



---

# **TIME SERIES ANALYSIS**

---

**Notes by Professor Dr. Athar Ali Khan,  
Department of Statistics & Operations Research  
Aligarh Muslim University, Aligarh**



**MOHAMMAD WASIQ**

# TIME SERIES ANALYSIS

## DSM - 3001

### Unit I

**Characteristics of Time Series :** The nature of time series data, time series statistical models, measures of dependence, stationary time series, estimation of correlation, vector-valued and multidimensional series. time series regression and exploratory data analysis: classical regression in the time series context, exploratory data analysis, smoothing in the time series context.

### Unit II

**ARIMA Models :** Autoregressive moving average models, difference equations, autocorrelation and partial autocorrelation, forecasting, estimation, integrated models for nonstationary data, building ARIMA Models, regression with autocorrelated errors, multiplicative seasonal ARIMA Models.

### Unit III

**Spectral Analysis and Filtering :** Cyclical behavior and periodicity, the spectral density, periodogram and discrete fourier transform, nonparametric spectral estimation, parametric spectral estimation, multiple series and cross-spectra, linear filters, GARCH Models, Long Memory ARMA and Fractional Differencing, Unit Root Testing.

### Unit IV

**State Space Models :** Linear Gaussian Model, filtering, smoothing, and forecasting, maximum likelihood estimation, missing data modifications, structural models: signal extraction and forecasting, state-space models with correlated errors, bootstrapping state space models.

# DSM-3001

## Time Series Analysis

Mohammad Wasiq

### Table of Contents

1	Syllabus .....	6
1.1	Unit — I .....	6
1.2	Unit — II .....	6
1.3	Unit — III .....	6
1.3.1	Unit — IV .....	6
1.3.2	Books .....	7
2	UNIT — I .....	7
3	Introduction to Time Series .....	7
3.1	Examples of Time Series .....	7
3.1.1	Annual Rainfall in Los Angeles .....	7
3.1.2	Time Series Plot of Los Angeles Annual Rainfall .....	8
3.1.3	Scatterplot of LA Rainfall versus Last Year's LA Rainfall .....	9
3.2	An Industrial Chemical Process .....	9
3.2.1	Time Series Plot of Color Property from a Chemical Process .....	9
3.2.2	Scatterplot of Color Value versus Previous Color Value .....	10
3.3	Annual Abundance of Canadian Hare .....	10
3.3.1	Abundance of Canadian Hare .....	10
3.3.2	Hare Abundance versus Previous Year's Hare Abundance .....	11
3.4	Monthly Average Temperatures in Dubuque, Iowa .....	12
3.4.1	Average Monthly Temperatures, Dubuque, Iowa .....	12
3.5	Monthly Oil Filter Sales .....	12
3.5.1	Monthly Oil Filter Sales .....	13
3.6	Monthly Oil Filter Sales with Special Plotting Symbols .....	13
4	Fundamental Concepts of Time Series .....	14
4.1	Time Series and Stochastic Processes .....	14
4.1.1	Means, Variances, and Covariances .....	14
4.2	The Random Walk .....	15
4.2.1	Means, Variances, and Covariances .....	15

4.3	A Moving Average.....	17
4.4	Stationarity .....	18
4.4.1	Weakly Stationarity .....	19
4.4.2	White Noise .....	20
4.4.3	Random Cosine Wave.....	20
4.5	Different Series.....	21
5	Trends.....	21
5.1	Regression Methods for Estimation of Trends.....	21
5.1.1	Linear and Quadratic Trends in Time.....	21
5.1.2	Cyclical or Seasonal Trends .....	23
5.1.3	Cosine Trends.....	25
5.2	Residual Analysis.....	27
5.2.1	Sample Autocorrelation Function .....	30
6	MODELS FOR STATIONARY TIME SERIES.....	32
6.1	Auto-Regressive Moving Average (ARMA) models.....	32
6.2	General Linear Processes .....	32
6.3	Moving Average Processes — $MA(q)$ .....	33
6.3.1	The First-Order Moving Average Process — $MA(1)$ .....	34
6.3.2	The Second-Order Moving Average Process — $MA(2)$ .....	35
6.3.3	General Moving Average Process — $MA(q)$ .....	38
6.4	Autoregressive Processes — $AR(p)$ .....	38
6.4.1	The Second-Order Autoregressive Process — $AR(2)$ .....	43
6.4.2	Varicance for the $AR(2)$ Model .....	45
6.5	General Auto-Regressive Process .....	45
6.6	Mixed Autoregressive Moving Average Model — $ARMA(p, q)$ .....	46
6.6.1	$ARMA(1, 1)$ Model .....	46
7	Models for Non-Stationary Time Series — $ARIMA$ .....	47
7.1	Stationarity Through Differencing — $ARIMA$ Models .....	48
7.1.1	Constant Terms in $ARIMA$ Models .....	50
7.2	Other Transformation.....	50
7.2.1	Log Transformation — $\log()$ .....	50
7.2.2	Percentage Changes and Logarithms — returns.....	51
7.2.3	Power Transformations .....	53
7.3	The Backshift Operator .....	54

7.3.1	Moving Average Model — $MA(\cdot)$ Model .....	54
7.3.2	Auto-Regressive Models — $AR(p)$ Models .....	55
7.3.3	$ARMA(p, q)$ Models .....	55
8	Parameter Estimation .....	56
8.1	The Method of Estimation .....	56
8.1.1	1. Methods of Moments .....	56
8.1.2	2. Methods of Least Square Estimation .....	58
8.1.3	Maximum Likelihood Estimation — MLE .....	59
8.2	Properties of the Estimates .....	60
8.2.1	Parameter Estimates For Simulated Data Using $AR(1)$ Model .....	61
8.2.2	Parameter Estimates For Simulated Data Using $AR(2)$ Model .....	61
8.2.3	Canadian Hare Abundance Series — Hare .....	62
8.2.4	Fitting of $ARIMA$ Model — $AR(3)$ .....	64
8.3	Bootstrapping $ARIMA$ Models .....	65
8.4	MODEL DIAGNOSTICS .....	66
8.4.1	Standardized Residuals from Log Oil Price $IMA(1,1)$ Mode .....	67
8.4.2	Normality of the Residuals .....	68
8.4.3	Auto-Correlation Residuals — Box-Ljung Test .....	70
9	FORECASTING .....	72
9.1	Minimum Mean Square Error Forecasting .....	72
9.2	Deterministic Trends .....	72
9.3	$ARIMA$ Forecasting .....	73
9.3.1	$AR(1)$ .....	73
10	Seasonal Models .....	75
10.1	Seasonal $ARIMA$ Models .....	76
10.2	Model Specification .....	76
10.2.1	Model Fitting .....	80
11	TIME SERIES MODELS OF HETEROSCEDASTICITY .....	80
11.1	Some Common Features of Financial Time Series .....	80
11.2	$GARCH$ Models .....	85
11.3	Maximum Likelihood Estimation .....	89
11.3.1	$GARCH(1, 1)$ .....	89
12	Sequential Solution for Linear Gaussian State-Space Model .....	96
12.1	Kalman Filter .....	96

12.1.1	Kalman Filtering.....	97
12.1.2	Algorithm — Kalman Filtering.....	97
12.1.3	Algorithm — Kalman Prediction.....	97
12.1.4	Algorithm — Kalman Smoothing.....	98
12.2	Example: Local-Level Model Case.....	98
12.2.1	Specification of Parameter Values.....	99
12.2.2	Execution of Filtering, Prediction and Smoothing.....	101
12.2.3	Diagnostic Checking Results.....	106

This course is taught by **Prof. Dr. Athar Ali Khan**

## **1 Syllabus**

### **1.1 Unit — I**

**Characteristics of Time Series :** The nature of time series data, time series statistical models, measures of dependence, stationary time series, estimation of correlation, vector-valued and multidimensional series. time series regression and exploratory data analysis: classical regression in the time series context, exploratory data analysis, smoothing in the time series context.

### **1.2 Unit — II**

**ARIMA Models :** Autoregressive moving average models, difference equations, autocorrelation and partial autocorrelation, forecasting, estimation, integrated models for nonstationary data, building ARIMA Models, regression with autocorrelated errors, multiplicative seasonal ARIMA Models.

### **1.3 Unit — III**

**Spectral Analysis and Filtering :** cyclical behavior and periodicity, the spectral density, periodogram and discrete fourier transform, nonparametric spectral estimation, parametric spectral estimation, multiple series and cross-spectra, linear filters, GARCH Models, Long Memory ARMA and Fractional Differencing, Unit Root Testing.

#### **1.3.1 Unit — IV**

**State Space Models :** Linear Gaussian Model, filtering, smoothing, and forecasting, maximum likelihood estimation, missing data modifications, structural models: signal extraction and forecasting, state-space models with correlated errors, bootstrapping state space models.

### **1.3.2 Books**

#### **1. Time Series Analysis With Applications in R**

**Second Edition**

**Jonathan D. Cryer , Kung-Sik Chan**

#### **2. Time Series Analysis for the State-Space Model with R|Stan**

**Junichiro Hagiwara**

**2021**

## **2 UNIT — I**

**Characteristics of Time Series :** The nature of time series data, time series statistical models, measures of dependence, stationary time series, estimation of correlation, vector-valued and multidimensional series. time series regression and exploratory data analysis: classical regression in the time series context, exploratory data analysis, smoothing in the time series context.

## **3 Introduction to Time Series**

Data obtained from observations collected sequentially over time are extremely common. In business, we observe weekly interest rates, daily closing stock prices, monthly price indices, yearly sales figures, and so forth. In meteorology, we observe daily high and low temperatures, annual precipitation and drought indices, and hourly wind speeds. In agriculture, we record annual figures for crop and livestock production, soil erosion, and export sales. In the biological sciences, we observe the electrical activity of the heart at millisecond intervals. In ecology, we record the abundance of an animal species.

The purpose of time series analysis is generally twofold: to understand or model the stochastic mechanism that gives rise to an observed series and to predict or forecast the future values of a series based on the history of that series and, possibly, other related series or factors.

### **3.1 Examples of Time Series**

#### **3.1.1 Annual Rainfall in Los Angeles**

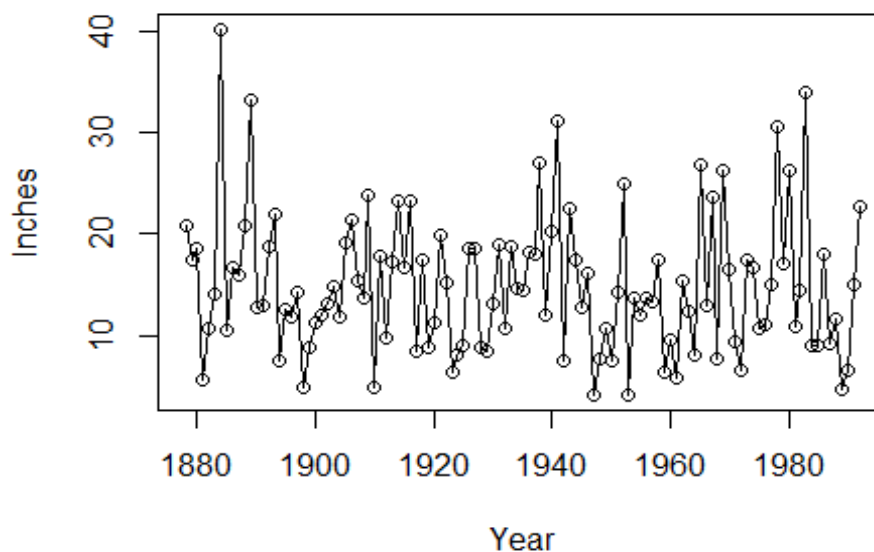
Exhibit 1.1 displays a time series plot of the annual rainfall amounts recorded in Los Angeles, California, over more than 100 years. The plot shows considerable variation in rainfall amount over the years—some years are low, some high, and many are in-between in value. The year 1883 was an exceptionally wet year for Los Angeles, While 1983 was quite dry. For analysis and modeling purposes we are interested in whether or not consecutive years are related in some way. If so, we might be able to use one year's rainfall

value to help forecast next year's rainfall amount. One graphical way to investigate that question is to pair up consecutive rainfall values and plot the resulting scatterplot of pairs.

### 3.1.2 Time Series Plot of Los Angeles Annual Rainfall

```
library(TSA)
```

```
##  
## Attaching package: 'TSA'  
  
## The following objects are masked from 'package:stats':  
##  
##   acf, arima  
  
## The following object is masked from 'package:utils':  
##  
##   tar  
  
data(larain)  
head(larain)  
  
## [1] 20.86 17.41 18.65  5.53 10.74 14.14  
  
plot(larain, ylab='Inches', xlab='Year' ,type='o')
```

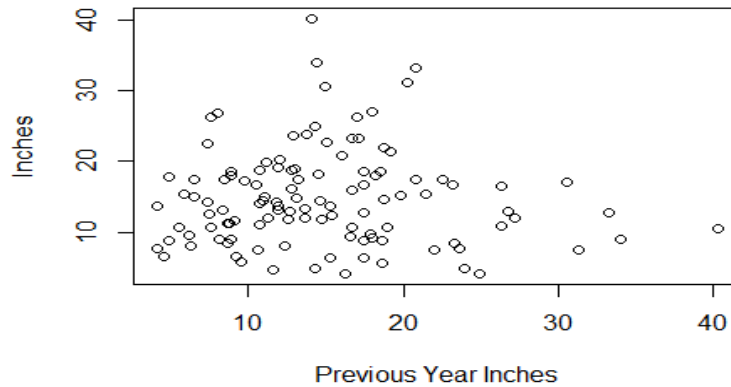


Above figure shows scatterplot for rainfall. For example, the point plotted near the lower right-hand corner shows that the year of extremely high rainfall, 40 inches in 1883, was followed by a middle of the road amount (about 12 inches) in 1884. The point near the top of the display shows that the 40 inch year was preceded by a much more typical year of about 15 inches.



### 3.1.3 Scatterplot of LA Rainfall versus Last Year's LA Rainfall

```
plot(y= larain, x= zlag(larain), ylab= 'Inches',  
     xlab= 'Previous Year Inches')
```



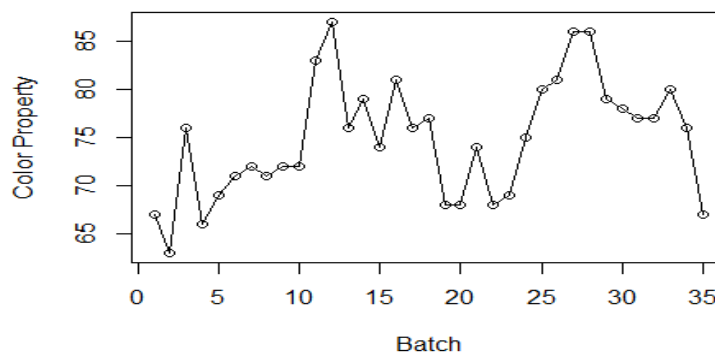
The main impression that we obtain from this plot is that there is little if any information about this year's rainfall amount from last year's amount. The plot shows no "trends" and no general tendencies. There is little correlation between last year's rainfall amount and this year's amount. From a modeling or forecasting point of view, this is not a very interesting time series.

## 3.2 An Industrial Chemical Process

As a second example, we consider a time series from an industrial chemical process. The variable measured here is a color property from consecutive batches in the process. Exhibit 1.3 shows a time series plot of these color values. Here values that are neighbors in time tend to be similar in size. It seems that neighbors are related to one another.

### 3.2.1 Time Series Plot of Color Property from a Chemical Process

```
data(color)  
plot(color, ylab= 'Color Property', xlab= 'Batch', type= 'o')
```

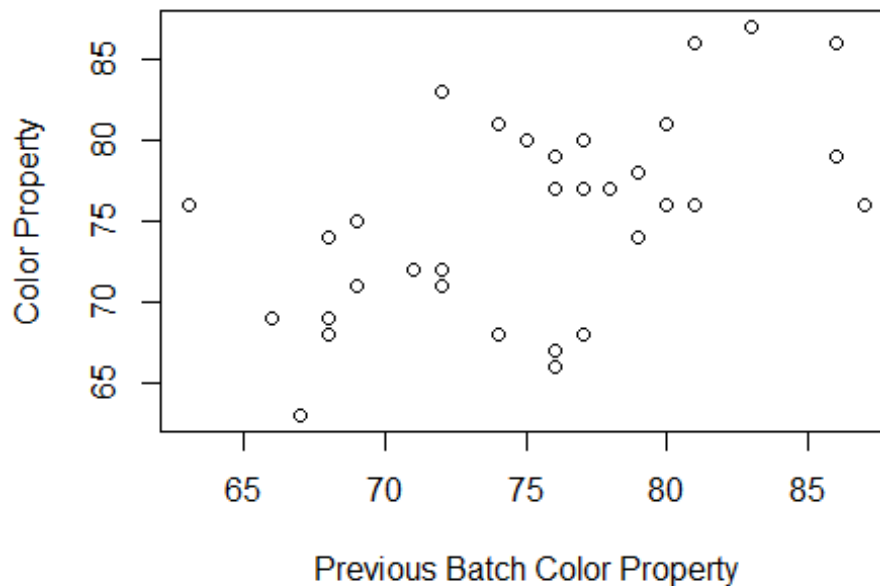


This can be seen better by constructing the scatterplot of neighboring pairs as we did with the first example.

Exhibit 1.4 displays the scatterplot of the neighboring pairs of color values. We see a slight upward trend in this plot—low values tend to be followed in the next batch by low values, middle-sized values tend to be followed by middle-sized values, and high values tend to be followed by high values. The trend is apparent but is not terribly strong. For example, the correlation in this scatterplot is about 0.6.

### 3.2.2 Scatterplot of Color Value versus Previous Color Value

```
plot(y= color, x= zlag(color), ylab= 'Color Property',  
      xlab= 'Previous Batch Color Property')
```

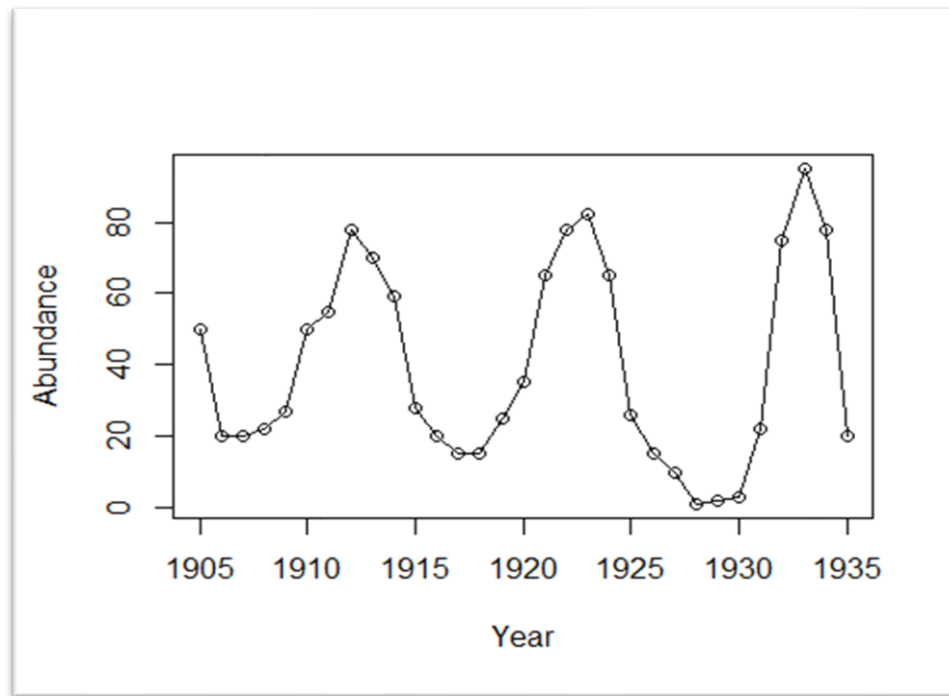


## 3.3 Annual Abundance of Canadian Hare

Our third example concerns the annual abundance of Canadian hare. Exhibit 1.5 gives the time series plot of this abundance over about 30 years. Neighboring values here are very closely related. Large changes in abundance do not occur from one year to the next. This neighboring correlation is seen clearly in Exhibit 1.6 where we have plotted abundance versus the previous year's abundance. As in the previous example, we see an upward trend in the plot—low values tend to be followed by low values in the next year, middle-sized values by middle-sized values, and high values by high values.

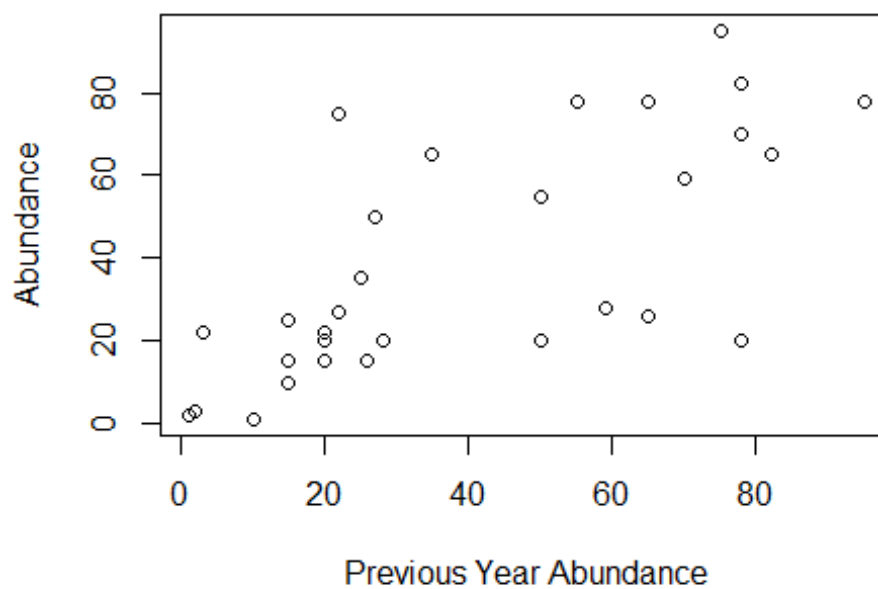
### 3.3.1 Abundance of Canadian Hare

```
data(hare)  
plot(hare, ylab= 'Abundance', xlab= 'Year', type= 'o')
```



### 3.3.2 Hare Abundance versus Previous Year's Hare Abundance

```
plot(y= hare, x= zlag(hare), ylab= 'Abundance',
      xlab= 'Previous Year Abundance')
```

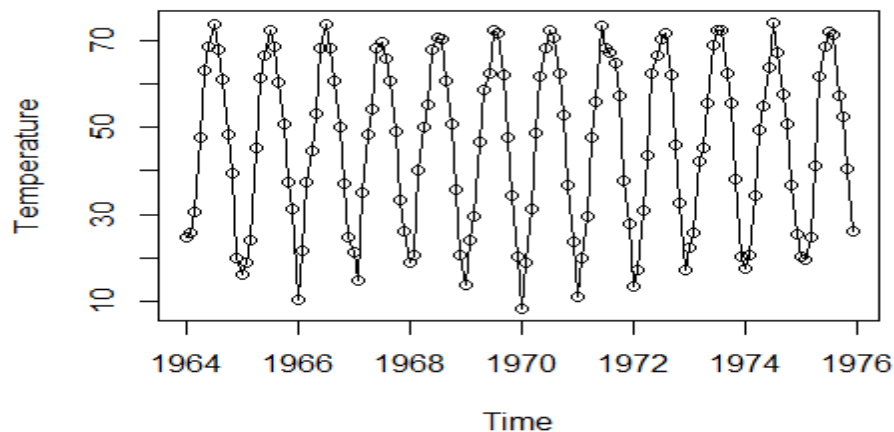


### 3.4 Monthly Average Temperatures in Dubuque, Iowa

The average monthly temperatures (in degrees Fahrenheit) over a number of years recorded in Dubuque, Iowa, are shown in Exhibit 1.7. This time series displays a very regular pattern called seasonality. Seasonality for monthly values occurs when observations twelve months apart are related in some manner or another. All Januarys and Februarys are quite cold but they are similar in value and different from the temperatures of the warmer months of June, July, and August, for example. There is still variation among the January values and variation among the June values. Models for such series must accommodate this variation while preserving the similarities. Here the reason for the seasonality is well understood—the Northern Hemisphere’s changing inclination toward the sun.

#### 3.4.1 Average Monthly Temperatures, Dubuque, Iowa

```
data(tempdub)
plot(tempdub, ylab= 'Temperature', type= 'o')
```

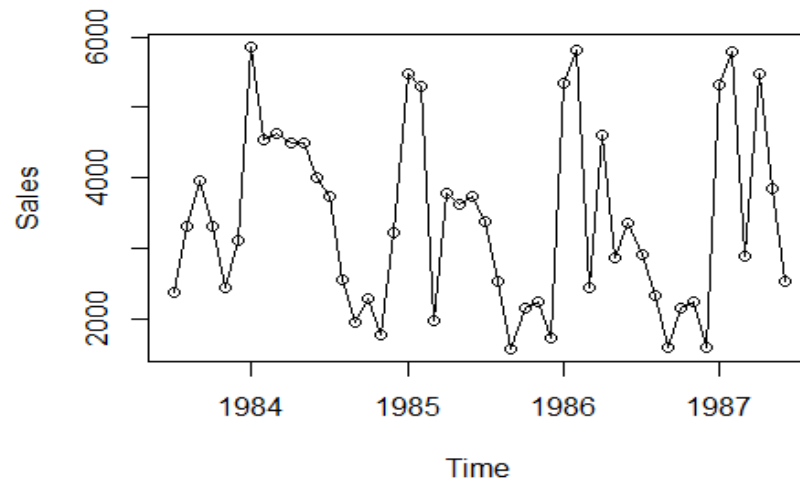


### 3.5 Monthly Oil Filter Sales

Our last example for this chapter concerns the monthly sales to dealers of a specialty oil filter for construction equipment manufactured by John Deere. When these data were first presented to one of the authors, the manager said, “There is no reason to believe that these sales are seasonal.” Seasonality would be present if January values tended to be related to other January values, February values tended to be related to other February values, and so forth. The time series plot shown in Exhibit 1.8 is not designed to display seasonality especially well. Exhibit 1.9 gives the same plot but amended to use meaningful plotting symbols. In this plot, all January values are plotted with the character J, all Februarys with F, all Marches with M, and so forth.† With these plotting symbols, it is much easier to see that sales for the winter months of January and February all tend to be high, while sales in September, October, November, and December are generally quite low. The seasonality in the data is much easier to see from this modified time series plot.

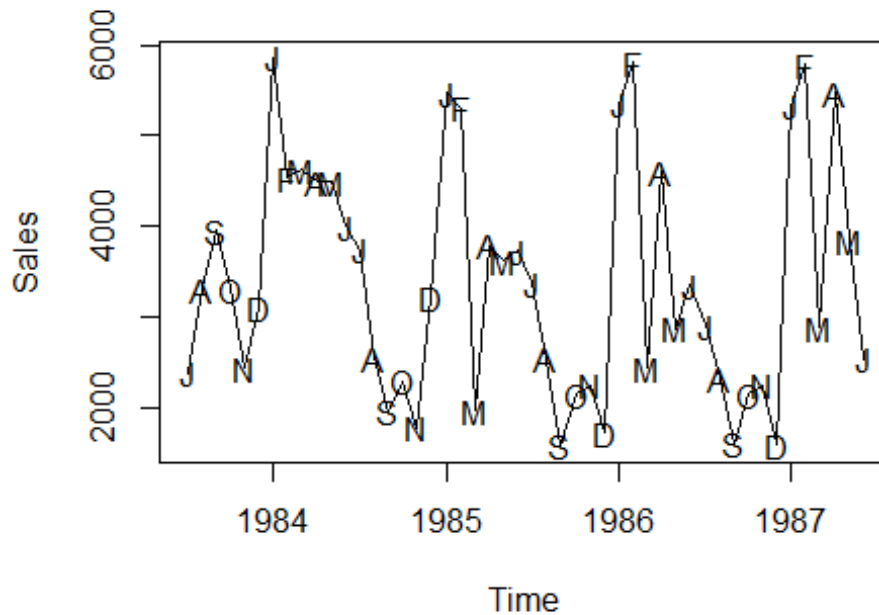
### 3.5.1 Monthly Oil Filter Sales

```
data(oilfilters)
plot(oilfilters, type= 'o', ylab= 'Sales')
```



### 3.6 Monthly Oil Filter Sales with Special Plotting Symbols

```
plot(oilfilters, type= 'l', ylab= 'Sales')
points(y= oilfilters, x= time(oilfilters),
       pch= as.vector(season(oilfilters)))
```



## 4 Fundamental Concepts of Time Series

We introduce the concepts of stochastic processes, mean and covariance functions, stationary processes, and autocorrelation functions.

### 4.1 Time Series and Stochastic Processes

The sequence of random variables  $Y_t: t = 0, \pm 1, \pm 2, \pm 3, \dots$  is called a stochastic process and serves as a model for an observed time series. It is known that the complete probabilistic structure of such a process is determined by the set of distributions of all finite collections of the  $Y$ 's. Fortunately, we will not have to deal explicitly with these multivariate distributions. Much of the information in these joint distributions can be described in terms of means, variances, and covariances. Consequently, we concentrate our efforts on these first and second moments. (If the joint distributions of the  $Y$ 's are multivariate normal distributions, then the first and second moments completely determine all the joint distributions.)

#### 4.1.1 Means, Variances, and Covariances

For a stochastic process  $Y_t: t = 0, \pm 1, \pm 2, \pm 3, \dots$ , the **mean function** is defined by

$$\mu_t = E(Y_t) \quad \text{for } t = 0, \pm 1, \pm 2, \dots$$

That is  $\mu_t$  is just the expected value of the process at time  $t$ . In general,  $\mu_t$  can be different at each time point  $t$ .

The **autocovariance** function,  $\gamma_{t,s}$ , is defined as

$$\gamma_{t,s} = \text{Cov}(Y_t, Y_s) \quad \text{for } t, s = 0, \pm 1, \pm 2, \dots$$

$$\text{where } \text{Cov}(Y_t, Y_s) = E((Y_t - \mu_t)(Y_s - \mu_s)) = E(Y_t Y_s) - \mu_t \mu_s$$

The **autocorrelation** function,  $\rho_{t,s}$ , is given by

$$\rho_{t,s} = \text{Corr}(Y_t, Y_s), \quad \text{for } t, s = 0, \pm 1, \pm 2, \dots$$

where

$$\text{corr}(y_t, y_s) = \frac{\text{cov}(y_t, y_s)}{\sqrt{\text{var}(y_t) \text{var}(y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t} \gamma_{s,s}}}$$

The following **important properties** follow from known results and our definitions:

- $\gamma_{t,t} = \text{Var}(Y_t) \Rightarrow \rho_{t,t} = 1$
- $\gamma_{t,s} = \gamma_{s,t} \Rightarrow \rho_{t,s} = \rho_{s,t}$
- $\gamma_{ts} \leq \gamma_{t,t} \gamma_{s,s} \Rightarrow \rho_{t,s} \leq 1$

Values of  $\rho_{t,s}$  near  $\pm 1$  indicate strong (linear) dependence, whereas values near zero indicate weak (linear) dependence. If  $\rho_{t,s} = 0$ , we say that  $Y_t$  and  $Y_s$  are uncorrelated.

To investigate the covariance properties of various time series models, the following result will be used repeatedly : If  $c_1, c_2, \dots, c_m$  and  $d_1, d_2, \dots, d_n$  are constants and  $t_1, t_2, \dots, t_m$  and  $s_1, s_2, \dots, s_n$  are time points, then

$$\text{cov} \left[ \sum_{i=1}^m c_i Y_{t_i}, \sum_{j=1}^n d_j Y_{s_j} \right] = \sum_{i=1}^m \sum_{j=1}^n c_i d_j \text{cov} (Y_{t_i}, Y_{s_j})$$

## 4.2 The Random Walk

Let  $e_1, e_2, \dots$  be a sequence of independent, identically distributed random variables each with zero mean and variance  $\sigma_e^2$ . The observed time series,  $Y_t: t = 1, 2, \dots$ , is constructed as follows:

$$\begin{aligned} Y_1 &= e_1 \\ Y_2 &= e_1 + e_2 \\ &\vdots \\ Y_t &= e_1 + e_2 + \dots + e_t \end{aligned}$$

Alternatively, we can write

$$Y_t = Y_{t-1} + e_t$$

with initial condition  $Y_1 = e_1$

### 4.2.1 Means, Variances, and Covariances

#### Mean

$$\mu_t = E(Y_t) = E(e_1 + e_2 + \dots + e_t)$$

$$\mu_t = E(e_1) + E(e_2) + \dots + E(e_t)$$

$$\mu_t = 0 + 0 + \dots + 0$$

$$\mu_t = 0$$

so that  $\mu_t = 0$  for all  $t$ .

We also have **variance**

$$\text{Var}(Y_t) = \text{Var}(e_1 + e_2 + \dots + e_t)$$

$$\text{Var}(Y_t) = \text{Var}(e_1) + \text{Var}(e_2) + \dots + \text{Var}(e_t)$$

$$\text{Var}(Y_t) = \sigma_e^2 + \sigma_e^2 + \dots + \sigma_e^2$$

$$Var(Y_t) = t \sigma_e^2$$

Notice that the process variance increases with linearly time  $t$ .

To investigate of a **covariance** function of a random walk. Suppose we have  $1 \leq t \leq s$ , then we have,

$$\gamma_{t,s} = Cov(Y_t, Y_s) = Cov(e_1 + e_2 + \dots + e_t, e_1 + e_2 + \dots + e_s)$$

$$cov \left[ \sum_{i=1}^m c_i Y_{t_i}, \sum_{j=1}^n d_j Y_{s_j} \right] = \sum_{i=1}^m \sum_{j=1}^n c_i d_j cov(Y_{t_i}, Y_{s_j})$$

we have,

$$cov(Y_t, Y_s) = \sum_{i=1}^m \sum_{j=1}^n cov(e_i, e_j)$$

However these covariances are 0 unless  $i = j$  in which case they are equal variance of  $e_i = \sigma_e^2$ .

There are exactly  $t$  of these so that  $\gamma_{t,s} = t \sigma_e^2$ .

Since  $\gamma_{t,s} = \gamma_{s,t}$ , this specifies the **autocovariance** function for all time points  $t$  and  $s$  and we can write

$$\gamma_{t,s} = t \sigma_e^2 \quad for \quad 1 \leq t \leq s$$

The autocorrelation function for the random walk is now easily obtained as

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t} \gamma_{s,s}}} = \sqrt{\frac{t}{s}} \quad for \quad 1 \leq t \leq s$$

The following numerical values help us understand the behavior of the random walk.

$$\rho_{1,2} = \sqrt{\frac{1}{2}} = 0.707 \quad \rho_{8,9} = \sqrt{\frac{8}{9}} = 0.943$$

$$\rho_{24,25} = \sqrt{\frac{24}{25}} = 0.980 \quad \rho_{1,25} = \sqrt{\frac{1}{25}} = 0.200$$

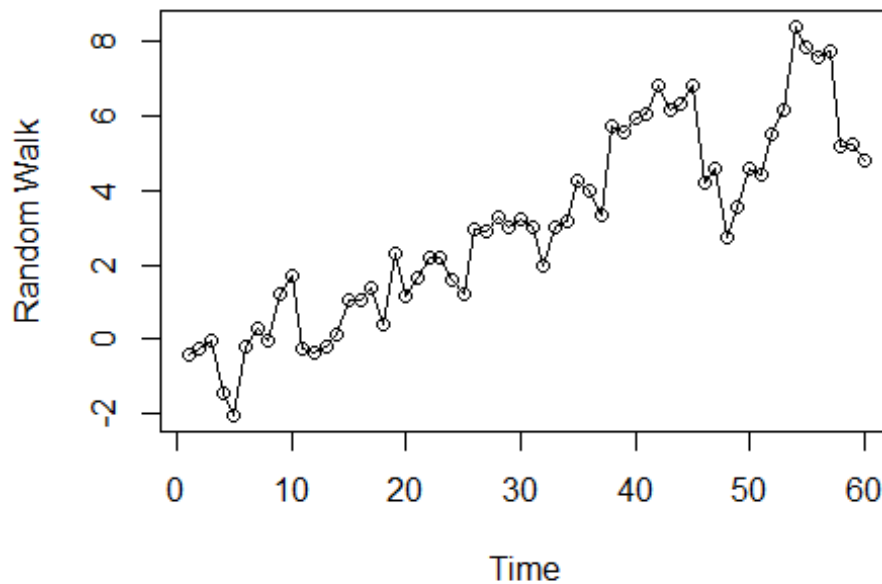
The values of  $Y$  at neighboring time points are more and more strongly and positively correlated as time goes by. On the other hand, the values of  $Y$  at distance time points(1,25) are less and less correlated.



```
require(TSA)
data("rwalk")
head(rwalk)

##           y
## [1,] -0.41433
## [2,] -0.26562
## [3,] -0.01265
## [4,] -1.43122
## [5,] -2.03763
## [6,] -0.21557

plot(rwalk, type= "o", ylab= "Random Walk")
```



### 4.3 A Moving Average

As a second example, suppose that  $Y_t$  is constructed as

$$Y_t = \frac{e_t + e_{t-1}}{2}$$

where the  $e$ 's are assumed to be independent and identically distributed with mean *zero* and variance  $\sigma_e^2$ . Here **mean**

$$\mu_t = E(Y_t) = E\left(\frac{e_t + e_{t-1}}{2}\right) = \frac{E(e_t) + E(e_{t-1})}{2} = 0$$

and **variance**

$$\text{Var}(Y_t) = \text{Var}\left(\frac{e_t + e_{t-1}}{2}\right) = \frac{1}{4} \text{Var}(e_t) + \text{Var}(e_{t-1})$$

$$\text{Var}(Y_t) = \frac{1}{4} 2 \sigma_e^2 = 0.5 \sigma_e^2$$

Also **Covariance**,

$$\begin{aligned} \text{Cov}(Y_t, Y_{t-1}) &= \text{Cov}\left(\frac{e_t + e_{t-1}}{2}, \frac{e_{t-1} + e_{t-2}}{2}\right) \\ &= \frac{\text{Cov}(e_t, e_{t-1}) + \text{Cov}(e_t, e_{t-2}) + \text{Cov}(e_{t-1}, e_{t-1}) + \text{Cov}(e_{t-1}, e_{t-2})}{4} \\ &= \frac{\text{Cov}(e_{t-1}, e_{t-1})}{4} \quad (\text{as all the other covariances are zero}) \end{aligned}$$

$$\text{Cov}(Y_t, Y_{t-1}) = 0.25 \sigma_e^2 \quad \text{for all } t$$

Furthermore

$$\text{Cov}(Y_t, Y_{t-2}) = \text{Cov}\left(\frac{e_t + e_{t-1}}{2}, \frac{e_{t-2} + e_{t-3}}{2}\right)$$

$$\text{Cov}(Y_t, Y_{t-2}) = 0 \quad \text{since the } e\text{'s are independent}$$

Similarly,  $\text{Cov}(Y_t, Y_{t-k}) = 0$  for  $k > 1$ , so we may write

$$\gamma_{t,s} = \begin{cases} 0.5 \sigma_e^2, & \text{for } |t - s| = 0 \\ 0.25 \sigma_e^2, & \text{for } |t - s| = 1 \\ 0, & \text{for } |t - s| > 1 \end{cases}$$

For the autocorrelation function, we have

$$\rho_{t,s} = \begin{cases} 1, & \text{for } |t - s| = 0 \\ 0.5, & \text{for } |t - s| = 1 \\ 0, & \text{for } |t - s| > 1 \end{cases}$$

$$\rho_{t,t-1} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t} \gamma_{s,s}}} = \frac{0.25 \sigma_e^2}{\sqrt{0.5 \sigma_e^2 \cdot 0.5 \sigma_e^2}} = \frac{1}{2} = 0.5$$

Notice that  $\rho_{2,1} = \rho_{3,2} = \rho_{4,3} = \rho_{9,8} = 0.5$ . Values of  $Y$  precisely one time unit apart have exactly the same correlation no matter where they occur in time. Furthermore,  $\rho_{3,1} = \rho_{4,2} = \rho_{t,t-2}$  and, more generally,  $\rho_{t,t-k}$  is the same for all values of  $t$ . This leads us to the important concept of stationarity.

#### 4.4 Stationarity

The basic idea of stationarity is that the probability laws that govern the behavior of the process do not change over time. In a sense, the process is in statistical equilibrium. Specifically, a process  $\{Y_t\}$  is said to be **strictly stationary** if the joint distribution of

$Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$  is the same as the joint distribution of  $Y_{t_1-k}, Y_{t_2-k}, \dots, Y_{t_n-k}$  for all choices of time points  $t_1, t_2, \dots, t_n$  and all choices of time lag  $k$ .

Thus, when  $n = 1$  the (univariate) distribution of  $Y_t$  is the same as that of  $Y_{t-k}$  for all  $t$  and  $k$ ; in other words, the  $Y$ 's are (marginally) identically distributed. It then follows that  $E(Y_t) = E(Y_{t-k})$  for all  $t$  and  $k$  so that the mean function is constant for all time.

Additionally,  $Var(Y_t) = Var(Y_{t-k})$  for all  $t$  and  $k$  so that the variance is also constant over time.

Setting  $n = 2$  in the stationarity definition we see that the bivariate distribution of  $Y_t$  and  $Y_s$  must be the same as that of  $Y_{t-k}$  and  $Y_{s-k}$  from which it follows that  $Cov(Y_t, Y_s) = Cov(Y_{t-k}, Y_{s-k})$  for all  $t, s$ , and  $k$ . Putting  $k = s$  and then  $k = t$ , we obtain

$$\begin{aligned}\gamma_{t,s} &= Cov(Y_{t-s}, Y_0) \\ &= Cov(Y_0, Y_{s-t}) \\ &= Cov(Y_0, Y_{|s-t|}) \\ \gamma_{t,s} &= \gamma_{0,|s-t|}\end{aligned}$$

That is, the covariance between  $Y_t$  and  $Y_s$  depends on time only through the time difference  $|t - s|$  and not otherwise on the actual times  $t$  and  $s$ . Thus, for a stationary process, we can simplify our notation and write

$$\gamma_k = Cov(Y_t, Y_{t-s})$$

and

$$\rho_k = Corr(Y_t, Y_{t-s})$$

Note also that

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

The general properties now become

- $\gamma_k = Cov(Y_t, Y_t) = Var(Y_t) \Rightarrow \rho_0 = 1$
- $\gamma_k = \gamma_{-k} \Rightarrow \rho_k = \rho_{-k}$
- $|\gamma_k| \leq \gamma_0 \Rightarrow |\rho_k| \leq 1$

If a process is strictly stationary and has finite variance, then the covariance function must depend only on the time lag  $k$ .

#### 4.4.1 Weakly Stationarity

A stochastic process  $\{Y_t\}$  is said to be **weakly** (or **second-order**) **stationary** if

1. The mean function is constant over time. & 2.  $\gamma_{t,t-k} = \gamma_{0,k}$  for all time  $t$  and lag  $k$

#### 4.4.2 White Noise

A very important example of a stationary process is the so-called **white noise** process, which is defined as a sequence of independent, identically distributed random variables  $\{e_t\}$ . Its importance stems not from the fact that it is an interesting model itself but from the fact that many useful processes can be constructed from white noise. The fact that  $\{e_t\}$  is strictly stationary is easy to see since as required. Also,  $\mu_t$

$$\begin{aligned} & Pr(e_{t_1} \leq x_1, e_{t_2} \leq x_2, \dots, e_{t_n} \leq x_n) \\ &= Pr(e_{t_1} \leq x_1)Pr(e_{t_2} \leq x_2) \dots Pr(e_{t_n} \leq x_n) \quad (\text{by independence}) \\ &= Pr(e_{t_1-k} \leq x_1)Pr(e_{t_2-k} \leq x_2) \dots Pr(e_{t_n-k} \leq x_n) \quad (\text{identical distribution}) \\ &= Pr(e_{t_1-k} \leq x_1, e_{t_2-k} \leq x_2, \dots, e_{t_n-k} \leq x_n) \quad (\text{by independence}) \end{aligned}$$

as required. Also,  $\mu_t = E(e_t)$  is constant and

$$\gamma_k = \begin{cases} \text{Var}(e_i), & \text{for } k = 0 \\ 0, & \text{for } k \neq 0 \end{cases}$$

Alternatively, we can write as required. Also,  $\mu_t = E(e_t)$  is constant and

$$\rho_k = \begin{cases} 1, & \text{for } k = 0 \\ 0, & \text{for } k \neq 0 \end{cases}$$

##### 4.4.2.1 Moving Average of White Noise

The moving average example where  $Y_t = (e_t + e_{t-1})/2$ , is another example of a stationary process constructed from white noise. In our new notation, we have for the moving average process that

$$\rho_k = \begin{cases} 1, & \text{for } k = 0 \\ 0.5, & \text{for } |k|=1 \\ 0, & \text{for } |k| \geq 2 \end{cases}$$

#### 4.4.3 Random Cosine Wave

As a somewhat different example, consider the process defined as follows :

$$Y_t = \cos \left[ 2\pi \left( \frac{t}{12} + \Phi \right) \right] \quad t = \pm 1, \pm 2, \dots$$

where  $\Phi \sim U(0,1)$

$$\begin{aligned} \mu_t &= E(Y_t) = E \left\{ \cos \left[ 2\pi \left( \frac{t}{12} + \Phi \right) \right] \right\} \\ \mu_t &= E(Y_t) = \int_0^1 \cos \left[ 2\pi \left( \frac{t}{12} + \phi \right) \right] d\phi \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi} \sin[2\pi \left(\frac{t}{12} + \phi\right)]_{\phi=0}^{\phi=1} \\
&= \frac{1}{2\pi} \left[ \sin\left(2\pi \frac{t}{12} + 2\pi\right) - \sin\left(2\pi \frac{t}{12}\right) \right] \\
\mu_t &= E(Y_t) = 0 \quad \text{for all } t \quad \because \sin(2\pi) = 0
\end{aligned}$$

Also

$$\begin{aligned}
\gamma_{t,s} &= E \left\{ \cos \left[ 2\pi \left( \frac{t}{12} + \phi \right) \right] \cos \left[ 2\pi \left( \frac{s}{12} + \phi \right) \right] \right\} \\
&= \int_0^1 \cos \left[ 2\pi \left( \frac{t}{12} + \phi \right) \right] \cos \left[ 2\pi \left( \frac{s}{12} + \phi \right) \right] d\phi \\
&= \frac{1}{2} \int_0^1 \left\{ \cos \left[ 2\pi \left( \frac{t-s}{12} \right) \right] \cos \left[ 2\pi \left( \frac{t+s}{12} + 2\phi \right) \right] \right\} d\phi \\
&= \frac{1}{2} \left\{ \cos \left[ 2\pi \left( \frac{t-s}{12} \right) \right] + \frac{1}{4\pi} \sin \left[ 2\pi \left( \frac{t+s}{12} + 2\phi \right) \right]_{\phi=0}^{\phi=1} \right\} \\
\gamma_{t,s} &= \frac{1}{2} \cos \left[ 2\pi \left( \frac{|t-s|}{12} \right) \right]
\end{aligned}$$

So the process is stationary with autocorrelation function

$$\rho_k = \cos \left( 2\pi \frac{k}{12} \right) \quad \text{for } k = 0, \pm 1, \pm 2, \dots$$

## 4.5 Different Series

Suppose that instead of analyzing  $\{Y_t\}$  directly, we consider the differences of successive  $Y$  values denoted by :

$$\nabla Y_t = Y_t - Y_{t-1} = e_t$$

So, that different t-series  $\{\nabla Y_t\}$  is stationary.

## 5 Trends

### 5.1 Regression Methods for Estimation of Trends

The classical statistical method of regression analysis may be readily used to estimate the parameters of common nonconstant mean trend models. We shall consider the most useful ones : **linear, quadratic, seasonal means** and **cosine trends**.

#### 5.1.1 Linear and Quadratic Trends in Time

Consider the deterministic time trend expressed as

$$\mu_t = \beta_0 + \beta_1 t$$

Thus, least squares method suggest that we will minimize

$$Q(\beta_0, \beta_1) = \sum_{t=1}^n [Y_t - (\beta_0 + \beta_1 t)]^2$$

The solution may be obtain

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^n (Y_t - \bar{Y})(t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2}$$

and

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{t}$$

where  $t = \frac{n(n+1)}{2n} = \frac{(n+1)}{2}, t = 1, 2, \dots, n$

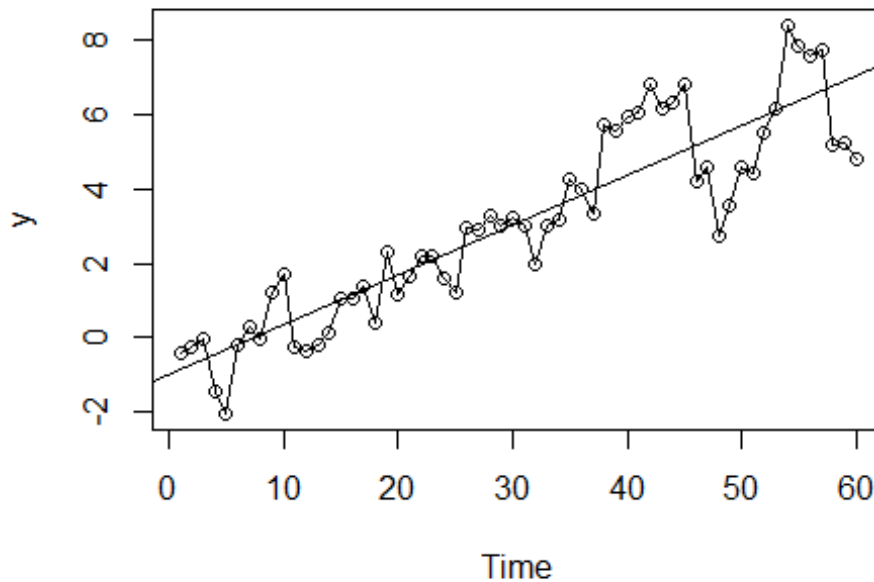
**Example :** Consider the random walk process that was shown in Exhibit 2.1. Suppose we (mistakenly) treat this as a linear time trend and estimate the slope and intercept by least-squares regression. Using statistical software we obtain Exhibit 3.1.

```
library(TSA)
data(rwalk)

model1<- lm(rwalk~ time(rwalk))
summary(model1)

##
## Call:
## lm(formula = rwalk ~ time(rwalk))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.70045 -0.79782  0.06391  0.63064  2.22128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.007888   0.297245  -3.391  0.00126 **
## time(rwalk)  0.134087   0.008475  15.822 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.137 on 58 degrees of freedom
## Multiple R-squared:  0.8119, Adjusted R-squared:  0.8086
## F-statistic: 250.3 on 1 and 58 DF, p-value: < 2.2e-16

plot(rwalk, type= 'o', ylab= 'y')
abline(model1) # add the fitted least squares line from model1
```



### 5.1.2 Cyclical or Seasonal Trends

Consider now modeling and estimating seasonal trends, such as for the average monthly temperature data in Exhibit 1.7. Here we assume that the observed series can be represented as

$$Y_t = \mu_t + X_t$$

where  $E(X_t) = 0$  for all  $t$ .

The most general assumption for  $\mu_t$  with monthly seasonal data is that there are 12 constants (parameters),  $\beta_1, \beta_2, \dots, \beta_{12}$ , giving the expected average temperature for each of the 12 months. We may write

$$\mu_t = \begin{cases} \beta_1, & \text{for } t=1,13,25,\dots \\ \beta_2, & \text{for } t=2,14,26,\dots \\ \vdots & \\ \beta_{12}, & \text{for } t=12,24,36,\dots \end{cases}$$

This is sometimes called a **seasonal means** model.

```
data(tempdub)
month<- season(tempdub) # period added to improve table display
model12<- lm(tempdub ~ month-1) # -1 removes the intercept term
summary(model12)
```

```
##
## Call:
## lm(formula = tempdub ~ month - 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.2750 -2.2479  0.1125  1.8896  9.8250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## monthJanuary      16.608      0.987   16.83  <2e-16 ***
## monthFebruary     20.650      0.987   20.92  <2e-16 ***
## monthMarch        32.475      0.987   32.90  <2e-16 ***
## monthApril        46.525      0.987   47.14  <2e-16 ***
## monthMay          58.092      0.987   58.86  <2e-16 ***
## monthJune         67.500      0.987   68.39  <2e-16 ***
## monthJuly         71.717      0.987   72.66  <2e-16 ***
## monthAugust       69.333      0.987   70.25  <2e-16 ***
## monthSeptember    61.025      0.987   61.83  <2e-16 ***
## monthOctober      50.975      0.987   51.65  <2e-16 ***
## monthNovember     36.650      0.987   37.13  <2e-16 ***
## monthDecember     23.642      0.987   23.95  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.419 on 132 degrees of freedom
## Multiple R-squared:  0.9957, Adjusted R-squared:  0.9953
## F-statistic: 2569 on 12 and 132 DF, p-value: < 2.2e-16
```

```
model3<- lm(tempdub~ month) # January is dropped automatically
summary(model3)
```

```
##
## Call:
## lm(formula = tempdub ~ month)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.2750 -2.2479  0.1125  1.8896  9.8250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      16.608      0.987  16.828  < 2e-16 ***
## monthFebruary      4.042      1.396   2.896  0.00443 **
## monthMarch        15.867      1.396  11.368  < 2e-16 ***
## monthApril        29.917      1.396  21.434  < 2e-16 ***
## monthMay          41.483      1.396  29.721  < 2e-16 ***
## monthJune         50.892      1.396  36.461  < 2e-16 ***
## monthJuly         55.108      1.396  39.482  < 2e-16 ***
## monthAugust       52.725      1.396  37.775  < 2e-16 ***
```



```
## monthSeptember    44.417      1.396   31.822 < 2e-16 ***
## monthOctober      34.367      1.396   24.622 < 2e-16 ***
## monthNovember     20.042      1.396   14.359 < 2e-16 ***
## monthDecember      7.033      1.396    5.039 1.51e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.419 on 132 degrees of freedom
## Multiple R-squared:  0.9712, Adjusted R-squared:  0.9688
## F-statistic: 405.1 on 11 and 132 DF,  p-value: < 2.2e-16
```

### 5.1.3 Cosine Trends

The seasonal means model for monthly data consists of 12 independent parameters and does not take the shape of the seasonal trend into account at all. For example, the fact that the March and April means are quite similar (and different from the June and July means) is not reflected in the model. In some cases, seasonal trends can be modeled economically with cosine curves that incorporate the smooth change expected from one time period to the next while still preserving the seasonality.

Consider the cosine curve with equation

$$\mu_t = \beta \cos(2\pi f t + \Phi)$$

We call  $\beta (> 0)$  the *amplitude*,  $f$  the *frequency*, and  $\Phi$  the *phase* of the curve. As  $t$  varies, the curve oscillates between a maximum of  $\beta$  and a minimum of  $-\beta$ . Since the curve repeats itself exactly every  $1/f$  time units,  $1/f$  is called the *period* of the cosine wave.  $\Phi$  serves to set the arbitrary origin on the time axis. For monthly data with time indexed as  $1, 2, \dots$ , the most important frequency is  $f = 1/12$ , because such a cosine wave will repeat itself every 12 months. We say that the period is 12.

Equation (3.3.4) is inconvenient for estimation because the parameters  $\beta$  and  $\Phi$  do not enter the expression linearly. Fortunately, a trigonometric identity is available that reparameterizes (3.3.4) more conveniently, namely

$$\beta \cos(2\pi f t + \Phi) = \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$$

where

$$\beta = \sqrt{\beta_1^2 + \beta_2^2}, \quad \Phi = \text{atan}(-\beta_2/\beta_1)$$

and conversely,

$$\beta_1 = \beta \cos(\Phi), \quad \beta_2 = \beta \sin(\Phi)$$

To estimate the parameters  $\beta_1$  and  $\beta_2$  with regression techniques, we simply use  $\cos(2\pi f t)$  and  $\sin(2\pi f t)$  as regressors or predictor variables.

The simplest such model for the trend would be expressed as

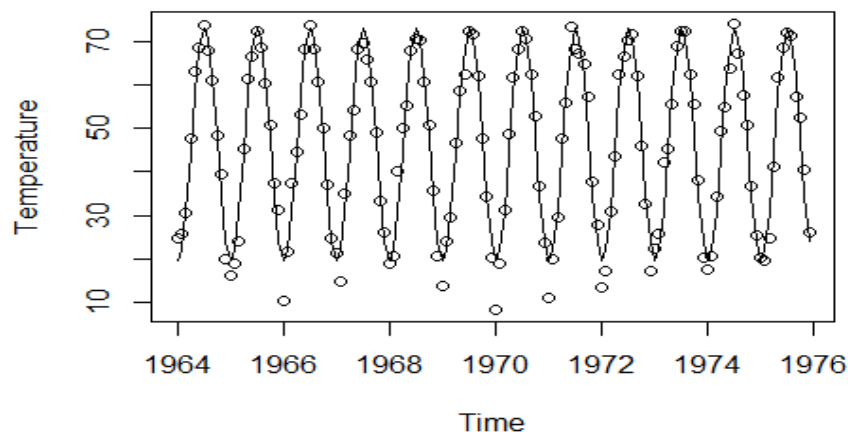
$$\mu_t = \beta_0 + \beta_1 \cos(2\pi ft) + \beta_2 \sin(2\pi ft)$$

Here the constant term,  $\beta_0$ , can be meaningfully thought of as a cosine with frequency zero.

```
har<- harmonic(tempdub, 1)
model4<- lm(tempdub~ har)
summary(model4)

##
## Call:
## lm(formula = tempdub ~ har)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.1580  -2.2756  -0.1457   2.3754  11.2671
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    46.2660     0.3088 149.816 < 2e-16 ***
## harcos(2*pi*t) -26.7079     0.4367 -61.154 < 2e-16 ***
## harsin(2*pi*t)  -2.1697     0.4367  -4.968 1.93e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.706 on 141 degrees of freedom
## Multiple R-squared:  0.9639, Adjusted R-squared:  0.9634
## F-statistic: 1882 on 2 and 141 DF, p-value: < 2.2e-16

# win.graph(width=4.875, height=2.5,pointsize=8)
plot(ts(fitted(model4), freq= 12, start= c(1964,1)),
      ylab= 'Temperature', type= 'l', ylim= range(c(fitted(model4), tempdub)))
points(tempdub)
```



*# ylim ensures that the y axis range fits the raw data and the fitted values*

## 5.2 Residual Analysis

As we have already noted, the unobserved stochastic component  $\{X_t\}$  can be estimated, or predicted, by the residual

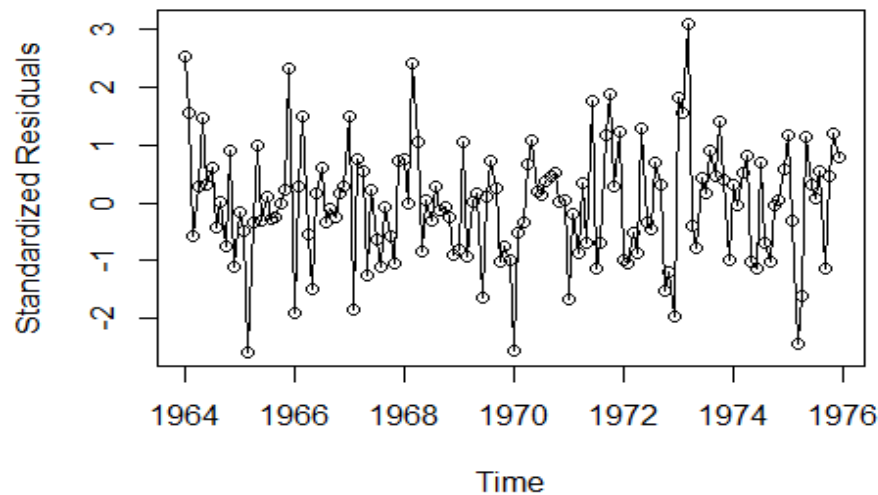
$$\hat{X} = Y_t - \hat{\mu}_t$$

We call  $\hat{X}_t$  the residual corresponding to the  $t^{th}$  observation. If the trend model is reasonably correct, then the residuals should behave roughly like the true stochastic component, and various assumptions about the stochastic component can be assessed by looking at the residuals. If the stochastic component  $\{Y_t\}$  is white noise, then the residuals should behave roughly like independent (normal) random variables with *zero*<sup>(0)</sup> mean and standard deviation  $s$ . Standardizing the residuals as  $\hat{X}/s$  may be consider for plotting.

```
library(TSA)
data("tempdub")
month1<- season(tempdub)
model13<- lm(tempdub~month1)
summary(model13)

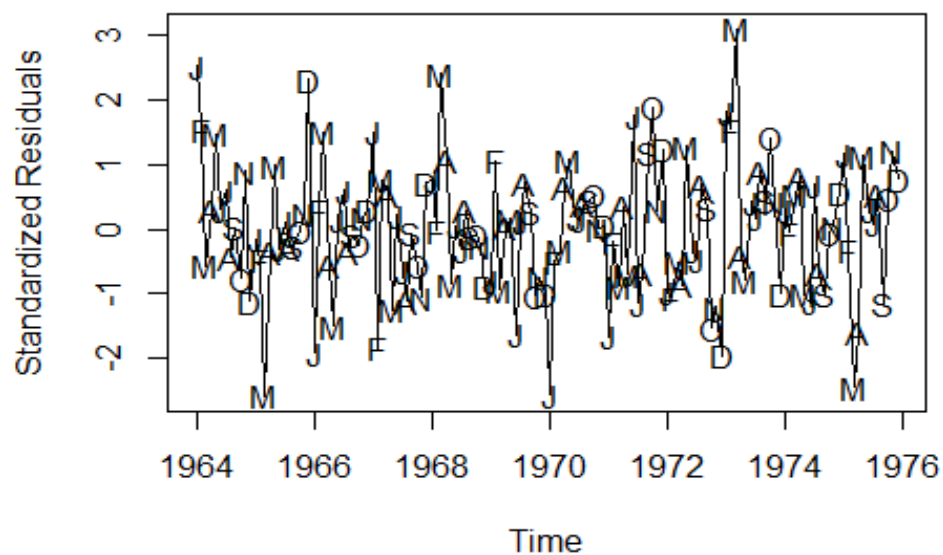
##
## Call:
## lm(formula = tempdub ~ month1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.2750 -2.2479  0.1125  1.8896  9.8250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    16.608     0.987   16.828 < 2e-16 ***
## month1February     4.042     1.396    2.896  0.00443 **
## month1March       15.867     1.396   11.368 < 2e-16 ***
## month1April       29.917     1.396   21.434 < 2e-16 ***
## month1May         41.483     1.396   29.721 < 2e-16 ***
## month1June        50.892     1.396   36.461 < 2e-16 ***
## month1July        55.108     1.396   39.482 < 2e-16 ***
## month1August      52.725     1.396   37.775 < 2e-16 ***
## month1September  44.417     1.396   31.822 < 2e-16 ***
## month1October     34.367     1.396   24.622 < 2e-16 ***
## month1November    20.042     1.396   14.359 < 2e-16 ***
## month1December     7.033     1.396    5.039 1.51e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.419 on 132 degrees of freedom
## Multiple R-squared:  0.9712, Adjusted R-squared:  0.9688
## F-statistic: 405.1 on 11 and 132 DF, p-value: < 2.2e-16
```

```
plot(y= rstudent(model3), x= as.vector(time(tempdub)),
     xlab= 'Time', ylab= 'Standardized Residuals', type= 'o')
```



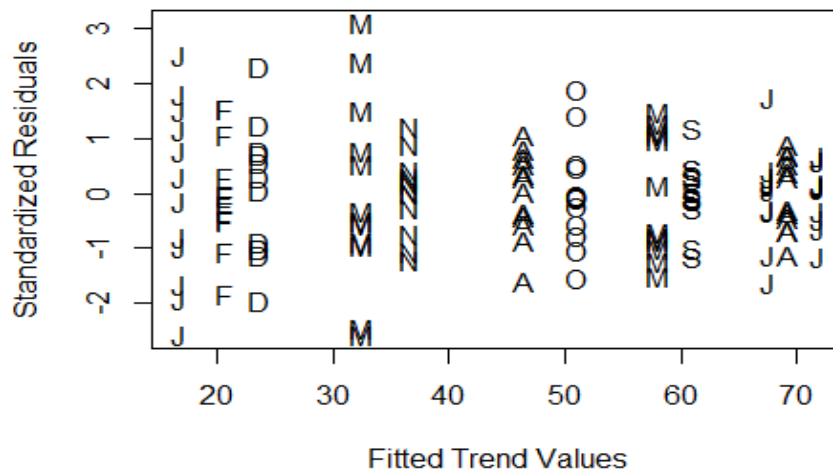
```
plot(y= rstudent(model3), x= as.vector(time(tempdub)), xlab= 'Time',
     ylab= 'Standardized Residuals', type= 'l')
```

```
points(y= rstudent(model3), x= as.vector(time(tempdub)),
       pch= as.vector(season(tempdub)))
```



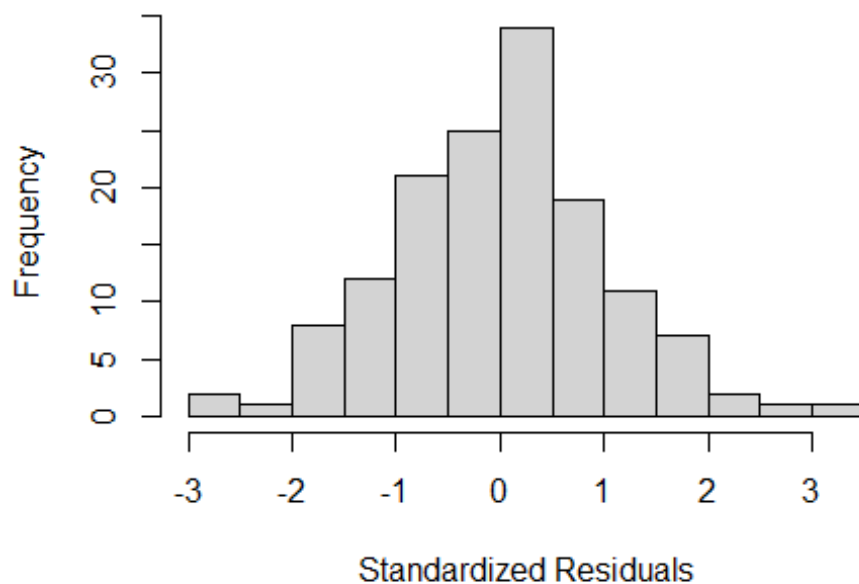
```
plot(y= rstudent(model3), x= as.vector(fitted(model3)),
     xlab= 'Fitted Trend Values', ylab= 'Standardized Residuals', type= 'n')

points(y= rstudent(model3), x= as.vector(fitted(model3)),
       pch= as.vector(season(tempdub)))
```



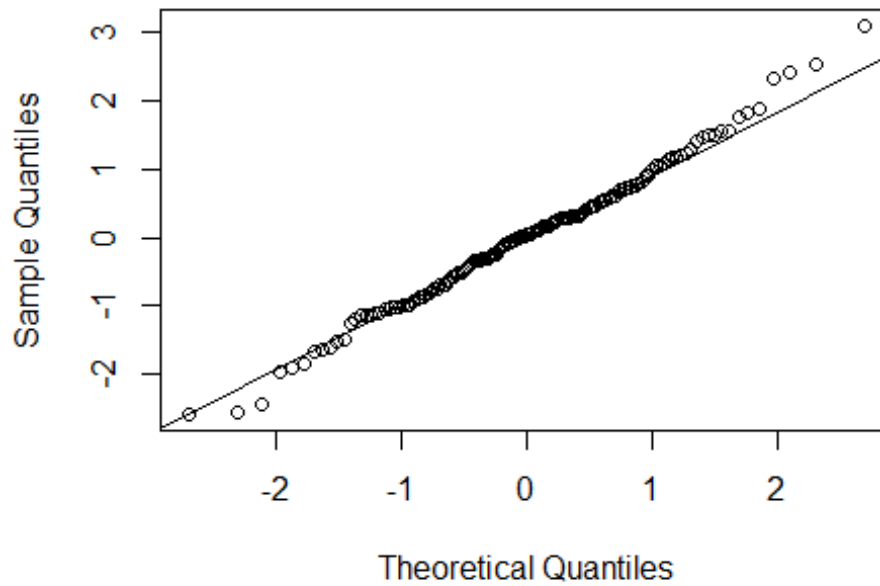
```
hist(rstudent(model3), xlab= 'Standardized Residuals', main= "Histogram of
Standardized Residuals from Seasonal Means Model")
```

### gram of Standardized Residuals from Seasonal Mea



```
qqnorm(rstudent(model3), main= " Q-Q Plot: Standardized Residuals of Seasonal Means Model")
qqline(rstudent(model3))
```

### Q-Q Plot: Standardized Residuals of Seasonal Means



#### 5.2.1 Sample Autocorrelation Function

Another very important diagnostic tool for examining dependence is the sample autocorrelation function. Consider any sequence of data  $Y_1, Y_2, \dots, Y_n$  — whether residuals, standardized residuals, original data, or some transformation of data. Tentatively assuming stationarity, we would like to estimate the autocorrelation function  $\rho_k$  for a variety of lags  $k = 1, 2, \dots$ . The obvious way to do this is to compute the sample correlation between the pairs  $k$  units apart in time. That is, among  $(Y_1, Y_{1+k}), (Y_2, Y_{2+k}), (Y_3, Y_{3+k}), \dots$ , and  $(Y_{n-k}, Y_n)$ . However, we modify this slightly, taking into account that we are assuming stationarity, which implies a common mean and variance for the series. With this in mind, we define the **sample autocorrelation function**,  $r_k$ , at lag  $k$  as

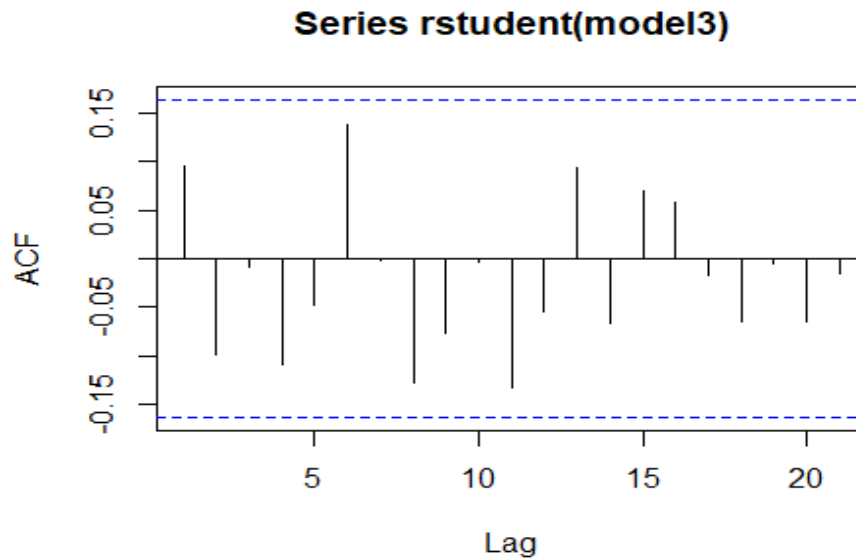
$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad \text{for } k = 1, 2, \dots$$

A plot of  $r_k$  versus lag  $k$  is often called a **correlogram**.

The hypotheses  $H_0: \rho_k = 0$  can be rejected at the usual significance levels. The horizontal dashed lines, which are placed at  $\pm 2/\sqrt{n}$  and none of these **auto-correlation** is cross those line.

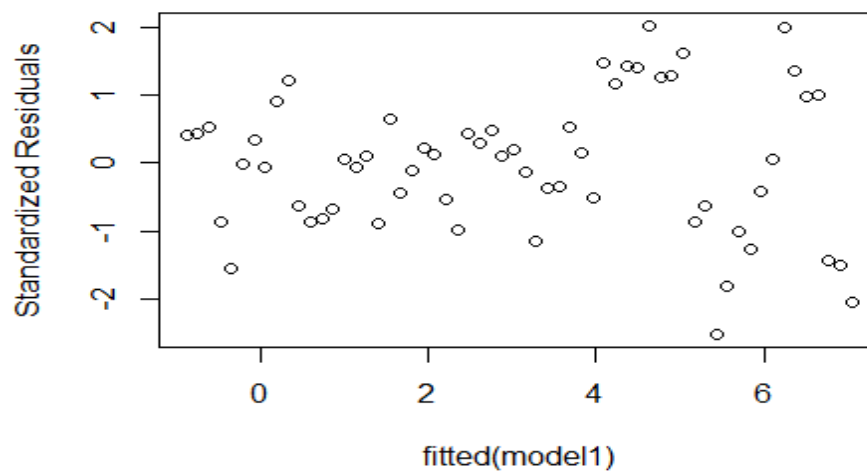
```
data("tempdub")
month1<- season(tempdub)
model3<- lm(tempdub~ month1)

acf(rstudent(model3))
```

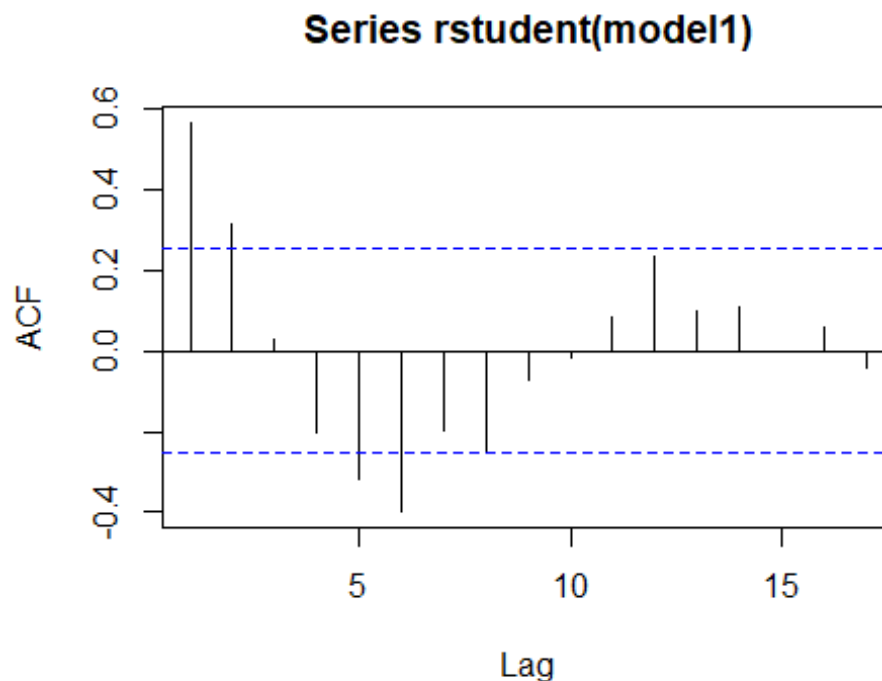


```
data(rwalk)
model1<- lm(rwalk~ time(rwalk), data= rwalk)

plot(y= rstudent(model1), x= fitted(model1), ylab= "Standardized Residuals",
type= "p")
```



```
acf(rstudent(model1))
```



we observed that the *lag* 1 and *lag* 2 autocorrelations exceed two standard errors above  $zero^{(0)}$  and the *lag* 5 and *lag* 6 autocorrelations more than two standard errors below zero. This is not what we expect from a white noise process.

## 6 MODELS FOR STATIONARY TIME SERIES

### 6.1 Auto-Regressive Moving Average (ARMA) models

These models have assumed great importance in modeling real-world processes.

### 6.2 General Linear Processes

A general linear process,  $\{Y_t\}$ , is one that can be represented as a weighted linear combination of present and past white noise terms as

$$Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots$$

with the assumption that  $\sum_{i=1}^{\infty} \psi_i^2 < \infty$

An important nontrivial example to which we will return often is the case where the  $\psi$ 's form an exponentially decaying sequence

$$\psi_j = \phi^j$$



where  $\phi$  is a number strictly between  $-1$  and  $+1$ . Then

$$Y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \dots$$

For this example.

$$E(Y_t) = E(e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \dots) = 0$$

so that  $\{Y_t\}$  has a constant mean of zero. Also

$$\begin{aligned} \text{Var}(Y_t) &= \text{Var}(e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \dots) \\ &= \text{Var}(e_t) + \phi^2 \text{Var}(e_{t-1}) + \phi^4 \text{Var}(e_{t-2}) + \dots \\ &= \sigma_e^2 (1 + \phi^2 + \phi^4 + \dots) \end{aligned}$$

$$\text{Var}(Y_t) = \frac{\sigma_e^2}{(1 - \phi^2)} \quad (\text{by summing a geometric series})$$

Furthermore,

$$\begin{aligned} \text{Cov}(Y_t, Y_{t-1}) &= \text{Cov}(e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \dots, e_{t-1} + \phi e_{t-2} + \phi^2 e_{t-3} + \dots) \\ &= \text{Cov}(\phi e_{t-1}, e_{t-1}) + \text{Cov}(\phi^2 e_{t-2}, \phi e_{t-2}) + \dots \\ &= \phi \sigma_e^2 + \phi^3 \sigma_e^2 + \phi^5 \sigma_e^2 + \dots \\ &= \phi \sigma_e^2 (1 + \phi^2 + \phi^4 + \dots) \end{aligned}$$

$$\text{Cov}(Y_t, Y_{t-1}) = \frac{\phi \sigma_e^2}{(1 - \phi^2)} \quad (\text{by summing a geometric series})$$

Thus

$$\text{cov}(Y_t, Y_{t-1}) = \frac{\left[ \frac{\phi \sigma_e^2}{1 - \phi^2} \right]}{\left[ \frac{\sigma_e^2}{1 - \phi^2} \right]} = \phi$$

In a similar manner, we can find  $\text{cov}(Y_t, Y_{t-1}) = \frac{\phi^k \sigma_e^2}{1 - \phi^2}$  and thus  $\text{cov}(Y_t, Y_{t-1}) = \phi^k$

It is important to note that the process defined in this way is stationary—the autocovariance structure depends only on time lag and not on absolute time.

### 6.3 Moving Average Processes — MA(q)

In this process, we change notation somewhat

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

We call such a series a **moving average of order q** and abbreviate the name to **MA(q)**.

### 6.3.1 The First-Order Moving Average Process — MA(1)

The model is

$$Y_t = e_t - \theta e_{t-1}$$

It's **Mean**

$$E(Y_t) = E(e_t - \theta e_{t-1})$$

$$E(Y_t) = E(e_t) - \theta E(e_{t-1})$$

$$E(Y_t) = 0$$

and **Variance**

$$Var(Y_t) = Var(e_t - \theta e_{t-1})$$

$$Var(Y_t) = Var(e_t) - \theta^2 Var(e_{t-1})$$

$$= \sigma_e^2 + \theta^2 \sigma_e^2$$

$$Var(Y_t) = \sigma_e^2(1 + \theta^2)$$

and **Covariance**

$$Cov(Y_t, Y_{t-1}) = Cov(e_t - \theta e_{t-1}, e_{t-1} - \theta e_{t-2})$$

$$= Cov(-\theta e_{t-1}, e_{t-1})$$

$$Cov(Y_t, Y_{t-1}) = -\theta \sigma_e^2$$

and

$$Cov(Y_t, Y_{t-2}) = Cov(e_t - \theta e_{t-1}, e_{t-2} - \theta e_{t-3})$$

$$Cov(Y_t, Y_{t-2}) = 0$$

Similarly,

$$Cov(Y_t, Y_{t-k}) = 0 \quad \text{for } k \geq 2$$

the process has no correlation beyond lag 1.

In summary, for an **MA(1)** model  $Y_t = e_t - \theta e_{t-1}$ ,

$$1. \quad E(Y_t) = 0$$

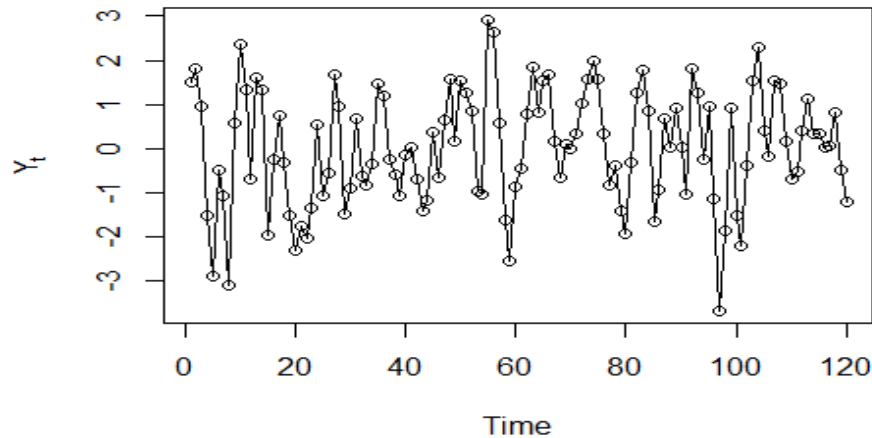
$$2. \quad \gamma_0 = Var(Y_t) = \sigma_e^2(1 + \theta^2)$$

$$3. \quad \gamma_1 = Cov(Y_t, Y_{t-1}) = -\theta \sigma_e^2$$

$$4. \quad \rho_1 = \frac{-\theta}{(1 + \theta^2)}$$

$$\gamma_k = \rho_k = 0 \quad \text{for } k \geq 2$$

```
library(TSA)
data(ma1.2.s)
plot(ma1.2.s, ylab= expression(Y[t]), type= 'o')
```



### 6.3.2 The Second-Order Moving Average Process — MA(2)

Consider the moving average process of order 2 :

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

It's **Mean**

$$E(Y_t) = E(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})$$

$$E(Y_t) = E(e_t) - \theta_1 E(e_{t-1}) - \theta_2 E(e_{t-2})$$

$$E(Y_t) = 0$$

and **Variance**

$$Var(Y_t) = Var(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})$$

$$Var(Y_t) = Var(e_t) - \theta_1^2 Var(e_{t-1}) - \theta_2^2 Var(e_{t-2})$$

$$= \sigma_e^2 + \theta_1^2 \sigma_e^2 + \theta_2^2 \sigma_e^2$$

$$\gamma_0 = Var(Y_t) = \sigma_e^2 (1 + \theta_1^2 + \theta_2^2)$$

and **Covariance**

$$\gamma_1 = Cov(Y_t, Y_{t-1}) = Cov(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, e_{t-1} - \theta_1 e_{t-2} - \theta_2 e_{t-3})$$

$$= Cov(-\theta_1 e_{t-1}, e_{t-1}) + Cov(-\theta_2 e_{t-2}, -\theta_1 e_{t-2})$$

$$= -\theta_1 Cov(e_{t-1}, e_{t-1}) + \theta_1 \theta_2 Cov(e_{t-2}, e_{t-2})$$

$$Cov(Y_t, Y_{t-1}) = [-\theta_1 + \theta_1\theta_2]\sigma_e^2$$

and

$$\begin{aligned}\gamma_2 &= Cov(Y_t, Y_{t-2}) = Cov(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, e_{t-2} - \theta_1 e_{t-3} - \theta_2 e_{t-4}) \\ &= Cov(-\theta_2 e_{t-2}, e_{t-2}) \\ \gamma_2 &= Cov(Y_t, Y_{t-2}) = -\theta_2 \sigma_e^2\end{aligned}$$

for an **MA(2)** process

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_k = 0 ; \quad k = 3, 4, \dots, \infty \quad k \geq 3$$

For specific case

$$Y_t = e_t - e_{t-1} + 0.6Y_{t-2}$$

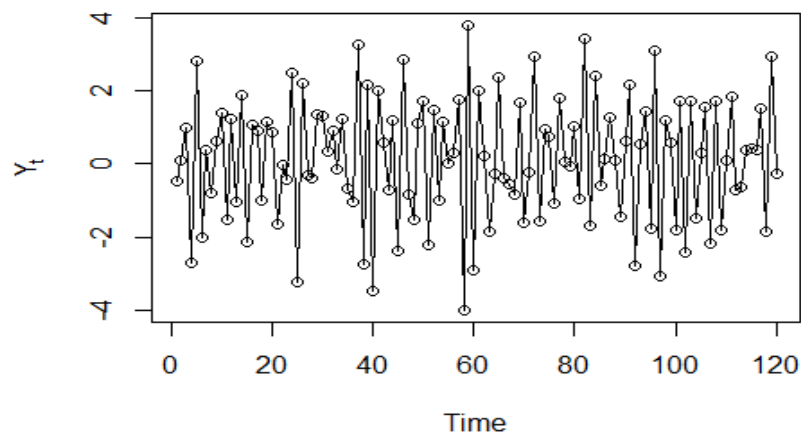
we have  $\theta_1 = 1, \theta_2 = 0.6$

$$\rho_1 = \frac{-1 + (1)(-0.6)}{1 + (1)^2 + (0.6)^2} = \frac{-1.6}{3.6} = -0.678$$

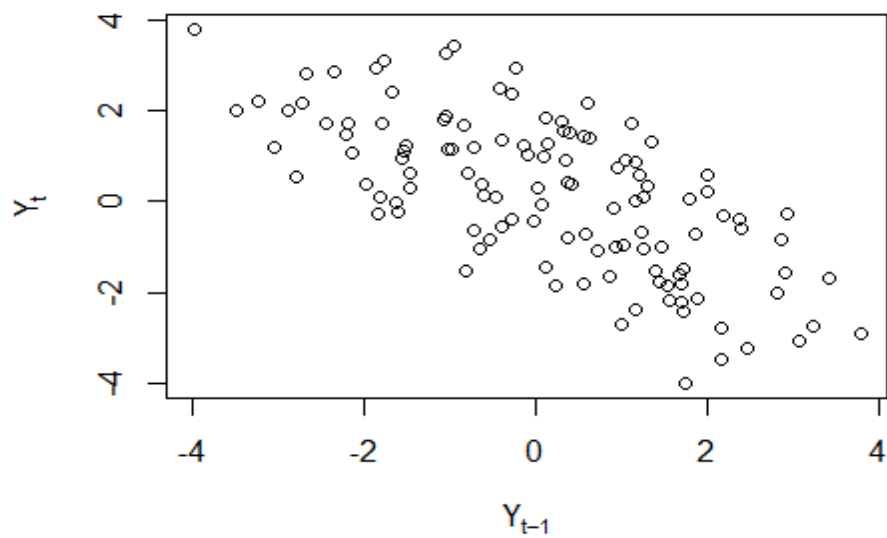
$$\rho_2 = \frac{0.6}{2.36} = 0.254$$

```
data(ma2.s)
plot(ma2.s, ylab= expression(Y[t]), type= 'o',
     main= "Time Plot of an MA(2) with 01=1 and 02=-0.6")
```

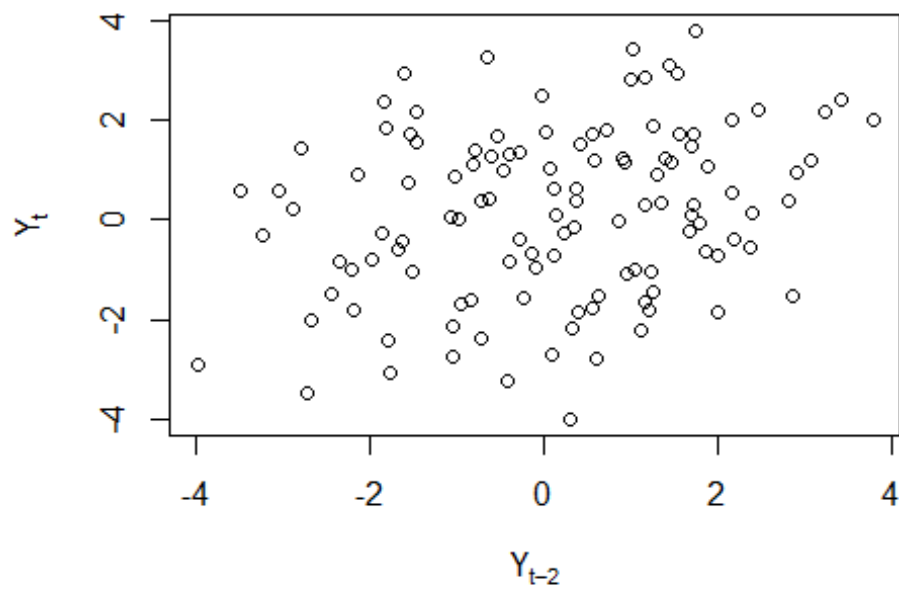
**Time Plot of an MA(2) with 01=1 and 02=-0.6**



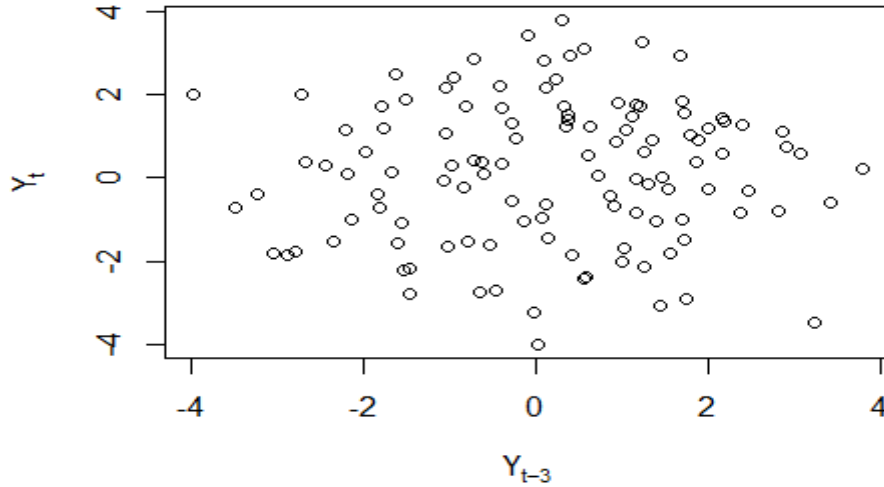
```
plot(y= ma2.s, x= zlag(ma2.s), ylab= expression(Y[t]),  
xlab= expression(Y[t-1]), type= "p") # rho1= -0.678
```



```
plot(y= ma2.s, x= zlag(ma2.s, 2), ylab= expression(Y[t]),  
xlab= expression(Y[t-2]), type= "p") # rho2= 0.254
```



```
plot(y= ma2.s, x= zlag(ma2.s, 3), ylab= expression(Y[t]),
xlab= expression(Y[t-3]), type= "p") # rho3= 0
```



### 6.3.3 General Moving Average Process — MA(q)

For general **MA(q)** process

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

**Mean :**

$$E(Y_t) = 0$$

**Variance :**

$$\gamma_0 = \text{Var}(Y_t) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2) \sigma_e^2$$

and

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \cdots + \theta_q \theta_{k-q}}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} & ; \text{for } k=1,2,\dots,q \\ 0 & ; \text{for } k > q \end{cases}$$

### 6.4 Autoregressive Processes — AR(p)

Autoregressive processes are as their name suggests—regressions on themselves. Specifically, a  $p^{\text{th}}$  — order **autoregressive process**  $\{Y_t\}$  satisfies the equation

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

## The First-Order Autoregressive Process — AR(1)

Assume the series is stationary and satisfies

$$Y_t = \phi Y_{t-1} + e_t \quad (1)$$

$$E(Y_t) = 0$$

$$\text{Var}(Y_t) = \gamma_0 = \text{Var}(\phi Y_{t-1} + e_t)$$

$$\text{Var}(Y_t) = \gamma_0 = \phi^2 \text{Var}(Y_{t-1}) + \text{Var}(e_t)$$

$$\text{Var}(Y_t) = \gamma_0 = \phi^2 \gamma_0 + \sigma_e^2$$

solving for  $\gamma_0$ , we get

$$\gamma_0 = \frac{\sigma_e^2}{1 - \phi^2}$$

Notice the immediate implication that  $\phi^2 < 1$  or that  $|\phi| < 1$ .

multiply both sides of equation (1) by  $Y_{t-k}$  ( $k = 1, 2, \dots$ ), we have

$$Y_t Y_{t-k} = \phi Y_{t-1} Y_{t-k} + e_t Y_{t-k}$$

$$E(Y_t Y_{t-k}) = \phi E(Y_{t-1} Y_{t-k}) + E(e_t Y_{t-k})$$

$$\gamma_k = \phi \gamma_{k-1}$$

Since the series is assumed to be stationary with zero mean, and since  $e_t$  is independent of  $Y_{t-k}$ .

Thus

$$E(e_t Y_{t-k}) = E(e_t) E(Y_{t-k}) = 0$$

$$\gamma_k = \phi \gamma_{k-1} \quad ; \text{ for } k=1, 2, 3, \dots$$

Setting  $k = 1$ , we get

$$\gamma_1 = \phi \gamma_0 = \phi \frac{\sigma_e^2}{1 - \phi^2}$$

Setting  $k = 2$ , we get

$$\gamma_2 = \phi \gamma_1 = \phi^2 \frac{\sigma_e^2}{1 - \phi^2}$$

In general

$$\gamma_k = \phi \gamma_{k-1} = \phi^k \frac{\sigma_e^2}{1 - \phi^2}$$

and thus

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k \quad ; \text{ for } k=1,2,3,\dots$$

for  $k = 0$

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$$

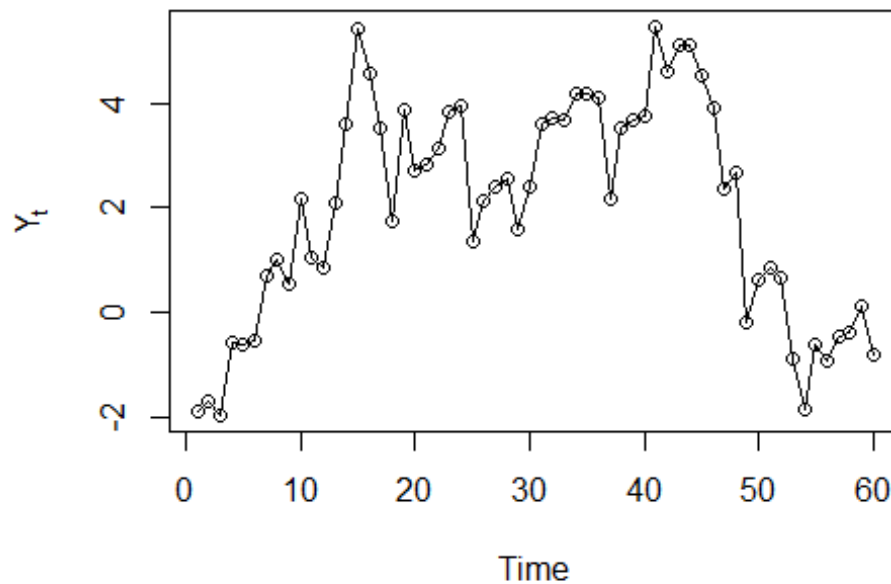
for  $k = 1$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \phi$$

Since  $|\phi| < 1$ , the magnitude of the autocorrelation function decreases exponentially as the number of *lags*,  $k$ , increases.

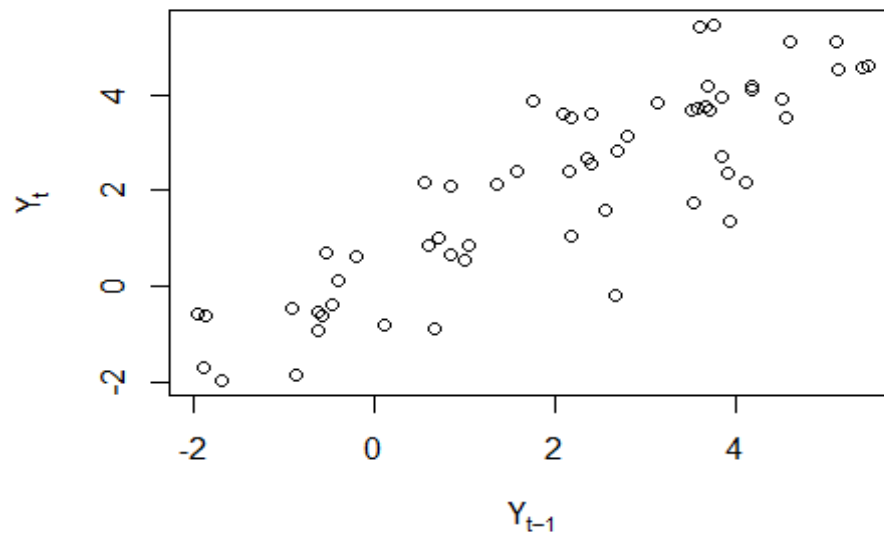
- If  $0 < \phi < 1$ , all correlations are positive.
- If  $-1 < \phi < 0$ , the *lag* 1 autocorrelation is negative ( $\rho_1 = \phi$ ) and the signs of successive autocorrelations alternate from positive to negative, with their magnitudes decreasing exponentially.

```
data(ar1.s)
plot(ar1.s, ylab= expression(Y[t]), type= "o")
```

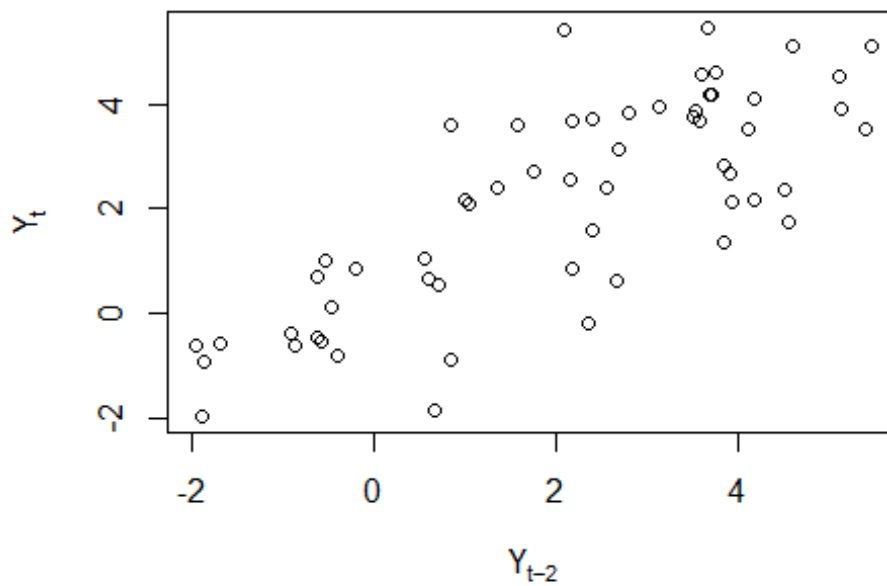




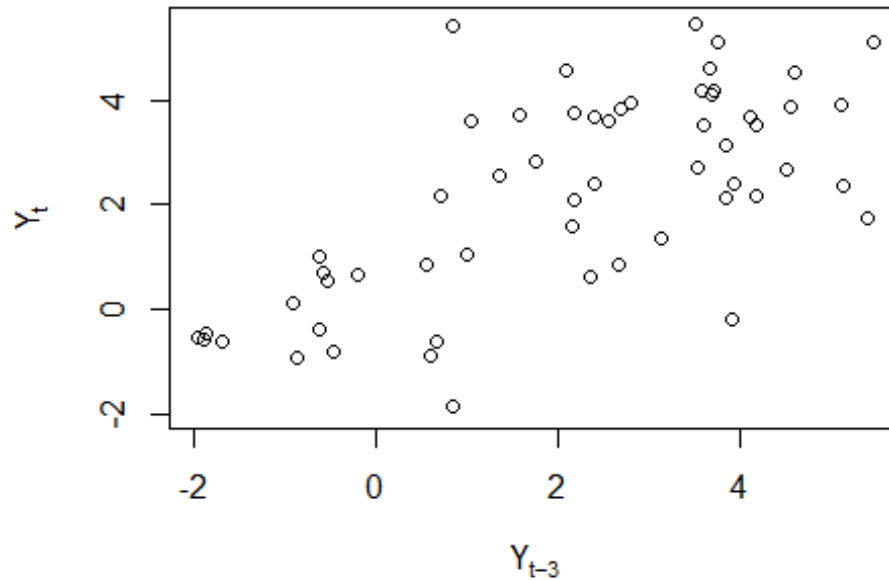
```
plot(ar1.s, x= zlag(ar1.s), ylab= expression(Y[t]), xlab= expression(Y[t-1]),
type= "p") # rho1= 0.9
```



```
plot(ar1.s, x= zlag(ar1.s, 2), type= "p", ylab= expression(Y[t]),
xlab = expression(Y[t-2])) # rho2= 0.9^2
```



```
plot(ar1.s, x= zlag(ar1.s, 3), type= "p", ylab= expression(Y[t]),
xlab= expression(Y[t-3])) # rho2= 0.9^3
```



#### 6.4.0.1 The General Linear Process Version of the AR(1) Model

**AR(1)** Model is

$$Y_t = \phi Y_{t-1} + e_t \quad \text{--- (1)}$$

If we replacing the equation by  $t - 1$

$$Y_{t-1} = \phi Y_{t-2} + e_{t-1} \quad \text{--- (2)}$$

Substituting  $Y_{t-1}$  in equation 1

$$Y_t = \phi(\phi Y_{t-2} + e_{t-1}) + e_t$$

$$Y_t = \phi^2 Y_{t-2} + \phi e_{t-1} + e_t$$

$$Y_t = e_t + \phi e_{t-1} + \phi^2 Y_{t-2}$$

If we repeat this substitution into the past, say  $k - 1$  times, we get

$$Y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots + \phi^{k-1} e_{t-k+1} + \phi^k Y_{t-k}$$

Assuming  $|\phi| < 1$  and letting  $k$  increase without bound, we should obtain the infinite series representation

$$Y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} +$$

This is the general form of the equation

$$Y_t = e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \dots$$

with

$$\Psi_j = \phi^j$$

**Stationarity of an AR(1) Process is that  $|\phi| < 1$ .**

#### **6.4.1 The Second-Order Autoregressive Process — AR(2)**

Now consider the series satisfying

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$$

where, as usual, we assume that  $e_t$  is independent of  $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$ . To discuss stationarity, we introduce the **AR characteristic polynomial**

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2$$

and the corresponding **AR characteristic equation**

$$1 - \phi_1 x - \phi_2 x^2 = 0$$

We recall that a quadratic equation always has two roots (possibly complex).

##### **6.4.1.1 Stationarity of the AR(2) Process**

In the second-order case, the roots of the quadratic characteristic equation are easily found to be

$$\frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$$

For stationarity, we require that these roots exceed 1 in absolute value, this will be true if and only if three conditions are satisfied :

- i.  $\phi_1 + \phi_2 < 1$
- ii.  $\phi_2 - \phi_1 < 1$
- iii.  $|\phi_1| < 1$

For **AR(1)** process, the AR Characteristic Equation is just

$$1 - \phi x = 0$$

with root  $1/\phi$  which exceed 1 in absolute value iff  $|\phi| < 1$ .

#### 6.4.1.2 The Autocorrelation Function for the AR(2) Process

To derive the autocorrelation function for the AR(2) case, we take the defining recursive relationship of equation

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$$

multiplying both the sides by  $Y_{t-k}$  and get expectations, we have

$$Y_t Y_{t-k} = \phi_1 Y_{t-1} Y_{t-k} + \phi_2 Y_{t-2} Y_{t-k} + e_t Y_{t-k}$$

$$E(Y_t Y_{t-k}) = \phi_1 E(Y_{t-1} Y_{t-k}) + \phi_2 E(Y_{t-2} Y_{t-k}) + E(e_t Y_{t-k})$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \quad \text{--- (A)}$$

Dividing equation (A) by  $\gamma_0$

$$\frac{\gamma_k}{\gamma_0} = \phi_1 \frac{\gamma_{k-1}}{\gamma_0} + \phi_2 \frac{\gamma_{k-2}}{\gamma_0}$$

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad \text{--- (B) for } k=1,2,\dots$$

Equations (A) and/or (B) are usually called the **Yule-Walker equations**

Setting  $k = 1$ , we get

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1}$$

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

and so

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

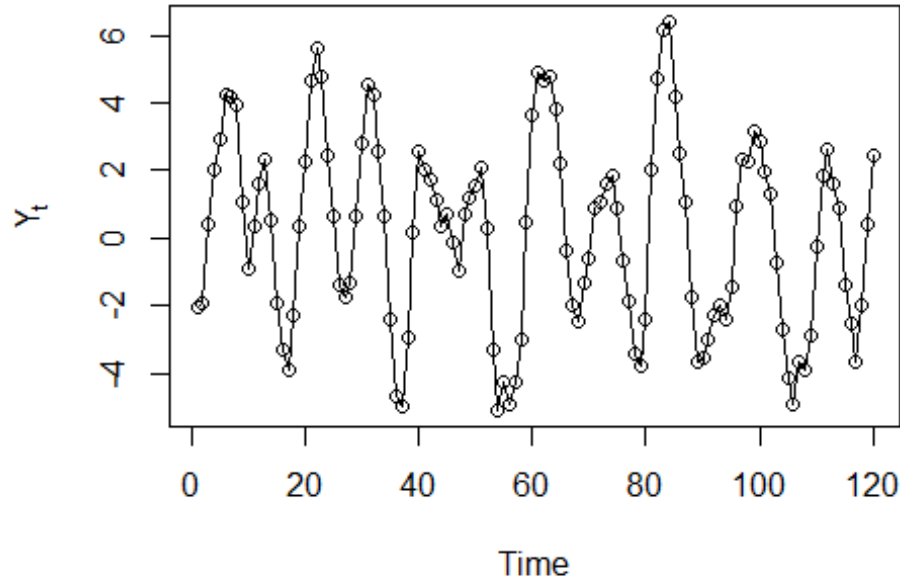
Using the now known values for  $\rho_1$  (and  $\rho_0$ ), equation(B) can be used with  $k = 2$  to obtain

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0$$

$$\rho_2 = \frac{\phi_1^2 + \phi_2(1 - \phi_2)}{(1 - \phi_2)}$$

Successive values of  $\rho_k$  may be easily calculated numerically from the recursive relationship of Equation (B).

```
data(ar2.s)
plot(ar2.s, ylab= expression(Y[t]), type= 'o')
```



#### 6.4.2 Varicance for the AR(2) Model

Consider the auto-regression

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$$

$$Var(Y_t) = \phi_1^2 Var(Y_{t-1}) + \phi_2^2 Var(Y_{t-2}) + 2\phi_1\phi_2 Cov(Y_{t-1}Y_{t-2}) + e_t$$

$$\gamma_0 = (\phi_1^2 + \phi_2^2)\gamma_0 + 2\phi_1\phi_2\gamma_1 + \sigma_e^2 \quad \text{--- (i)}$$

Setting  $k = 1$  in equation(A) gives a second linear equation for  $\gamma_0$  and  $\gamma_1$ ,  $\gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1$ , which can be solved simultaneously with euation  $i$  to obtain

$$\gamma_0 = \frac{(1 - \phi_2)\sigma_e^2}{(1 - \phi_2)(1 - \phi_1^2 - \phi_2^2) - 2\phi_2\phi_1^2}$$

$$\gamma_0 = \left(\frac{1 - \phi_2}{1 + \phi_2}\right) \frac{\sigma_e^2}{(1 - \phi_2)^2 - \phi_1^2}$$

#### 6.5 General Auto-Regressive Process

Consider the  $p^{th}$ -order autoregressive model

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

with AR characteristic polynomial

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p$$

and corresponding AR characteristic equation

$$1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0$$

The necessary condition for a stationary is

$$\phi_1 + \phi_2 + \dots + \phi_p < 1$$

and

$$|\phi_p| < 1$$

## 6.6 Mixed Autoregressive Moving Average Model — ARMA(p, q)

If we assume that the series is partly autoregressive and partly moving average, we obtain a quite general time series model. In general, if

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

we say that  $\{Y_t\}$  is a mixed **autoregressive moving average** process of orders  $p$  and  $q$ , respectively. We abbreviate the name to  $ARMA(p, q)$ .

### 6.6.1 ARMA(1, 1) Model

The defining equation can be written

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1} \quad \text{--- (ii)}$$

To derive Yule-Walker type equations, we first note that

$$E(e_t Y_t) = E[e_t(\phi Y_{t-1} + e_t - \theta e_{t-1})]$$

$$E(e_t Y_t) = \sigma_e^2$$

and

$$E(e_{t-1} Y_t) = E[e_{t-1}(\phi Y_{t-1} + e_t - \theta e_{t-1})]$$

$$E(e_{t-1} Y_t) = \phi \sigma_e^2 - \theta \sigma_e^2$$

$$E(e_{t-1} Y_t) = (\phi - \theta) \sigma_e^2$$

If we multiply Equation ii by  $Y_{t-k}$  and take expectations, we have

$$Y_t Y_{t-k} = \phi Y_{t-1} Y_{t-k} + e_t Y_{t-k} - \theta e_{t-1} Y_{t-k}$$

$$E(Y_t Y_{t-k}) = \phi E(Y_{t-1} Y_{t-k}) + E(e_t Y_{t-k}) - \theta E(e_{t-1} Y_{t-k})$$

For  $k = 0$

$$\gamma_0 = \phi\gamma_1 + \sigma_e^2 - \theta(\phi - \theta)\sigma_e^2$$

$$\gamma_0 = \phi\gamma_1 + [1 - \theta(\phi - \theta)]\sigma_e^2$$

$$\gamma_1 = \phi\gamma_1 - \theta\sigma_e^2$$

$$\gamma_k = \phi\gamma_{k-1} \quad \text{for } k \geq 2$$

Solving the first two equations yields

$$\gamma_0 = \frac{1 - 2\theta\phi + \theta^2}{1 - \phi^2} \sigma_e^2$$

and solving the simple recursion gives

$$\frac{(1 - \theta\phi)(\phi - \theta)}{1 - 2\theta\phi + \theta^2} \phi^{k-1} \quad \text{for } k \geq 1$$

**Remark :** Note that this autocorrelation function decays exponentially as the lag  $k$  increases. The damping factor is  $\phi$ , but the decay starts from initial value  $\rho_1$ , which also depends on  $\theta$ . This is in contrast to the AR(1) autocorrelation, which also decays with damping factor  $\phi$  but always from initial value  $\rho_0 = 1$ .

**For example,** if  $\phi = 0.8$  and  $\theta = 0.4$ , then

- $\rho_1 = 0.523$ ,
- $\rho_2 = 0.418$ ,
- $\rho_3 = 0.335$ ,
- and so on.

## 7 Models for Non-Stationary Time Series — ARIMA

Any time series without a constant mean over time is nonstationary. Models of the form

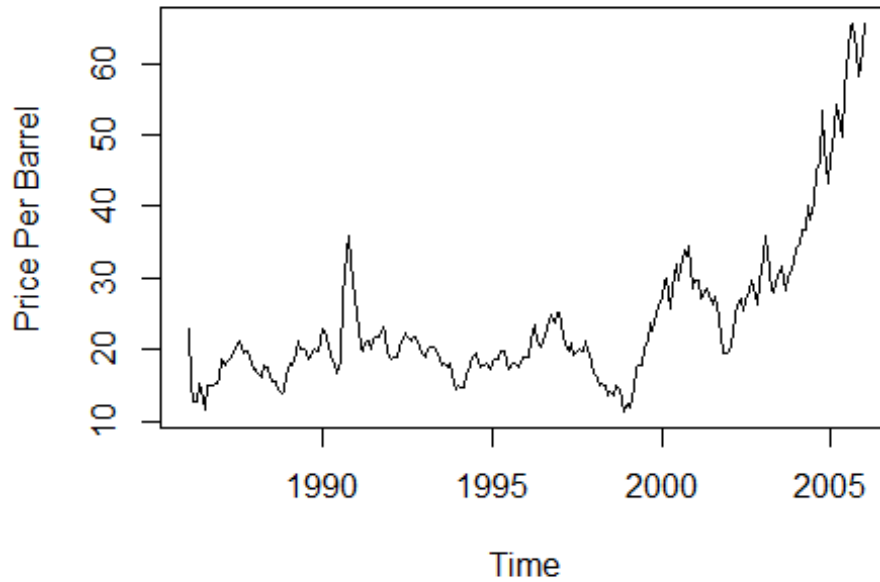
$$Y_t = \mu_t + X_t$$

where  $\mu_t$  is a non-constant mean function and  $X_t$  is a zero-mean.

As an example consider the monthly price of a barrel of crude oil from January 1986 through January 2006.

```
library(TSA)
data("oil.price")
plot(oil.price, ylab= "Price Per Barrel", type= "l",
main= "Non-Stationary Model")
```

## Non-Stationary Model



The series displayed considerable variation specially since 2001 and a stationary model doesn't seem to be reasonable.

### 7.1 Stationarity Through Differencing — ARIMA Models

A time series  $\{Y_t\}$  is said to follow an **integrated autoregressive moving average** model if the  $d^{th}$  difference  $W_t = \nabla^d Y_t$  is a stationary ARMA process. If  $\{W_t\}$  follows an **ARMA(p,q) model**, we say that  $\{Y_t\}$  is an **ARIMA(p,d,q) process**. Fortunately, for practical purposes, we can usually take  $d = 1$  or at most 2.

Consider then an **ARIMA(p,1,q) process**. With  $W_t = Y_t - Y_{t-1}$ , we have

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

or, in terms of the observed series,

$$\begin{aligned} & Y_t - Y_{t-1} \\ = & \phi_1 (Y_{t-1} - Y_{t-2}) + \phi_2 (Y_{t-2} - Y_{t-3}) + \cdots + \phi_p (Y_{t-p} - Y_{t-p-1}) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots \\ & \quad - \theta_q e_{t-q} \end{aligned}$$

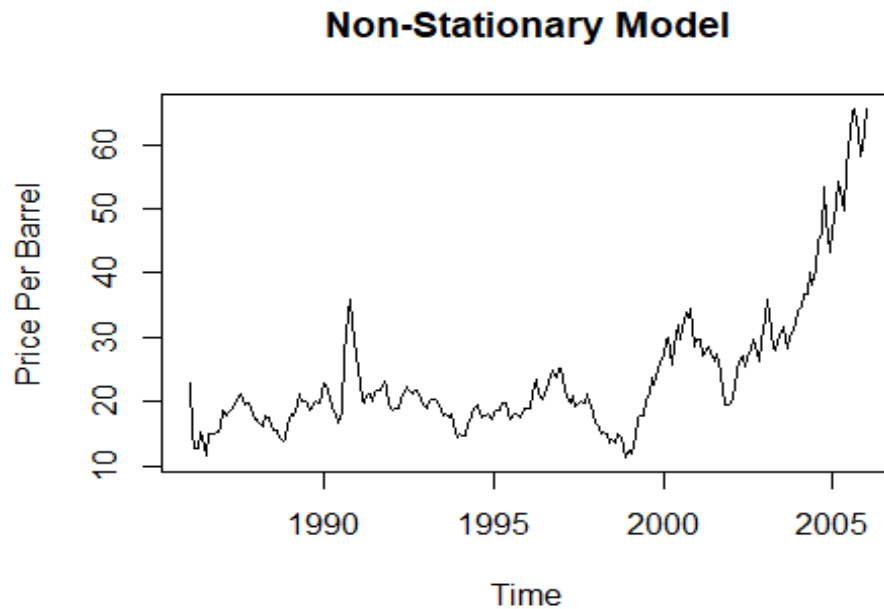
which we may rewrite as

$$\begin{aligned} & Y_t \\ = & (1 + \phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + (\phi_3 - \phi_2)Y_{t-3} + \cdots + (\phi_p - \phi_{p-1})Y_{t-p} - \phi_p Y_{t-p-1} + e_t \\ & \quad - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \end{aligned}$$

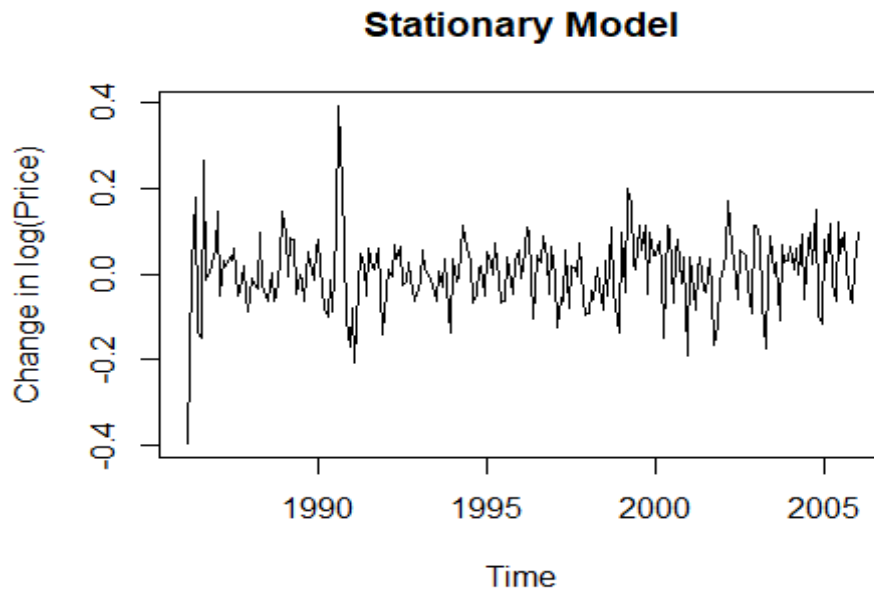
We call this the **difference equation form** of the model.



```
data("oil.price")
plot(oil.price, ylab= "Price Per Barrel", type= "l",
main= "Non-Stationary Model")
```



```
plot(diff(log(oil.price)), ylab= "Change in log(Price)", type= "l",
main= "Stationary Model")
```



It is seen that difference are stationary.

### 7.1.1 Constant Terms in ARIMA Models

For an  $ARIMA(p, d, q)$  model,  $\nabla^d Y_t = W_t$  is a stationary  $ARMA(p, q)$  process. Our standard assumption is that stationary models have a zero mean; that is, we are actually working with deviations from the constant mean. A nonzero constant mean,  $\mu$ , in a stationary ARMA model  $\{W_t\}$  can be accommodated in either of two ways. We can assume that

$$W_t - \mu = \phi_1(W_{t-1} - \mu) + \phi_2(W_{t-2} - \mu) + \cdots + \phi_p(W_{t-p} - \mu) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Alternatively, we can introduce a constant term  $\theta_0$  into the model as follows :

$$W_t = \theta_0 + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Taking expected values on both sides of the latter expression, we find that

$$\begin{aligned} E(W_t) &= \theta_0 + \phi_1 E(W_{t-1}) + \phi_2 E(W_{t-2}) + \cdots + \phi_p E(W_{t-p}) + E(e_t) - \theta_1 E(e_{t-1}) - \theta_2 E(e_{t-2}) - \cdots \\ &\quad - \theta_q E(e_{t-q}) \end{aligned}$$

$$\mu = \theta_0 + \phi_1 \mu + \phi_2 \mu + \cdots + \phi_p \mu + 0 + 0 \cdots + 0$$

$$\mu = \theta_0 + (\phi_1 + \phi_2 + \cdots + \phi_p) \mu$$

$$\mu = \frac{\theta_0}{1 - (\phi_1 + \phi_2 + \cdots + \phi_p)}$$

or, conversely, that

$$\theta_0 = \mu[1 - (\phi_1 + \phi_2 + \cdots + \phi_p)]$$

## 7.2 Other Transformation

### 7.2.1 Log Transformation — $\log()$

We have seen how differencing can be a useful transformation for achieving stationarity. However, the logarithm transformation is also a useful method in certain circumstances.

We frequently encounter series where increased dispersion seems to be associated with higher levels of the series—the higher the level of the series, the more variation there is around that level and conversely.

Specifically, suppose that  $Y_t > 0$  for all  $t$  and that

$$E(Y_t) = \mu_t \quad \text{and} \quad \sqrt{\text{Var}(Y_t)} = \mu_t \sigma$$

Then

$$E[\log(Y_t)] \approx \log(\mu_t) \quad \text{and} \quad \text{Var}(\log(Y_t)) \approx \sigma^2$$

These results follow from taking expected values and variances of both sides of the (Taylor) expansion

$$\log(Y_t) \approx \log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t}$$

$$E(\log(Y_t)) \approx E(\log(\mu_t)) + \frac{E(Y_t - \mu_t)}{\mu_t}$$

$$E(\log(Y_t)) \approx \log(\mu_t)$$

$$\text{Var}(\log(Y_t)) \approx \text{Var}(\log(\mu_t)) + \frac{1}{\mu_t^2} \text{Var}(Y_t - \mu_t)$$

$$\text{Var}(\log(Y_t)) \approx 0 + \frac{1}{\mu_t^2} \mu_t^2 \sigma^2$$

$$\text{Var}(\log(Y_t)) \approx \sigma^2$$

In other words, if standard deviation of the series is proportional to the level of the series, then transforming to logarithms will produce a series with approximately constant variance over time.

### 7.2.2 Percentage Changes and Logarithms — returns

Suppose  $Y_t$  tends to have relatively stable percentage changes from one time period to the next. Specifically, assume that

$$Y_t = (1 + X_t)Y_{t-1}$$

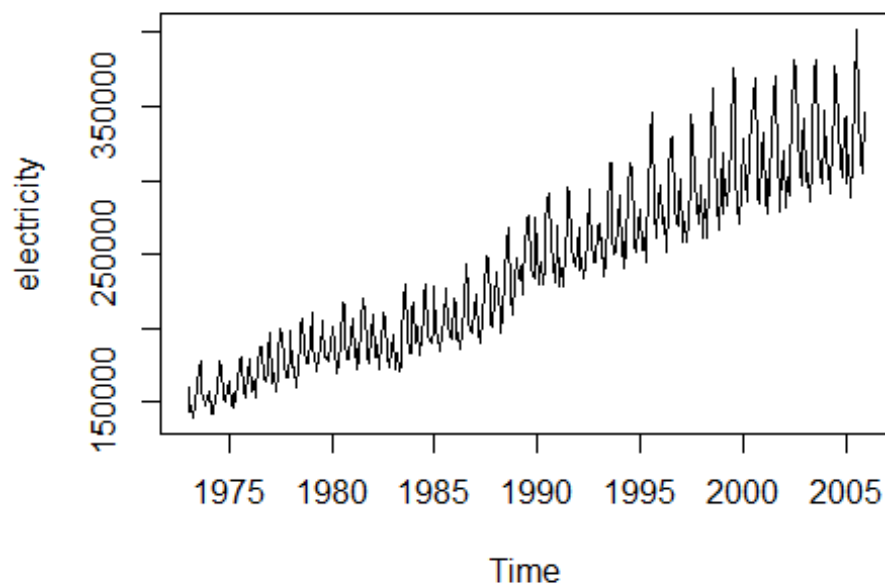
where  $100X_t$  is the percentage change (possibly negative) from  $Y_{t-1}$  to  $Y_t$ . Then

$$\log(Y_t) - \log(Y_{t-1}) = \log\left(\frac{Y_t}{Y_{t-1}}\right) = \log(1 + X_t)$$

$$\Delta[\log(Y_t)] \approx X_t \quad \text{if } |X_t| < 0.20$$

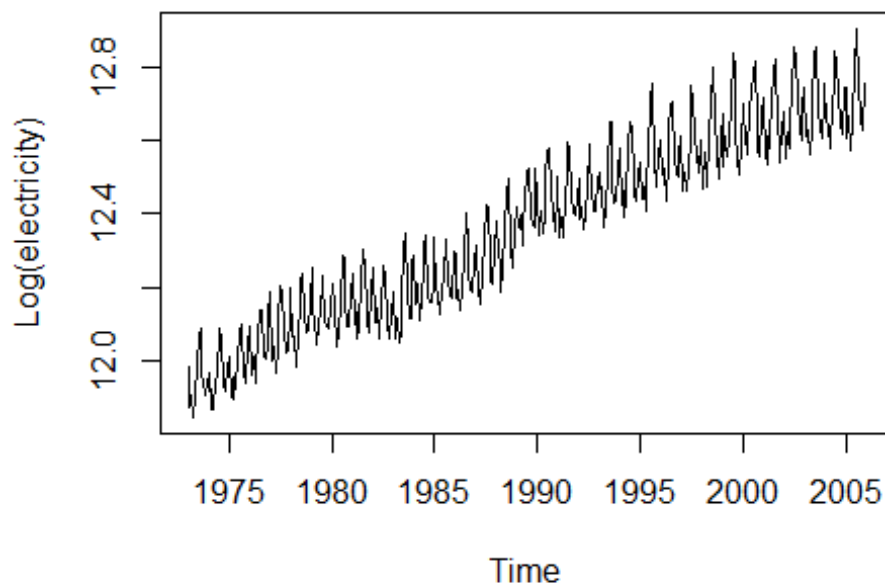
will be relatively stable and perhaps well-modeled by a stationary process. Notice that we take logs first and then compute first differences — the order does matter. In financial literature, the differences of the (natural) logarithms are usually called **returns**.

```
data(electricity)
plot(electricity)
```

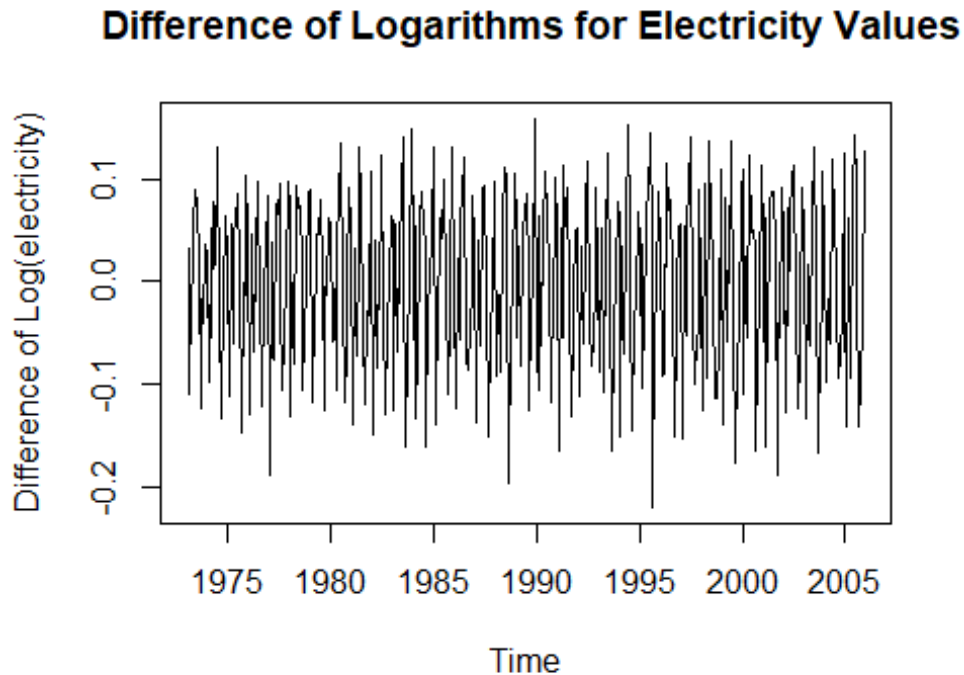


```
plot(log(electricity), ylab= "Log(electricity)",  
main= "Logarithms of Electricity Values")
```

### Logarithms of Electricity Values



```
plot(diff(log(electricity)), ylab= "Difference of Log(electricity)",
main= " Difference of Logarithms for Electricity Values")
```



### 7.2.3 Power Transformations

A flexible family of transformations, the **power transformations**, was introduced by Box and Cox (1964). For a given value of the parameter  $\lambda$ , the transformation is defined by

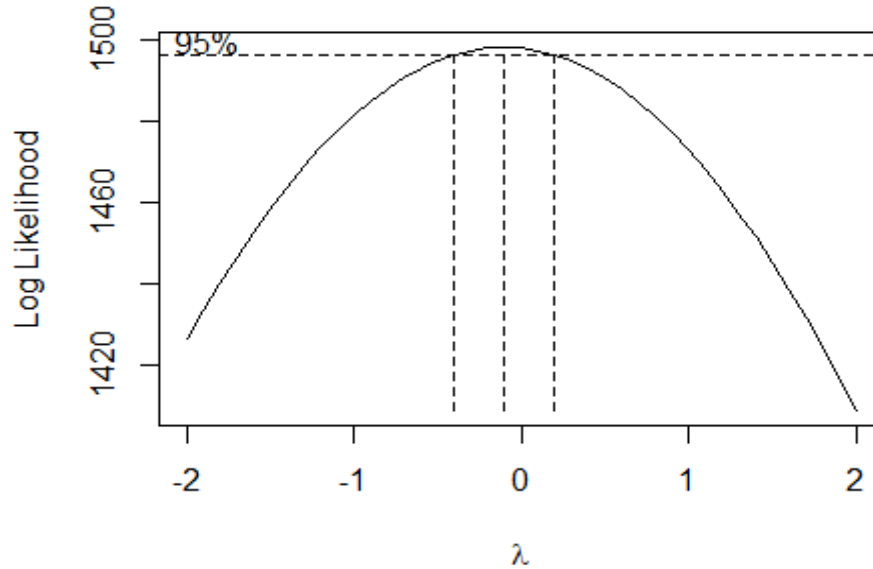
$$g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log(x) & \text{for } \lambda = 0 \end{cases}$$

$$\therefore \lim_{\lambda \rightarrow 0} \frac{x^\lambda - 1}{\lambda} = \log(x)$$

Note that  $\lambda = \frac{1}{2}$  produce square root transformation useful for Poisson data and  $\lambda = -1$  corresponds to reciprocal transformation.

Now the question arises how to get the value of  $\lambda$  which decides what transformation is to be used.

**BoxCox.ar(electricity)**



when  $\lambda = 0 \Rightarrow$  logarithmic transformation is required. This suggests logarithmic transformation for the data.

when  $\lambda = 1 \Rightarrow$  No transformation.

### 7.3 The Backshift Operator

The Backshift Operator is linear since

$$B(aY_t + bX_t + c) = aBY_t + bBX_t + c$$

#### 7.3.1 Moving Average Model — MA() Model

Consider the MA(1) Model. In terms of  $B$ , we can write as

$$Y_t = e_t - \theta e_{t-1}$$

$$= (1 - \theta B)e_t$$

$$Y_t = \theta(B)e_t$$

where  $\theta(B)$  is the MA characteristic polynomial estimated at  $B$ .

Since  $BY_t$  is itself a time series, it is meaningful to consider  $BBY_t$ . But clearly  $BBY_t = BY_{t-1} = Y_{t-2}$ , and we can write

$$B^2Y_t = Y_{t-2}$$

More generally, we have

$$B^m Y_t = Y_{t-m}$$

for any positive integer  $m$ . For a general  $MA(q)$  model, we can then write

$$\begin{aligned} Y_t &= e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \\ &= e_t - \theta B e_t - \theta B^2 e_t - \cdots - \theta_q B^q e_t \\ Y_t &= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) e_t \end{aligned}$$

OR

$$Y_t = \theta(B) e_t$$

where, again,  $\theta(B)$  is the MA characteristic polynomial evaluated at  $B$ .

### 7.3.2 Auto-Regressive Models — AR(p) Models

For auto-regressive models  $AR(p)$ , we first move all of the terms involving  $Y$  to the left-hand side

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p} = e_t$$

and then write

$$Y_t - \phi_1 B Y_t - \phi_2 B^2 Y_t - \cdots - \phi_p B^p Y_t = e_t$$

or

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t = e_t$$

which can be expressed as

$$\phi(B) Y_t = e_t$$

where  $\phi(B)$  is the AR characteristic polynomial evaluated at  $B$ .

### 7.3.3 ARMA(p, q) Models

Combining the two, the general ARMA(p,q) model may be written compactly as

$$\phi(B) Y_t = \theta(B) e_t$$

Differencing can also be conveniently expressed in terms of  $B$ . We have

$$\nabla Y_t = Y_t - Y_{t-1} = Y_t - B Y_t$$

$$\nabla Y_t = (1 - B) Y_t$$

with second differences given by

$$\nabla^2 Y_t = (1 - B)^2 Y_t$$

Effectively,  $\nabla = (1 - B)$  and  $\nabla^2 = (1 - B)^2$ .

The general  $ARIMA(p, d, q)$  model is expressed concisely as

$$\phi(B)(1 - B)^d Y_t = \theta(B)e_t$$

## 8 Parameter Estimation

This chapter deals with the problem of estimating the parameters of an  $ARIMA$  model based on the observed time series  $Y_1, Y_2, \dots, Y_n$ . With regard to non-stationarity, since the  $d^{th}$  difference of the observed series is assumed to be a stationary  $ARMA(p, q)$  process, we need only concern ourselves with the problem of estimating the parameters in such stationary models.

### 8.1 The Method of Estimation

1. Methods of Moments
2. Methods of Least Square Estimation
3. Methods of Maximum Likelihood Estimation
4. Bayesian Methods.

#### 8.1.1 1. Methods of Moments

The method consists of equating sample moments to corresponding theoretical moments and solving the resulting equations to obtain estimates of any unknown parameters.

##### 8.1.1.1 Autoregressive Models — $AR(p)$ Model

Consider first the  $AR(1)$  case. For this process, we have the simple relationship  $\rho_1 = \phi$ .

In the method of moments,  $\rho_1$  is equated to  $r_1$ , the *lag* 1 sample autocorrelation. Thus we can estimate  $\phi$  by

$$\hat{\phi} = r_1$$

Now consider the  $AR(2)$  case. The relationships between the parameters  $\phi_1$  and  $\phi_2$  and various moments are given by the *Yule-Walker* equations

$$\rho_1 = \phi_1 + \rho_1 \phi_2$$

$$\rho_2 = \rho_1 \phi_1 + \phi_2$$

Thus method of moments gives

$$r_1 = \phi_1 + r_1 \phi_2$$

$$r_2 = r_1 \phi_1 + \phi_2$$

which are then solved to obtain



$$\widehat{\phi}_1 = \frac{r_1(1 - r_1)}{1 - r_1^2}$$

and

$$\widehat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$$

The general  $AR(p)$  Model

$$\phi_1 + r_1\phi_2 + r_2\phi_3 + \cdots + r_{p-1}\phi_p = r_1$$

$$r_1\phi_1 + \phi_2 + r_1\phi_3 + \cdots + r_{p-2}\phi_p = r_2$$

$$r_{p-1}\phi_1 + r_{p-2}\phi_2 + r_{p-3}\phi_3 + \cdots + \phi_p = r_p$$

The estimates obtained in this way are also called **Yule-Walker estimates**.

#### **8.1.1.2 Moving Average Models — $MA(q)$ Model**

$MA(q)$  are quite complicated using method of moments.

#### **8.1.1.3 Mixed Model — $ARMA(1, 1)$ Model**

We consider only the  $ARMA(1,1)$  case. Recall Equation

$$\rho_k = \frac{(1 - \theta\phi)(\phi - \theta)}{1 - 2\theta\phi + \theta^2} \phi^{k-1} \quad \text{for } k \geq 1$$

Noting that  $\rho_2/\rho_1 = \phi$ , we can first estimate  $\phi$  as

$$\widehat{\phi} = \frac{r_2}{r_1}$$

Having done so, we can then use

$$r_1 = \frac{(1 - \theta\widehat{\phi})(\widehat{\phi} - \theta)}{1 - 2\theta\widehat{\phi} + \theta^2}$$

then we shall solve it for  $\theta$  to get  $\widehat{\theta}$ .

#### **8.1.1.4 Estimates of the Noise Variance**

In all cases, we can first estimate the process variance,  $\gamma_0 = \text{Var}(Y_t)$ , by the sample variance

$$s^2 = \frac{1}{n-1} \sum_{t=1}^n (Y_t - \bar{Y})^2$$

and use the relationship from  $AR(p)$

$$\widehat{\sigma_e^2} = (1 - \widehat{\phi}_1 r_1 - \widehat{\phi}_2 r_2 - \dots - \widehat{\phi}_p r_p) s^2$$

In particular, for an  $AR(1)$  process,

$$\widehat{\sigma_e^2} = (1 - r_1^2) s^2 \quad \{\because \widehat{\phi}_1 = r_1\}$$

For the  $MA(q)$  case, we have, using Equation

$$\widehat{\sigma_e^2} = \frac{s^2}{1 + \widehat{\theta}_1^2 + \widehat{\theta}_2^2 + \dots + \widehat{\theta}_q^2}$$

For the  $ARMA(1,1)$  process

$$\widehat{\sigma_e^2} = \frac{1 - \widehat{\phi}^2}{1 - 2\widehat{\phi}\widehat{\theta} + \widehat{\theta}^2} s^2$$

### 8.1.2 2. Methods of Least Square Estimation

Consider auto-regressive model

$$\begin{aligned} Y_t - \mu &= \phi(Y_{t-1} - \mu) + e_t \\ \Rightarrow e_t &= [Y_t - \mu - \phi(Y_{t-1} - \mu)] \\ \Rightarrow e_t^2 &= [Y_t - \mu - \phi(Y_{t-1} - \mu)]^2 \\ \Rightarrow \sum_{t=2}^n e_t^2 &= \sum_{t=2}^n [Y_t - \mu - \phi(Y_{t-1} - \mu)]^2 \\ &\Rightarrow \sum_{t=2}^n e_t^2 = S_c(\phi, \mu) \end{aligned}$$

We have to minimize this  $S_c(\mu, \phi)$  with respect to  $\mu$  and  $\phi$

$$\begin{aligned} \frac{\partial S_c(\mu, \phi)}{\partial \mu} &= 0 \quad \text{we have} \\ \frac{\partial S_c(\mu, \phi)}{\partial \mu} &= \sum_{t=2}^n 2 [Y_t - \mu - \phi(Y_{t-1} - \mu)](-1 + \phi) = 0 \\ \Rightarrow \hat{\mu} &= \frac{1}{(n-1)(1-\phi)} \left[ \sum_{t=2}^n Y_t - \phi \sum_{t=2}^n Y_{t-1} \right] \end{aligned}$$

Now for large  $n$

$$\frac{1}{(n-1)} \sum_{t=2}^n Y_t \approx \frac{1}{(n-1)} \sum_{t=2}^n Y_{t-1} \approx \bar{Y}$$

Thus, regardless of the value of  $\phi$

$$\hat{\mu} \approx \frac{1}{1 - \phi} (\bar{Y} - \phi \bar{Y}) = \bar{Y}$$

Consider now the minimization

$$S_c(\mu, \phi) = \sum_{t=2}^n [(Y_t - \bar{Y}) - \phi(Y_{t-1} - \bar{Y})]^2$$

differentiate with respect to  $\phi$ , we have

$$\frac{\partial S_c(\phi, \mu)}{\partial \phi} = \sum_{t=2}^n -2[(Y_t - \bar{Y}) - \phi(Y_{t-1} - \bar{Y})](Y_{t-1} - \bar{Y}) = 0$$

$$\hat{\phi} = \frac{\sum_{t=2}^n (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=2}^n (Y_{t-1} - \bar{Y})^2}$$

Except for one term missing in the denominator, namely  $(Y_n - \bar{Y})^2$ , this is the same as  $r_1$ . The lone missing term is negligible for stationary processes, and thus the least squares and method-of-moments estimators are nearly identical, especially for large samples.

### 8.1.3 Maximum Likelihood Estimation — MLE

- **Likelihood Function** : When joint density of  $Y_1, Y_2, \dots, Y_n$  is treated as a function of parameters, it is termed as likelihood function.
- **Maximum Likelihood Estimator** : The estimate of the parameters which maximizes likelihood function is termed as maximum likelihood estimator.

#### 8.1.3.1 Estimation of AR(1) Model

We know that  $e_t \sim N(0, \sigma_e^2)$ . Thus, the probability density function for each  $e_t$  is then

$$(2\pi\sigma_e^2)^{-1/2} \exp\left(\frac{-e_t^2}{2\sigma_e^2}\right) \quad \text{for } -\infty < e_t < \infty$$

by independence, the joint pdf for  $e_2, e_3, \dots, e_n$  is

$$(2\pi\sigma_e^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma_e^2} \sum_{t=2}^n e_t^2\right)$$

Now consider

$$Y_2 - \mu = \phi(Y_1 - \mu) + e_2$$

$$Y_3 - \mu = \phi(Y_2 - \mu) + e_3$$

⋮

$$Y_n - \mu = \phi(Y_{n-1} - \mu) + e_n$$

Therefore

$$f(Y_2, Y_3, \dots, Y_n | Y_1) = (2\pi\sigma_e^2)^{-(n-1)/2} \times \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{t=2}^n [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 \right\}$$

We know that

$$Y_1 \sim N \left( \mu, \frac{\sigma_e^2}{1 - \phi^2} \right)$$

Therefore

$$L(\phi, \mu, \sigma_e^2) = (2\pi\sigma_e^2)^{-n/2} (1 - \phi^2)^{1/2} \exp \left[ -\frac{1}{2\sigma_e^2} S(\phi, \mu) \right]$$

where,

$$S(\phi, \mu) = \sum_{t=2}^n [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1 - \phi^2)(Y_1 - \mu)^2$$

The function  $S(\phi, \mu)$  is called the **unconditional sum-of-squares function**.

The Likelihood function

$$l(\phi, \mu, \sigma_e^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_e^2) + \frac{1}{2} \log(1 - \phi^2) - \frac{1}{2\sigma_e^2} S(\phi, \mu)$$

- This function could be maximize using `optim()` function of R.
- Note also that  $S(\phi, \mu) = S_c(\phi, \mu) + (1 - \phi^2)(Y_1 - \mu)^2$

## 8.2 Properties of the Estimates

For large  $n$ , the estimators are approximately unbiased and normally distributed.

The variances and correlations are as follows :

$$\begin{aligned} AR(1): Var(\hat{\phi}) &\approx \frac{1 - \phi^2}{n} \\ AR(2): \begin{cases} Var(\hat{\phi}_1) \approx Var(\hat{\phi}_2) \approx \frac{1 - \phi_2^2}{n} \\ Corr(\hat{\phi}_1, \hat{\phi}_2) \approx -\frac{\phi_1}{1 - \phi_2} = -\rho_1 \end{cases} \\ MA(1): Var(\hat{\theta}) &\approx \frac{1 - \theta^2}{n} \end{aligned}$$

$$MA(2): \begin{cases} Var(\widehat{\theta}_1) \approx Var(\widehat{\theta}_2) \approx \frac{1 - \theta_2^2}{n} \\ Corr(\widehat{\theta}_1, \widehat{\theta}_2) \approx -\frac{\theta_1}{1 - \theta_2} \end{cases}$$

### 8.2.1 Parameter Estimates For Simulated Data Using AR(1) Model

```
data(ar1.s)
ar(ar1.s, order.max= 1, AIC=F , method= 'yw')

##
## Call:
## ar(x = ar1.s, order.max = 1, method = "yw", AIC = F)
##
## Coefficients:
##      1
## 0.8314
##
## Order selected 1  sigma^2 estimated as  1.382

ar(ar1.s, order.max= 1, AIC= F, method= 'ols')

##
## Call:
## ar(x = ar1.s, order.max = 1, method = "ols", AIC = F)
##
## Coefficients:
##      1
## 0.857
##
## Intercept: 0.02499 (0.1308)
##
## Order selected 1  sigma^2 estimated as  1.008

ar(ar1.s, order.max= 1, AIC= F, method= 'mle')

##
## Call:
## ar(x = ar1.s, order.max = 1, method = "mle", AIC = F)
##
## Coefficients:
##      1
## 0.8924
##
## Order selected 1  sigma^2 estimated as  1.041
```

### 8.2.2 Parameter Estimates For Simulated Data Using AR(2) Model

```
data(ar1.2.s)
ar(ar1.2.s, order.max= 1, AIC= F, method= 'yw')

##
## Call:
```

```

## ar(x = ar1.2.s, order.max = 1, method = "yw", AIC = F)
##
## Coefficients:
##      1
## 0.4699
##
## Order selected 1  sigma^2 estimated as  0.9198

ar(ar1.2.s, order.max= 1, AIC= F, method= 'ols')

##
## Call:
## ar(x = ar1.2.s, order.max = 1, method = "ols", AIC = F)
##
## Coefficients:
##      1
## 0.4731
##
## Intercept: -0.006084 (0.1237)
##
## Order selected 1  sigma^2 estimated as  0.9024

ar(ar1.2.s, order.max= 1, AIC= F, method= 'mle')

##
## Call:
## ar(x = ar1.2.s, order.max = 1, method = "mle", AIC = F)
##
## Coefficients:
##      1
## 0.4654
##
## Order selected 1  sigma^2 estimated as  0.8875

```

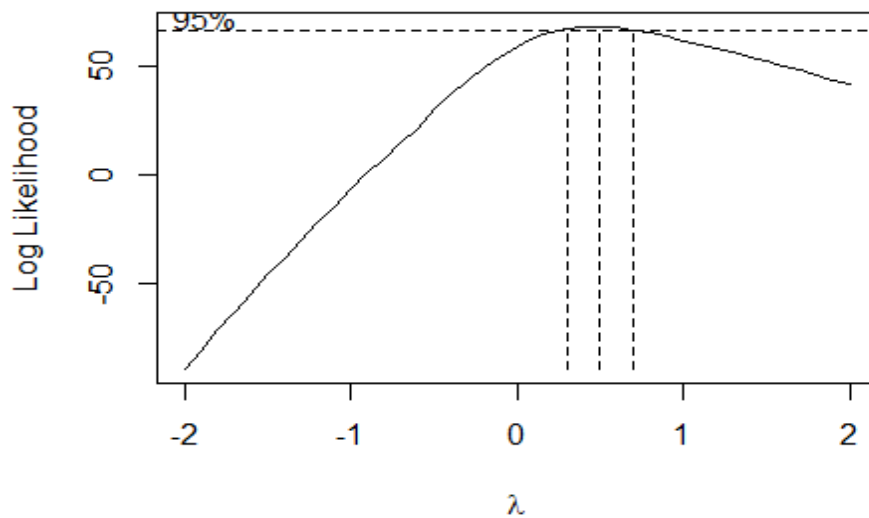
### 8.2.3 Canadian Hare Abundance Series — Hare

Consider the canadian hare abundance series as before we base our modelling on the square root of original abundance number based on partial auto-correlation function we see  $AR(3)$  Model was suitable.

```

data("hare")
BoxCox.ar(hare)

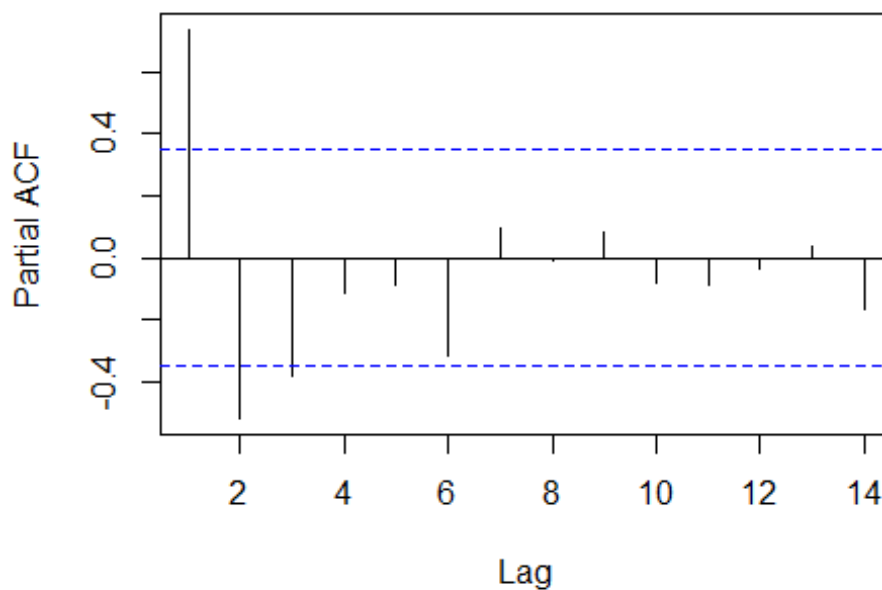
```



This output suggest should be square root transformation and using `BoxCox.ar()`. Suggest that it should be square root transformation.

```
pacf(hare^0.5)
```

### Series hare<sup>0.5</sup>



$\Rightarrow p = 3 \Rightarrow AR(3)$  Model should be fitted for hare data.

### 8.2.4 Fitting of ARIMA Model — AR(3)

```
data("hare")
arima(sqrt(hare), order= c(3, 0, 0))

##
## Call:
## arima(x = sqrt(hare), order = c(3, 0, 0))
##
## Coefficients:
##          ar1          ar2          ar3  intercept
##       1.0519   -0.2292   -0.3931      5.6923
## s.e.  0.1877    0.2942    0.1915     0.3371
##
## sigma^2 estimated as 1.066:  log likelihood = -46.54,  aic = 101.08
```

Here we see that  $\hat{\phi}_1 = 1.0519$ ,  $\hat{\phi}_2 = -0.2292$ , and  $\hat{\phi}_3 = -0.3930$ . We also see that the estimated noise variance is  $\hat{\sigma}_e^2 = 1.066$  and  $\text{ratio(z-ratio)} \frac{\hat{\phi}_1}{se(\hat{\phi}_1)} = \frac{1.0519}{0.1877} = 3.6041 \Rightarrow$  *significant*,  $\frac{\hat{\phi}_2}{se(\hat{\phi}_2)} = \frac{-0.2292}{0.2942} = -0.779 \Rightarrow$  *not – significant* and  $\frac{\hat{\phi}_3}{se(\hat{\phi}_3)} = \frac{-0.3931}{0.1915} = -2.0527 \Rightarrow$  *significant*

The estimated model would be written

$$\begin{aligned} & \sqrt{Y_t} - 5.6923 \\ &= 1.0519(\sqrt{Y_{t-1}} - 5.6923) - 0.2292(\sqrt{Y_{t-2}} - 5.6923) - 0.3930(\sqrt{Y_{t-3}} - 5.6923) + e_t \end{aligned}$$

or

$$\sqrt{Y_t} = 3.25 + 1.0519\sqrt{Y_{t-1}} - 0.2292\sqrt{Y_{t-2}} - 0.3930\sqrt{Y_{t-3}} + e_t$$

### Example of MA(1) Model

As a last example, we return to the oil price series. The sample ACF shown in the following which suggested an MA(1) model on the differences of the logs of the prices.

```
data(oil.price)

arima(log(oil.price), order= c(0,1,1), method= 'CSS') #Conditional Sum of Square

##
## Call:
## arima(x = log(oil.price), order = c(0, 1, 1), method = "CSS")
##
## Coefficients:
##          ma1
##       0.2731
## s.e.  0.0681
##
## sigma^2 estimated as 0.006731:  part log likelihood = 259.58
```



```

arima(log(oil.price), order= c(0,1,1), method= 'ML') # Maximum Likelihood
##
## Call:
## arima(x = log(oil.price), order = c(0, 1, 1), method = "ML")
##
## Coefficients:
##          ma1
##          0.2956
## s.e.      0.0693
##
## sigma^2 estimated as 0.006689:  log likelihood = 260.29,  aic = -518.58

```

### 8.3 Bootstrapping ARIMA Models

```

data("hare")
res<- arima(sqrt(hare), order= c(3,0,0))

coef.cond.norm<- arima.boot(res, cond.boot= T, is.normal= T, B= 1000,
init= sqrt(hare))
# coef.cond.norm
dim(coef.cond.norm)

## [1] 1000    5

head(coef.cond.norm)

##          ar1          ar2          ar3 intercept noise var
## [1,] 0.9102461 -0.31808454 -0.3101608  5.594744 1.4890163
## [2,] 0.7751742 -0.09333595 -0.5347781  5.636021 0.8256838
## [3,] 0.8826921  0.22369476 -0.6788125  5.630259 0.5112204
## [4,] 1.2444489 -0.54086888 -0.1983053  5.139973 0.9153553
## [5,] 1.1723716 -0.54049810 -0.2186630  6.005468 0.7560832
## [6,] 1.0775729 -0.29477058 -0.3207675  5.891328 0.9038140

temp<- apply(coef.cond.norm, 2, function(x) {quantile(x, c(0.025, 0.975))})
temp

##          ar1          ar2          ar3 intercept noise var
## 2.5%  0.6156069 -0.6860157 -0.66475591  5.028042 0.5678981
## 97.5% 1.3072532  0.2514345  0.01014008  6.383945 1.4784757

temp_mean<- apply(coef.cond.norm, 2, mean)
temp_mean

##          ar1          ar2          ar3 intercept noise var
## 0.9394614 -0.1636588 -0.3819844  5.7281329 0.9721643

temp_median<- apply(coef.cond.norm, 2, median)
temp_median

##          ar1          ar2          ar3 intercept noise var
## 0.9345375 -0.1396426 -0.3942623  5.7238108 0.9517358

```

```
signif(temp, 3)
```

```
##          ar1      ar2      ar3 intercept noise var
## 2.5%  0.616 -0.686 -0.6650      5.03      0.568
## 97.5% 1.310  0.251  0.0101      6.38      1.480
```

```
round(temp, 3)
```

```
##          ar1      ar2      ar3 intercept noise var
## 2.5%  0.616 -0.686 -0.665      5.028      0.568
## 97.5% 1.307  0.251  0.010      6.384      1.478
```

```
round(rbind(temp_median, temp), 3)
```

```
##          ar1      ar2      ar3 intercept noise var
## temp_median 0.935 -0.140 -0.394      5.724      0.952
## 2.5%          0.616 -0.686 -0.665      5.028      0.568
## 97.5%          1.307  0.251  0.010      6.384      1.478
```

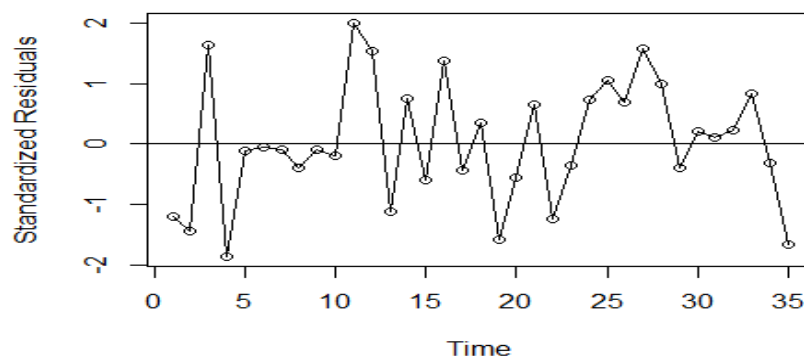
## 8.4 MODEL DIAGNOSTICS

```
data(color)
```

```
m1.color<- arima(color, order= c(1,0,0)); m1.color
```

```
##
## Call:
## arima(x = color, order = c(1, 0, 0))
##
## Coefficients:
##          ar1 intercept
##          0.5705    74.3293
## s.e.  0.1435     1.9151
##
## sigma^2 estimated as 24.83:  log likelihood = -106.07,  aic = 216.15
```

```
plot(rstandard(m1.color), ylab= 'Standardized Residuals', type ='o')
abline(h= 0)
```



### 8.4.1 Standardized Residuals from Log Oil Price IMA(1,1) Mode

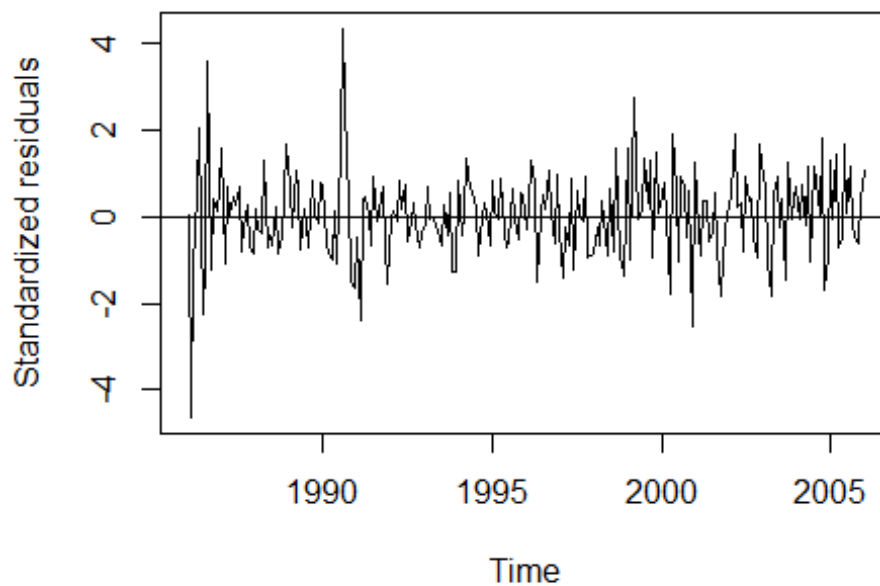
#### IMA : Integrated Moving Average

```
data(oil.price)

m1.oil<- arima(log(oil.price), order= c(0,1,1))
m1.oil

##
## Call:
## arima(x = log(oil.price), order = c(0, 1, 1))
##
## Coefficients:
##          ma1
##          0.2956
## s.e.    0.0693
##
## sigma^2 estimated as 0.006689:  log likelihood = 260.29,  aic = -518.58

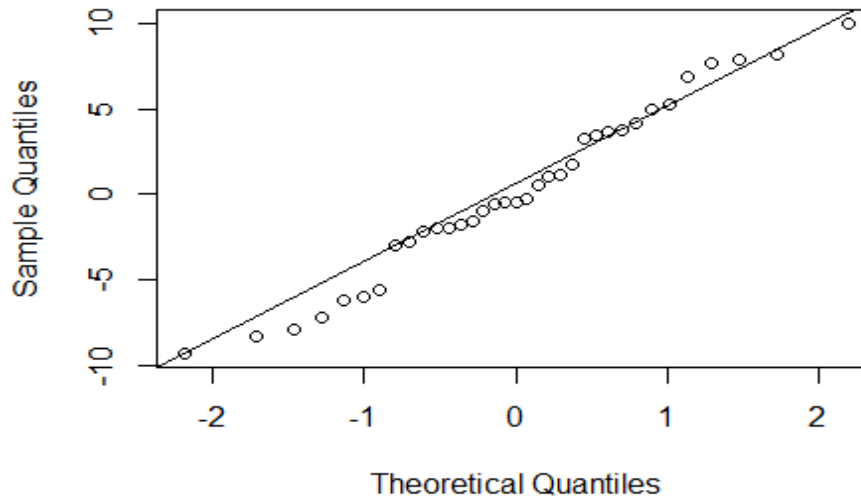
plot(rstandard(m1.oil), ylab= 'Standardized residuals', type= 'l')
abline(h= 0)
```



#### 8.4.2 Normality of the Residuals

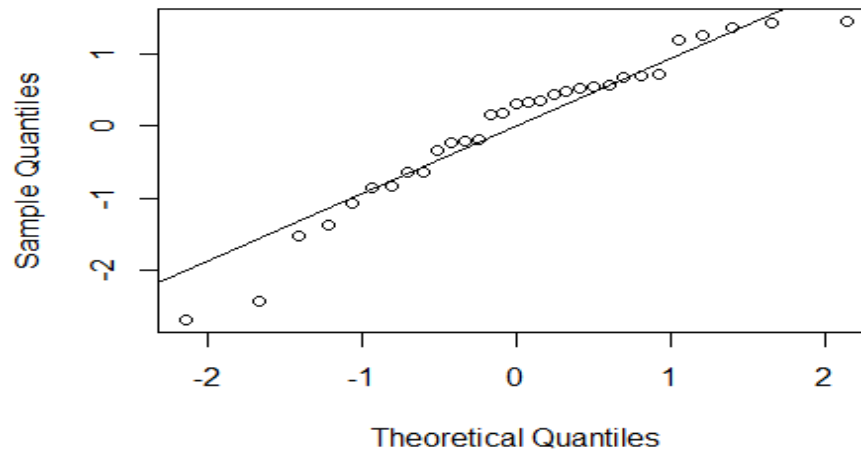
```
data("color")  
m1.color<- arima(color, order= c(1,0,0))  
  
qqnorm(residuals(m1.color), main= "QQ Plot of Residuals from AR(1) Model")  
qqline(residuals(m1.color))
```

**QQ Plot of Residuals from AR(1) Model**



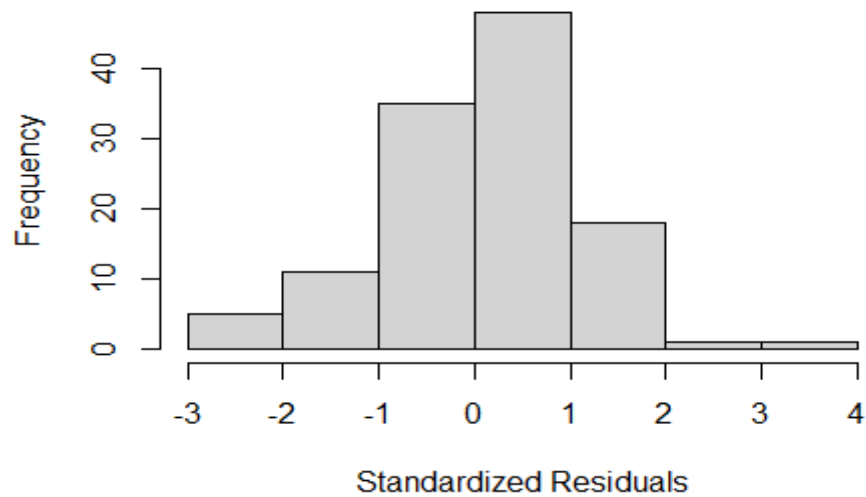
```
data(hare)  
m1.hare<- arima(sqrt(hare), order= c(3,0,0))  
  
qqnorm(residuals(m1.hare), main= "QQ Plot of Residuals from AR(3) Model")  
qqline(residuals(m1.hare))
```

**QQ Plot of Residuals from AR(3) Model**



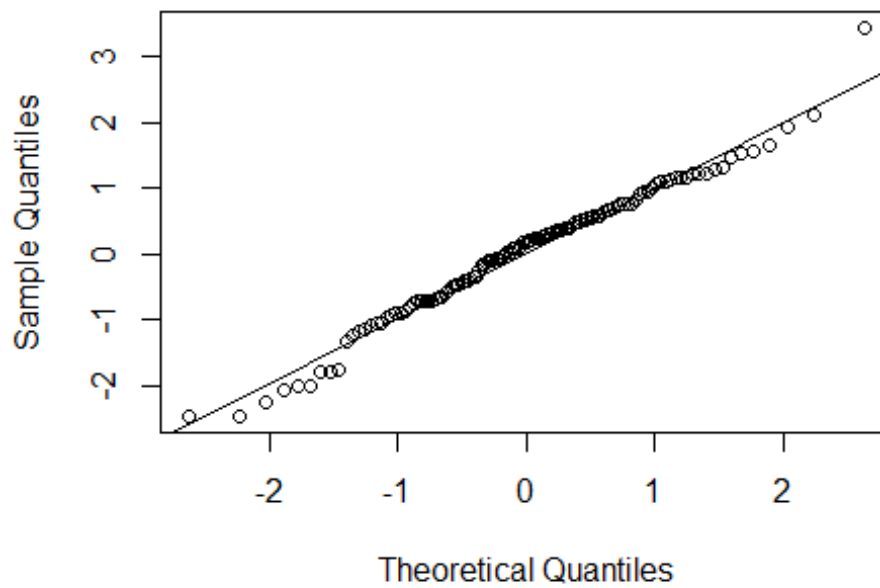
```
hist(window(rstandard(m1.co2), start= c(1995,2)),  
xlab= 'Standardized Residuals')
```

**listogram of window(rstandard(m1.co2), start = c(199**



```
# QQnormal  
qqnorm(window(rstandard(m1.co2), start=c(1995,2)))  
qqline(window(rstandard(m1.co2), start=c(1995,2)))
```

**Normal Q-Q Plot**



```
shapiro.test(window(rstandard(m1.co2), start= c(1995,2)))
## Shapiro-Wilk normality test
##
## data: window(rstandard(m1.co2), start = c(1995, 2))
## W = 0.98205, p-value = 0.1134
```

Here we again see the one outlier in the upper tail, but the Shapiro-Wilk test of normality has a test statistic of  $W = 0.982$ , leading to a p-value of 0.11, and normality is not rejected at any of the usual significance levels.

### 8.4.3 Auto-Correlation Residuals — Box-Ljung Test

Box and Ljung or Box-Pierce proposed a test statistics defined as

$$Q = n(n+2) \left( \frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \dots + \frac{\hat{r}_k^2}{n-k} \right) \sim \chi_{k-p-q}^2$$

```
acf(residuals(m1.color), plot= F)$acf
```

```
## , , 1
```

```
##          [,1]
## [1,] -0.05138241
## [2,]  0.03224346
## [3,]  0.04749703
## [4,]  0.02088157
## [5,] -0.01729829
## [6,] -0.01924314
## [7,] -0.17341651
## [8,] -0.28136877
## [9,]  0.06127493
## [10,] -0.27045217
## [11,] -0.05123367
## [12,] -0.02712861
## [13,]  0.18344079
## [14,] -0.04423987
## [15,]  0.18564229
```

```
signif(acf(residuals(m1.color), plot= F)$acf[1:6], 2) # display the first 6
acf values to 2 significant digits
```

```
## [1] -0.051  0.032  0.047  0.021 -0.017 -0.019
```

The Ljung-Box test statistic with  $K = 6$  is equal to

$$Q = 35(35+2) \left( \frac{(-0.051)^2}{35-1} + \frac{(0.032)^2}{35-2} + \frac{(0.047)^2}{35-3} + \frac{(0.021)^2}{35-4} + \frac{(-0.017)^2}{35-5} + \frac{(-0.019)^2}{35-6} \right) \approx 0.28$$

$$Q \approx 0.28 \sim \chi_{6-1-0}^2 \Rightarrow Q \approx 0.28 \sim \chi_5^2$$

Now compute the  $p$  – value

```
p_value<- pchisq(0.28, df= 5, lower.tail= F)
p_value
```

```
## [1] 0.9980024
```

The  $p$  – value of 0.998, so we have no evidence to reject the null hypothesis that the error terms are uncorrelated

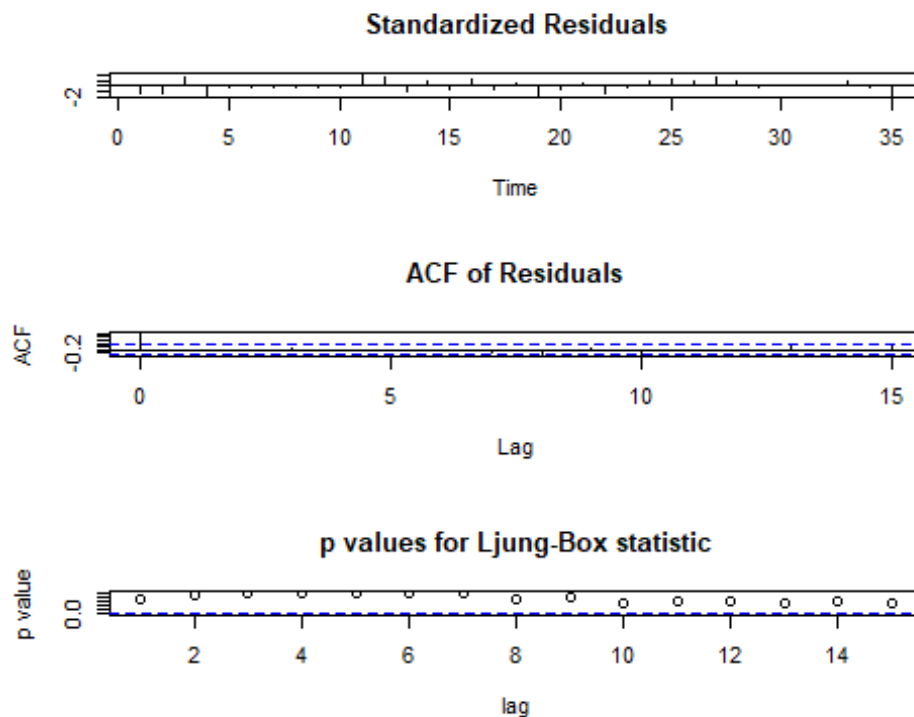
**OR** Since p-value is not lesser than 0.05 this implies that we don't reject  $H_0$

```
LB.test(m1.co2, lag=24)
```

```
##
## Box-Ljung test
##
## data: residuals from m1.co2
## X-squared = 25.587, df = 22, p-value = 0.2698
```

The **Ljung-Box** test for this model gives a chi-squared value of 25.59 with 22 degrees of freedom, leading to a p-value of 0.27—a further indication that the model has captured the dependence in the time series.

```
tsdiag(m1.color, gof= 15, omit.initial= F)
```



## 9 FORECASTING

One of the primary objectives of building a model for a time series is to be able to forecast the values for that series at future times. Of equal importance is the assessment of the precision of those forecasts.

### 9.1 Minimum Mean Square Error Forecasting

Based on the available history of the series up to time  $t$ , namely  $Y_1, Y_2, \dots, Y_{t-1}, Y_t$ , we would like to forecast the value of  $Y_{t+l}$  that will occur  $l$  time units into the future. We call time  $t$  the forecast origin and  $l$  the lead time for the forecast, and denote the forecast itself as  $\hat{Y}_{t(l)}$ .

The minimum mean square error forecast is given by

$$\hat{Y}_{t(l)} = E(Y_{t+l} | Y_1, Y_2, \dots, Y_t)$$

### 9.2 Deterministic Trends

Consider once more the deterministic trend model

$$Y_t = \mu_t + X_t$$

with  $E(X_t) = 0$  and  $V(X_t) = \gamma_0$  then,

$$\begin{aligned}\hat{Y}_{t(l)} &= E(\mu_{t+l} + X_{t+l} | Y_1, Y_2, \dots, Y_t) \\ &= E(\mu_{t+l} | Y_1, Y_2, \dots, Y_t) + E(X_{t+l} | Y_1, Y_2, \dots, Y_t) \\ \hat{Y}_{t(l)} &= \mu_{t+l} + E(X_{t+l}) \\ \hat{Y}_{t(l)} &= \mu_{t+l} + 0\end{aligned}$$

or

$$\hat{Y}_{t(l)} = \mu_{t+l} \quad \text{--- (1)}$$

For the linear trend case,  $\mu_t = \beta_0 + \beta_1 t$ , the forecast is

$$\hat{Y}_{t(l)} = \beta_0 + \beta_1(t + l)$$

For **seasonal model**, we have

$$\mu_t = \mu_{t+12}$$

Our forecast is

$$\hat{Y}_{t(l)} = \mu_{t+12+l} = \hat{Y}_{t(l+12)} =$$

Thus the forecast will also be periodic, as desired.

The **Forecast Error**,  $(e_t(l))$  is given by



$$\begin{aligned}
e_t(l) &= Y_{t+l} - \hat{Y}_{t(l)} \\
&= \mu_{t+l} + X_{t+l} - \mu_{t+l} \\
e_t(l) &= X_{t+l}
\end{aligned}$$

so that

$$E(e_t(l)) = E(X_{t+l}) = 0$$

That is, the forecasts are **unbiased**. Also

$$Var(e_t(l)) = Var(X_{t+l}) = \gamma_0$$

is the forecast error variance for all lead times  $l$ .

**Example :** The cosine trend model for the average monthly temperature series was estimated as

$$\hat{\mu}_t = 46.26660 + (-2.1697)(2\pi t) + (-2.1697)\sin(2\pi t)$$

Here time is measured in years with a starting value of January 1964, frequency  $f = 1$  per year and the final observed value is for December 1975. To forecast the June 1976 temperature value, we use  $t = 1976.41667$  as the time value and obtain

$$\begin{aligned}
\hat{\mu}_t &= 46.2660 + (-26.7079)\cos(2\pi(1976.41667)) + (-2.1697)\sin(2\pi(1976.41667)) \\
\hat{\mu}_t &= 68.3^\circ F
\end{aligned}$$

### 9.3 ARIMA Forecasting

#### 9.3.1 AR(1)

We shall first illustrate many of the ideas with the simple AR(1) Model

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t \quad \text{--- (2)}$$

Consider the problem of forecasting one time unit into the future. Replacing  $t$  by  $t + 1$  in the above expression, we have

$$Y_{t+1} - \mu = \phi(Y_t - \mu) + e_{t+1} \quad \text{--- (3)}$$

Given  $Y_1, Y_2, \dots, Y_{t-1}, Y_t$ , we take the conditional expectations of both sides of Equation (3) and obtain

$$\hat{Y}_t(l) - \mu = \phi E[E(Y_t | Y_1, Y_2, \dots, Y_t) - \mu] + E(e_{t+1} | Y_1, Y_2, \dots, Y_t)$$

Now, from the properties of conditional expectation, we have conditional expectation of

$$E(Y_t | Y_1, Y_2, \dots, Y_t) = Y_t$$

Also, since  $e_{t+l}$  is independent of  $Y_1, Y_2, \dots, Y_{t-1}, Y_t$ , we obtain

$$E(e_{t+1}|Y_1, Y_2, \dots, Y_t) = E(e_{t+l}) = 0$$

Thus,

$$\hat{Y}_t(l) = \phi(Y_t - \mu) + \mu$$

Similarly, we consider a general lead time  $l$

$$\begin{aligned}\hat{Y}_t(l) &= \mu + \phi(\hat{Y}_t(l-1) - \mu) \\ &= \phi[\hat{Y}_t(l-1) - \mu] + \mu \\ &= \phi[\phi(\hat{Y}_t(l-2) + \mu - \mu)] + \mu \\ &=: \\ \hat{Y}_t(l) &= \phi^l(Y_t - \mu) + \mu\end{aligned}$$

**Numerical Illustration :**

```
data("color")
m1.color<- arima(color, order= c(1,0,0))
m1.color

##
## Call:
## arima(x = color, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##      0.5705    74.3293
## s.e.  0.1435     1.9151
##
## sigma^2 estimated as 24.83:  log likelihood = -106.07,  aic = 216.15
```

Thus  $\hat{\phi} = 0.5705$  and  $\hat{\mu} = 74.3293$ . The lost property observed is 67. So,

$$\hat{Y}_t(1) = 74.3293 + (0.5705)(67 - 74.3293) = 70.14793$$

For lead time 2

$$\hat{Y}_t(2) = 74.3293 + (0.5705)(70.14793 - 74.3293) = 71.94383$$

Alternatively,

$$\hat{Y}_t(2) = 74.3293 + (0.5705)^2(67 - 74.3293) = 71.92823$$

At lead 5

$$\hat{Y}_t(5) = 74.3293 + (0.5705)^5(67 - 74.3293) = 73.88636$$

Similarly,

$$\hat{Y}_t(10) = 74.3293 + (0.5705)^{10}(67 - 74.3293) = 74.30253$$

which is very nearly  $\mu (= 74.3293)$ .

Thus, we have  $|\phi| < 1$ , then we have

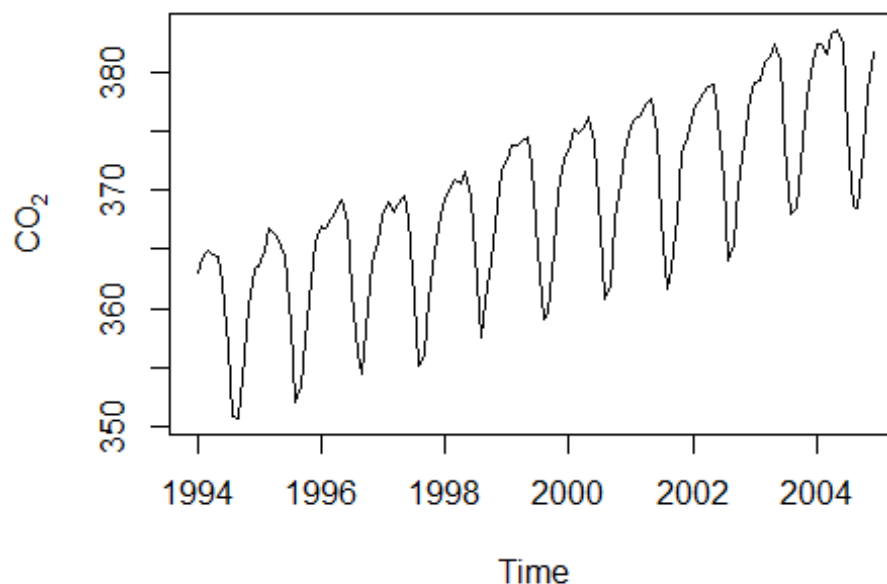
$$\hat{Y}_t(l) \approx \mu \quad \text{for large } l$$

## 10 Seasonal Models

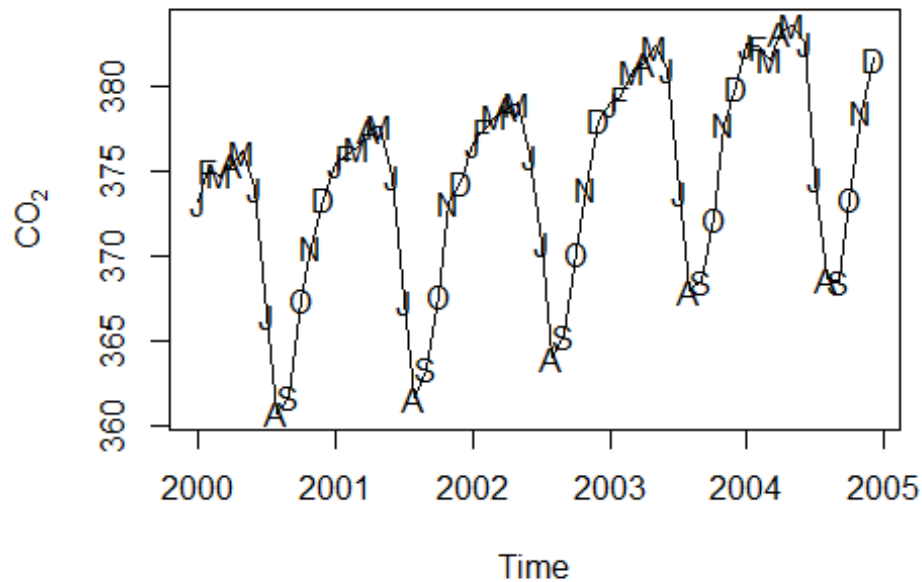
In earlier case we saw how seasonal deterministic trends might be modeled. However in many areas time series are used particularly businesses and economics. The assumption of any deterministic trend is quite suspect even though cyclical trend are very common in such series.

Here is an example of labels of  $CO_2$  are monitor at several sights around the world to investigate atmospheric changes. One of the sight is alert (North- West Territories Canada) near the archic circle.

```
data("co2")
plot(co2, ylab= expression(CO[2]))
```



```
plot(window(co2, start= c(2000,1)), ylab= expression(CO[2]))
Month<- c('J','F','M','A','M','J','J','A','S','O','N','D')
points(window(co2, start= c(2000,1)), pch= Month)
```



## 10.1 Seasonal ARIMA Models

We begin by studying stationary models and then consider nonstationary generalizations. We let  $s$  denote the known seasonal period; for monthly series  $s = 12$  and for quarterly series  $s = 4$ .

Consider the time series generated according to

$$Y_t = e_t - \theta e_{t-12}$$

Notice that

$$\text{Cov}(Y_t, Y_{t-1}) = \text{Cov}(e_t - \theta e_{t-12}, e_{t-1} - \theta e_{t-13}) = 0$$

but that

$$\text{Cov}(Y_t, Y_{t-12}) = \text{Cov}(e_t - \theta e_{t-12}, e_{t-12} - \theta e_{t-24})$$

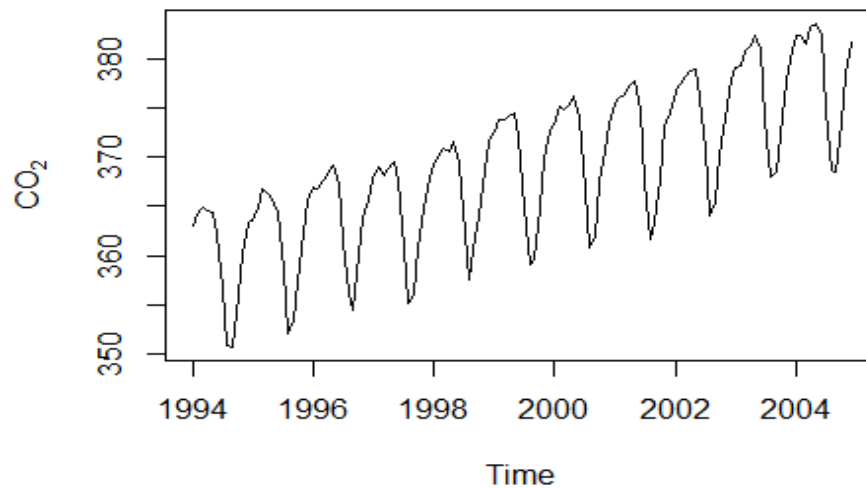
$$\text{Cov}(Y_t, Y_{t-12}) = -\theta \sigma_e^2$$

It is easy to see that such a series is stationary and has nonzero autocorrelations only at *lag* 12.

## 10.2 Model Specification

As always, a careful inspection of the time series plot is the first step.

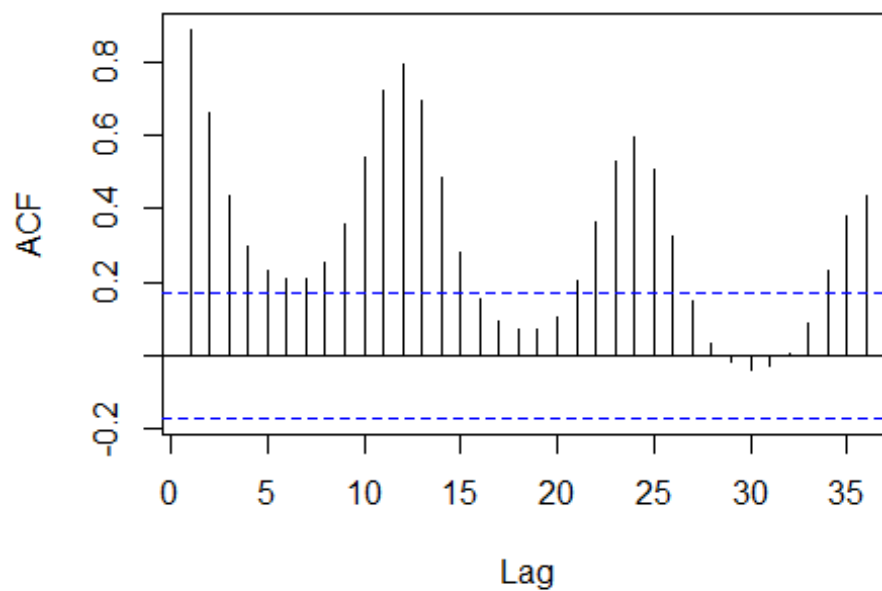
```
plot(co2, ylab= expression(CO[2]))
```



The above display displays monthly carbon dioxide  $CO_2$  levels in northern Canada. The upward trend alone would lead us to specify a nonstationary model.

```
acf(as.vector(co2), lag.max= 36)
```

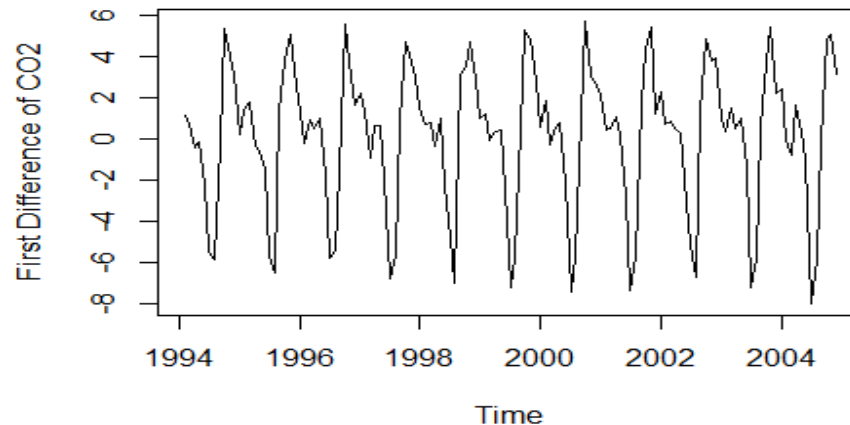
### Series as.vector(co2)



The above plot shows sample auto-correlation function in that series. The seasonal auto-correlation relationships are shown quite prominently in this display.

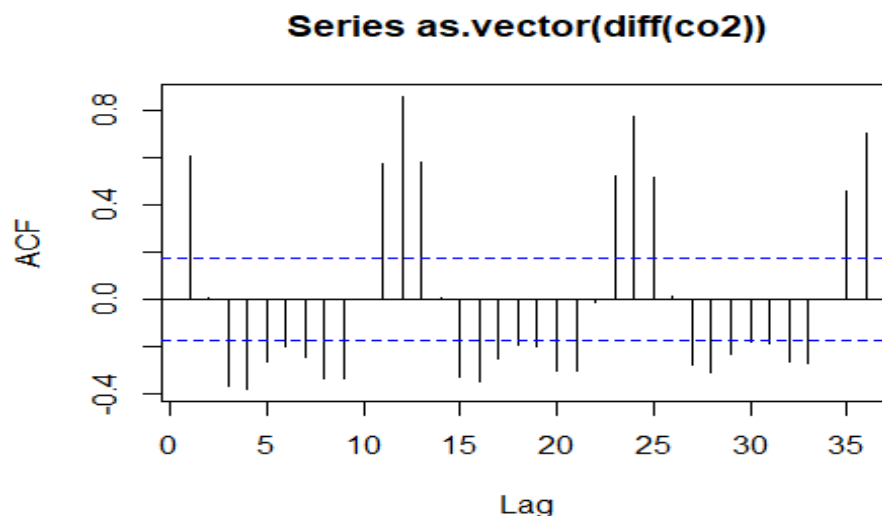
Notice the strong correlation at lags 12, 24, 36, and so on. In addition, there is substantial other correlation that needs to be modeled.

```
plot(diff(co2), ylab= 'First Difference of CO2', xlab= 'Time')
```



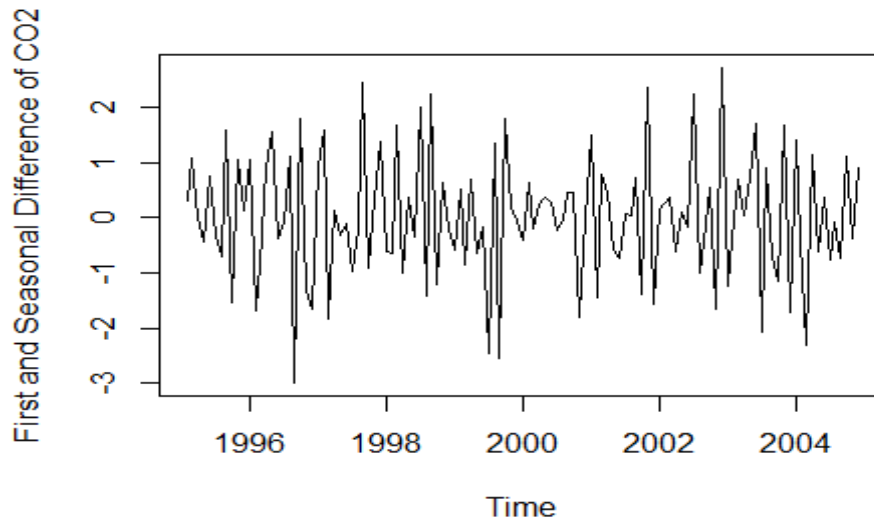
The general upward trend has now disappeared but the strong seasonality is still present, as evidenced by the behavior shown in the above figure. Perhaps seasonal differencing will bring us to a series that may be modeled parsimoniously.

```
acf(as.vector(diff(co2)), lag.max= 36)
```

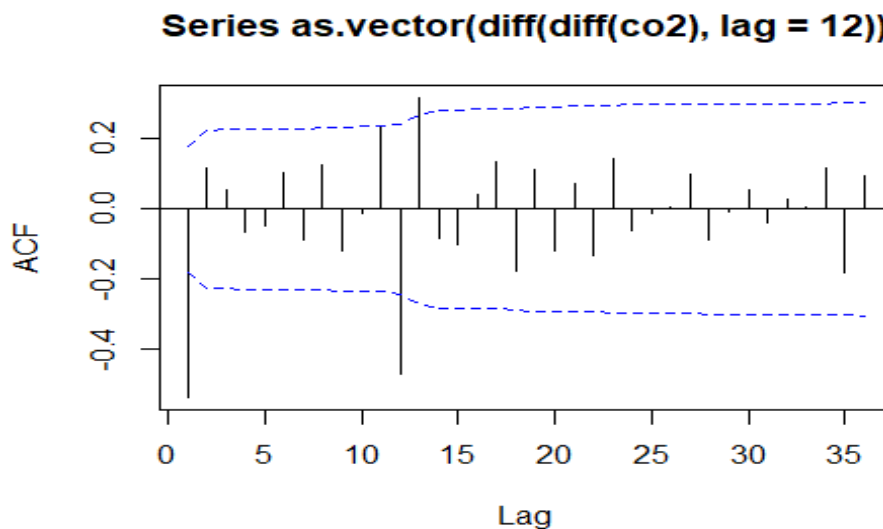


From above plot it shows strong seasonality, which could be removed by plotting seasonal differences.

```
plot(diff(diff(co2), lag= 12), xlab= 'Time', ylab= 'First and Seasonal
Difference of CO2')
```



```
acf(as.vector(diff(diff(co2), lag= 12)), lag.max= 36, ci.type= 'ma')
```



confirms that very little autocorrelation remains in the series after these two differences have been taken. This plot also suggests that a simple model which incorporates the *lag* 1 and *lag* 12 autocorrelations might be adequate.

We will consider specifying the multiplicative, seasonal  $ARIMA(0,1,1) \times (0,1,1)_{12}$  model.

$$\nabla_{12}\nabla Y_t = e_t - \theta e_{t-1} - \theta e_{t-12} + \theta\theta e_{t-13}$$

### 10.2.1 Model Fitting

```
m1.co2<- arima(co2, order= c(0,1,1), seasonal= list(order= c(0,1,1),
period= 12)) ; m1.co2
## Call:
## arima(x = co2, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1),
period = 12))
##
## Coefficients:
##          ma1      sma1
##      -0.5792  -0.8206
## s.e.    0.0791   0.1137
##
## sigma^2 estimated as 0.5446:  log likelihood = -139.54,  aic = 283.08
```

## 11 TIME SERIES MODELS OF HETEROSCEDASTICITY

The models discussed so far concern the conditional mean structure of time series data. However, more recently, there has been much work on modeling the conditional variance structure of time series data—mainly motivated by the needs for financial modeling.

Let  $Y_t$  be a time series of interest. The conditional variance of  $Y_t$  given the past  $Y$  values,  $Y_{t-1}, Y_{t-2}, \dots$ , measures the uncertainty in the deviation of  $Y_t$  from its conditional mean  $E(Y_t|Y_{t-1}, Y_{t-2}, \dots)$ . If  $Y_t$  follows some ARIMA model, the (one-step ahead) conditional variance is always equal to the noise variance for any present and past values of the process. Indeed, the constancy of the conditional variance is true for predictions of any fixed number of steps ahead for an ARIMA process. In practice, the (one-step-ahead) conditional variance may vary with the current and past values of the process, and, as such, the conditional variance is itself a random process, often referred to as the conditional variance process. For example, daily returns of stocks are often observed to have larger conditional variance following a period of violent price movement than a relatively stable period. The development of models for the conditional variance process with which we can predict the variability of future values based on current and past data is the main concern of the present chapter.

### 11.1 Some Common Features of Financial Time Series

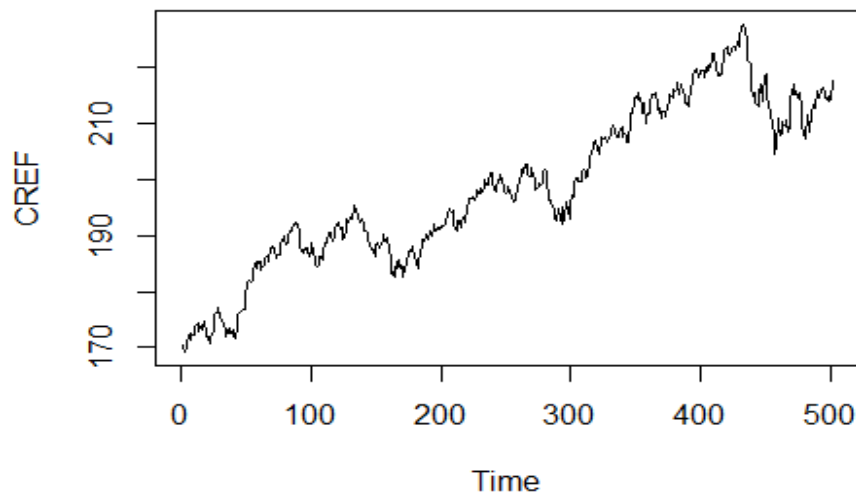
As an example of financial time series, we consider the daily values of a unit of the CREF stock fund over the period from August 26, 2004 to August 15, 2006. The CREF stock fund is a fund of several thousand stocks and is not openly traded in the stock market. Since stocks are not traded over weekends or on holidays, only on so-called trading days, the CREF data do not change over weekends and holidays. For simplicity, we will analyze the data as if they were equally spaced. Exhibit 12.1 shows the time series plot of the CREF data. It shows a generally increasing trend with a hint of higher variability with higher level of the stock value. Let  $p_t$  be the time series of, say, the daily price of some financial asset. The (continuously compounded) return on the  $t^{th}$  day is defined as

$$r_t = \log(p_t) - \log(p_{t-1})$$



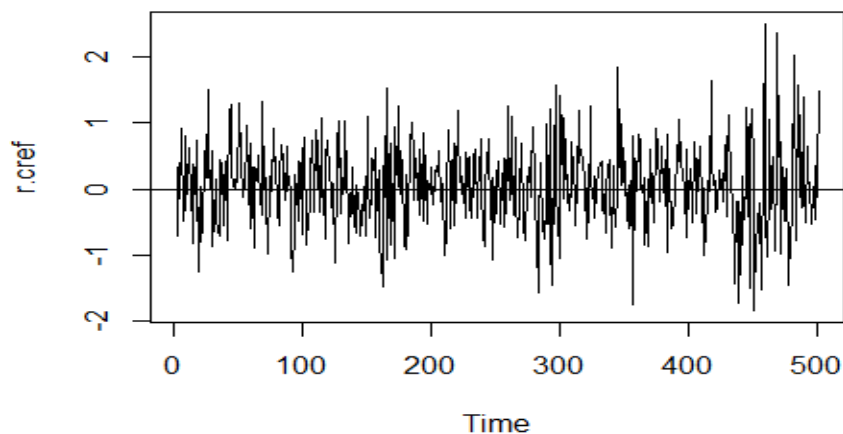
Sometimes the returns are then multiplied by 100 so that they can be interpreted as percentage changes in the price. The multiplication may also reduce numerical errors as the raw returns could be very small numbers and render large rounding errors in some calculations.

```
data(CREF)
plot(CREF)
```



```
r.cref<- diff(log(CREF))*100
plot(r.cref,
main= "Daily CREF Stock Returns: August 26, 2004 to August 15,2006")
abline(h= 0)
```

**ily CREF Stock Returns: August 26, 2004 to August 1**



```
mean(r.cref)
```

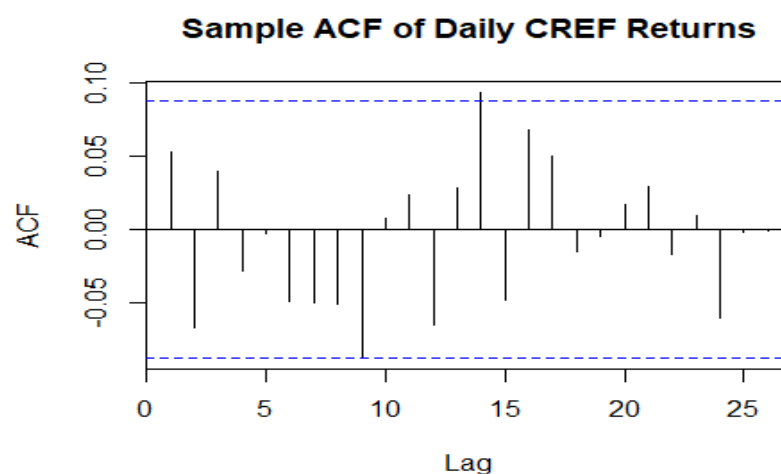
```
## [1] 0.04929517
```

```
sd(r.cref)/sqrt(length(r.cref))
```

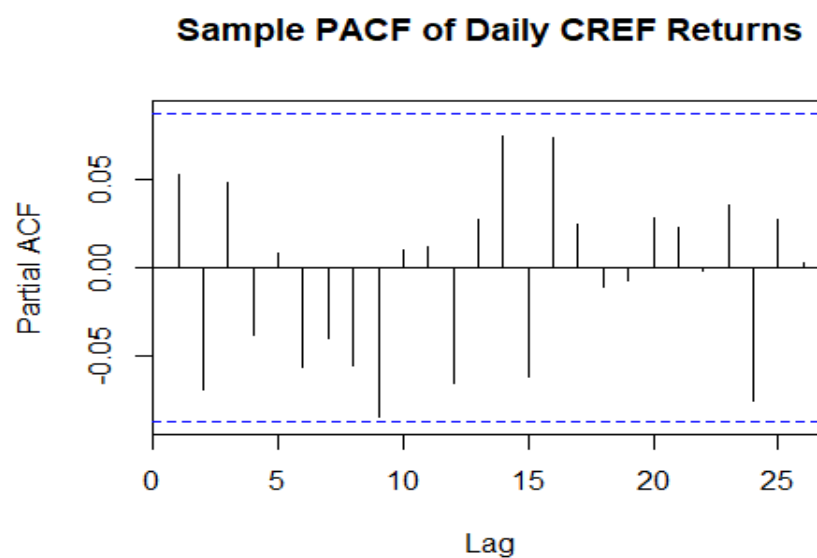
```
## [1] 0.02884742
```

Form above plot it is clear that returns were more volatile over some time periods and became very volatile toward the end of the study period. The average CREF return equals 0.0493 with a standard error of 0.02885.

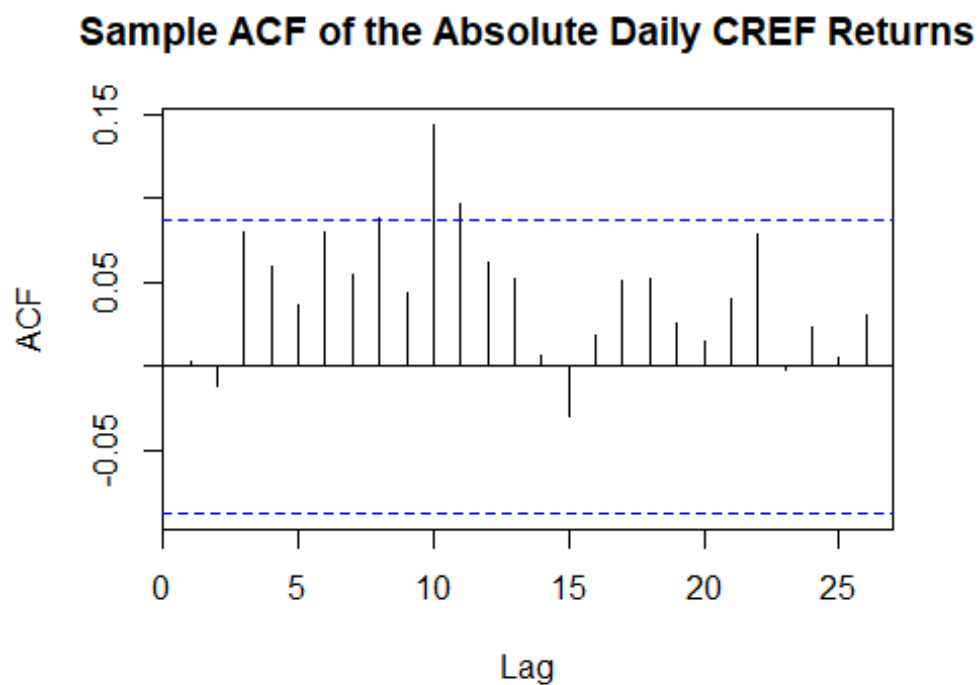
```
acf(r.cref, main= "Sample ACF of Daily CREF Returns")
```



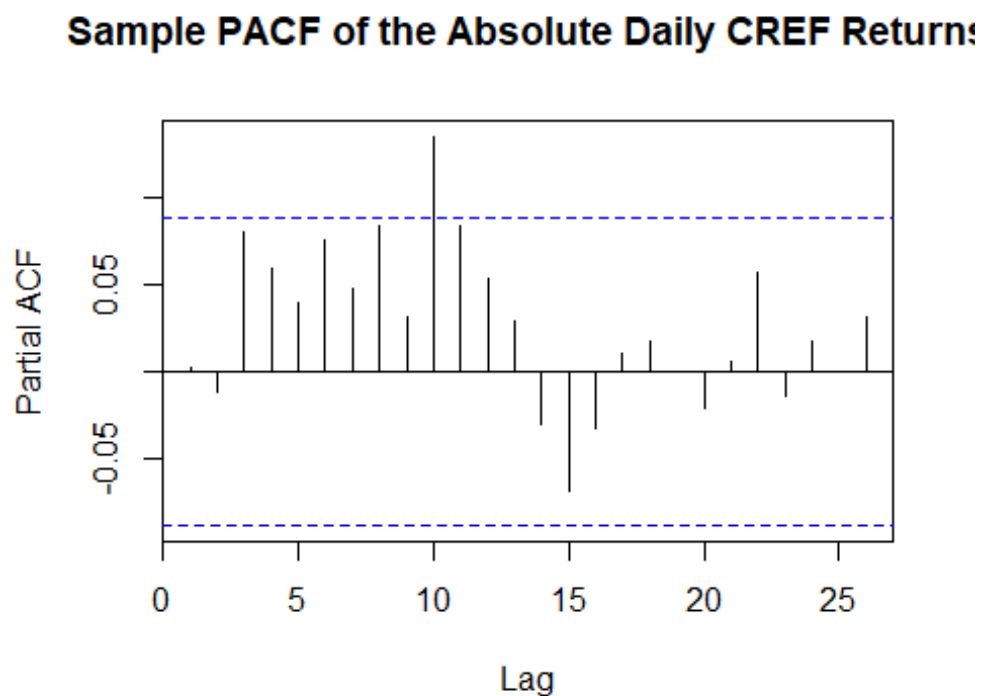
```
pacf(r.cref, main= "Sample PACF of Daily CREF Returns")
```



```
acf(abs(r.cref), main= "Sample ACF of the Absolute Daily CREF Returns")
```

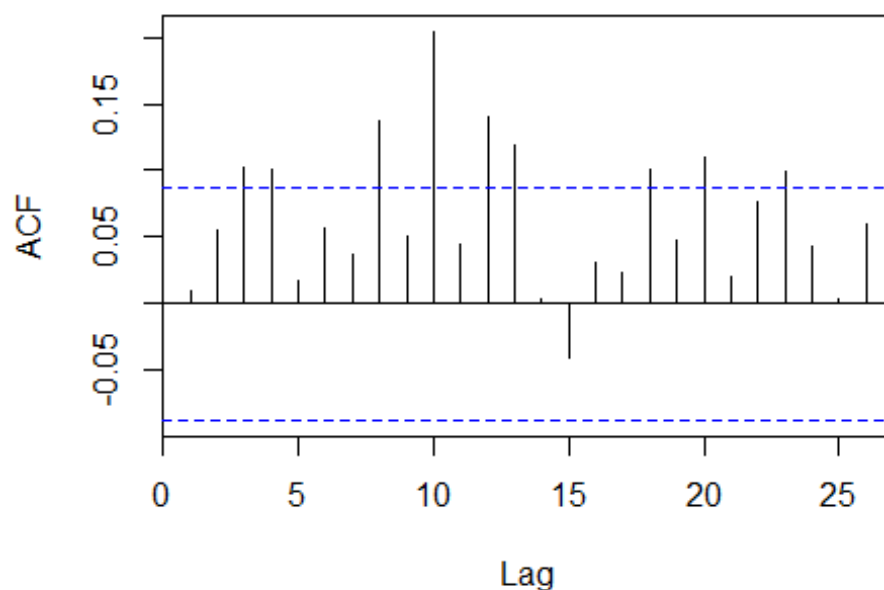


```
pacf(abs(r.cref), main= "Sample PACF of the Absolute Daily CREF Returns")
```



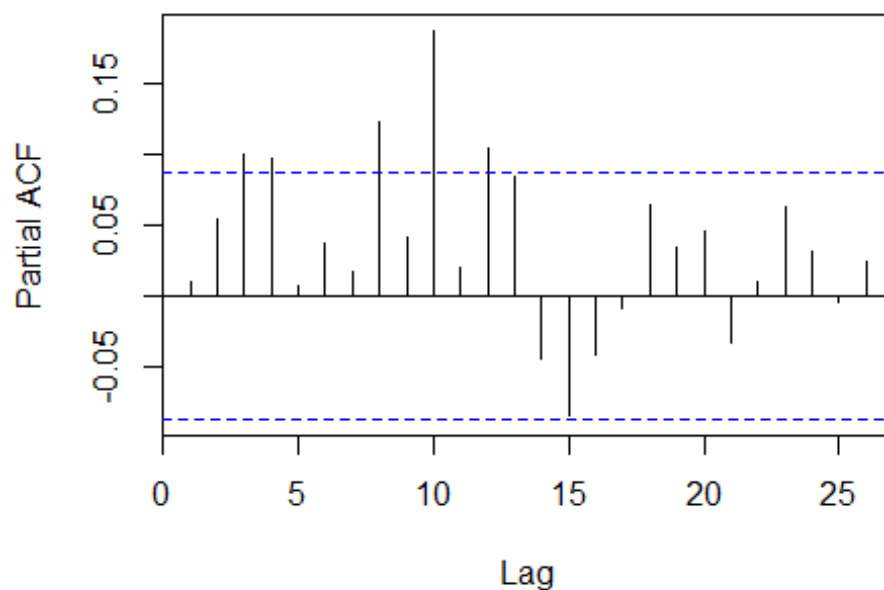
```
acf(r.cref^2, main= "Sample ACF of the Squared Daily CREF Returns")
```

### Sample ACF of the Squared Daily CREF Returns



```
pacf(r.cref^2, main= "Sample PACF of the Squared Daily CREF Returns")
```

### Sample PACF of the Squared Daily CREF Returns



However, the volatility clustering observed in the CREF return data gives us a hint that they may not be independently and identically distributed—otherwise the variance would be constant over time. This is the first occasion in our study of time series models where we need to distinguish between series values being uncorrelated and series values being independent. If series values are truly independent, then nonlinear instantaneous transformations such as taking logarithms, absolute values, or squaring preserves independence. However, the same is not true of correlation, as correlation is only a measure of linear dependence. Higher-order serial dependence structure in data can be explored by studying the autocorrelation structure of the absolute returns (of lesser sampling variability with less mathematical tractability) or that of the squared returns (of greater sampling variability but with more manageability in terms of statistical theory). If the returns are independently and identically distributed, then so are the absolute returns (as are the squared returns), and hence they will be white noise as well. Hence, if the absolute or squared returns admit some significant autocorrelations, then these autocorrelations furnish some evidence against the hypothesis that the returns are independently and identically distributed. Indeed, the sample ACF and PACF of the absolute returns and those of the squared returns in Exhibits 12.5 through 12.8 display some significant autocorrelations and hence provide some evidence that the daily CREF returns are not independently and identically distributed.

## 11.2 GARCH Models

The  $ARCH(q)$  model, proposed by Engle (1982), by specifying that

$$\sigma_{t|t-1}^2 = \omega + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \cdots + \alpha_q r_{t-q}^2$$

Here,  $q$  is referred to as the  $ARCH$  order. Another approach, proposed by Bollerslev (1986) and Taylor (1986), introduces  $p$  lags of the conditional variance in the model, where  $p$  is referred to as the  $GARCH$  order. The combined model is called the generalized autoregressive conditional heteroscedasticity,  $GARCH(p, q)$ , model.

$$\sigma_{t|t-1}^2 = \omega + \beta_1 \sigma_{t-1|t-2}^2 + \cdots + \beta_p \sigma_{t-p|t-p-1}^2 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \cdots + \alpha_q r_{t-q}^2$$

In terms of the backshift  $B$  notation, the model can be expressed as

$$(1 - \beta_1 B - \cdots - \beta_p B^p) \sigma_{t|t-1}^2 = \omega + (\alpha_1 B + \cdots + \alpha_q B^q) r_t^2$$

Now simulate data from a  $GARCH(1,1)$  model with standard normal innovations and parameter values  $\omega = 0.02$ ,  $\alpha = 0.05$ , and  $\beta = 0.9$ . with model

$$\sigma_{t|t-1}^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1|t-2}^2$$

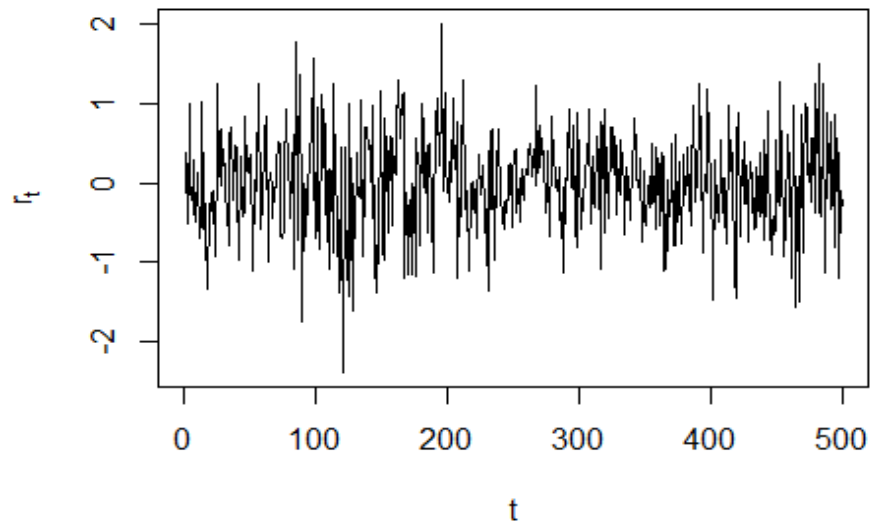
**library(TSA)**

**set.seed(1234567)**

**garch11.sim<- garch.sim(alpha= c(0.02,0.05), beta= 0.9, n= 500)**

**plot(garch11.sim, type='l', ylab= expression(r[t]), xlab= 't',  
mai= "Simulated GARCH(1,1) Process")**

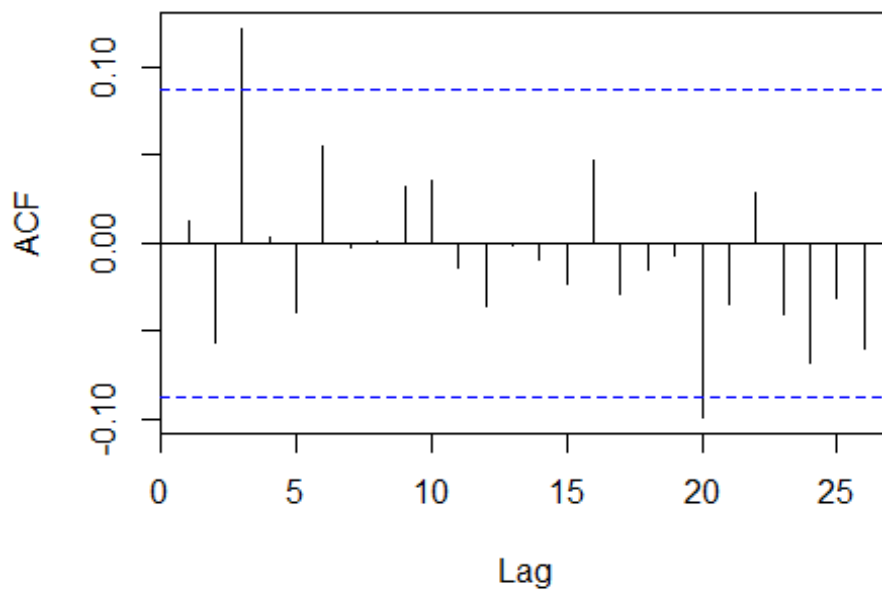
### Simulated GARCH(1,1) Process



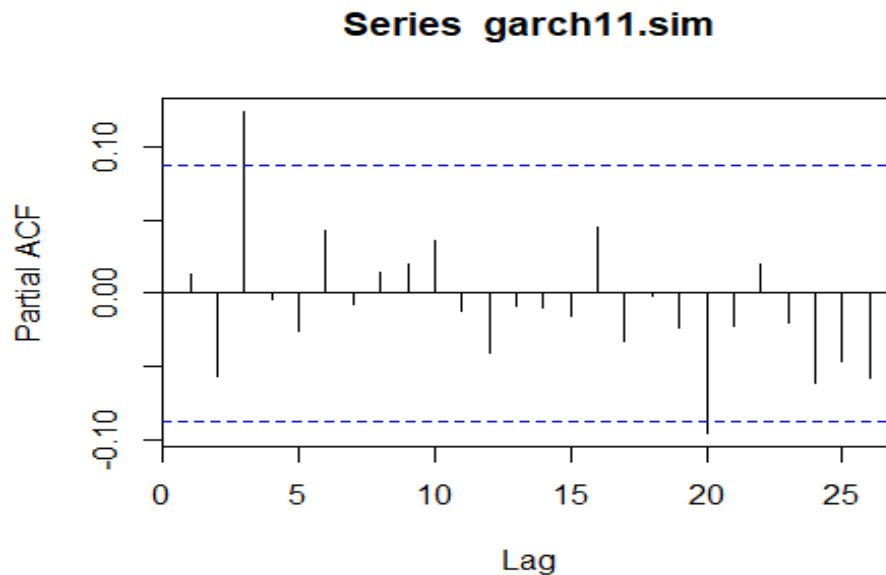
Volatility clustering is evident in the plot, as large(small) fluctuations are usually succeeded by large(small) fluctuations.

```
acf(garch11.sim)
```

### Series garch11.sim



```
pacf(garch11.sim)
```

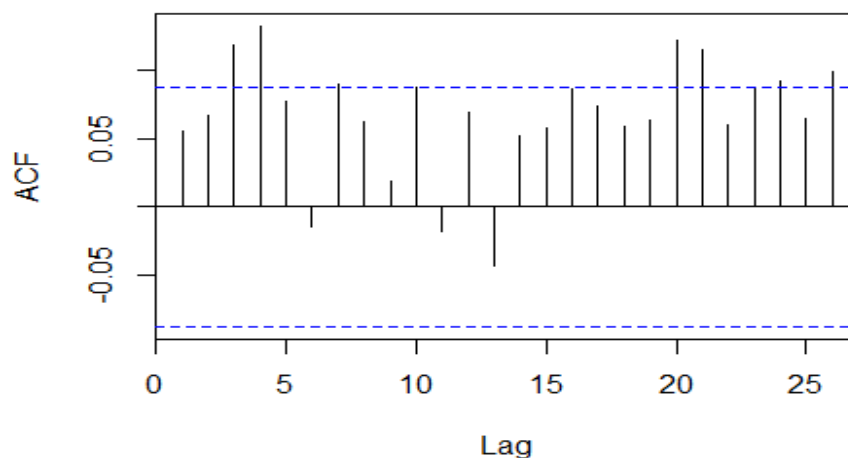


Except for *lags* 3 and *lag* 20, which are mildly significant, the sample ACF and PACF of the simulated data, do not show significant correlations. Hence, the simulated process seems to be basically serially uncorrelated as it is.

Now we shall plot sample ACF and PACF of the absolute values and the squares of the simulated data.

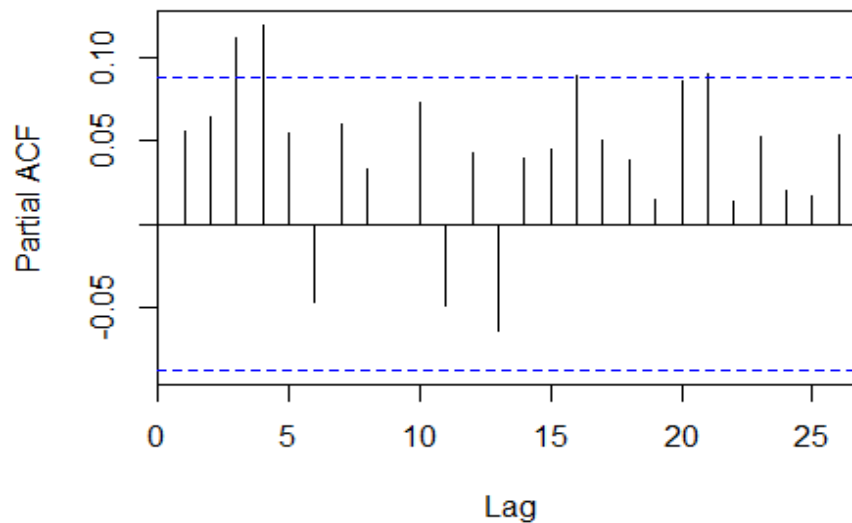
```
acf(abs(garch11.sim),  
main= "Sample ACF of the Absolute Values of GARCH(1,1) Process")
```

**Sample ACF of the Absolute Values of GARCH(1,1) Process**



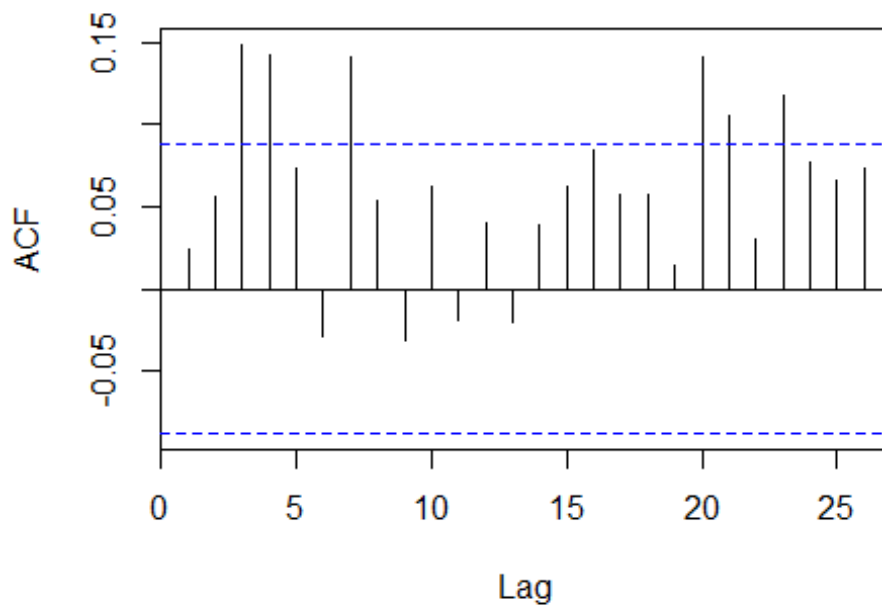
```
pacf(abs(garch11.sim),  
main= "Sample PACF of the Absolute Values of GARCH(1,1) Process")
```

### Sample PACF of the Absolute Values of GARCH(1,1) Process



```
acf(garch11.sim^2,  
main= "Sample ACF of the Squared Values of GARCH(1,1) Process")
```

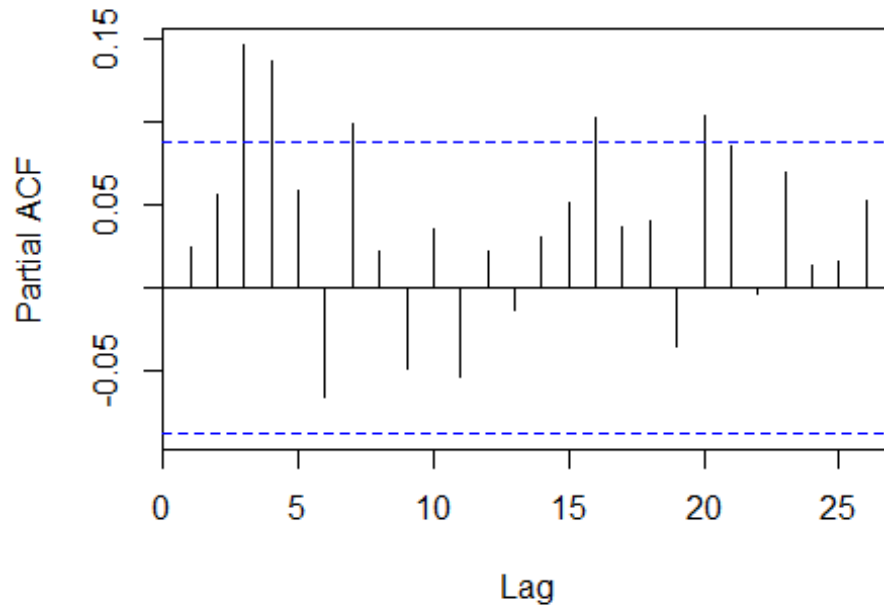
### Sample ACF of the Squared Values of GARCH(1,1) Process





```
pacf(garch11.sim^2,
main= "Sample PACF of the Squared Values of GARCH(1,1) Process")
```

### Sample PACF of the Squared Values of GARCH(1,1) Process



These plots indicate the existence of significant autocorrelation patterns in the absolute and squared data and indicate that the simulated process is in fact serially dependent. Interestingly, the *lag* 1 auto-correlations are not significant in any of these last four plots.

## 11.3 Maximum Likelihood Estimation

### 11.3.1 GARCH(1, 1)

For  $GARCH(1,1)$  given the parameters  $\omega$ ,  $\alpha$  and  $\beta$  the conditional variances can be computed by the formula

$$\sigma_{t|t-1}^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1|t-2}^2$$

for  $t \geq 2$  with likelihood

$$L(\omega, \alpha, \beta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \{ \log(\sigma_{t-1|t-2}^2) + r_t^2 / \sigma_{t|t-1}^2 \}$$

We shall maximize this function with respect to  $\omega$ ,  $\alpha$ ,  $\beta$  using the function  $GARCH$  — `garch()` of **tseries** package

```
library(tseries)
```

```
g1<- garch(garch11.sim, order= c(2,2))
```

```
##
```

```
## ***** ESTIMATION WITH ANALYTICAL GRADIENT *****
```

```
##
```

```
##
```

```
##      I      INITIAL X(I)      D(I)
```

```
##
```

```
##      1      3.294364e-01      1.000e+00
```

```
##      2      5.000000e-02      1.000e+00
```

```
##      3      5.000000e-02      1.000e+00
```

```
##      4      5.000000e-02      1.000e+00
```

```
##      5      5.000000e-02      1.000e+00
```

```
##
```

```
##      IT      NF      F      RELDF      PRELDF      RELDX      STPPAR      D*STEP
```

```
NPRELDF
```

```
##      0      1      2.661e+01
```

```
##      1      4      2.641e+01      7.54e-03      1.69e-02      3.7e-02      5.1e+02      3.0e-02
```

```
4.31e+00
```

```
##      2      5      2.609e+01      1.21e-02      1.87e-02      4.3e-02      2.0e+00      3.0e-02
```

```
1.50e+01
```

```
##      3      6      2.573e+01      1.40e-02      1.46e-02      3.3e-02      2.0e+00      3.0e-02
```

```
6.06e+00
```

```
##      4      9      2.279e+01      1.14e-01      1.06e-01      3.4e-01      1.8e+00      2.4e-01
```

```
6.46e+00
```

```
##      5      11      2.225e+01      2.39e-02      2.80e-02      6.9e-02      2.0e+00      4.7e-02
```

```
2.82e+02
```

```
##      6      12      2.157e+01      3.03e-02      3.31e-02      5.4e-02      2.0e+00      4.7e-02
```

```
3.50e+02
```

```
##      7      13      2.111e+01      2.17e-02      2.66e-02      5.4e-02      2.0e+00      4.7e-02
```

```
1.76e+02
```

```
##      8      15      2.105e+01      2.88e-03      1.23e-02      1.9e-02      2.0e+00      1.7e-02
```

```
3.37e+01
```

```
##      9      16      2.080e+01      1.15e-02      1.42e-02      2.0e-02      2.0e+00      1.7e-02
```

```
5.77e+00
```

```
##      10      17      2.056e+01      1.15e-02      2.28e-02      4.4e-02      2.0e+00      3.5e-02
```

```
4.10e+00
```

```
##      11      20      2.052e+01      1.88e-03      3.51e-03      2.3e-03      6.1e+00      2.7e-03
```

```
1.16e+01
```

```
##      12      21      2.049e+01      1.79e-03      1.79e-03      2.5e-03      2.2e+00      2.7e-03
```

```
1.86e+01
```

```
##      13      22      2.042e+01      3.35e-03      3.41e-03      6.1e-03      2.0e+00      5.4e-03
```

```
1.91e+01
```

```
##      14      25      2.002e+01      1.98e-02      2.64e-02      5.0e-02      2.0e+00      5.4e-02
```

```
1.85e+01
```

```
##      15      29      1.997e+01      2.07e-03      3.54e-03      1.8e-03      4.2e+00      2.0e-03
```

```
8.73e-01
```

```
##      16      30      1.994e+01      1.59e-03      1.61e-03      2.1e-03      2.7e+00      2.0e-03
```

5.52e-01								
##	17	31	1.989e+01	2.73e-03	2.79e-03	3.8e-03	2.0e+00	4.0e-03
5.91e-01								
##	18	34	1.958e+01	1.53e-02	2.11e-02	3.6e-02	1.9e+00	3.7e-02
5.99e-01								
##	19	35	1.930e+01	1.45e-02	2.02e-02	3.3e-02	1.9e+00	3.7e-02
6.31e-01								
##	20	36	1.907e+01	1.21e-02	3.55e-02	2.9e-02	1.8e+00	3.7e-02
3.27e-01								
##	21	37	1.865e+01	2.17e-02	3.50e-02	2.5e-02	1.9e+00	3.7e-02
4.75e-01								
##	22	39	1.857e+01	4.21e-03	5.14e-03	2.5e-03	2.0e+00	3.7e-03
1.23e-01								
##	23	42	1.847e+01	5.60e-03	5.64e-03	7.2e-03	1.6e+00	9.6e-03
1.21e-01								
##	24	44	1.826e+01	1.13e-02	1.17e-02	1.7e-02	1.3e+00	1.9e-02
1.66e-01								
##	25	47	1.825e+01	3.99e-04	4.53e-04	2.1e-04	2.9e+00	3.8e-04
2.11e-01								
##	26	51	1.808e+01	9.32e-03	1.40e-02	1.3e-02	3.3e+00	2.3e-02
2.30e-01								
##	27	53	1.803e+01	2.97e-03	5.38e-03	3.1e-03	2.0e+00	4.6e-03
1.77e-01								
##	28	54	1.798e+01	2.94e-03	3.01e-03	4.0e-03	2.0e+00	4.6e-03
1.72e-01								
##	29	56	1.797e+01	6.34e-04	6.77e-04	7.7e-04	2.2e+00	9.2e-04
1.66e-01								
##	30	57	1.795e+01	1.10e-03	1.12e-03	1.6e-03	2.0e+00	1.8e-03
1.62e-01								
##	31	59	1.792e+01	1.49e-03	1.59e-03	3.0e-03	2.0e+00	3.7e-03
1.58e-01								
##	32	61	1.791e+01	5.48e-04	5.63e-04	6.0e-04	2.0e+00	7.3e-04
1.30e-01								
##	33	62	1.790e+01	6.99e-04	7.34e-04	1.1e-03	2.0e+00	1.5e-03
1.22e-01								
##	34	64	1.789e+01	2.65e-04	2.50e-04	2.3e-04	2.3e+00	2.9e-04
1.18e-01								
##	35	66	1.789e+01	5.43e-05	5.50e-05	5.2e-05	1.5e+01	5.9e-05
1.17e-01								
##	36	68	1.789e+01	1.03e-04	1.03e-04	1.0e-04	4.4e+00	1.2e-04
1.19e-01								
##	37	70	1.789e+01	2.08e-05	2.03e-05	2.0e-05	4.4e+02	2.3e-05
1.19e-01								
##	38	72	1.789e+01	4.14e-05	4.05e-05	4.1e-05	4.7e+01	4.7e-05
1.28e-01								
##	39	74	1.789e+01	8.27e-06	8.07e-06	8.1e-06	9.6e+02	9.4e-06
1.28e-01								
##	40	76	1.789e+01	1.65e-06	1.61e-06	1.6e-06	4.8e+03	1.9e-06
1.30e-01								
##	41	78	1.789e+01	3.30e-07	3.23e-07	3.3e-07	2.4e+04	3.8e-07

1.30e-01								
##	42	80	1.789e+01	6.61e-07	6.45e-07	6.5e-07	3.0e+03	7.5e-07
1.30e-01								
##	43	82	1.789e+01	1.32e-07	1.29e-07	1.3e-07	6.0e+04	1.5e-07
1.30e-01								
##	44	84	1.789e+01	2.64e-07	2.58e-07	2.6e-07	7.6e+03	3.0e-07
1.30e-01								
##	45	86	1.789e+01	5.29e-07	5.16e-07	5.2e-07	3.8e+03	6.0e-07
1.30e-01								
##	46	88	1.789e+01	1.06e-07	1.03e-07	1.0e-07	7.6e+04	1.2e-07
1.30e-01								
##	47	90	1.789e+01	2.11e-08	2.06e-08	2.1e-08	3.8e+05	2.4e-08
1.30e-01								
##	48	92	1.789e+01	4.23e-08	4.13e-08	4.2e-08	4.7e+04	4.8e-08
1.30e-01								
##	49	94	1.789e+01	8.46e-09	8.26e-09	8.3e-09	9.4e+05	9.6e-09
1.30e-01								
##	50	96	1.789e+01	1.69e-08	1.65e-08	1.7e-08	1.2e+05	1.9e-08
1.30e-01								
##	51	98	1.789e+01	3.38e-09	3.30e-09	3.3e-09	2.4e+06	3.8e-09
1.30e-01								
##	52	100	1.789e+01	6.77e-10	6.61e-10	6.7e-10	1.2e+07	7.7e-10
1.30e-01								
##	53	102	1.789e+01	1.35e-09	1.32e-09	1.3e-09	1.5e+06	1.5e-09
1.30e-01								
##	54	104	1.789e+01	2.71e-10	2.64e-10	2.7e-10	3.0e+07	3.1e-10
1.30e-01								
##	55	106	1.789e+01	5.41e-10	5.28e-10	5.3e-10	3.7e+06	6.2e-10
1.30e-01								
##	56	108	1.789e+01	1.08e-10	1.06e-10	1.1e-10	7.4e+07	1.2e-10
1.30e-01								
##	57	111	1.789e+01	8.66e-10	8.46e-10	8.5e-10	2.3e+06	9.9e-10
1.30e-01								
##	58	114	1.789e+01	1.73e-11	1.69e-11	1.7e-11	4.6e+08	2.0e-11
1.30e-01								
##	59	116	1.789e+01	3.46e-11	3.38e-11	3.4e-11	5.8e+07	3.9e-11
1.30e-01								
##	60	119	1.789e+01	6.92e-13	6.76e-13	6.8e-13	1.2e+10	7.9e-13
1.30e-01								
##	61	121	1.789e+01	1.38e-13	1.35e-13	1.4e-13	5.8e+10	1.6e-13
1.30e-01								
##	62	124	1.789e+01	1.11e-12	1.08e-12	1.1e-12	1.8e+09	1.3e-12
1.30e-01								
##	63	127	1.789e+01	2.30e-14	2.16e-14	2.2e-14	3.6e+11	2.5e-14
1.30e-01								
##	64	129	1.789e+01	4.39e-14	4.33e-14	4.4e-14	4.5e+10	5.0e-14
1.30e-01								
##	65	131	1.789e+01	9.33e-15	8.66e-15	8.7e-15	9.0e+11	1.0e-14
1.30e-01								
##	66	133	1.789e+01	1.71e-14	1.73e-14	1.7e-14	1.1e+11	2.0e-14

```

1.26e-01
##      67  135  1.789e+01  3.69e-14  3.46e-14  3.5e-14  5.6e+10  4.0e-14
1.31e-01
##      68  137  1.789e+01 -5.59e+08  6.93e-15  7.0e-15  1.1e+12  8.1e-15
1.31e-01
##
## ***** FALSE CONVERGENCE *****
##
## FUNCTION      1.788786e+01  RELDX      6.980e-15
## FUNC. EVALS    137      GRAD. EVALS    68
## PRELDF        6.927e-15  NPRELDF    1.307e-01
##
##      I      FINAL X(I)      D(I)      G(I)
##
##      1      1.835300e-02      1.000e+00      -3.499e+00
##      2      9.975762e-16      1.000e+00      5.063e+00
##      3      1.135831e-01      1.000e+00      1.355e+01
##      4      3.368745e-01      1.000e+00      -2.752e+00
##      5      5.099761e-01      1.000e+00      -2.577e+00

summary(g1)

##
## Call:
## garch(x = garch11.sim, order = c(2, 2))
##
## Model:
## GARCH(2,2)
##
## Residuals:
##      Min      1Q    Median      3Q      Max
## -3.346827 -0.631881  0.008473  0.736112  3.202344
##
## Coefficient(s):
##      Estimate Std. Error t value Pr(>|t|)
## a0 1.835e-02  1.515e-02  1.211  0.2257
## a1 9.976e-16  4.723e-02  0.000  1.0000
## a2 1.136e-01  5.855e-02  1.940  0.0524 .
## b1 3.369e-01  3.696e-01  0.911  0.3621
## b2 5.100e-01  3.575e-01  1.426  0.1538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
## Jarque Bera Test
##
## data: Residuals
## X-squared = 0.41859, df = 2, p-value = 0.8112
##
##

```

```
## Box-Ljung test
##
## data: Squared.Residuals
## X-squared = 0.005298, df = 1, p-value = 0.942
```

None of the above coefficients are significant. This is obvious because data was simulated from **GARCH(1, 1)** Model.

Now fit **GARCH(1,1)** Model

```
g2<- garch(garch11.sim, order= c(1,1))

##
## ***** ESTIMATION WITH ANALYTICAL GRADIENT *****
##
##      I      INITIAL X(I)      D(I)
##
##      1      3.706160e-01      1.000e+00
##      2      5.000000e-02      1.000e+00
##      3      5.000000e-02      1.000e+00
##
##      IT      NF      F      RELDF      PRELDF      RELDX      STPPAR      D*STEP
NPRELDF
##      0      1      2.773e+01
##      1      5      2.773e+01      2.16e-04      4.35e-04      4.2e-03      7.0e+02      4.2e-03
1.52e-01
##      2      6      2.772e+01      2.97e-04      3.53e-04      5.2e-03      2.0e+00      4.2e-03
2.85e-01
##      3      7      2.770e+01      5.63e-04      6.26e-04      8.8e-03      2.0e+00      8.3e-03
1.97e-01
##      4      11      2.652e+01      4.27e-02      1.90e-02      3.8e-01      9.4e-01      2.7e-01
1.56e-01
##      5      12      2.385e+01      1.01e-01      1.53e-01      4.4e-01      2.0e+00      5.3e-01
4.52e+01
##      6      15      2.356e+01      1.24e-02      1.10e-01      6.4e-03      1.3e+01      1.0e-02
3.35e+00
##      7      16      2.290e+01      2.81e-02      2.74e-02      6.5e-03      2.0e+00      1.0e-02
9.28e+00
##      8      18      2.060e+01      1.00e-01      1.90e-01      4.7e-02      6.2e+00      7.5e-02
1.05e+00
##      9      20      1.952e+01      5.22e-02      4.29e-02      3.5e-02      2.0e+00      5.9e-02
5.80e+00
##      10     22      1.809e+01      7.32e-02      8.53e-02      6.2e-02      2.0e+00      1.2e-01
3.74e+01
##      11     26      1.785e+01      1.32e-02      2.55e-02      7.9e-04      3.4e+00      1.7e-03
1.17e+01
##      12     30      1.765e+01      1.13e-02      1.15e-02      6.9e-03      2.0e+00      1.4e-02
9.52e+00
##      13     31      1.746e+01      1.08e-02      2.06e-02      1.4e-02      2.0e+00      2.8e-02
1.10e+00
```

```

##      14      34  1.744e+01  9.37e-04  2.19e-03  1.4e-04  1.1e+01  2.8e-04
1.37e-01
##      15      35  1.744e+01  1.00e-04  1.03e-04  1.4e-04  2.0e+00  2.8e-04
3.18e-02
##      16      36  1.744e+01  7.34e-05  9.97e-05  2.7e-04  2.0e+00  5.6e-04
2.24e-02
##      17      37  1.743e+01  2.06e-04  3.04e-04  5.1e-04  2.0e+00  1.1e-03
1.40e-02
##      18      38  1.743e+01  1.16e-04  1.59e-04  4.8e-04  1.9e+00  1.1e-03
3.80e-03
##      19      39  1.743e+01  4.11e-05  7.69e-05  5.1e-04  1.3e+00  1.1e-03
2.31e-04
##      20      40  1.743e+01  8.47e-06  1.29e-05  1.3e-04  0.0e+00  3.0e-04
1.29e-05
##      21      41  1.743e+01  2.90e-06  1.57e-05  9.8e-05  0.0e+00  2.4e-04
1.57e-05
##      22      42  1.743e+01  7.81e-07  7.47e-07  9.0e-06  0.0e+00  2.1e-05
7.47e-07
##      23      43  1.743e+01  2.69e-07  4.49e-08  1.6e-06  0.0e+00  4.1e-06
4.49e-08
##      24      44  1.743e+01 -2.34e-08  5.16e-11  1.9e-07  0.0e+00  4.6e-07
5.16e-11
##
## ***** RELATIVE FUNCTION CONVERGENCE *****
##
## FUNCTION      1.743183e+01  RELDX      1.897e-07
## FUNC. EVALS      44      GRAD. EVALS      24
## PRELDF      5.158e-11  NPRELDF      5.158e-11
##
##      I      FINAL X(I)      D(I)      G(I)
##
##      1      7.574525e-03      1.000e+00      9.582e-03
##      2      4.718358e-02      1.000e+00      -5.087e-04
##      3      9.353769e-01      1.000e+00      9.737e-04

summary(g2)

##
## Call:
## garch(x = garch11.sim, order = c(1, 1))
##
## Model:
## GARCH(1,1)
##
## Residuals:
##      Min      1Q    Median      3Q      Max
## -3.307030 -0.637977  0.009156  0.741977  3.019441
##
## Coefficient(s):
##      Estimate Std. Error t value Pr(>|t|)

```

```
## a0  0.007575    0.007590    0.998    0.3183
## a1  0.047184    0.022308    2.115    0.0344 *
## b1  0.935377    0.035839    26.100    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##  Jarque Bera Test
##
## data:  Residuals
## X-squared = 0.82911, df = 2, p-value = 0.6606
##
##  Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.53659, df = 1, p-value = 0.4638
```

Hence,  $\alpha$  and  $\beta$  are significant because data has been generated from **GARCH(1,1)**.

## 12 Sequential Solution for Linear Gaussian State-Space Model

This chapter describe the sequential estimation method for the case where the observations are obtained sequentially in the linear gaussian state space model. This solution is called **Kalman Filter**.

### 12.1 Kalman Filter

The optional sequential estimation method for the linear gaussian state space model is referred to as Kalman Filter named after Kalman. The word optional have means to minimize the MSE between point estimates and true values of the data to be estimated.

**Model :**

$$X_t = G_t x_{t-1} + w_t \quad \therefore \{w_t \sim N(0, W_t)\}$$

$$y_t = F_t x_t + v_t \quad \therefore \{v_t \sim N(0, V_t)\}$$

where

- $G_t$  is  $p \times p$  state transition matrix.
- $F_t$  is  $1 \times p$  observation matrix.
- $W_t$  is  $p \times p$  covariance matrix of the state noise.
- $V_t$  is variance of the observation noise.

Regarding the prior distribution, we also assume that  $x_0 \sim N(m_0, C_0)$  where  $m_0$  is the  $p$  dimensional mean vector and  $C_0$  is the  $p \times p$  covariance matrix.

Note that listing all the parameter in this linear gaussian state yield  $\theta = \{G_t, F_t, W_t, V_t, m_0, C_0\}$  model.



### 12.1.1.1 Kalman Filtering

We define filtering distribution  $N(m_t, C_t)$  for the mean vector  $m_t$  and variance covariance matrix  $C_t$ .

One step ahead predictive distribution  $N(a_t, R_t)$  for mean  $a_t$  and variance covariance matrix  $R_t$ .

One step ahead predictive likelihood  $N(f_t, Q_t)$  for mean  $f_t$  and variance covariance matrix  $Q_t$ .

### 12.1.2 Algorithm — Kalman Filtering

0. Filtering distribution at time point  $t - 1$ :  $m_{t-1}, C_{t-1}$
1. Update procedure at time point  $t$ 
  - One step ahead predictive distribution
    - (Mean)  $a_t \leftarrow G_t m_{t-1}$
    - (Covariance)  $R_t \leftarrow G_t C_{t-1} G_t^\top + W_t$
  - One step ahead predictive likelihood :
    - (Mean)  $f_t \leftarrow F_t a_t$
    - (Covariance)  $Q_t \leftarrow F_t R_t F_t^\top + V_t$
  - Kalman Gain
    - $K_t \leftarrow R_t F_t^\top Q_t^{-1}$
  - State Update
    - (Mean)  $m_t \leftarrow a_t + K_t [y_t - f_t]$
    - (Covariance)  $C_t = [I - K_t F_t] R_t$
2. Filtering distribution at time point  $t$ :  $m_t, C_t$ .

### 12.1.3 Algorithm — Kalman Prediction

It shows predictive distribution  $\sim N(a_t(k), R_t(k))$

0. The  $k - 1$  step for obtaining the  $k$  step ahead predictive distribution at time point  $t + (k - 1)$ :  $a_t(k - 1), R_t(k - 1)$
1. Updating procedure at time point  $t + k$ 
  - the  $k$  steps ahead predictive distribution
    - (Mean)  $a_t(k) \leftarrow G_{t+k} a_t(k - 1)$
    - (Covariance)  $R_t(k) \leftarrow G_{t+k} R_t(k - 1) G_{t+k}^\top + W_{t+k}$
2. The  $k$  step ahead predictive distribution at time point  $t + k$ :  $a_t(k), R_t(k)$

### 12.1.4 Algorithm — Kalman Smoothing

Smoothing distribution  $\sim N(s_t, S_t)$  for the mean vector  $s_t$  and covariance matrix  $S_t$

0. Smoothing distribution at time  $t + 1: s_{t+1}, S_{t+1}$
1. Update procedure at time  $t$ 
  - Smoothing Gain  $A_t \leftarrow C_t G_{t+1}^\top R_{t+1}^{-1}$
  - State Update
    - (Mean)  $s_t \leftarrow m_t + A_t [s_{t+1} - a_{t+1}]$
    - (Covariance)  $S_t \leftarrow C_t + A_t [S_{t+1} - R_{t+1}] A_t^\top$
2. Smoothing distribution at time  $t: s_t, S_t$

## 12.2 Example: Local-Level Model Case

The local-level model is defined as follows :

$$x_t = x_{t-1} + w_t, \quad w_t \sim N(0, W)$$

$$y_t = x_t + v_t, \quad v_t \sim N(0, V)$$

The above model corresponds to the linear Gaussian state-space model that has setting  $p = 1, G_t = [1], F_t = [1], W_t = W, V_t = V$

The local-level model is similar to the  $AR(1)$  model. The difference lies in the  $AR(1)$  coefficient  $\phi$  for the state  $x_{t-1}$ , the coefficient of the local-level model is one. Such a coefficient of the non-stationary time series for the state known as a **random walk**. For this reason, the local-level model is also called a *random walk plus noise model*.

### Define local-level model

```
# Preprocessing
library(dlm)

# Setting of state space model
mod<- dlmModPoly(order= 1)

# Confirmation of model contents
str(mod)

## List of 10
## $ m0 : num 0
## $ C0 : num [1, 1] 1e+07
## $ FF : num [1, 1] 1
## $ V : num [1, 1] 1
## $ GG : num [1, 1] 1
## $ W : num [1, 1] 1
## $ JFF: NULL
## $ JV : NULL
## $ JGG: NULL
```

```
## $ JW : NULL
## - attr(*, "class")= chr "dlm"
```

### 12.2.1 Specification of Parameter Values

In this example, the parameters  $W$  and  $V$  are unknown, we must specify them. We attempt to use the maximum likelihood method for *Nile* data as it has sufficient data.

```
# Specification of parameter values in Local-Level model

# User-defined function to define and build a model
build_dlm<- function(par){
  mod$W[1, 1]= exp(par[1])
  mod$V[1, 1]= exp(par[2])

  return(mod)
}

# Maximum Likelihood Estimation of parameters, confirming the search results
with three different initial values
lapply(list(c(0,0), c(1, 10), c(20, 3)),
  function(parms){
    dlmMLE(y= Nile, parm= parms, build= build_dlm)
  })

## [[1]]
## [[1]]$par
## [1] 7.291951 9.622437
##
## [[1]]$value
## [1] 549.6918
##
## [[1]]$counts
## function gradient
##      34      34
##
## [[1]]$convergence
## [1] 0
##
## [[1]]$message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
##
##
## [[2]]
## [[2]]$par
## [1] 7.291992 9.622452
##
## [[2]]$value
## [1] 549.6918
##
```

```

## [[2]]$counts
## function gradient
##      22      22
##
## [[2]]$convergence
## [1] 0
##
## [[2]]$message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
##
##
## [[3]]
## [[3]]$par
## [1] 7.291948 9.622437
##
## [[3]]$value
## [1] 549.6918
##
## [[3]]$counts
## function gradient
##      40      40
##
## [[3]]$convergence
## [1] 0
##
## [[3]]$message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"

# MLE of parameters, including the Hessian matrix in return value
fit_dlm<- dlmMLE(y= Nile, parm= c(0, 0), build= build_dlm, hessian= T)

# Find the a(symptotic) standard error in the MLE from the Hessian matrix
using the delta method
exp(fit_dlm$par)*sqrt(diag(solve(fit_dlm$hessian)))

## [1] 1280.170 3145.999

# Set the maximum likelihood estimates of parameters in the model
mod<- build_dlm(fit_dlm$par)

# Confirmation of the results
mod

## $FF
##      [,1]
## [1,]      1
##
## $V
##      [,1]
## [1,] 15099.8
##

```

```
## $GG
##      [,1]
## [1,]    1
##
## $W
##      [,1]
## [1,] 1468.432
##
## $m0
## [1] 0
##
## $C0
##      [,1]
## [1,] 1e+07
```

$$H = - \begin{bmatrix} \frac{\partial^2 l(W, V)}{\partial W \partial W} & \frac{\partial^2 l(W, V)}{\partial W \partial V} \\ \frac{\partial^2 l(W, V)}{\partial V \partial W} & \frac{\partial^2 l(W, V)}{\partial V \partial V} \end{bmatrix}$$

We use the delta method to obtain the standard error of the parameters after exponential transformation.

The basic idea of the delta method method lies in Taylor expansions for the statistics of interest, these are simply treated at order. Thus,

$$\begin{aligned} Var[exp(\theta)] &\approx Var \left[ exp(\hat{\theta}) + \frac{exp'(\hat{\theta})}{1!} (\theta - \hat{\theta}) \right] \\ &= exp^2(\hat{\theta}) Var(\theta - \hat{\theta}) \\ Var[exp(\theta)] &= exp^2(\hat{\theta}) Var(\theta) \end{aligned}$$

Hence

$$exp(\text{maximum likelihood estimates of a parameter}) \times \sqrt{diag(solve(H))}$$

We have obtain SE of  $W$  and  $V$ , repectively as 1280.170 and 3145.999.

### 12.2.2 Execution of Filtering, Prediction and Smoothing

We examine the results through the plots with the *mean*  $\pm$  1.96 *se*.

#### Filtering

The operation of Kalman filtering is based on Kalman Filtering Algorithm. We perform the processing using library **d1m**

```
# Filtering Process
d1mFiltered_obj<- d1mFilter(y= Nile, mod= mod)
```

```

# Confirmation of the results
str(dlmFiltered_obj, max.level= 1)

## List of 9
## $ y : Time-Series [1:100] from 1871 to 1970: 1120 1160 963 1210 1160
1160 813 1230 1370 1140 ...
## $ mod:List of 10
## ..- attr(*, "class")= chr "dlm"
## $ m : Time-Series [1:101] from 1870 to 1970: 0 1118 1140 1072 1117 ...
## $ U.C:List of 101
## $ D.C: num [1:101, 1] 3162.3 122.8 88.9 76 70 ...
## $ a : Time-Series [1:100] from 1871 to 1970: 0 1118 1140 1072 1117 ...
## $ U.R:List of 100
## $ D.R: num [1:100, 1] 3162.5 128.6 96.8 85.1 79.8 ...
## $ f : Time-Series [1:100] from 1871 to 1970: 0 1118 1140 1072 1117 ...
## - attr(*, "class")= chr "dlmFiltered"

# Find the mean and SD of the filtering distribution
m<- dropFirst(dlmFiltered_obj$m)

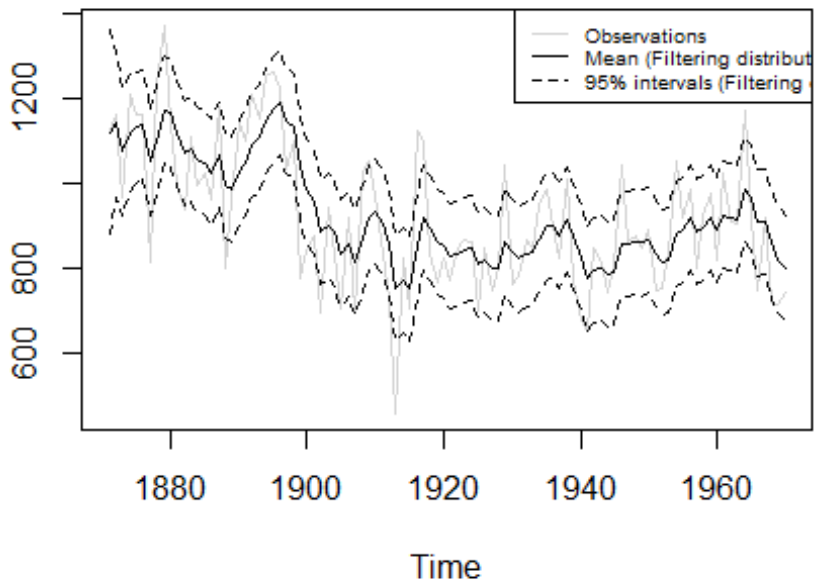
m_sdev<- sqrt(
  dropFirst(as.numeric(
    dlmSvd2var(dlmFiltered_obj$U.C, dlmFiltered_obj$D.C)
  ))
)

# Find 2.5% and 97.5% values for 95% interval of the filtering distribution
m_quant<- list(m+qnorm(0.025, sd= m_sdev),
  m+qnorm(0.975, sd= m_sdev))

# Plot results
ts.plot(cbind(Nile, m, do.call("cbind", m_quant)),
  col= c("lightgray", "black", "black", "black"),
  lty= c("solid", "solid", "dashed", "dashed"))

# Legend
legend(legend= c("Observations", "Mean (Filtering distribution)", "95%
intervals (Filtering distribution)"),
  lty= c("solid", "solid", "dashed"),
  col= c("lightgray", "black", "black"),
  x= "topright", text.width= 32, cex= 0.6)

```



## Prediction

The operation of Kalman prediction is based on Kalman Prediction Algorithm

```
# Prediction processing
dlmForecasted_obj <- dlmForecast(mod = dlmFiltered_obj, nAhead = 10)

# Confirmation of the results
str(dlmForecasted_obj, max.level = 1)

## List of 4
## $ a: Time-Series [1:10, 1] from 1971 to 1980: 798 798 798 798 798 ...
## ..- attr(*, "dimnames")=List of 2
## $ R:List of 10
## $ f: Time-Series [1:10, 1] from 1971 to 1980: 798 798 798 798 798 ...
## ..- attr(*, "dimnames")=List of 2
## $ Q:List of 10

# Find the mean and standard deviation of the predictive distribution
a <- ts(data = dlmForecasted_obj$a, start = c(1971, 1))

a_sdev <- sqrt(
  as.numeric(
    dlmForecasted_obj$R
  )
)
```

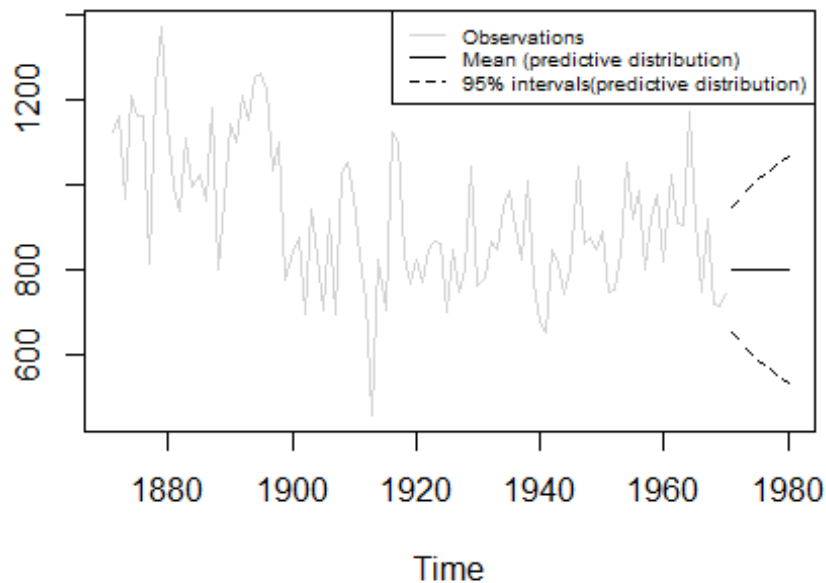
```

# Find 2.5% and 97.5% values for 95% intervals of the predictive distribution
a_quant <- list(a + qnorm(0.025, sd = a_sdev), a + qnorm(0.975, sd = a_sdev))

# Plot results
ts.plot(cbind(Nile, a, do.call("cbind", a_quant)),
        col = c("lightgray", "black", "black", "black"),
        lty = c("solid", "solid", "dashed", "dashed"))

# Legend
legend(legend = c("Observations", "Mean (predictive distribution)", "95%
intervals(predictive distribution)"),
       lty = c("solid", "solid", "dashed"),
       col = c("lightgray", "black", "black"),
       x = "topright", cex = 0.6)

```



**Smoothing** The operation of Kalman prediction is based on Kalman Smoothing Algorithm

```

# Smoothing Processing
d1mSmoothed_obj<- dlmSmooth(y= Nile, mod= mod)

# Confirmation of the results
str(d1mSmoothed_obj, max.level= 1)

```



```

## List of 3
## $ s : Time-Series [1:101] from 1870 to 1970: 1111 1111 1111 1105 1113
...
## $ U.S:List of 101
## $ D.S: num [1:101, 1] 74.1 63.5 56.9 53.1 50.9 ...

# Find the mean and SD of the predictive distribution
s<- dropFirst(dlmSmoothed_obj$s)

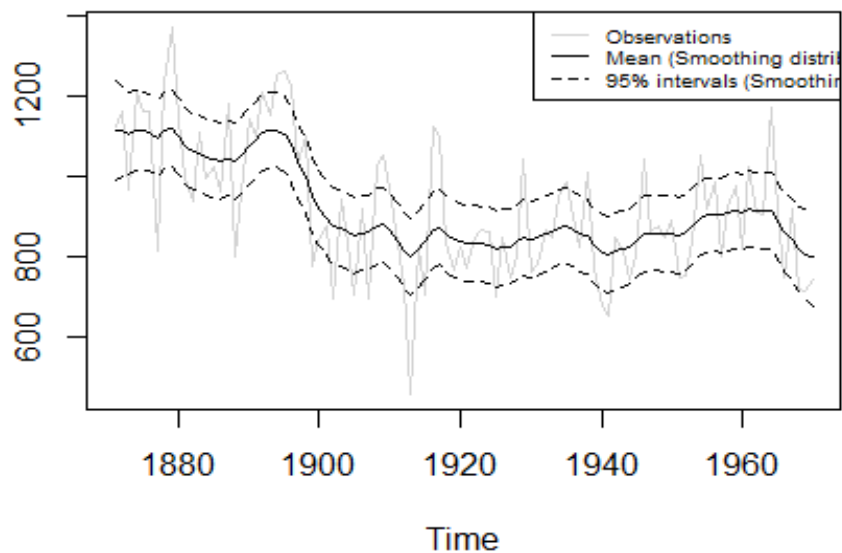
s_sdev<- sqrt(
  dropFirst(as.numeric(
    dlmSvd2var(dlmSmoothed_obj$U.S, dlmSmoothed_obj$D.S)
  ))
)

# Find 2.5% and 97.5% values for 95% interval of the smoothing distribution
s_quant<- list(s+qnorm(0.025, sd= s_sdev),
               s+qnorm(0.975, sd= s_sdev))

# Plot results
ts.plot(cbind(Nile, s, do.call("cbind", s_quant)),
        col= c("lightgray", "black", "black", "black"),
        lty= c("solid", "solid", "dashed", "dashed"))

# Legend
legend(legend= c("Observations", "Mean (Smoothing distribution)", "95%
intervals (Smoothing distribution)"),
       lty= c("solid", "solid", "dashed"),
       col= c("lightgray", "black", "black"),
       x= "topright", text.width= 32, cex= 0.6)

```



### 12.2.3 Diagnostic Checking Results

Finally, we explain diagnostic checking for the results. We use the likelihood and prediction error as diagnostic tools.

#### Likelihood

```
# Likelihood in Linear Gaussian State space model
# Calculation of "negative" log-likelihood
dlmLL(y= Nile, mod= mod)

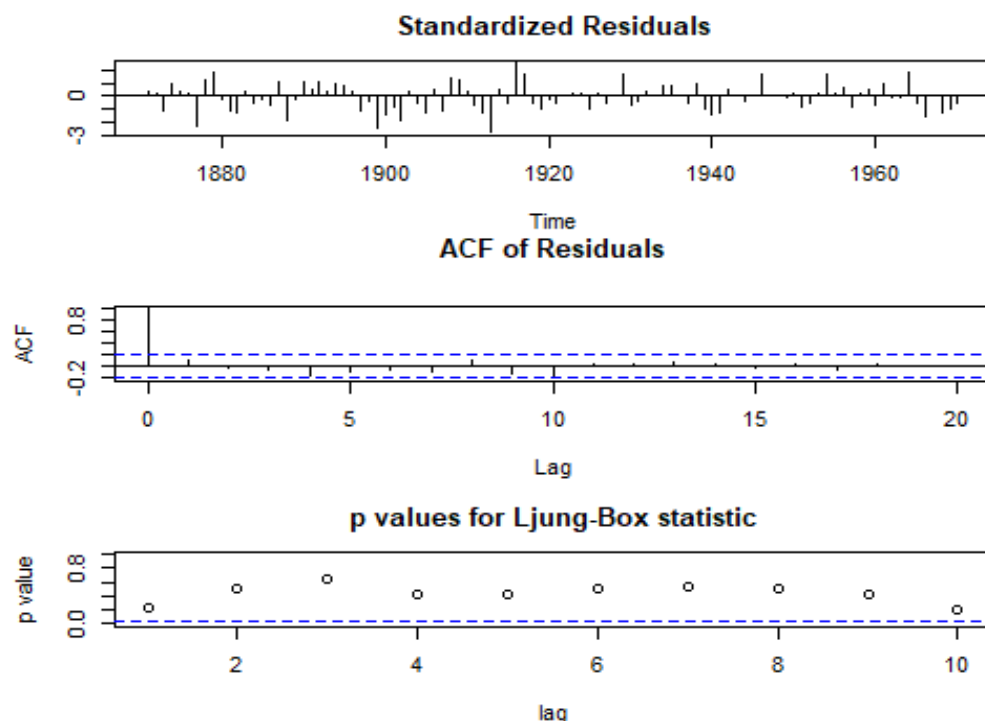
## [1] 549.6918
```

#### Innovation (Prediction Error / Residuals)

The difference between observations and the mean of the one-step-ahead predictive likelihood, i.e.,  $e_t = y_t - f_t$ , is called *Innovation*, *Prediction Error* or *Residuals*. If the model is appropriate, the innovations theoretically follow an independent and identical normal distribution. Note that  $e_t/\sqrt{Q_t}$  becomes *normalized innovations*.

```
# < Model Diagnosis Using Innovations >
# Adjust display area
oldpar<- par(no.readonly= T)
par(oma= c(0,0,0,0))
par(mar= c(4,4,3,1))

# Confirmation of Auto-correlation
tsdiag(object= dlmFiltered_obj)
```



```
par(oldpar)
```

```
# Confirm Normality
```

```
# Get standardized innovation
```

```
e<- residuals(object= dlmFiltered_obj, sd= F)
```

```
e
```

```
## Time Series:
```

```
## Start = 1871
```

```
## End = 1970
```

```
## Frequency = 1
```

```
## [1] 0.353882059 0.234347637 -1.132356160 0.920990056 0.293678981  
0.208635337 -2.253572772 1.259032474
```

```
## [9] 1.892476740 -0.217341987 -1.168902405 -1.274143563 0.285525795 -  
0.598907228 -0.257811351 -0.607000658
```

```
## [17] 1.087882179 -1.857139370 -0.253479119 1.082258192 0.514608573  
1.143623092 0.420243462 1.004779540
```

```
## [25] 0.806186369 0.312247137 -1.094920935 -0.314871097 -2.502167297 -  
1.374265341 -0.770458633 -1.818880663
```

```
## [33] 0.380717419 -0.466439568 -1.261597906 0.573222303 -1.140514421  
1.449290868 1.271364973 0.367565138
```

```
## [41] -0.692068900 -1.238867224 -2.789292186 0.519416299 -0.469282598  
2.568373713 1.743294475 -0.589405024
```

```
## [49] -0.905820813 -0.266834820 -0.564863230 0.122437752 0.222128119  
0.148887241 -1.033513562 0.266628842
```

```
## [57] -0.508263319 -0.010258329 1.692519755 -0.717199228 -0.372431518  
0.312263613 0.089544673 0.755407088
```

```
## [65] 0.832415412 0.004006761 -0.519616041 0.928982856 -0.984249216 -
```

```

1.383364390 -1.202138127  0.491394148
## [73]  0.123305797 -0.397331863  0.119827082  1.753036709  0.030865019
0.120167580 -0.093067758  0.224410162
## [81] -0.852741827 -0.590230865  0.187451524  1.614486965  0.263740703
0.667105819 -0.827838843  0.271075485
## [89]  0.561004302 -0.703558407  0.912596736 -0.125338671 -0.126711287
1.781342993 -0.491841367 -1.517108371
## [97]  0.093301043 -1.332051596 -1.004275772 -0.554991814

# Display Q-Q Plot
qqnorm(e)
qqline(e)

```

**Normal Q-Q Plot**

