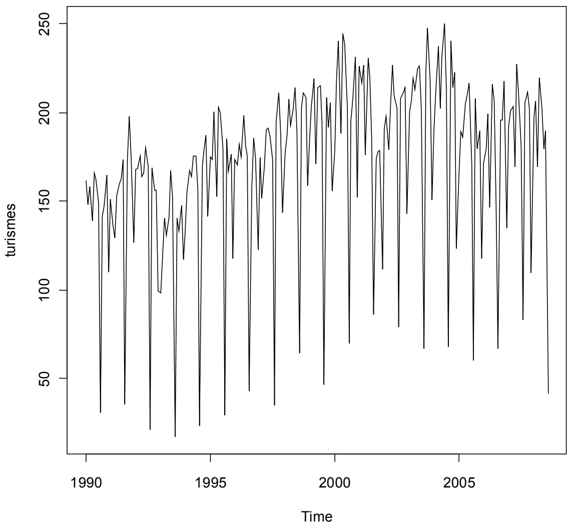
Transformations:

ACF and PACF of W_t:

Identification of the Model:

Possible models:

Cars



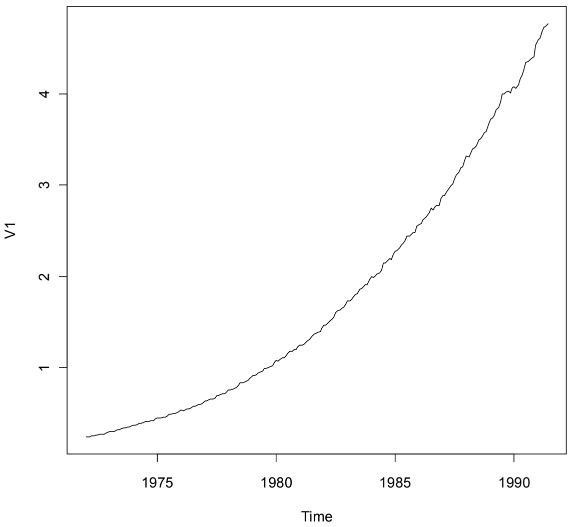
Transformations:

ACF and PACF of W_t:

Identification of the Model:

Possibles model:

ALP



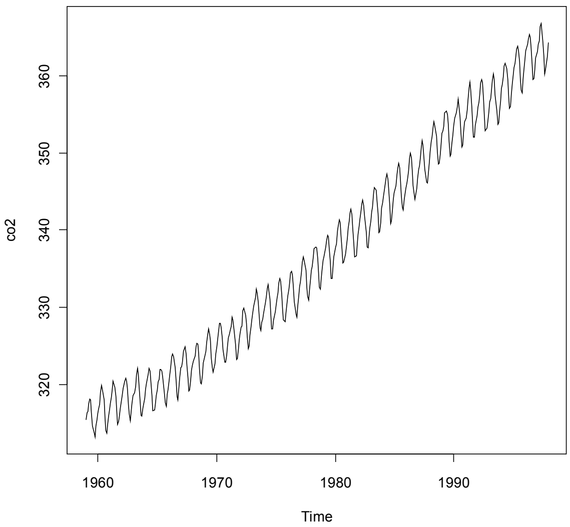
Transformations:

ACF and PACF of W_t:

Identification of the Model:

Possibles model:

CO2



Transformations:

ACF and PACF of W_t:

Identification of the Model:

Possible models:

Practices: Session 3

15) Practical cases: Estimation and Validation for *gnpsh*

The aim is to discuss the choice of the best model among the proposed in the identification phase, analysing the following features:

Residuals:

Verify if they behave like a normal, if they have constant variance and if there is independence between them

Determine the possible presence of outliers.

Adequacy Measures:

AIC (Akaike Information Criteria) AIC$=-2logLik+2(n.param+1)$

BIC (Bayesian Information Criteria)

Estimation of the residual Variance

Correction of the model:

Compare the theoretical and sample ACF and PACF

Comparison with other models:

Infinite order AR/MA expression (weights ψ and π)

Analysis of the forecasts:

Verify long term forecasts of the model without some of the last observations

Do the same discussion for the step by step forecasts

Stability of the model

Re-estimation of the model after having removed some observation to test the effect on the parameters.

SERIES: GNPSH...

MODEL 1

Model: *ln(X_t)~ARIMA(0,1,2) no zero mean (without outliers)*
Formulation: $W_t=(1-B)\ln(X_t)$ $W_t - \mu = (1+\theta_1 B+\theta_2 B^2)\, Z_t$
Amount of parameters: *3*
Software used: *R*
Instructions:

```
d1lngnpsh<-diff(log(gnpsh),lag=1)
gnp.arima1<-arima(d1lngnpsh,order=c(0,0,2),include.mean=T)
```

Regular Component	Value	St.Err.	|T-ratio|	Seasonal Component	Value	St.Err.	|T-ratio|
AR ϕ				AR Φ			
MA θ	*0.3121* *0.2714*	*0.0736* *0.0678*	*4.24* *4.00*	MA Θ			
Mean μ	*0.0077*	*0.0012*	*6.41*	Residual Var. σ_z^2	*9.504e-05*		

Estimated Model:

$$(1-B)\ln(X_t)=0.0077+(1+0.3121B+0.2714B^2)Z_t$$

AIC	*-1122.29*	BIC	

Observations:

MODEL 2

Model: *ln(X_t)~ARIMA(3,1,0) no zero mean (without outliers)*
Formulation: $W_t=(1-B)\ln(X_t)$ $(1+\phi_1 B+\phi_2 B^2+\phi_3 B^3)(W_t - \mu) = Z_t$
Amount of parameters: *4*
Software used: *R*
Instructions:

```
d1lngnpsh<-diff(log(gnpsh),lag=1)
gnp.arima1<-arima(d1lngnpsh,order=c(3,0,0),include.mean=T)
```

Regular Component	Value	St.Err.	|T-ratio|	Seasonal Component	Value	St.Err.	|T-ratio|
AR ϕ	*0.3480* *0.1793* *-0.1423*	*0.0745* *0.0778* *0.0745*	*4.67* *2.30* *-1.91*	AR Φ			
MA θ				MA Θ			
Mean μ	*0.0077*	*0.0012*	*6.41*	Residual Var. σ_z^2	*9.427e-05*		

Estimated Model:

$$(1-0.348B-0.1793B^2+0.1423B^3)(W_t - 0.0077) = Z_t$$

AIC	*-1121.69*	BIC	

Observations:
The coefficient phi_3 has a t-ratio barely significant

MODEL 3

Proposed Model: ***ln(Xt)~ARIMA(1,1,2) no zero mean (without outliers)***
Formulation: $W_t=(1-B)\ln(X_t)$ $(1+\phi_1 B)(W_t-\mu)=(1+\theta_1 B+\theta_2 B^2)\,Z_t$
Amount of parameters: ***4***
Software used: ***R***
Instructions:

```
d1lngnpsh<-diff(log(gnpsh),lag=1)
gnp.arima1<-arima(d1lngnpsh,order=c(1,0,2),include.mean=T)
```

Regular Component	Value	St.Err.	|T-ratio|	Seasonal Component	Value	St.Err.	|T-ratio|
AR ϕ	***0.2671***	***0.1810***	***1.47***	AR Φ			
MA θ	***0.0660*** ***0.2362***	***0.1749*** ***0.0864***	***0.38*** ***2.79***	MA Θ			
Mean μ	***0.0077***	***0.0013***	***5.92***	Residual Var. σ_z^2	***9.42e-05***		

Estimated Model:

$$(1-0.2671B)(W_t-0.0077)=(1+0.0660B+0.2362B^2)Z_t$$

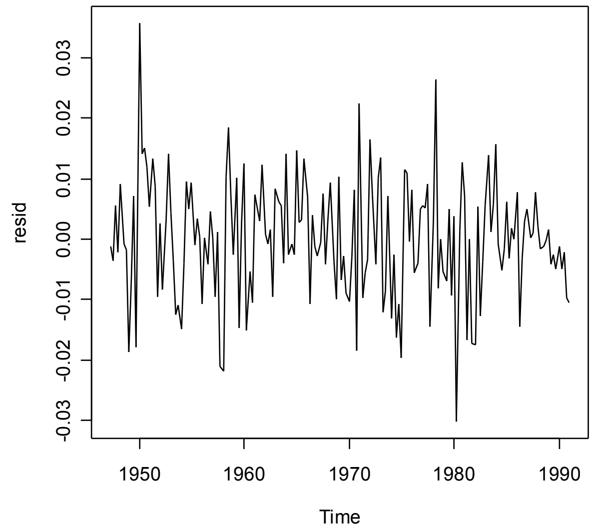
AIC	*-1121.8*	BIC	

Observations:
Two of the parameters are not significant

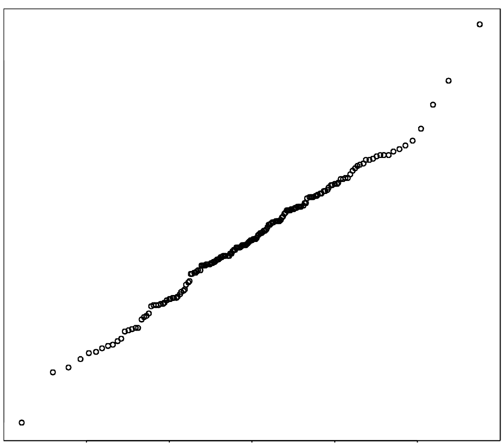
Gnpsh Case

Proposed Model: $ln(X_t)$~$ARIMA(0,1,2)$ $\quad W_t=(1-B)ln(X_t) \quad W_t-0.0077=(1+0.3121B+0.2714B^2)Z_t$

a) Analysis of the residuals: Variance (possibility of heteroscedasticity and outliers)

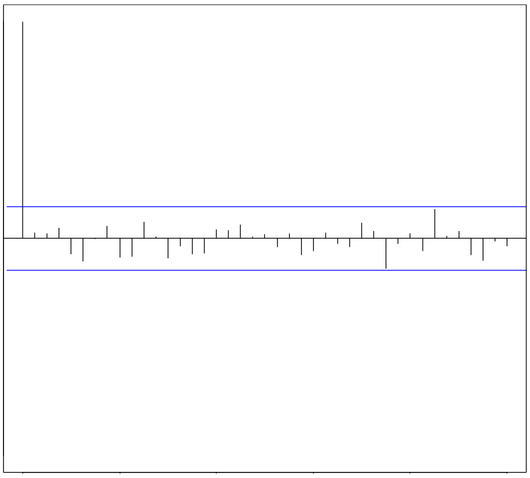
`plot(gnp.arima1$residuals)`

Interpretation: The residuals look quite homogeneous except at the end where the variance decreases. Maybe some of the last residues correspond to outliers.

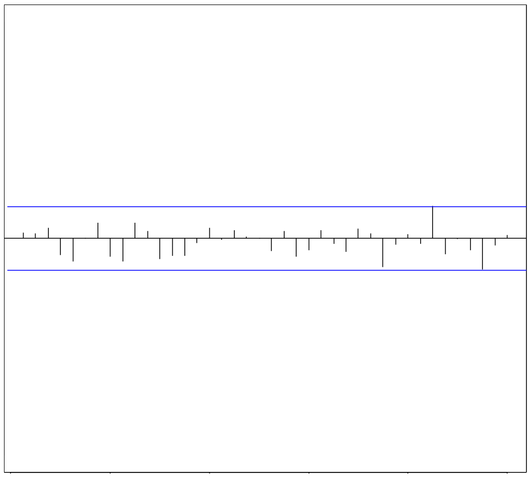
b) Analysis of the residuals: Normality

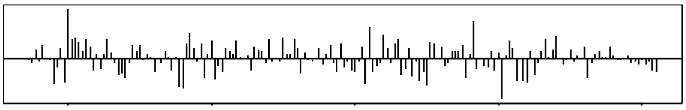
`qqnorm(gnp.arima1$residuals)`

Interpretation: The normality plot shows a quite normal behaviour of the data except at the extremes what suggest again the possible presence of outliers.

c) Analysis of the residuals: Independence

```
acf(gnp.arima1$residuals,ylim=c(-1,1),lag.max=40)
pacf(gnp.arima1$residuals,ylim=c(-1,1),lag.max=40)
tsdiag(gnp.arima1)
```



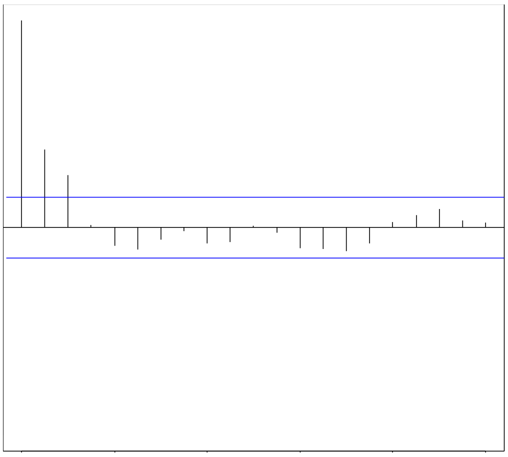


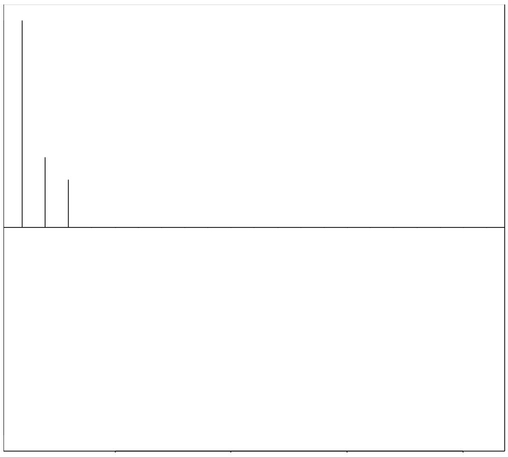


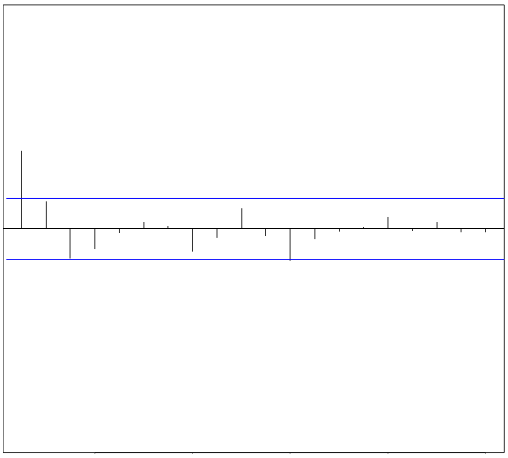
Interpretation:
None of the lags of the ACF or the PACF is significant thus the residuals are independent. The p-values of the Ljung-Box statistics show that the ACF is compatible with a white noise .

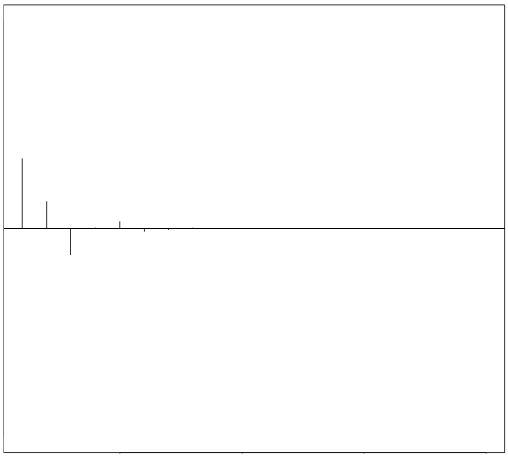
d) Correspondence between the estimated model and the sample ACF and PACF

```
dades<-d1lngnpsh
model<-gnp.arima1$model

acf(dades, ylim=c(-1,1) ,lag.max=20 ,main="Mostral")
plot(ARMAacf(model$phi,model$theta,lag.max=20),ylim=c(-1,1), type="h”)
abline(h=0)

pacf(dades,ylim=c(-1,1) ,lag.max=20, main="Mostral")
plot(ARMAacf(model$phi,model$theta,lag.max=20,pacf=T),ylim=c(-1,1),type="h")
abline(h=0)
```









Interpretation

The sample ACF and PACF are very similar to the theoretical ones. One only lag (12) is significant in the sample PACF but not in the model.

e) Expression of the model as an infinite AR/MA

```
ARMAtoMA(ar=model$phi,ma=model$theta,lag.max=20)

 [1] 0.312100275 0.271432116 0.000000000 0.000000000 0.000000000 0.000000000
 [7] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
[13] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
[19] 0.000000000 0.000000000
```

As a MA(∞)

$$W_t = Z_t + 0.3121 Z_{t-1} + 0.2714 Z_{t-2}$$

```
-ARMAtoMA(ar=-model$theta,ma=-model$phi,lag.max=20)

 [1]  3.121003e-01  1.740255e-01 -1.390275e-01 -3.845612e-03  3.893673e-02
 [6] -1.110834e-02 -7.101763e-03  5.231623e-03  2.948556e-04 -1.512055e-03
[11]  3.918795e-04  2.881146e-04 -1.962893e-04 -1.694160e-05  5.856671e-05
[16] -1.368019e-05 -1.162729e-05  7.342125e-06  8.645418e-07 -2.262712e-06
```

As an AR(∞)

$$W_t = 0.3121 W_{t-1} + 0.1740 W_{t-2} - 0.13908 W_{t-3} - 0.0038 W_{t-4} + 0.0389 W_{t-5} + \ldots + Z_t$$

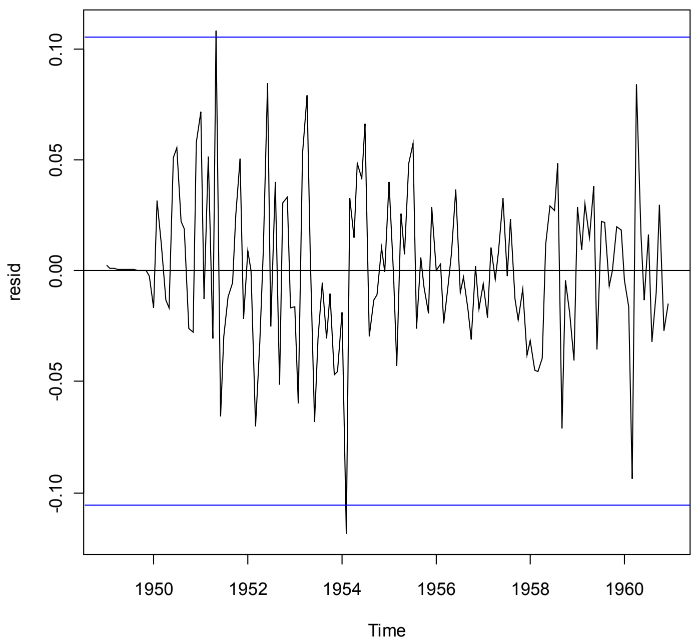
Conclusions:

The hypothesis seem adequate but it should be considered to study the outliers.

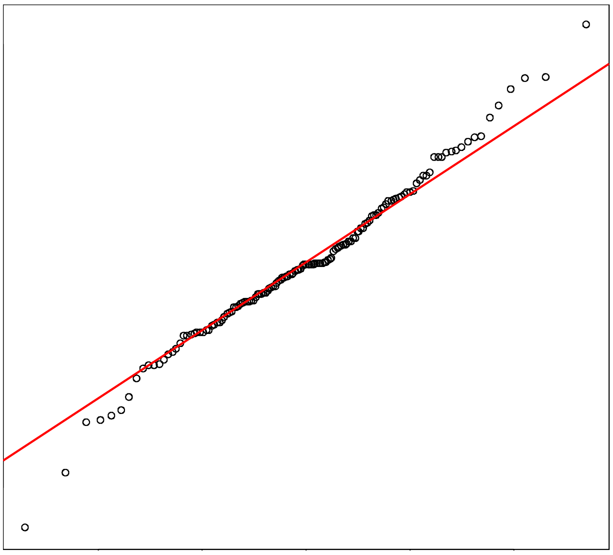
AirPassengers Case **_ARIMA(0,1,1)(0,1,1)$_{12}$ Model_**

```
> mean(d1d12lnairpass)
[1] 0.0002908799

> airpass.arima<-arima(d1d12lnairpass, order=c(0,0,1), seasonal=list(order=c(0,0,1),
period=12),include.mean=T)
Call:
arima(x = d1d12lnairpass, order = c(0, 0, 1), seasonal = list(order = c(0, 0, 1),
period = 12), include.mean = T)

Coefficients:
          ma1     sma1  intercept
      -0.4021  -0.5577     -2e-04
s.e.   0.0897   0.0732      1e-03

sigma^2 estimated as 0.001348:  log likelihood = 244.71,  aic = -481.42

> airpass.arima<-arima(lnairpass, order=c(0,1,1), seasonal=list(order=c(0,1,1),
period=12))

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

Coefficients:
          ma1     sma1
      -0.4018  -0.5569
s.e.   0.0896   0.0731

sigma^2 estimated as 0.001348:  log likelihood = 244.7,  aic = -483.4
```

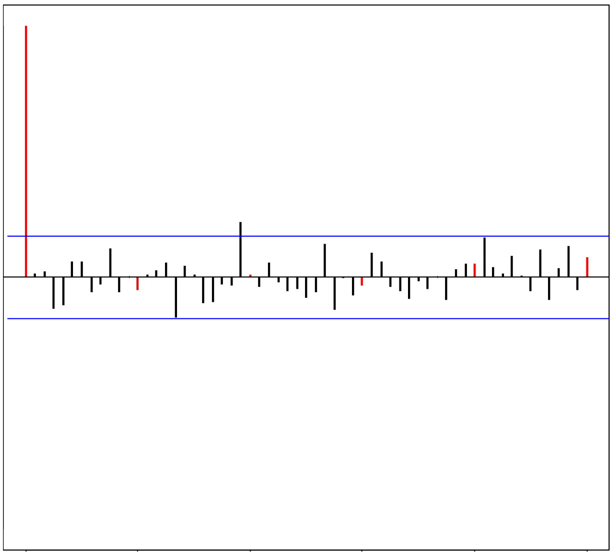
```
> resid=airpass.arima$residuals
> plot(resid)
> abline(h=0)
> abline(h=c(-3,3)*sd(resid),col=4,lty=3)
```

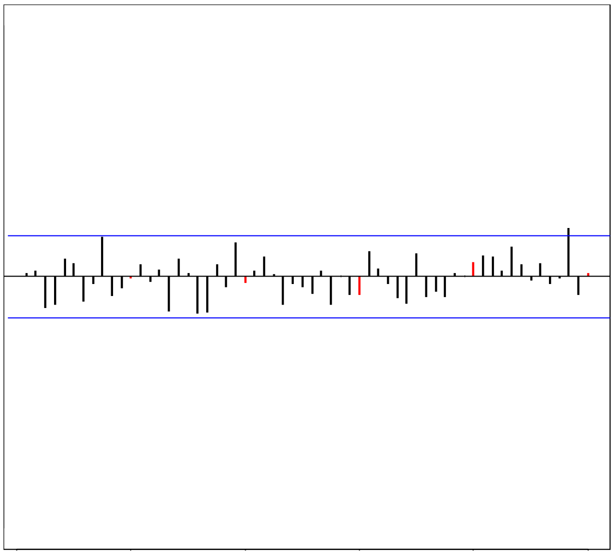


```
> qqnorm(resid)
> qqline(resid,col=2,lwd=2)
```

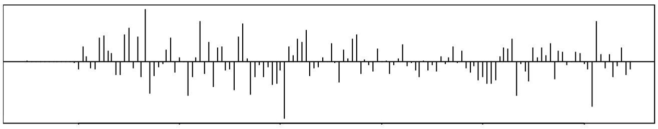


```
> acf(resid,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,11)), lwd=2)
> win.graph()
> pacf(resid,ylim=c(-1,1),lag.max=60,col=c(rep(1,11),2), lwd=2)
```





```
> tsdiag(airpass.arima,gof.lag=60)
```



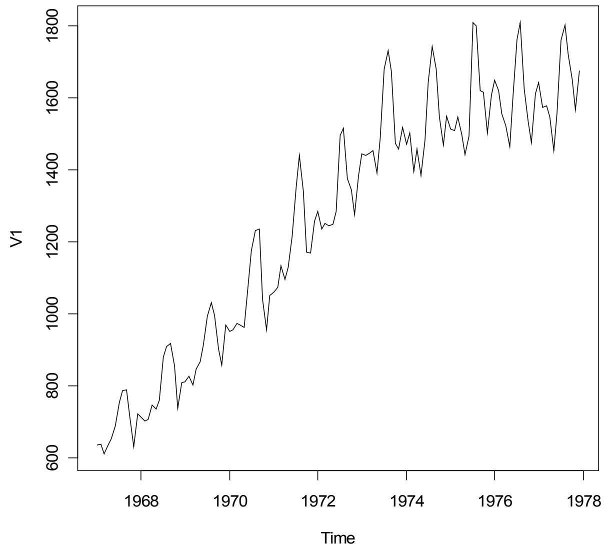
17) Exercise 3.1: GESA Case

GESA Case (G:\PST\CASOS\GESA\GESA.DAT)

The complete study of the series GESA can be found in chapter 6 of the notes.

Source: Martí M., Prat A., Hernández C. (1978)

Top monthly energy consumption in the GESA company, from January 1967 until December 1977.
The value of the six first observations are: 179.0, 170.1, 168.5, 156.6, 154.8, 166.9



Practices: Session 4

18) Practical Cases: Study of the stability of the model

GESA Case

The complete study of the series GESA can be found in chapter 6 of the notes.

Stability of the model:

To verify the stability of the mode, we estimate it without the last 12 observation and compare the parameters obtained in this way with the previous ones.

```
gesa<-ts(read.table("c:\\Gesa.dat"),start=1967,frequency=12)
lngesa<-log(gesa)
lngesa.arima<-arima(lngesa,order=c(0,1,1),   seasonal=list(order=c(0,1,1),period=12))
Call:
arima(x = lngesa, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:
         ma1     sma1
      -0.4964  -0.5842
s.e.   0.0878   0.0787

sigma^2 estimated as 0.001073:  log likelihood = 235.34,  aic = -464.69
```

$$(1-B)(1-B^{12})\log X_t = (1-0.4964B)(1-0.5842B^{12})Z_t \qquad Z_t \sim N(0,\sigma^2 = 0.001073)$$

Without the last 12 observations

```
gesa3<-window(gesa,start=1967,end=c(1976,12))
lngesa2<-log(gesa2)
lngesa2.arima<-arima(lngesa2,order=c(0,1,1), seasonal=list(order=c(0,1,1),period=12))
Call:
arima(x = lngesa2, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 
12))

Coefficients:
          ma1     sma1
      -0.5050  -0.5791
s.e.   0.0895   0.0875

sigma^2 estimated as 0.001146:  log likelihood = 207.83,  aic = -409.67
```

$$(1-B)(1-B^{12})\log X_t = (1-0.5050B)(1-0.57912B^{12})Z_t \qquad Z_t \sim N(0,\sigma^2 = 0.001146)$$

Conclusion: The model seems stable because the coefficients obtained are pretty similar.

Gnpsh Case

We will remove the last 9 observations(years 1989-1991), to evaluate the stability of the proposed models and their forecasting capacity:

```
gnpsh2<-ts(gnpsh[1:168],start= 1947,freq=4)
d1lngnpsh2<-diff(log(gnpsh2), lag=1)
```

- **ARIMA (0,1,2) Model with a constant** for the complete series:

```
gnpsh.arima<-arima(d1lngnpsh,order=c(0,0,2),include.mean=T)

Call:
arima(x = d1lngnpsh, order = c(0, 0, 2), include.mean = T)

Coefficients:
         ma1     ma2  intercept
      0.3121  0.2714     0.0077
s.e.  0.0736  0.0678     0.0012

sigma^2 estimated as 9.504e-05:  log likelihood = 565.14,  aic = -1122.29
```

$$(1-B)(\log X_t - 0.0077) = (1+0.3121B+0,2714B^2)Z_t \qquad Z_t \sim N(0,\sigma^2 = 9.504 \cdot 10^{-5})$$

- **ARIMA (0,1,2) Model with a constant** for the complete series without the last observations

```
gnpsh2.arima<-arima(d1lngnpsh2,order=c(0,0,2),include.mean=T)

Call:
arima(x = d1lngnpsh2, order = c(0, 0, 2), include.mean = T)

Coefficients:
         ma1     ma2  intercept
      0.3044  0.2667     0.0080
s.e.  0.0753  0.0699     0.0012

sigma^2 estimated as 9.84e-05:  log likelihood = 533.35,  aic = -1058.69
```

$$(1-B)(\log X_t - 0.0080) = (1+0.3044B+0.2667B^2)Z_t \qquad Z_t \sim N(0,\sigma^2 = 9.84 \cdot 10^{-5})$$

19) Calculation of the long term forecasts and their variances

Gnpsh Case

Forecasts for the series W_t= *(1-B)log(Gnpsh)*: MA(2) Model with a constant

```
gnpsh.arima<-arima(d1lngnpsh,order=c(0,0,2),include.mean=T)

Call: arima(x = d1lngnpsh, order = c(0, 0, 2), include.mean = T)

Coefficients:
         ma1     ma2  intercept
      0.3121  0.2714     0.0077
s.e.  0.0736  0.0678     0.0012

sigma^2 estimated as 9.504e-05:  log likelihood = 565.14,  aic = -1122.29
```

```
model<-gnpsh.arima$model
```

Autoregressive Part

```
model$phi
numeric(0)
```

Moving Average Part

```
model$theta
0.3121003 0.2714321
```

Mean Estimation

```
cte<-coef(gnpsh.arima)[["intercept"]]
0.007679398
```

Variance of the residuals

```
varz<- gnpsh.arima$sigma2
9.504103e-05
```

AR(∞) Expression

```
Wheights.pi<--ARMAtoMA(ar=-model$theta,ma=-model$phi,lag.max=16)

Wheights.pi
 [1]  3.121003e-01  1.740255e-01 -1.390275e-01 -3.845612e-03  3.893673e-02 -1.110834e-02
 [7] -7.101763e-03  5.231623e-03  2.948556e-04 -1.512055e-03  3.918795e-04  2.881146e-04
[13] -1.962893e-04 -1.694160e-05  5.856671e-05 -1.368019e-05
```

$$(W_t - \mu) = 0.3121(W_{t-1} - \mu) + 0.17403(W_{t-2} - \mu) - 0.13903(W_{t-3} - \mu) - 0.00385(W_{t-4} - \mu) +$$
$$+ 0.03894(W_{t-5} - \mu) - 0.01111(W_{t-6} - \mu) + \ldots + Z_t$$

MA(∞) Expression

```
Wheights.psi<-ARMAtoMA(ar=model$phi,ma=model$theta,lag.max=16)

 [1] 0.312100 0.271432 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
[10] 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
```

$$(W_t - \mu) = Z_t + 0.3121 Z_{t-1} + 0.2714 Z_{t-2}$$

Last Observations:

```
d1lngnpsh
                :               :               :               :
1986  1.590876e-02 -4.449408e-03  2.100583e-03  5.699720e-03
1987  1.266847e-02  1.027965e-02  1.006104e-02  1.589055e-02
1988  1.253864e-02  8.931855e-03  6.539045e-03  6.730620e-03
1989  8.919729e-03  4.020532e-03  4.249032e-03  8.495797e-04
1990  4.200967e-03  1.078777e-03  3.579549e-03 -3.988769e-03
1991 -6.500125e-03
```

Long term forecasts:

For *h=1*

$$\widetilde{W}_{t+1|t} = \hat{\mu} + 0.3121(W_t - \hat{\mu}) + 0.17403(W_{t-1} - \hat{\mu}) - 0.13903(W_{t-2} - \hat{\mu})$$
$$- 0.00385*(W_{t-3} - \hat{\mu}) + 0.03894*(W_{t-4} - \hat{\mu}) + \ldots$$

$$\widetilde{W}_{t+1|t} = 0.0077 + 0.3121(-0.0650 - 0.0077) + 0.17403(-0.0039 - 0.0077) +$$
$$- 0.13903(0.0036 - 0.0077) - 0.00385(0.0018 - 0.0077) + 0.03894(0.0042 - 0.0077) + \ldots$$

```
p1<-cte+sum(weights.pi*(d1lngnpsh[176:161]-cte))
0.001765155
```

For *h=2*

$$\widetilde{W}_{t+2|t} = \hat{\mu} + 0.3121(\widetilde{W}_{t+1|t} - \hat{\mu}) + 0.17403(W_t - \hat{\mu}) - 0.13903(W_{t-1} - \hat{\mu})$$
$$- 0.00385(W_{t-2} - \hat{\mu}) + 0.03894(W_{t-3} - \hat{\mu}) + \ldots$$

$$\widetilde{W}_{t+2|t} = 0.0077 + 0.3121(0.00176 - 0.0077) + 0.17403(-0.0650 - 0.0077) +$$
$$- 0.13903(-0.0039 - 0.0077) - 0.00385(0.0036 - 0.0077) + 0.03894(0.0018 - 0.0077) + \ldots$$

```
p2<-cte+sum(weights.pi*(c(p1,d1lngnpsh[176:162])-cte))
0.00481258
```

For *h=3*

$$\widetilde{W}_{t+3|t} = \hat{\mu} + 0.3121(\widetilde{W}_{t+2|t} - \hat{\mu}) + 0.17403(\widetilde{W}_{t+1|t} - \hat{\mu}) - 0.13903(W_t - \hat{\mu})$$
$$- 0.00385(W_{t-1} - \hat{\mu}) + 0.03894(W_{t-2} - \hat{\mu}) + \ldots$$

$$\widetilde{W}_{t+3|t} = 0.0077 + 0.3121(0.00481 - 0.0077) + 0.17403(0.00176 - 0.0077) +$$
$$- 0.13903(-0.0650 - 0.0077) - 0.00385(-0.0039 - 0.0077) + 0.03894(0.0036 - 0.0077) + \ldots$$

```
p3<-cte+sum(weights.pi*(c(p2,p1,d1lngnpsh[176:163])-cte))
0.007679416
```

From now on the forecasts are constant.

Long term forecasts standard deviation

For *h=1*

$$\mathrm{var}(\widetilde{W}_{t+1|t} - \mu) = \mathrm{var}(Z_t)$$

$$\mathrm{var}(\widetilde{W}_{t+1|t}) = 0.00009504$$

```
s1<-sqrt(varz)
0.009709289
```

For *h=2*

$$\mathrm{var}(\widetilde{W}_{t+2|t} - \mu) = \mathrm{var}(Z_t) + 0.3121^2\,\mathrm{var}(Z_t)$$

$$\mathrm{var}(\widetilde{W}_{t+2|t}) = 0.00009504(1 + 0.3121^2)$$

```
s2<-sqrt(varz*(1+sum((weights.psi[1]^2))))
0.01021267
```

For *h=3*

$$\mathrm{var}(\widetilde{W}_{t+3|t} - \mu) = \mathrm{var}(Z_t) + 0.3121^2\,\mathrm{var}(Z_t) + 0.2714^2\,\mathrm{var}(Z_t)$$

$$\mathrm{var}(\widetilde{W}_{t+3|t}) = 0.00009504(1 + 0.3121^2 + 0.2714^2)$$

```
s3<-sqrt(varz*(1+sum((weights.psi[1:2]^2))))
0.01054992
```

From now on the standard deviation is constant, because ψ_{t+k}=0 if k>2

Predict function of R:

```
predict(gnpsh.arima,n.ahead=6)

$pred
            Qtr1        Qtr2        Qtr3        Qtr4
1991             0.001765109 0.004812581 0.007679398
1992 0.007679398 0.007679398 0.007679398

$se
            Qtr1        Qtr2        Qtr3        Qtr4
1991             0.009748899 0.010212671 0.010549921
1992 0.010549921 0.010549921 0.010549921
```

Forecasts for the series ***W_t= (1-B)log(Gnpsh)*****: AR(3) Model with a constant**

```
gnpsh.arima<-arima(d1lngnpsh,order=c(3,0,0),include.mean=T)
Call:
arima(x = d1lngnpsh, order = c(3, 0, 0), include.mean = T)

Coefficients:
         ar1     ar2      ar3  intercept
      0.3480  0.1793  -0.1423     0.0077
s.e.  0.0745  0.0778   0.0745     0.0012

sigma^2 estimated as 9.427e-05:  log likelihood = 565.84,  aic = -1121.69
```

```
model<-gnpsh.arima$model
```

<u>Autoregressive Part</u>

```
model$phi
0.3480210  0.1793063 -0.1422637
```

<u>Moving Average Part</u>

```
model$theta
numeric(0)
```

<u>Estimation of the mean</u>

```
cte<-coef(gnpsh.arima)[["intercept"]]
0.007680634
```

<u>Residual Variance</u>

```
varz<- gnpsh.arima$sigma2
9.427e-05
```

<u>AR(∞) Expression</u>

```
Weights.pi<--ARMAtoMA(ar=-model$theta,ma=-model$phi,lag.max=16)

Weights.pi
[1]  0.348021  0.179306 -0.142264  0.000000  0.000000  0.000000  0.000000  0.000000
[9]  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
```

$$(W_t - \mu) = 0.3480(W_{t-1} - \mu) + 0.1793(W_{t-2} - \mu) - 0.1423(W_{t-3} - \mu) + Z_t$$

<u>MA(∞) Expression</u>

```
Weights.psi<-ARMAtoMA(ar=model$phi,ma=model$theta,lag.max=16)
Weights.psi
 [1]  3.480210e-01  3.004249e-01  2.469275e-02  1.295089e-02 -3.380482e-02 -1.295550e-02
 [7] -1.241264e-02 -1.833661e-03 -1.020720e-03  1.081850e-03  4.543486e-04  4.973168e-04
[13]  1.006362e-04  5.955817e-05 -3.197797e-05 -1.476673e-05
```

$$(W_t - \mu) = Z_t + 0.3480Z_{t-1} + 0.3004Z_{t-2} + 0.0247Z_{t-3} + 0.0129Z_{t-4} - 0.0338Z_{t-5} + ...$$

<u>Last Observations</u>

```
d1lngnpsh
                :              :              :              :
1986  1.590876e-02 -4.449408e-03  2.100583e-03  5.699720e-03
1987  1.266847e-02  1.027965e-02  1.006104e-02  1.589055e-02
1988  1.253864e-02  8.931855e-03  6.539045e-03  6.730620e-03
1989  8.919729e-03  4.020532e-03  4.249032e-03  8.495797e-04
1990  4.200967e-03  1.078777e-03  3.579549e-03 -3.988769e-03
1991 -6.500125e-03
```

Long term forecasts

For *h=1*

$$\widetilde{W}_{t+1|t} = \hat{\mu} + 0.3480(W_t - \hat{\mu}) + 0.1793(W_{t-1} - \hat{\mu}) - 0.1423(W_{t-2} - \hat{\mu})$$

$$\widetilde{W}_{t+1|t} = 0.0077 + 0.3480(-0.0650 - 0.0077) + 0.1793(-0.0039 - 0.0077) - 0.1423(0.0036 - 0.0077)$$

```
p1<-cte+sum(weights.pi*(d1lngnpsh[176:161]-cte))
0.001236471
```

For *h=2*

$$\widetilde{W}_{t+2|t} = \hat{\mu} + 0.3480(\widetilde{W}_{t+1|t} - \hat{\mu}) + 0.1793(W_t - \hat{\mu}) - 0.1423(W_{t-1} - \hat{\mu})$$

$$\widetilde{W}_{t+2|t} = 0.0077 + 0.3480(0.0012 - 0.0077) + 0.1793(-0.0650 - 0.0077) - 0.1423(-0.0039 - 0.0077)$$

```
p2<-cte+sum(weights.pi*(c(p1,d1lngnpsh[176:162])-cte))
0.004555365
```

For *h=3*

$$\widetilde{W}_{t+3|t} = \hat{\mu} + 0.3480(\widetilde{W}_{t+2|t} - \hat{\mu}) + 0.1793(\widetilde{W}_{t+1|t} - \hat{\mu}) - 0.14226(W_t - \hat{\mu})$$

$$\widetilde{W}_{t+3|t} = 0.0077 + 0.3480(0.0046 - 0.0077) + 0.1793(0.0012 - 0.0077) - 0.1423(-0.0650 - 0.0077)$$

```
p3<-cte+sum(weights.pi*(c(p2,p1,d1lngnpsh[176:163])-cte))
0.007454904
```

For *h=4*

$$\widetilde{W}_{t+4|t} = \hat{\mu} + 0.3480(\widetilde{W}_{t+3|t} - \hat{\mu}) + 0.1793(\widetilde{W}_{t+2|t} - \hat{\mu}) - 0.1423(\widetilde{W}_{t+1|t} - \hat{\mu})$$

$$\widetilde{W}_{t+3|t} = 0.0077 + 0.3480(0.0074 - 0.0077) + 0.1793(0.0046 - 0.0077) - 0.1423(0.0012 - 0.0077)$$

```
p4<-cte+sum(weights.pi*(c(p3,p2,p1,d1lngnpsh[176:164])-cte))
0.007958466
```

Long term forecasts standard deviation

For *h=1*

$$\operatorname{var}(\widetilde{W}_{t+1|t} - \mu) = \operatorname{var}(Z_t)$$

$$\operatorname{var}(\widetilde{W}_{t+1|t}) = 0.00009427$$

```
s1<-sqrt(varz)
0.009709289
```

For *h=2*

$$\operatorname{var}(\widetilde{W}_{t+2|t} - \mu) = \operatorname{var}(Z_t) + 0.3480^2 \operatorname{var}(Z_t)$$

$$\operatorname{var}(\widetilde{W}_{t+2|t}) = 0.00009427(1 + 0.3480^2)$$

```
s2<-sqrt(varz*(1+sum((weights.psi[1]^2))))
0.010280475
```

For *h=3*

$$\operatorname{var}(\widetilde{W}_{t+3|t} - \mu) = \operatorname{var}(Z_t) + 0.3480^2 \operatorname{var}(Z_t) + 0.3004^2 \operatorname{var}(Z_t)$$

$$\operatorname{var}(\widetilde{W}_{t+3|t}) = 0.00009427(1 + 0.3480^2 + 0.3004^2)$$

```
s3<-sqrt(varz*(1+sum((weights.psi[1:2]^2))))
0.010686278
```

For *h=4*

$$\operatorname{var}(\widetilde{W}_{t+4|t} - \mu) = \operatorname{var}(Z_t) + 0.3480^2 \operatorname{var}(Z_t) + 0.3004^2 \operatorname{var}(Z_t) + 0.0247^2 \operatorname{var}(Z_t)$$

$$\operatorname{var}(\widetilde{W}_{t+4|t}) = 0.00009427(1 + 0.3480^2 + 0.3004^2 + 0.0247^2)$$

```
s3<-sqrt(varz*(1+sum((weights.psi[1:2]^2))))
0.010688967
```

Predict function of R:

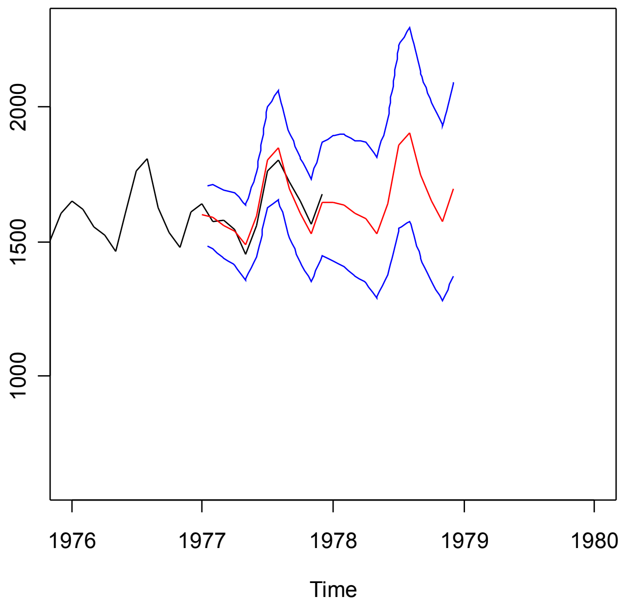
```
predict(gnpsh.arima,n.ahead=6)

$pred
            Qtr1        Qtr2        Qtr3        Qtr4
1991             0.001236471 0.004555365 0.007454904
1992 0.007958466 0.008181463 0.007936864

$se
            Qtr1        Qtr2        Qtr3        Qtr4
1991             0.009709289 0.010280475 0.010686278
1992 0.010688967 0.010689707 0.010694745
```

GESA Case

Long Term Forecasts Evaluation

Long term forecasts are calculated with the model obtained without the last observations. The first 12 forecasts can be compared with the observation removed before:

```
pred2<-predict(lngesa.arima2,n.ahead=24)
tl<-pred2$pred-1.96*pred2$se
tu<-pred2$pred+1.96*pred2$se
ts.plot(gesa,exp(tl),exp(tu),exp(pred2$pred),lty=c(1,2,2,1),
      col=c("black","blue","blue","red"),xlim=c(1976,1980))
```



Conclusion: It can be appreciate that the forecasts are situated inside the confident interval and they are quite similar to the actual observation of the series.

Some of the measures of the forecasting capacity are::

Relative Forecast Mean Squared Error $FMSE = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{X_{t+i} - \tilde{X}_{t+i|t}}{X_{t+i}}\right)^2$

Relative Forecast Absolute Squared Error: $FASE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{X_{t+i} - \tilde{X}_{t+i|t}}{X_{t+i}}\right|$

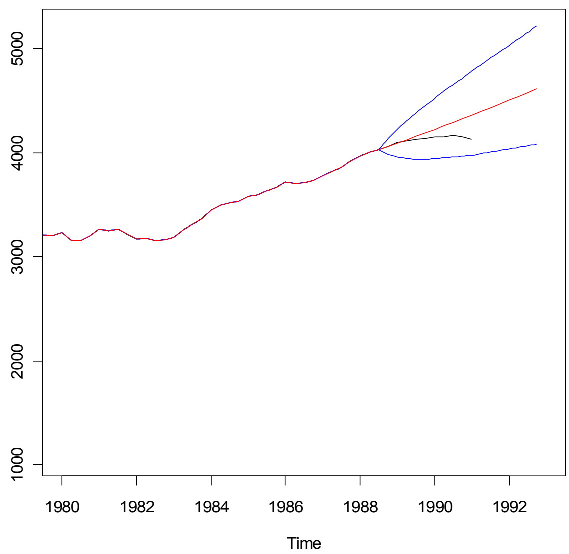
```
obs <- gesa[121:132]
prev <- exp(pred2$prev[1:12])
gesa.EQM<- sum((obs-prev)^2/obs)
 [1] 8.54129
gesa.EAM<- sum(abs(obs-prev)/obs)
 [1] 0.2321764
```

Gnpsh Case

Forecast and calculation of the MSE and the MAE:

(in this case, the forecasts are for the differenced series because it is the one that was fitted: d1lngnpsh. Thus it is necessary to undo the differencing to calculate the the forecasts for the original series)

```
pred1<-predict(gnpsh2.arima,n.ahead=16)
pr<-cumsum(c(lngnpsh[1],d1lngnpsh2,pred1$pred))

model<-gnpsh2.arima$model
varZ<-gnpsh2.arima$sigma
ma<-ARMAtoMA(ar=model$phi,ma=model$theta,lag.max=16)
se<-c(rep(0,167),sqrt(cumsum(cumsum(c(1,ma))^2)*varZ))

tl<-pr-1.96*se
tu<-pr+1.96*se

tl<-ts(exp(tl),start=1947,freq=4)
pr<-ts(exp(pr),start=1947,freq=4)
tu<-ts(exp(tu),start=1947,freq=4)

ts.plot(gnpsh,tl,tu,pr,lty=c(1,2,2,1),col=c("black","blue","blue","red"),xlim=c(1980,19
95))
```




```
obs <- gnpsh[169:177]
prev <- pr[169:177]
gnpsh.EQM<- sum((obs-prev)^2/obs)
 [1] 28.88466
gnpsh.EAM<- sum(abs(obs-prev)/obs)
 [1] 0.1937523
```

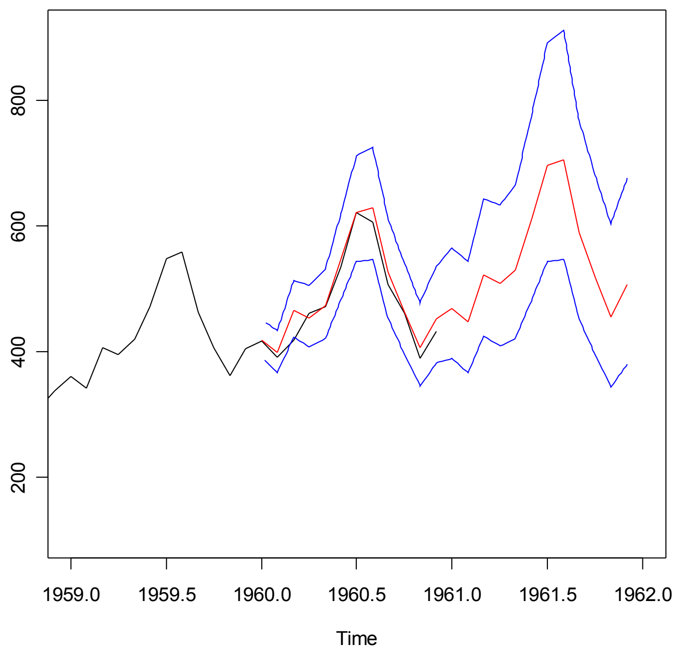
a) Do the same process to evaluate the forecasts for the model ARIMA(3,1,0) with a constant and compare the stability and forecast capacity of both models.

AirPassengers Case

```
airpass<-ts(airpass,start= 1949,freq=12)
lnairpass<-log(airpass)
```

- **ARIMA (0,1,1)(0,1,1)$_{12}$ Model** for the complete series

```
air.arima<-arima(lnairpass,order=c(0,1,1),seasonal=list(order=c(0,1,1),freq=12))
Call:
arima(x = lnairpass, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1),
    freq = 12))

Coefficients:
          ma1     sma1
      -0.4018  -0.5569
s.e.   0.0896   0.0731

sigma^2 estimated as 0.001348:  log likelihood = 244.7,  aic = -483.4

pred1<-predict(air2.arima,n.ahead=24)
pr<-pred1$pred
se<-pred1$se

tl<-pr-1.96*se
tu<-pr+1.96*se

tl<-ts(exp(tl),start=1960,freq=12)
pr<-ts(exp(pr),start=1960,freq=12)
tu<-ts(exp(tu),start=1960,freq=12)

ts.plot(airpass,tl,tu,pr,lty=c(1,2,2,1),col=c("black","blue","blue","red"),xlim=c(1949,
1962))
```



21) Exercise: Selection of the best model. G*npsh* and *AirPassengers* cases

Analyse the stability and calculate the long term forecasts for the proposed models for all the series studied before. Fill the table:

Gnpsh Case

	Model	σ^2	loglik	AIC	EQM	EAM
1	ARIMA(0,1,2)	$9.504 \cdot 10^{-5}$	565.14	-1122.29	28.88466	0.1937523
2	ARIMA(3,1,0)	$9.427 \cdot 10^{-5}$	565.84	-1121.69		
3	ARIMA(1,1,2)					
4						

AirPassengers Case

	Model	σ^2	loglik	AIC	EQM	EAM
1	ARIMA(3,1,0) $(0,1,1)_{12}$	0.001348	244.76	-479.52		
2	ARIMA(1,1,0) $(0,1,1)_{12}$	0.001367	243.74	-481.49		
3	ARIMA(0,1,1) $(0,1,1)_{12}$	0.001348	244.7	-483.4		
4						

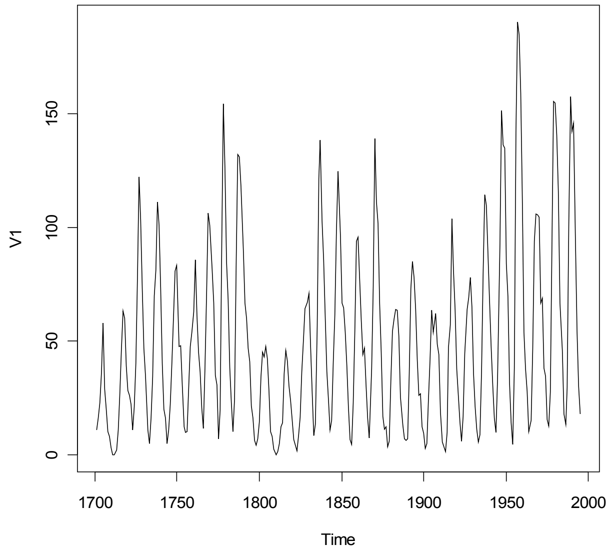
Discuss which model is considered the best forecasting model for each series, explaining in each case the reasons for the answer.

Practices: Session 5

22) Practical Cases: Series with a cyclic component. SUNSPOT case.

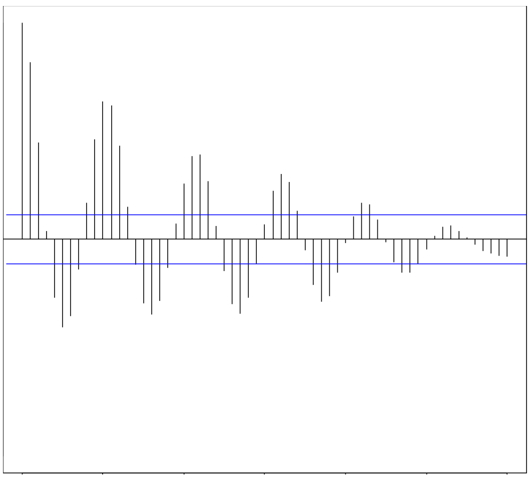
Sunspot.year Case

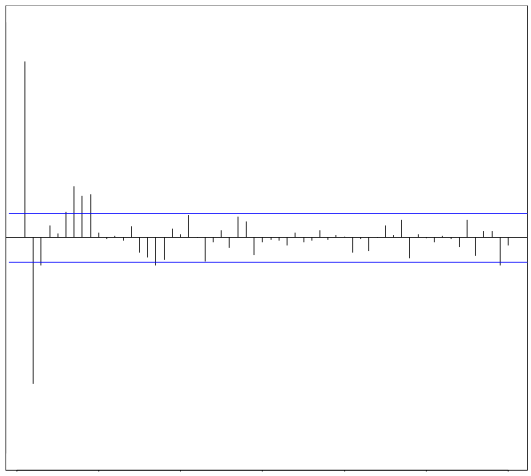
Sunspots, data recollected by Wölfer, 289 annual observations between 1700 and 1988. A historical reference of the Wölfer sunspot numbers from 1770 until 1995 can be found in the file **ssn_vals**



a) Even thought the series has a seasonal component, the value of period varies between 10 and 11. If the wave pattern has a non-constant period, the component is called cyclic instead of seasonal.
   Try the *decompose* function with period freq=10 and freq=11. The residuals have still a cyclic behaviour.

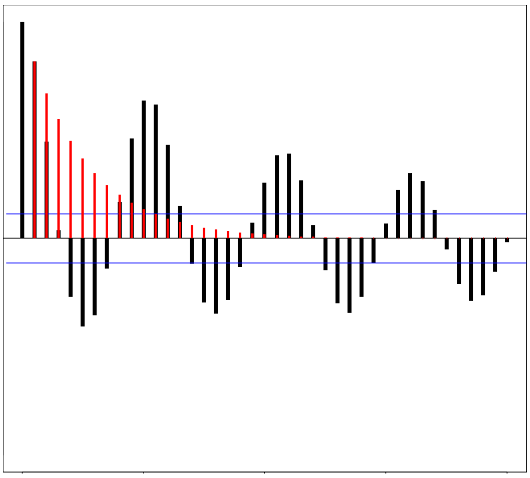
   Discuss the shape of the ACF and the PACF of the series.

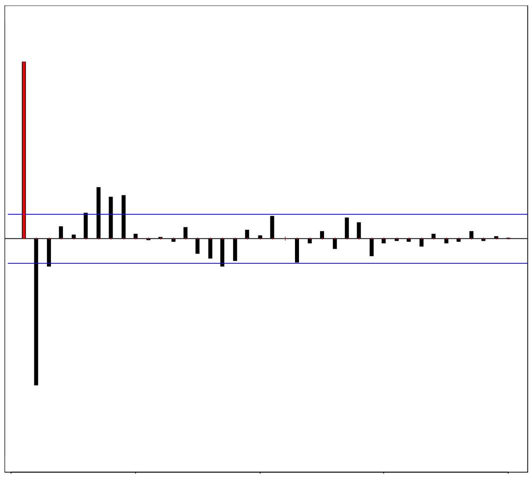




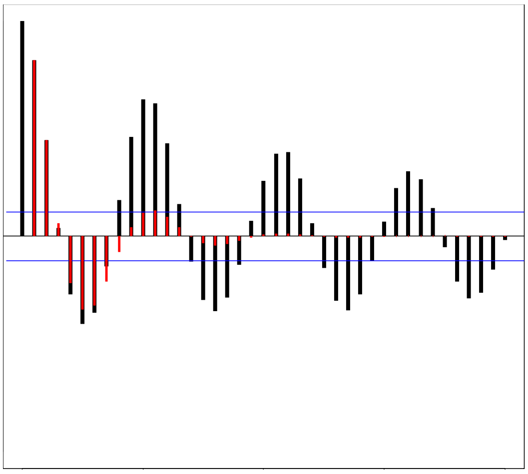
b) The R function used to estimate the AR(p) models using the Yule-Walker method is called *ar*. The aim of this exercise is to fit the model with AR models with increasing order and compare the ACF and the PACF of the original data and the AR(p) adjusted model. It is advisable to check the residuals at each step. What is the conclusion?

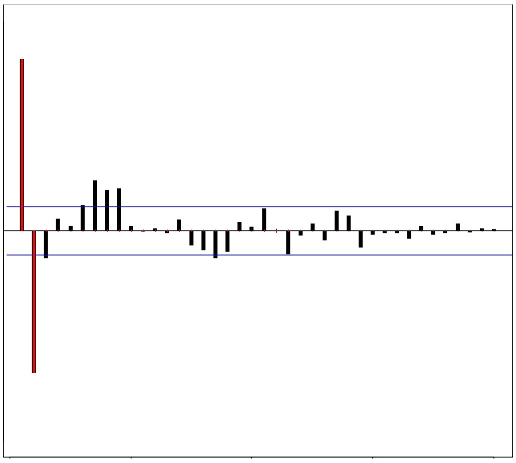
```
sun.ar<-ar(sunspots,aic=F,order.max=1)

Call:
ar(x = sunspots, aic = F, order.max = 1)

Coefficients:
     1
0.8195

Order selected 1  sigma^2 estimated as  537.9

acf(sunspots, ylim=c(-1,1) ,lag.max=40,lwd=4)
lines(ARMAacf(ar=sun.ar$ar,NULL,lag.max=40)[-1],type="h",col=2,lwd=2)
win.graph()
pacf(sunspots, ylim=c(-1,1) ,lag.max=40,lwd=4)
lines(ARMAacf(ar=sun.ar$ar,NULL,lag.max=40,pacf=T),type="h",col=2,lwd=2)
```



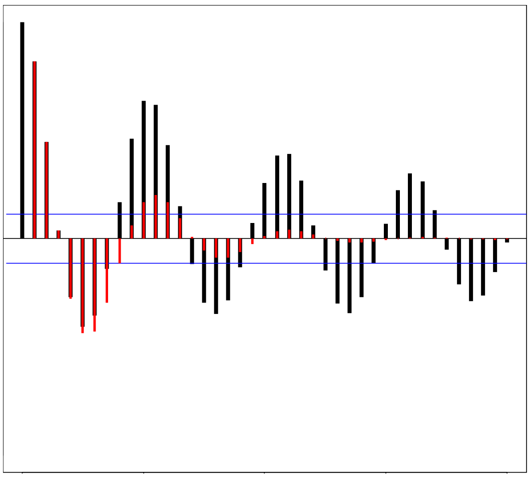


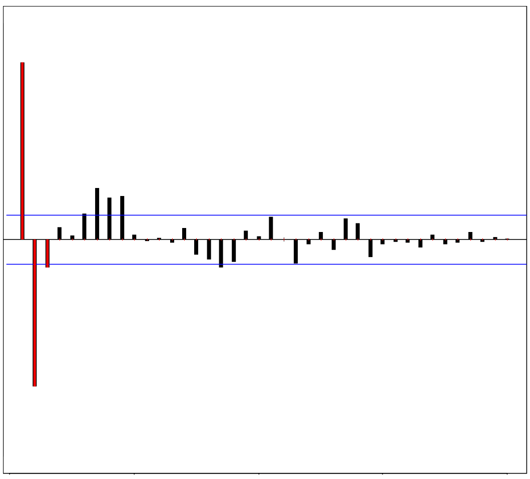
```
sun.ar<-ar(sunspots,aic=F,order.max=2)

Call:
ar(x = sunspots, aic = F, order.max = 2)

Coefficients:
      1        2
 1.3740  -0.6765

Order selected 2  sigma^2 estimated as  292.7
```





```
sun.ar<-ar(sunspots,aic=F,order.max=3)

Call:
ar(x = sunspots, aic = F, order.max = p)

Coefficients:
      1        2        3
 1.2866  -0.4992  -0.1291

Order selected 3  sigma^2 estimated as  288.8
```





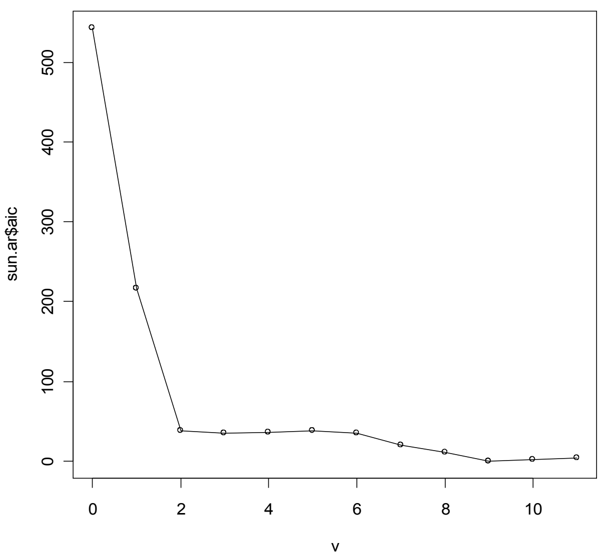
Fill the following table:

p	ϕ_{p1}	ϕ_{p2}	ϕ_{p3}	ϕ_{p4}	ϕ_{p5}	ϕ_{p6}	ϕ_{p7}	ϕ_{p8}	ϕ_{p9}	ϕ_{p10}	ϕ_{p11}	V_p
1	0.819	---	---	---	---	---	---	---	---	---	---	537.9
2	1.374	-0.677	---	---	---	---	---	---	---	---	---	292.7
3	1.287	-0.499	-0.129	---	---	---	---	---	---	---	---	288.8
4					---	---	---	---	---	---	---	
5						---	---	---	---	---	---	
6							---	---	---	---	---	
7								---	---	---	---	
8									---	---	---	
9										---	---	
10											---	
11												

What do the parameters ϕ_{pp} of the diagonal of the table correspond to?

c) If the function is called with the parameter AIC=T, it returns the order with lower AIC

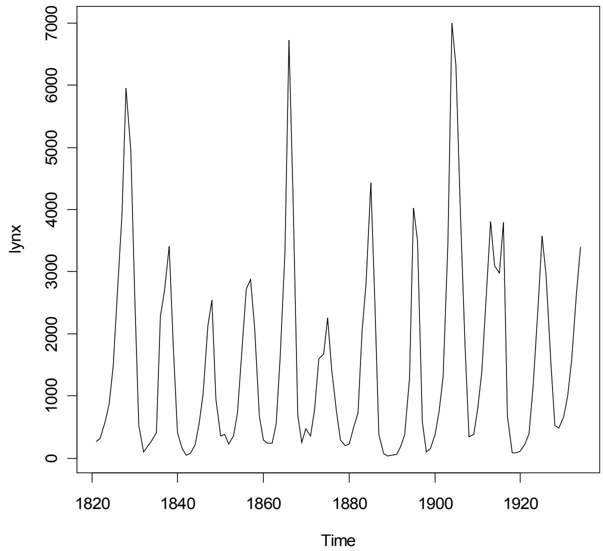
```
ar(sunspots,aic=T,order.max=11)

Call:
ar(x = sunspots, aic = T, order.max = 11)

Coefficients:
      1        2        3        4        5        6        7        8        9
 1.1743  -0.3986  -0.1656   0.0993  -0.0055  -0.0661   0.0752  -0.0496   0.2019

Order selected 9  sigma^2 estimated as  251.5

sun.ar<-ar(sunspots,aic=T,order.max=11)
v<-0:11
plot(sun.ar$aic~v,type="o")
```



23) Exercise: LYNX Case

Lynx Case

Annual amount of lynxes captured in the period 1821-1934 in Canada (Brockwell & Davis ,1991)



- Analyse the cyclic component and fit the series with an AR(p) model.

24) Cyclic components associated to complex roots: gnpsh case

```
d1lngnpsh<-diff(log(gnpsh),lag=1)
gnp.arima1<-arima(d1lngnpsh,order=c(3,0,0),include.mean=T)


Call:
arima(x = d1lngnpsh, order = c(3, 0, 0), include.mean = T)

Coefficients:
         ar1     ar2      ar3  intercept
      0.3480  0.1793  -0.1423     0.0077
s.e.  0.0745  0.0778   0.0745     0.0012

sigma^2 estimated as 9.427e-05:  log likelihood = 565.84,  aic = -1121.69
```

Estimated model:

$$(1-0.348B-0.1793B^2+0.1423B^3)(W_t-0.0077) = Z_t$$

Polynomial roots:

```
polyroot(c(1,-gnp.arima1$model$phi))
[1]  1.590262+1.063874i -1.920146+0.000000i  1.590262-1.063874i
```

The model has two conjugated complex roots

Modules and arguments:

```
Mod(polyroot(c(1,-gnp.arima1$model$phi)))
[1] 1.913312 1.920146 1.913312

Arg(polyroot(c(1,-gnp.arima1$model$phi)))
[1]  0.5896113  3.1415927 -0.5896113
```

All the root modules are greater than 1 and therefore they are outside the unit circle. Thus the model is stationary. Obviously the real root has argument π.

The polynomial can be decomposed (using Ruffini, for example):

$$(1-0.348B-0.1793B^2+0.1423B^3) = (1-0.52B)(1-0.87B+0.27B^2)$$

Calculating the following quotient for the AR(2) model:

$$k = \frac{2\pi}{|Arg(z_i)|} = \frac{2\pi}{\arccos(\phi_1 / 2\sqrt{-\phi_2})}$$

we will obtain the serris cycles mean value.

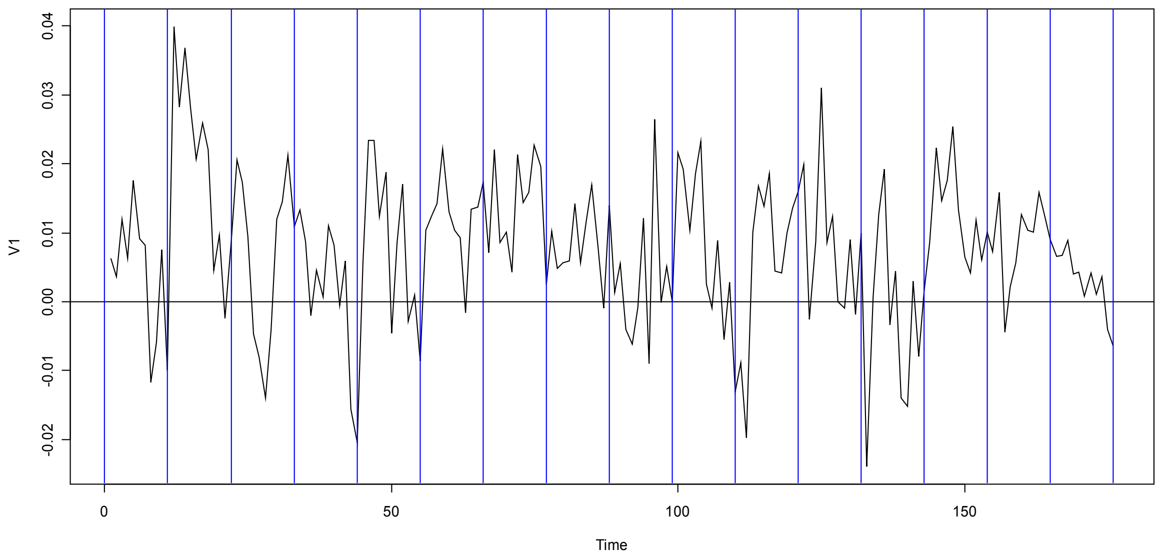
In this case:

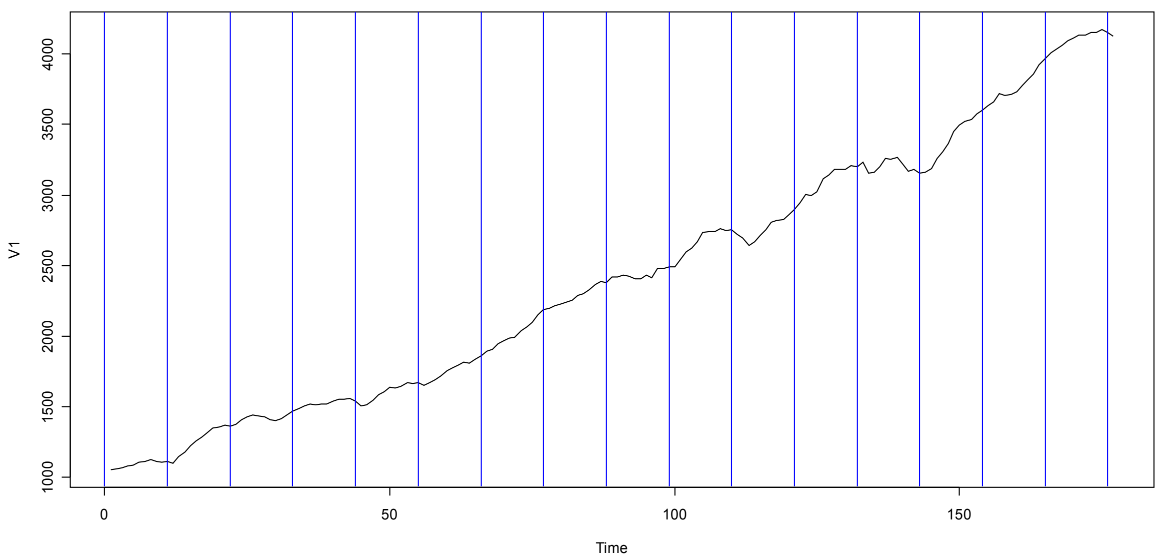
$$k = \frac{2\pi}{0.5896} = \frac{2\pi}{\arccos(0.87 / 2\sqrt{-(-0.27)})} = 10.7$$

The autocorrelation function take the following values:

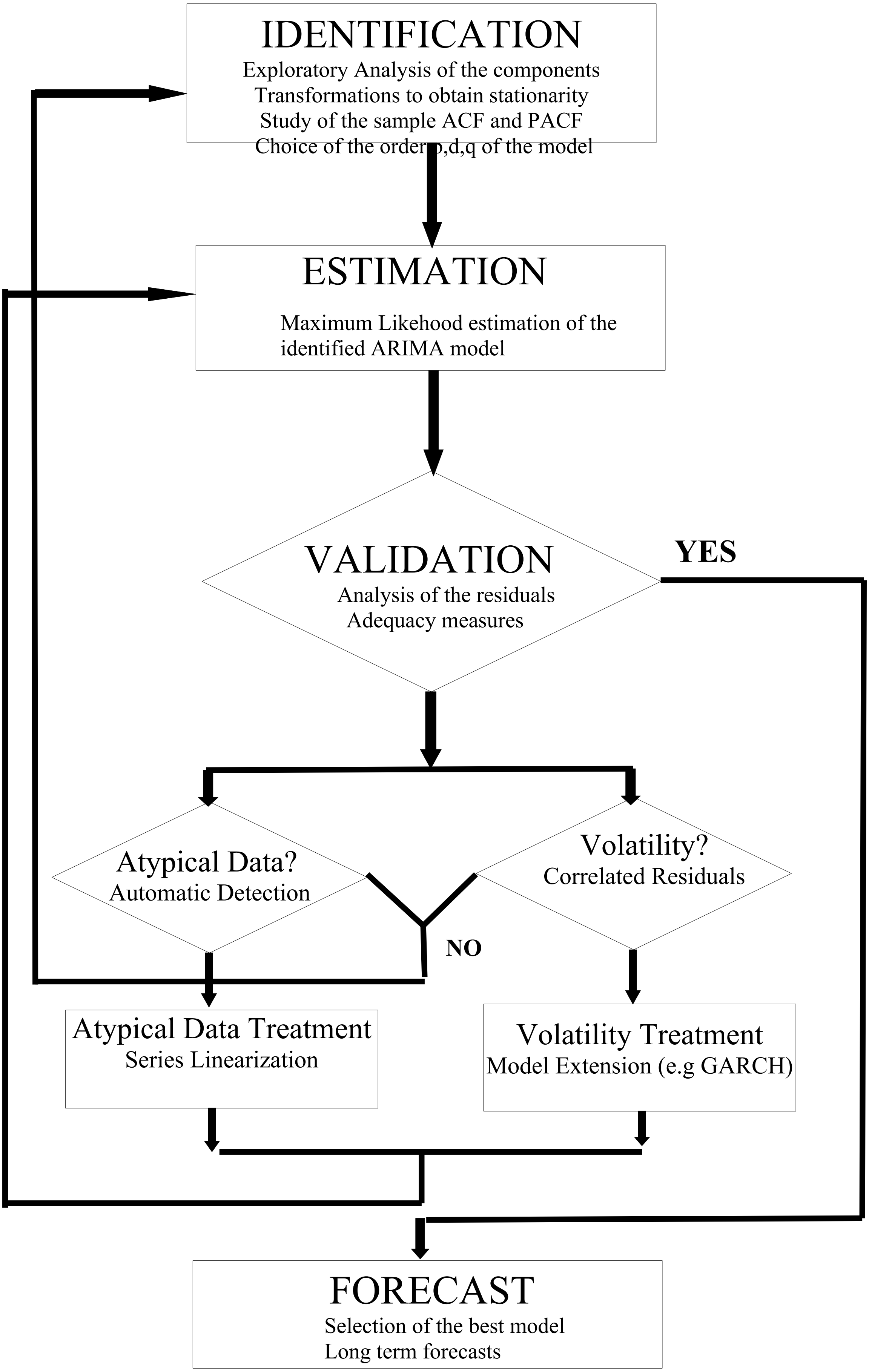
```
ar2.mod=c(0.87,-0.27)
> ARMAacf(ar2.mod,lag.max=40)
            0             1             2             3             4             5 
 1.000000e+00  6.850394e-01  3.259843e-01  9.864567e-02 -2.194016e-03 -2.854312e-02 
            6             7             8             9            10            11 
-2.424013e-02 -1.338227e-02 -5.097741e-03 -8.218212e-04  6.614057e-04  7.973147e-04 
           12            13            14            15            16            17 
 5.150842e-04  2.328483e-04  6.350530e-05 -7.619439e-06 -2.377534e-05 -1.862730e-05 
           18            19            20            21            22            23 
-9.786408e-06 -3.484804e-06 -3.894494e-07  6.020761e-07  6.289575e-07  3.846325e-07 
           24            25            26            27            28            29 
 1.648118e-07  3.953545e-08 -1.010334e-08 -1.946447e-08 -1.420619e-08 -7.103978e-09 
           30            31            32            33            34            35 
-2.344790e-09 -1.218928e-10  5.270464e-10  4.914415e-10  2.852515e-10  1.154796e-10 
           36            37            38            39            40 
 2.344937e-11 -1.077855e-11 -1.570867e-11 -1.075633e-11 -5.116669e-12 
```

The quarterly increments have the following evolution, the marks separate periods of eleven quarters. Cyclic components with a 3 year period approximately can be appreciated in the original series.





Extended Box-Jenkins Methodology



Practices: Session 6

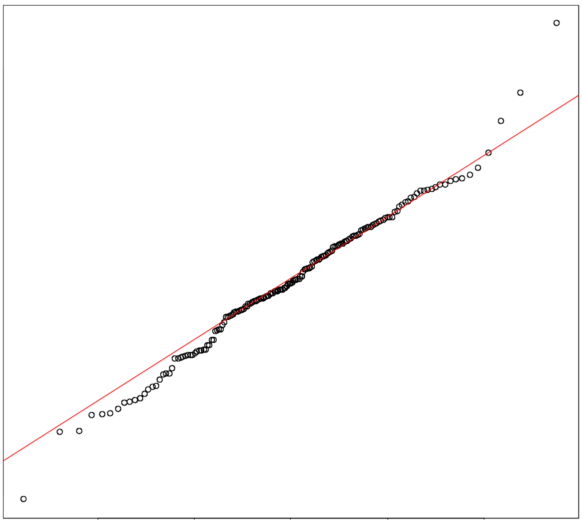
25) Practical Cases: Outliers treatment with R. GNPSH and TSGGB cases

1. Upload the **atipics2.r** code that contains the *outdetec* and *lineal functions*

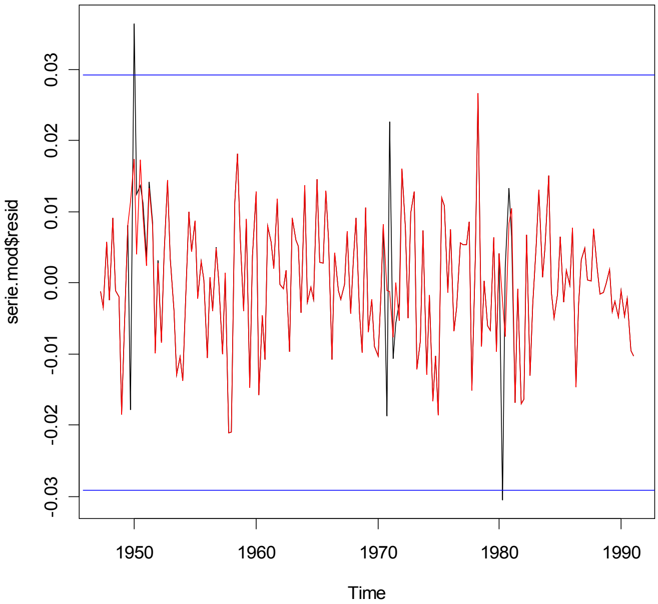
GNPSH Case

a) Study of the outliers for the ***ARIMA(3,1,0)*** model.

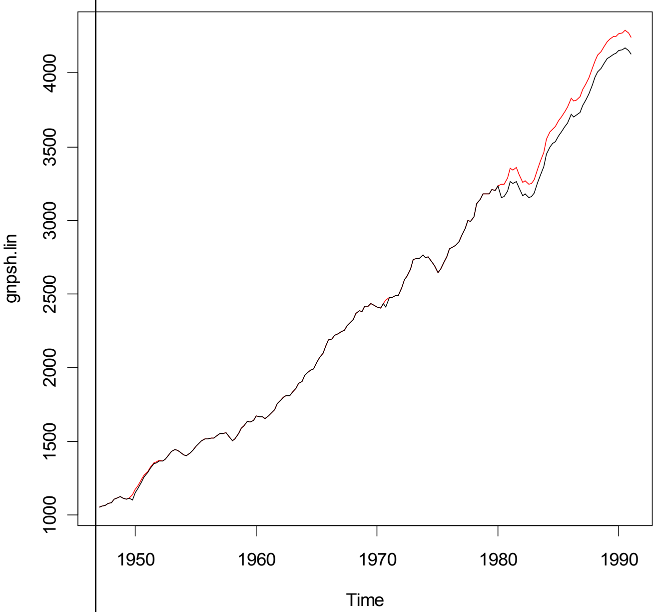
```
serie.mod<-arima(d1lngnpsh,order=c(3,0,0))


Call:
arima(x = d1lngnpsh, order = c(3, 0, 0))

Coefficients:
         ar1     ar2      ar3  intercept
      0.3480  0.1793  -0.1423     0.0077
s.e.  0.0745  0.0778   0.0745     0.0012

sigma^2 estimated as 9.427e-05:  log likelihood = 565.84,  aic = -1121.69
```

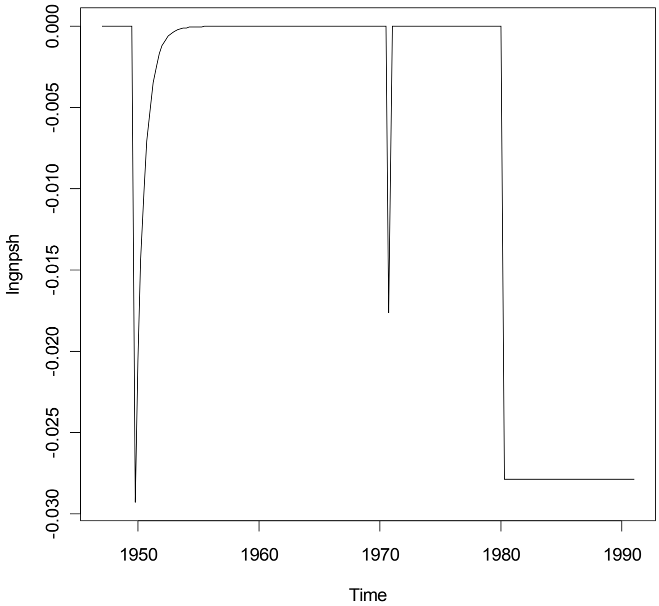


```
mod.atip<-outdetec(serie.mod, dif=c(1,0), crit=3.2, LS=T)

  Obs type_detected      W_coeff ABS_L_Ratio
1  11            TC -0.02928480    3.877097
2  95            AO -0.01766164    3.369289
3 133            LS -0.02788030    3.455208

plot(serie.mod$resid)
abline(h=3*sd(serie.mod$resid),col=4,lty=2)
abline(h=-3*sd(serie.mod$resid),col=4,lty=2)
lines(mod.atip$resid,col=2)
```




```
lngnpsh.lin<-c(lngnpsh[1],lineal(lngnpsh[-1],mod.atip$atip))
gnpsh.lin<-ts(exp(lngnpsh.lin),start=1947,freq=4)

plot(gnpsh.lin,col=2)
lines(gnpsh)

plot(lngnpsh-log(gnpsh.lin))
```



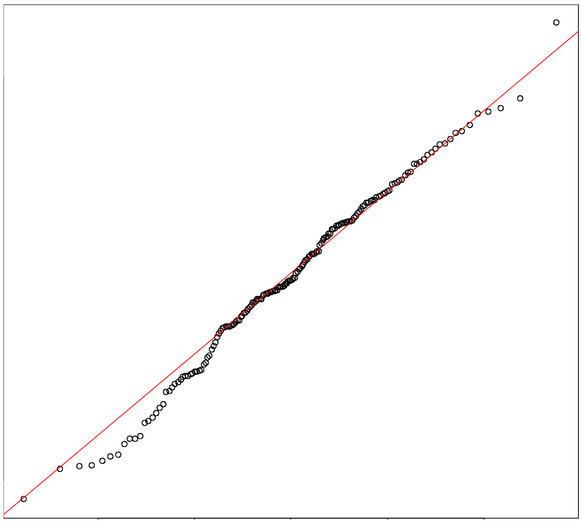



```
Call:
arima(x = d1lngnpsh.lin, order = c(3, 0, 0))
Coefficients:
         ar1     ar2      ar3  intercept
      0.4542  0.0841  -0.1209     0.0078
s.e.  0.0747  0.0820   0.0748     0.0011

sigma^2 estimated as 7.534e-05:  log likelihood = 585.55,  aic = -1161.1
```

Model:

$$
\boxed{\begin{array}{l}(1-0.45B+0.08B^{2}-0.12B^{3})(\nabla(\ln X_t-0.029I_1t-0.018I_2t-0.028I_3t)-0.0078)=Z_t \\ Z_t \sim N(0{,}7.5e-5)\end{array}}
$$

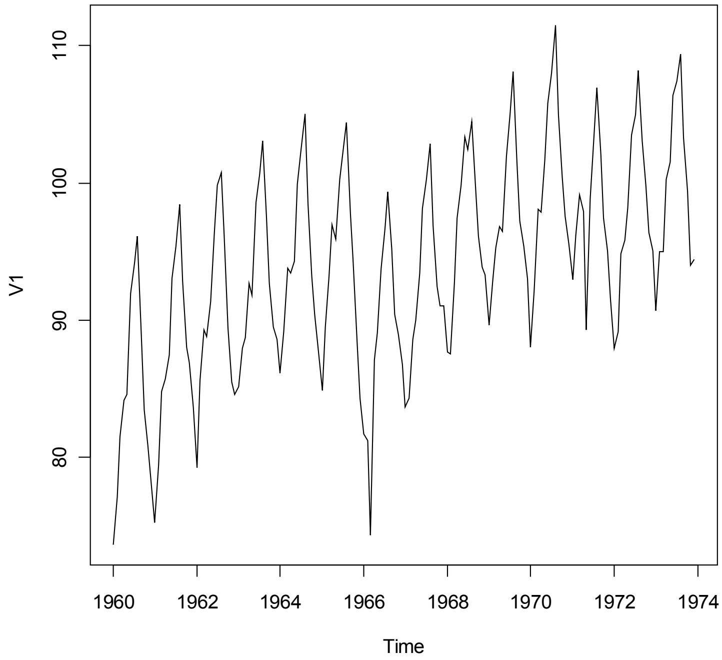
$$
AO_1(4/49):\quad I_1t=\begin{cases}1 & t=12\\ 0 & t\neq 12\end{cases}\qquad AO_2(4/70):\quad I_2t=\begin{cases}1 & t=96\\ 0 & t\neq 96\end{cases}\qquad LS(2/80):\quad I_3t=\begin{cases}1 & t\geq 134\\ 0 & t<134\end{cases}
$$



TSGGB Case

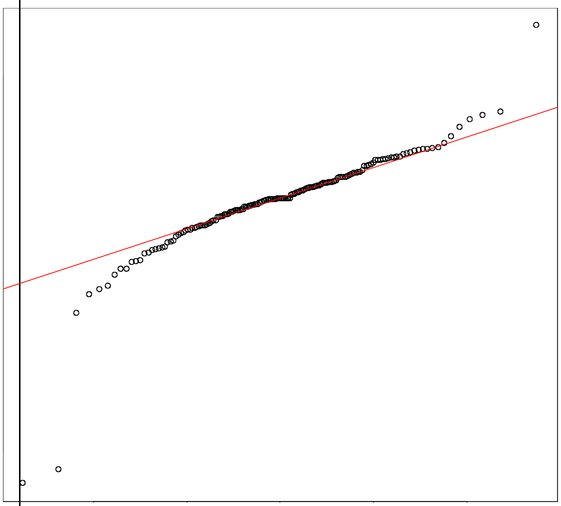
Monthly amount of vehicles driving on the Golden Gate Bridge between January 1967 and December 1980 (168 observations). The example is included in the manual STATGRAPHICS.

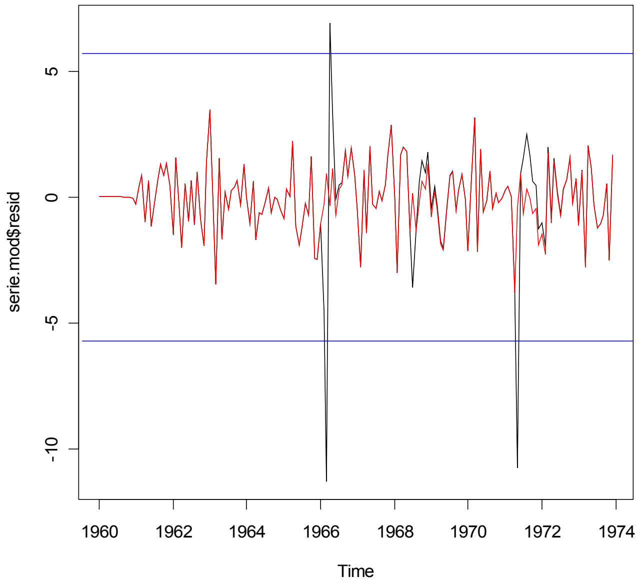
b) Study of the outliers for the ***ARIMA(0,1,1) (0,1,1)$_{12}$*** model



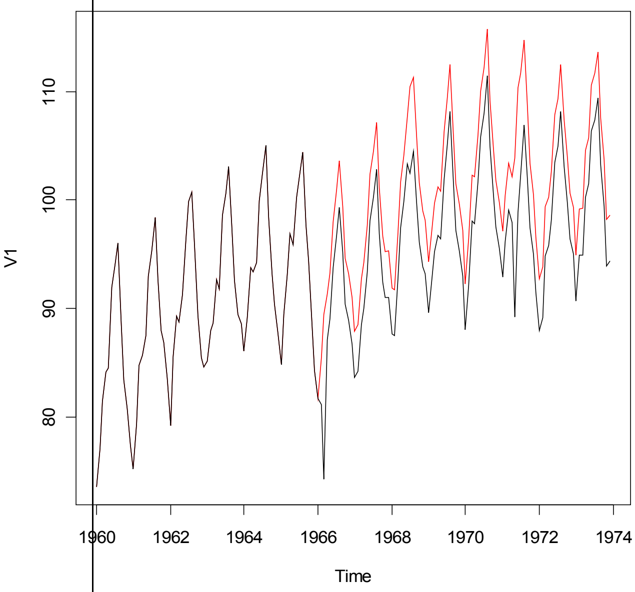

```
serie.mod<-arima(tsggb,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12))

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

Coefficients:
          ma1      sma1
      -0.3005   -1.0000
s.e.   0.0975    0.1041

sigma^2 estimated as 3.892:  log likelihood = -341.09,  aic = 688.17
```

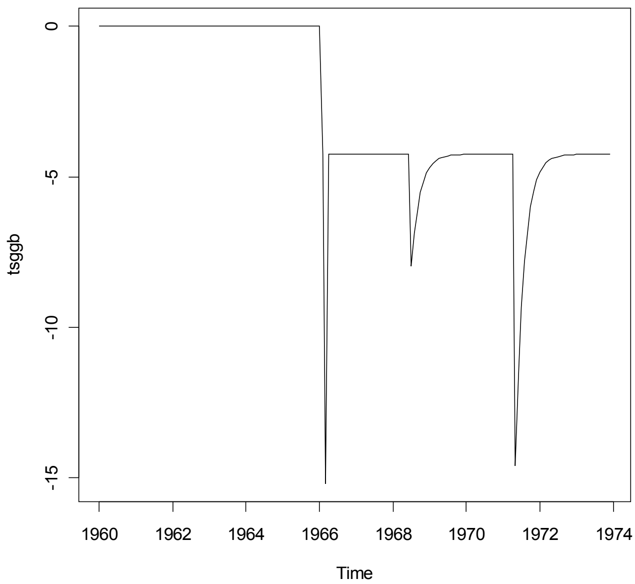
```
qqnorm(serie.mod$resid)
qqline(serie.mod$resid,col=2)


mod.atip<-outdetec(serie.mod, dif=c(1,12), crit=3.1, LS=T)
mod.atip$atip

  Obs type_detected    W_coeff ABS_L_Ratio
1  75            AO -10.925506    8.572121
2 137            TC -10.342355    8.323282
3  74            LS  -4.249713    3.468008
4 103            TC  -3.726305    3.197389

plot(serie.mod$resid)
lines(mod.atip$resid,col=2)
abline(h=3*sd(serie.mod$resid),col=4,lty=2)
abline(h=-3*sd(serie.mod$resid),col=4,lty=2)
```






```
tsggb.lin<-ts(lineal(tsggb,mod.atip$atip),start=1960,freq=12)
plot(tsggb.lin,col=2)
lines(tsggb)
plot(tsggb-tsggb.lin)
```



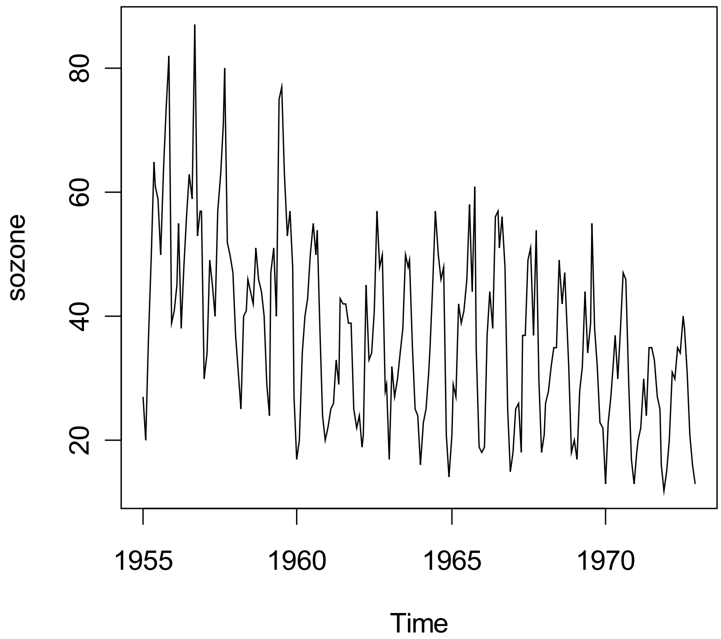



```
serielin.mod<-arima(tsggb.lin,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12))
Call:
arima(x = tsggb.lin, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1),
    period = 12))

Coefficients:
          ma1     sma1
      -0.3125  -0.8975
s.e.   0.0985   0.1083

sigma^2 estimated as 1.922:  log likelihood = -280.14,  aic = 566.29
```

26) Exercise: SOZONE Case

SOZONE Case

Monthly average of the hourly ozone concentration observations (pphm), from January of 1955 until December of 1972.



There are two interventions affecting the pollution in the city centre of Los Angeles:

- In 1960, the traffic was diverted to the Golden State Freeway, and a new law determined the maximum proportion of reactive hydrocarbons allowed in the gasoline.
- In the 1966 came into effect new regulations that forced a change in the design of the engines in order to reduce the production of pollution by the new cars.

- Fit an ARIMA model and study the outliers.

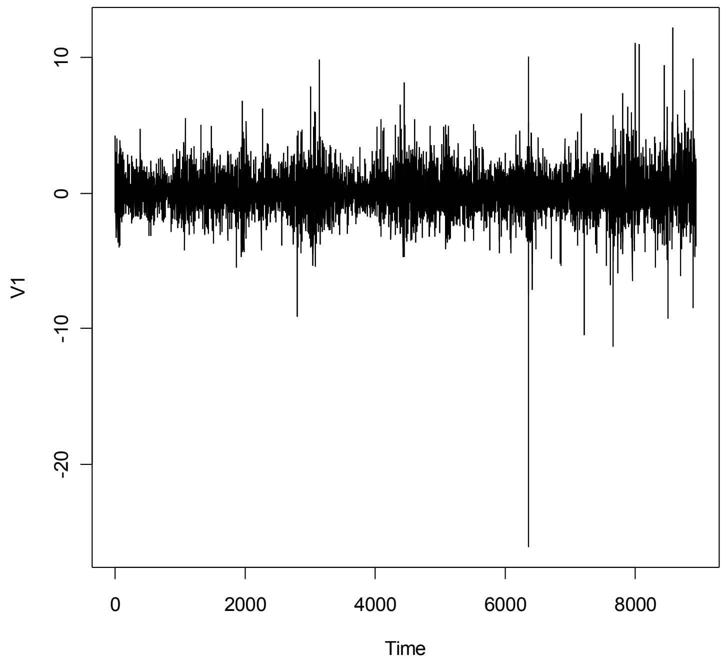
Practices: Session 7

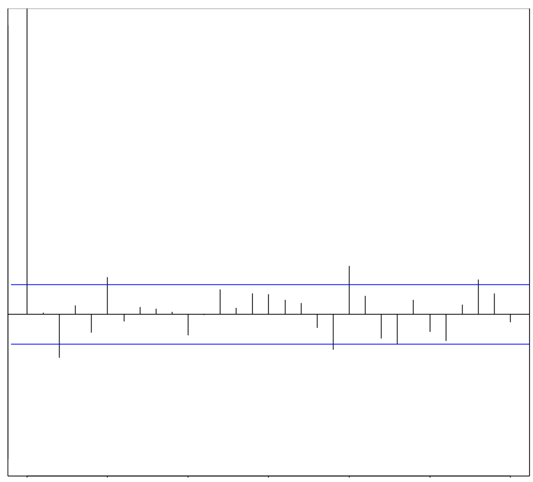
27) Practical cases: Series with volatility. IBM Case

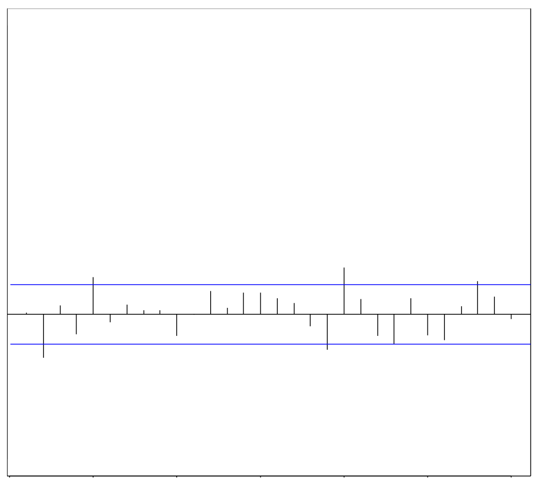
IBM Case:

Daily yields of the IBM shares $100(logX_t-logX_{t-1})$

IBM Series: Application of the Box-Jenkins methodology to construct a forecasting model from an ARIMA model:







The model is simplified by assuming that the non-significant lags are zero, so we only consider the first 5 lags and it can be assumed that the model is an AR(5).

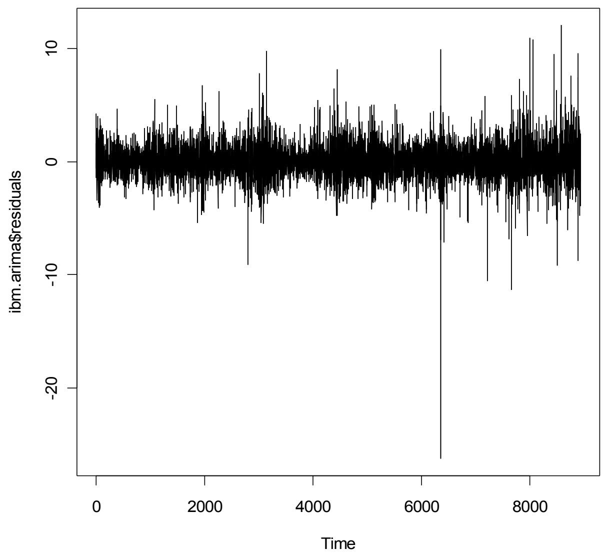
```
ibm.arima<-arima(ibm,order=c(5,0,0))

Call:
arima(x = ibm, order = c(5, 0, 0))

Coefficients:
         ar1      ar2     ar3      ar4     ar5  intercept
      0.0021  -0.0308  0.0071  -0.0137  0.0259     0.0393
s.e.  0.0106   0.0106  0.0106   0.0106  0.0106     0.0155

sigma^2 estimated as 2.189:  log likelihood = -16182.75,  aic = 32379.5


ibm.arima<-arima(ibm,order=c(5,0,0),fixed=c(0,NA,0,0,NA,NA))
Call:
arima(x = ibm, order = c(5, 0, 0), fixed = c(0, NA, 0, 0, NA, NA))

Coefficients:
      ar1      ar2  ar3  ar4     ar5  intercept
        0  -0.0304    0    0  0.0257     0.0393
s.e.    0   0.0106    0    0  0.0106     0.0156

sigma^2 estimated as 2.189:  log likelihood = -16183.83,  aic = 32375.65
```

Estimated model:

$$(1 + 0.0308B^2 - 0.0257B^5)(R_t - 0.0393) = Z_t \qquad Z_t \sim N(0,2.189^2)$$

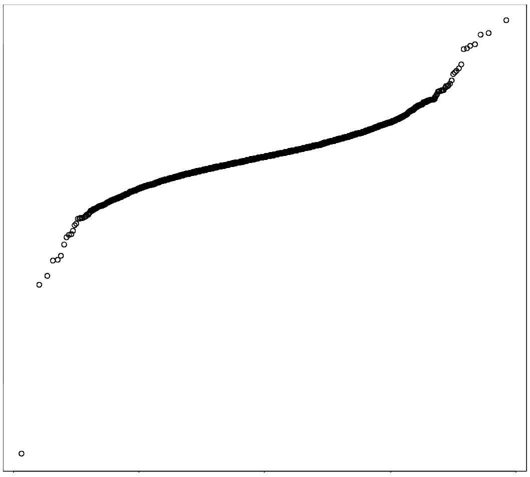
Validation of the process by analysing the residuals:

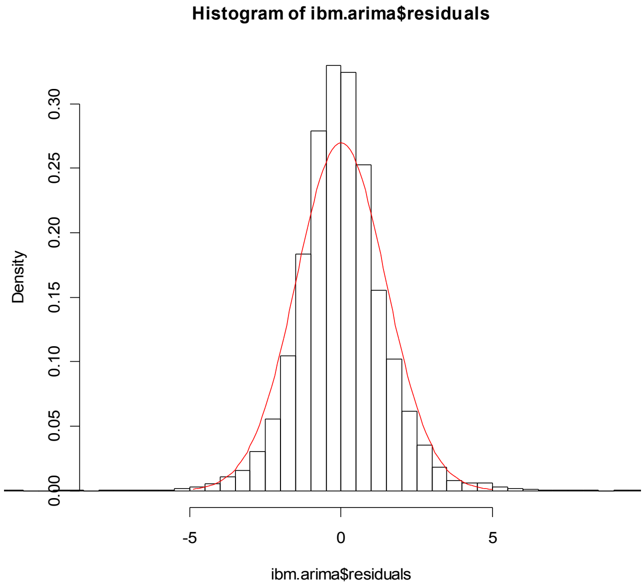
```
plot(ibm.arima$residuals)
```



Residuals plot: It can be appreciated that there are some periods with larger variance than others so the variance is non constant. There are also some outliers.

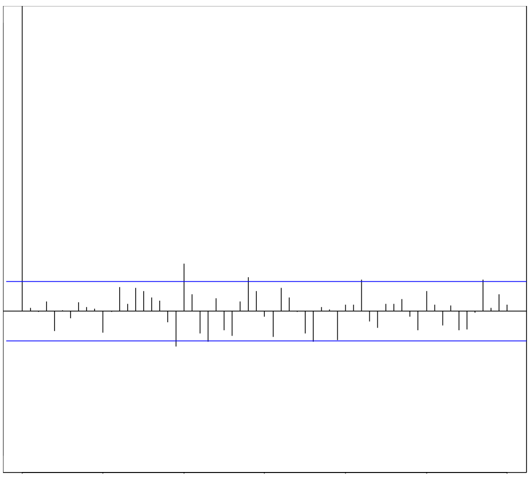
```
qqnorm(ibm.arima$residuals)
hist(ibm.arima$residuals,xlim=c(-8,8),breaks=60,prob=T)
x<--8+(0:100)*16/100
y<-dnorm(x,mean(ibm.arima$residuals),sd(ibm.arima$residuals))
lines(x,y,col=2)
```

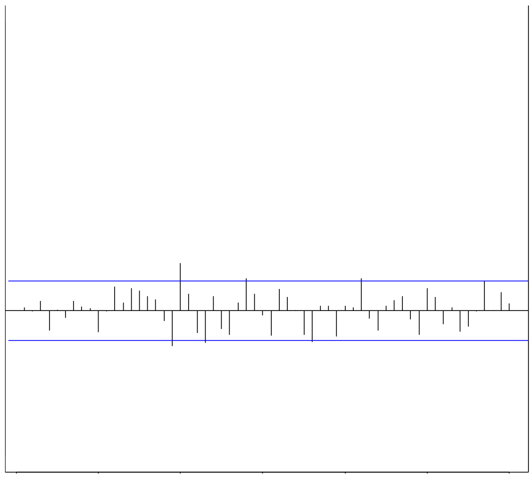




Normality of the residuals: The residuals behave normally but they have heavy tails (Kurtosis is too large to be a normal distribution)

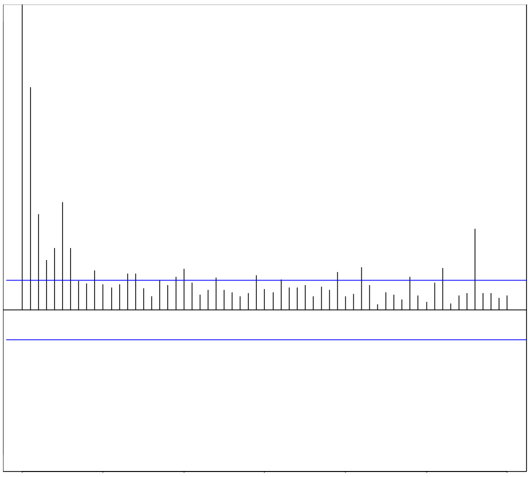
```
acf(ibm.arima$residuals,ylim=c(-0.10,0.2),lag.max=60)
pacf(ibm.arima$residuals,ylim=c(-0.10,0.2),lag.max=60)
```

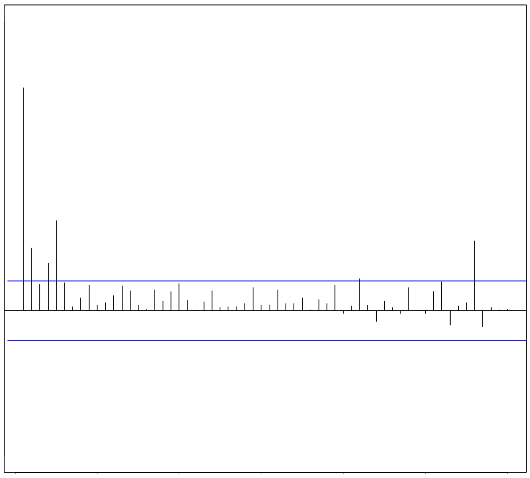




ACF and PACF of the residuals: there is an autocorrelation pattern at the first lags.

```
acf(ibm.arima$residuals^2,ylim=c(-0.10,0.2),lag.max=60)
pacf(ibm.arima$residuals^2,ylim=c(-0.10,0.2),lag.max=60)
```

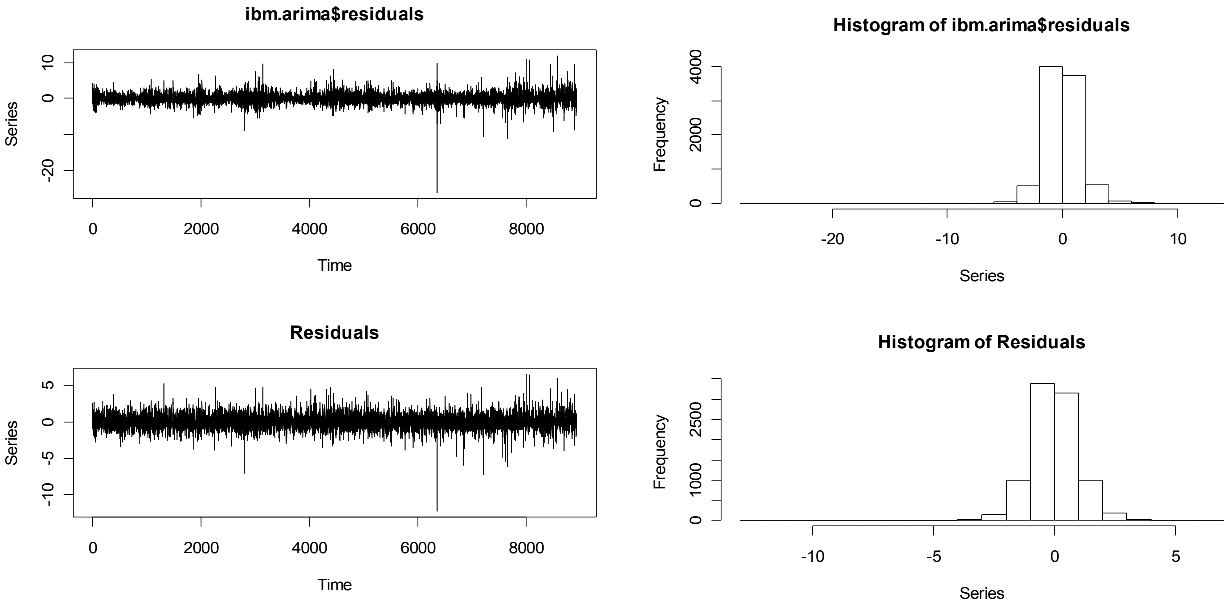




Squared ACF and PACF of the residuals: there are correlation between the squared residuals hence there is volatility.

Estimation of a GARCH model to fit the conditioned heteroscedasticity:

```
ibm.garch<- garch(ibm.arima$residuals,order=c(1,1))

Call:
garch(x = ibm.arima$residuals, order = c(1, 1))

Coefficient(s):
      a0        a1        b1
0.02999   0.06734   0.92205

summary(ibm.garch)

Call:
garch(x = ibm.arima$residuals, order = c(1, 1))

Model:
GARCH(1,1)

Residuals:
      Min         1Q     Median         3Q        Max
-12.31929   -0.61200   -0.02158    0.57027    6.56388

Coefficient(s):
    Estimate   Std. Error   t value  Pr(>|t|)
a0  0.029986     0.003172     9.454    <2e-16 ***
a1  0.067343     0.001966    34.261    <2e-16 ***
b1  0.922048     0.003123   295.265    <2e-16 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1



        Box-Ljung test

data:  Squared.Residuals
X-squared = 3.19, df = 1, p-value = 0.07409
```



Estimated Model:

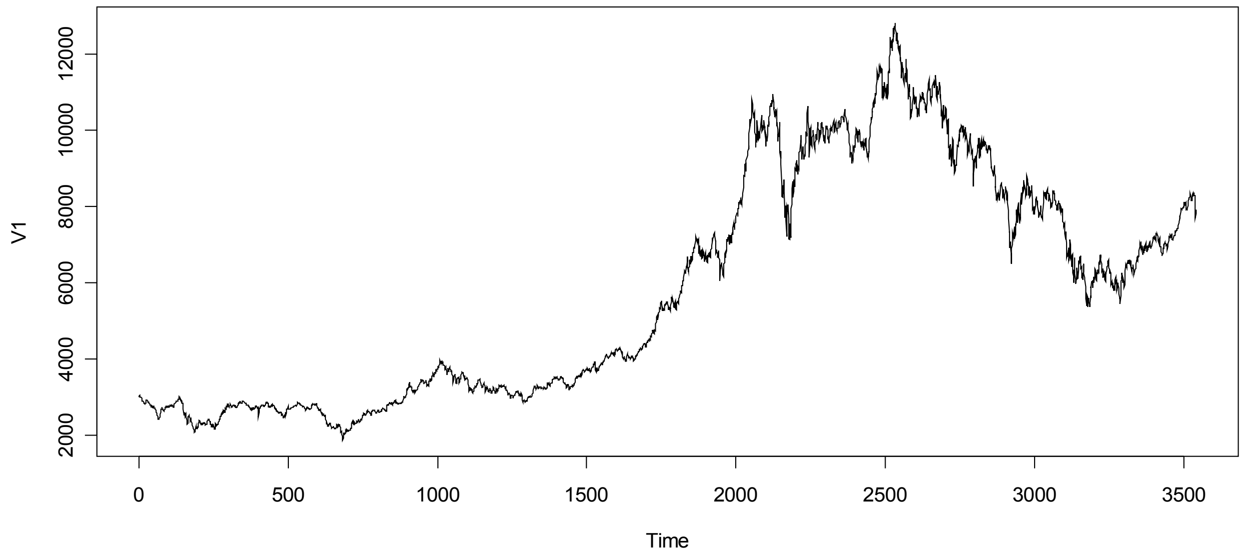
$$(1 + 0.0308B^2 - 0.0257B^5)(R_t - 0.0393) = Z_t \qquad Z_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = 0.02999 + 0.06734\frac{Z_{t-1}^2}{2.189} + 0.92205\sigma_{t-1}^2$$

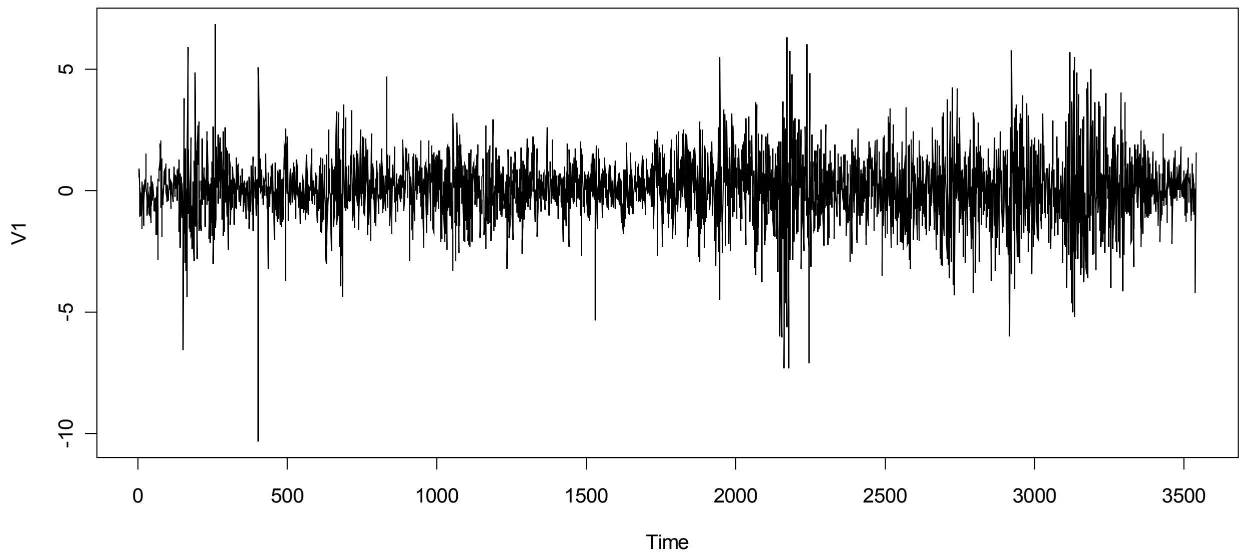
28) Exercise: IBEX35 Case

IBEX35 Case

Closing value of the Madrid's stock market index benchmark: IBEX 35 29/12/1989 to 17/03/2004 (3.540 observations)

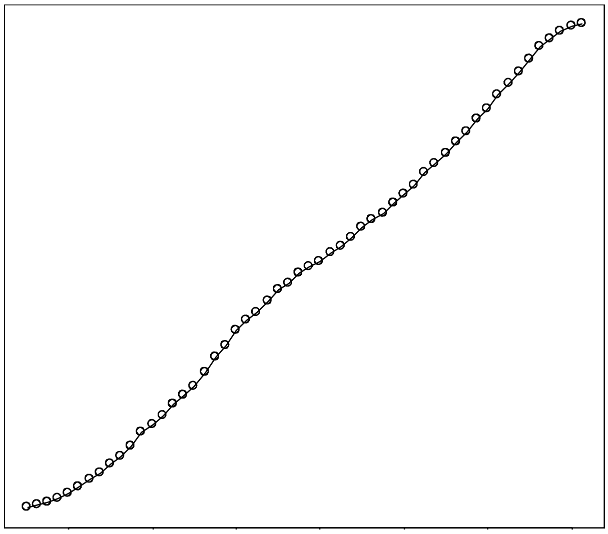


Yields of the IBEX35 index: $100(\log X_t - \log X_{t-1}) = 100(1-B)\log(X_t)$

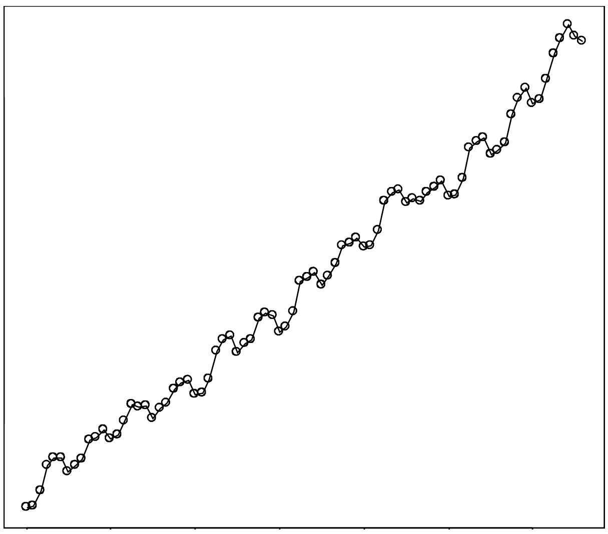


- Fit a long AR model and study its volatility.

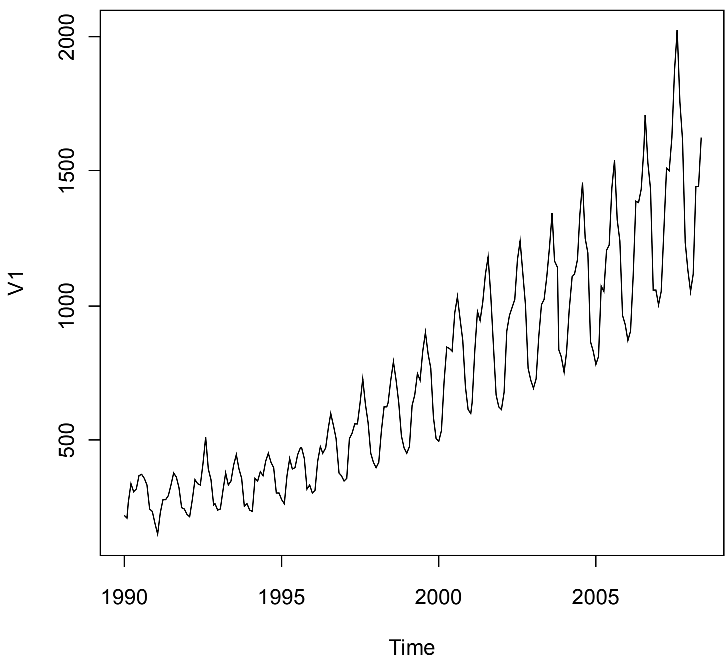
Appendix I: Proposed Series



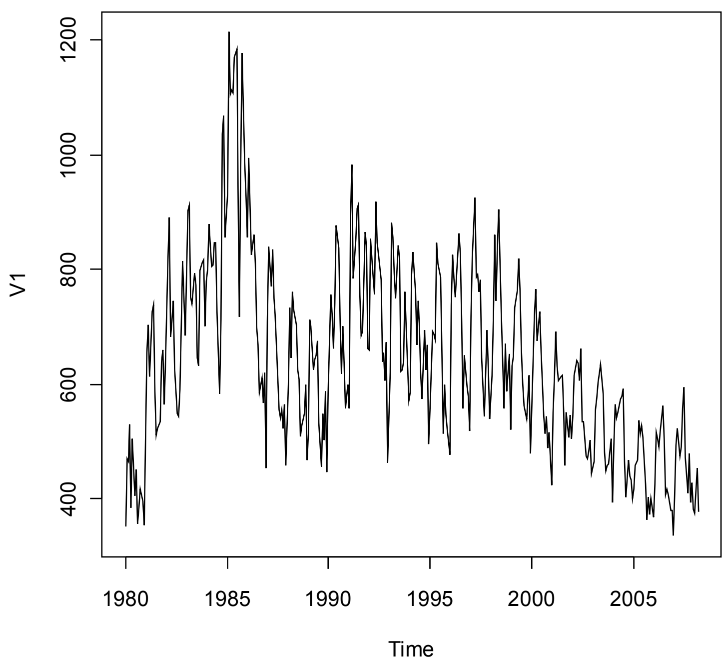
PIBsp
http://www.ine.es/
INEbase/Economy/National accounts/Quarterly Spanish National Acounts. Base 2000.
Producto Interior bruto a precios de mercado.Oferta (Indices de vol.).Datos corregidos de estacionalidad y calendario.
1T1955 a 2T2008: 54 datos trimestrales.
http://www.ine.es/jaxi/tabla.do?type=pcaxis&path=/t38/p604/a2000/l0/&file=0600002.px&L=1



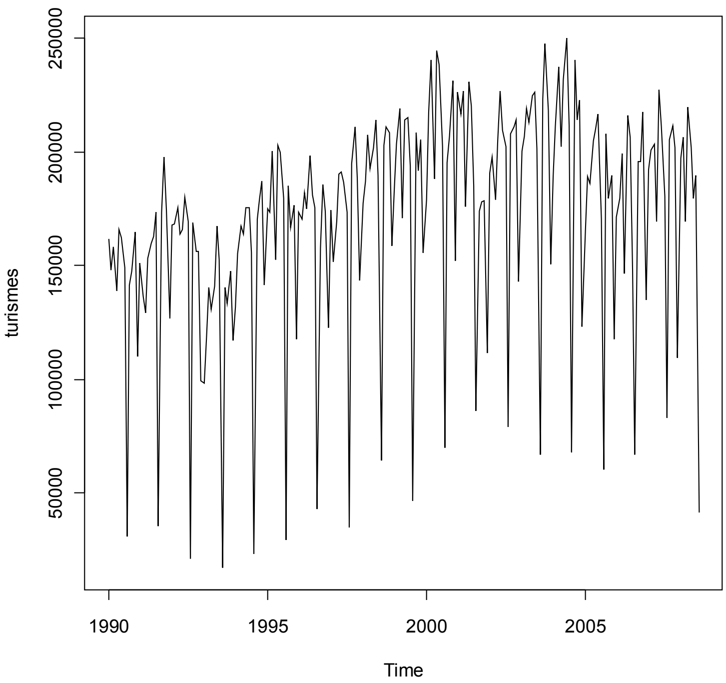
IPCsp
http://www.ine.es/
INEbase / Sociedad / Nivel, calidad y condiciones de vida / Índice de precios de consumo. Base 2006
consumer price index . National indexes: general and COICOP groups. General Index. 1/2002 to 8/2008: 80 monthly observations.



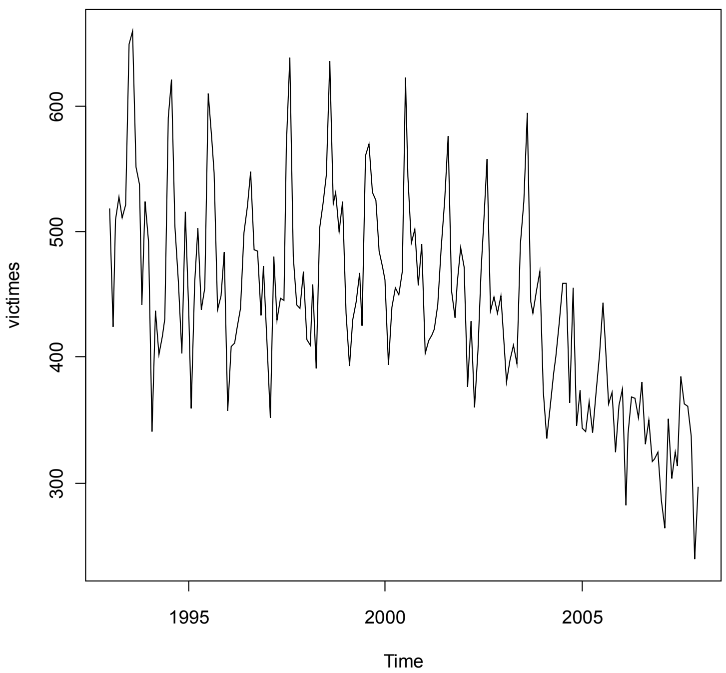
AirBCN
http://www.fomento.es/mfom/lang_castellano/
Información Estadística / Boletín On-line / Aviación Civil / 4.2 Trafico por aeropuertos. Barcelona
Passengers in miles of international flights. 1/1990 to 5/2008: 228 monthly data.



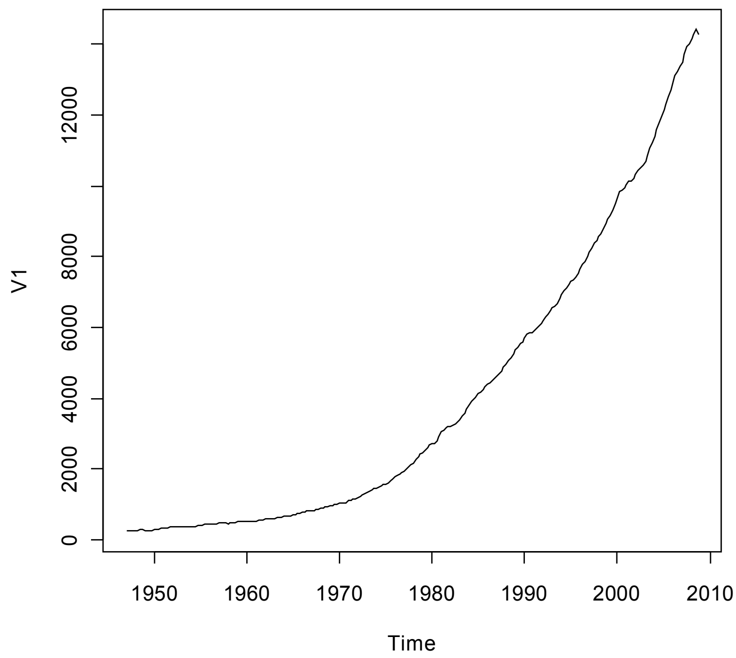
Tuberc
http://www.isciii.es/jsps/centros/epidemiologia/boletinesSemanal.jsp
Centro Nacional de Epidemiología > Vigilancia Epidemiólogica > Boletines
Weekly epidemiological Bulletin: N semanal: New respiratory tuberculosis cases in Spain.
1cs/1980 a 3cs/2008: 367 four weekly data (1 year = 13 x four weeks)



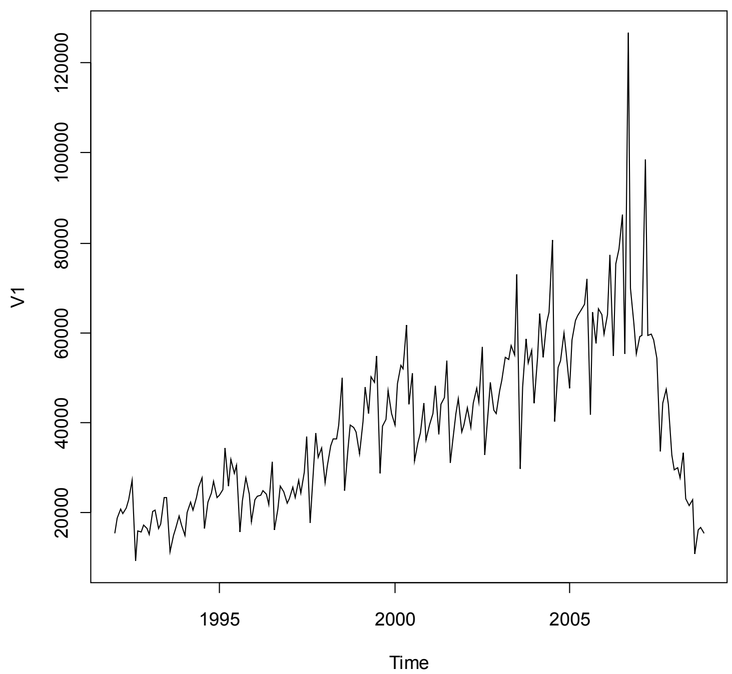
Cars
http://www.ine.es/
INEbase / Industria, Energía, Construcción / Industria / Fabricación de vehículos. Turismos
Monthly amount of cars made in Spain.
1/1990 to 8/2008: 224 monthly observations



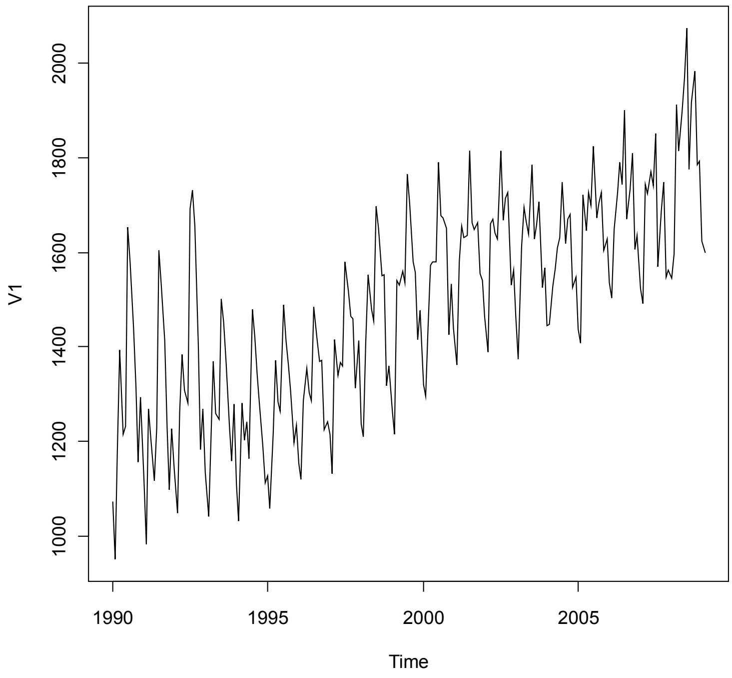
Victims
http://www.dgt.es/portal/es/seguridad_vial/estadistica/accidentes_30dias/series_historicas_accidentes/
Monthly amount of deaths by traffic accidents in Spain
1/1993 to 12/2007: 180 monthly obserbations



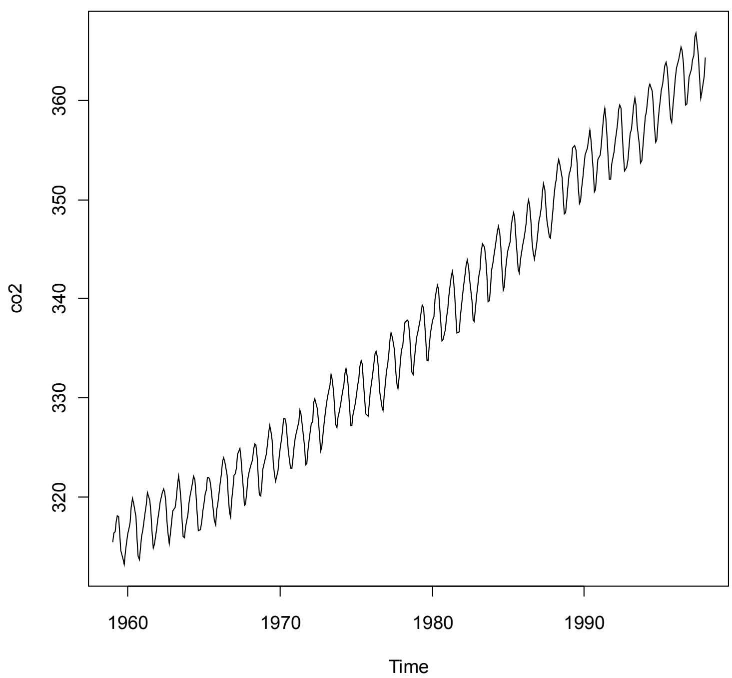
rgnp08
http://www.bea.gov/
Gross Domestic Product (GDP) / Current-dollar and "real" GDP
Gross Domestic Product of the US in thousands of dollars. Seasonally adjusted quarterly data.
Q1 1947 to Q4 2008: 248 quarterly observations



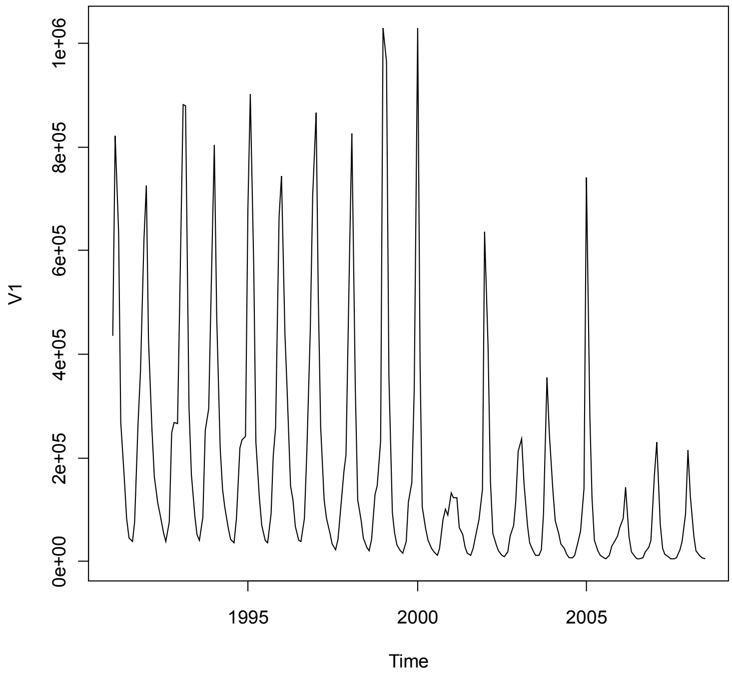
Acommodation
http://www.meh.es/
Estadísticas e Informes / Indicadores económicos / Base de dados de series de coyuntura económica / Indicadores de Producción y Demanda Nacional / Construcción / EOE. Numero de viviendas total obra nueva
Amount of new houses in Spain.
1/1992 to 11/2008: 203 monthly observations



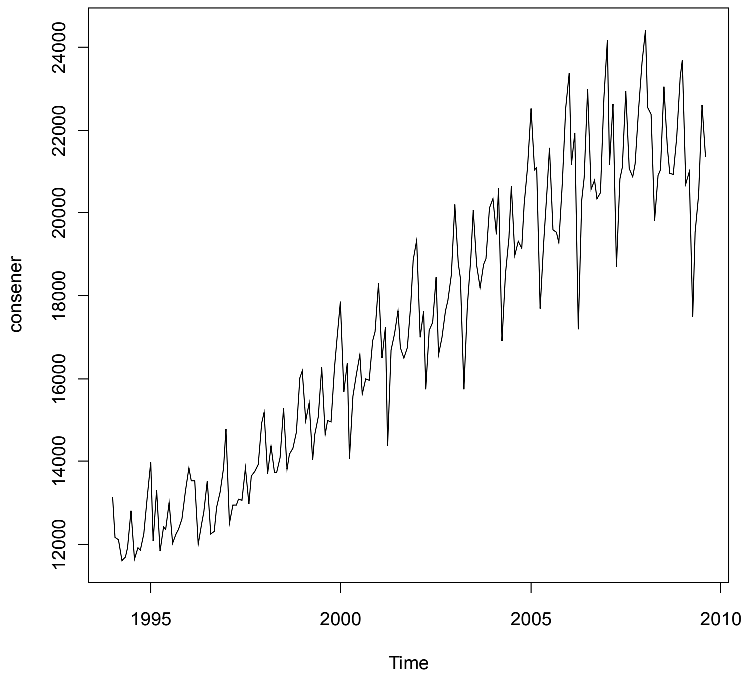
Renfe
http://www.meh.es/
Estadísticas e Informes / Indicadores económicos / Base de dados de series de coyuntura económica / Indicadores de Producción y Demanda Nacional / Servicios / Transporte RENFE. Pasajeros
Amount of monthly passengers using RENFE
1/1990 to 2/2009: 230 monthly observations



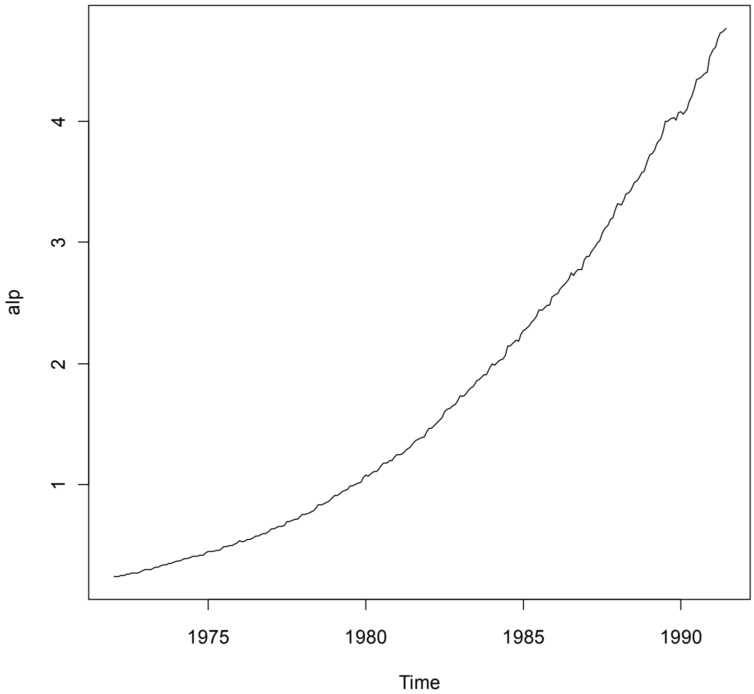
co2
ftp://cdiac.esd.ornl.gov/pub/maunaloa-co2/maunaloa.co2. In R.
CO2 concentration, expressed in ppm (parts per milion), in the atmosphere near the Mauna vulcano in Loa (Hawaii)
1/1959 to 12/1997: 468 monthly observations



Grip
INE / Sociedad / Salud / Enfermedades de Declaración Obligatoria / Gripe
Monthly mount of notified flu cases in Spain
1/1991 to 7/2008: 211 monthly observations

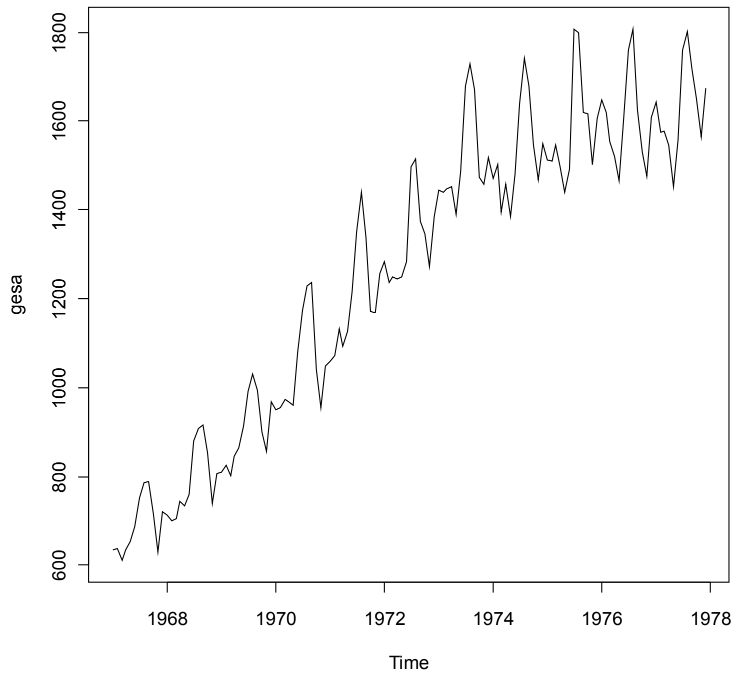


consener
http://www.meh.es/
Estadísticas e Informes / Indicadores económicos / Base de dados de series de coyuntura económica / Indicadores de Producción y Demanda Nacional / Otros indicadores de actividad / Consumo de energía eléctrica
Energy consumption in Spain measured at the power stations (available energy)
1/1994 to 8/2009: 188 monthly observations.



ALP
http://www.meh.es/
Spanish monetary aggregation (Public liquid assets)

1/1972 to 6/1991: 234 monthly observations



GESA
Top monthly energy consumption in the GESA company

1/1967 to12/1977. 132 monthly observations

Appendix II: Basic R Instructions

`#...` Write comments

```
# This is a comment and it will not be executed
```

`help(), ?...` R commands help

```
> help(ls)
> ?ls
```

`search()` List of the declared libraries

```
> search()
[1] ".GlobalEnv"      "package:methods"  "package:stats"    "package:graphics" "package:utils"
[6] "Autoloads"       "package:base"
```

`ls(name, pos = -1...)` List of the files in a library

```
> ls(,3)
  [1] "acf"                "acf2AR"              "add.scope"              "add1"
  [5] "addmargins"         "aggregate"           "aggregate.data.frame"   "aggregate.default"
  [9] "aggregate.ts"       "AIC"                 "alias"                  "anova"
 [13] "anova.glm"          "anova.glmlist"       "anova.lm"               "anova.lmlist"
         :                    :                     :                              :
Main functions for time series:
 [1] "acf"                 "acf2AR"              "ar"                     "ar.burg"
 [5] "ar.mle"              "ar.ols"              "ar.yw"                  "arima"
 [9] "arima.sim"           "arima0"              "arima0.diag"            "ARMAacf"
[13] "ARMAtoMA"            "bandwidth.kernel"    "Box.test"               "ccf"
[17] "cpgram"              "decompose"           "df.kernel"              "diffinv"
[21] "embed"               "filter"              "HoltWinters"            "is.tskernel"
[25] "KalmanForecast"      "KalmanLike"          "KalmanRun"              "KalmanSmooth"
[29] "kernapply"           "kernel"              "lag"                    "lag.plot"
[33] "makeARIMA"           "monthplot"           "na.contiguous"          "pacf"
[37] "pacf.mts"            "plot.spec"           "plot.spec.coherency"    "plot.spec.phase"
[41] "PP.test"             "spec.ar"             "spec.pgram"             "spec.taper"
[45] "spectrum"            "stl"                 "StructTS"               "toeplitz"
[49] "ts.intersect"        "ts.plot"             "ts.union"               "tsdiag"
[53] "tsSmooth"
```

`read.table(file, header = FALSE, sep = ""...)` Read text files and assign the content in a file

```
gnpsh<-read.table("c:\\gnpsh.dat")
```

`ts(data = x, start = 1, frequency = 1...)` Create a "ts" file indicating the period start

```
gnpsh<-ts(gnpsh,start=1947,frequency=4)
```

`plot(x,...)` Plot an file

```
plot(gnpsh)
```

Arithmetic Operations benchmark

```
x1<-abs(x)       #absolute value
x1<-x1+x2        #sum
x1<-x1-x2        #subtraction
x1<-x1*x2        #product
x1<-x1/x2        #division
x2<-log(x1)      #logarithm
x3<-sqrt(x1)     #square root
x4<-exp(x1)      #exponential
x5<-sin(x1)      #sinus
x6<-cos(x1)      #cosine
x7<-tan(x1)      #tangent
x8<-x1^0.4       #power
```

Statistical Functions Statistical functions

```
m<-mean(x)               #mean
md<-median(x)            #median
M<-max(x)                #maximum
m<-min(x)                #minimum
v<-var(x)                #variance
s<-sd(x)                 #standard deviation
x2<-quantile(x,0.25)     #percentiles
summary(x)                #position measures for numerical variables
```

acf(x, lag.max, type = c("corr", "cov", part"), plot = TRUE ...)

Calculation and plot of the autocorrelation, autocovariance or partial autocorrelation of a series

```
acf(lngnpsh,ylim=c(-1,1))
```

pacf(x, lag.max, plot ...)

Calculation and plot of the partial autocorrelation of a series

```
pacf(lngnpsh,ylim=c(-1,1))
```

diff(x, lag = 1...)

Differentiation with the indicated lag of a series

```
d1lngnpsh<-diff(lngnpsh, lag=1)
```

decompose(x, type = c("add", "mult")...)

Decomposition of a series in three components: trend, seasonality and random noises

```
decompose(lnairpass)
```

names(x)

Obtaining the field that has a determined file

```
aux<-decompose(lnairpass)
names(aux)
[1] "seasonal" "trend"    "random"   "figure"   "type"
```

x$... Obtaining the field that has a determined file

```
x$figure
 [1] -0.08790254 -0.11205814  0.02752720 -0.01539480 -0.01243591  0.11240967
 [7]  0.21445629  0.20790531  0.06326555 -0.07703544 -0.21786902 -0.10286816
```

rnorm(n, mean=0, sd=1) Generation of n random values with a Normal distribution

```
x<-rnorm(100,0,1)
```

arima.sim(model, n, rand.gen = rnorm ...) Simulation of a series with an ARIMA model

```
ts.sim <- arima.sim(list(order = c(1,1,0), ar = 0.7), n = 200)
```

ARMAacf(ar, ma, lag.max, pacf = FALSE) Obtaining the theoretical ACF and PACF of an ARMA model

```
ARMAacf(ar=c(1.0, -0.25), lag.max = 10)
```

ARMAtoMA(ar, ma, lag.max) Expression as a MA of infinite order

```
ARMAtoMA(ar=c(1.0, -0.25), lag.max = 10)
[1] 1.00000000 0.75000000 0.50000000 0.31250000 0.18750000 0.10937500
[7] 0.06250000 0.03515625 0.01953125 0.01074219
```

arima(x, order = c(0, 0, 0), seasonal = list(order = c(0, 0, 0), period = NA), include.mean = TRUE, fixed = NULL) Estimation of an ARIMA model

```
airpass.arima <-arima(lnairpass,order=c(0,1,1),seasonal=list(order=c(0,1,1), period=12)
Call:
arima(x = lnairpass, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1),
    period = 12))

Coefficients:
          ma1     sma1
      -0.4018  -0.5569
s.e.   0.0896   0.0731

sigma^2 estimated as 0.001348:  log likelihood = 244.7,  aic = -483.4

names(airpass.arima)
 [1] "coef"      "sigma2"    "var.coef"  "mask"      "loglik"    "aic"
 [7] "arma"      "residuals" "call"      "series"    "code"      "n.cond"
[13] "model"
```