Time Series Analysis

GU4261/GR5261 Statistical Methods in Finance | Columbia University

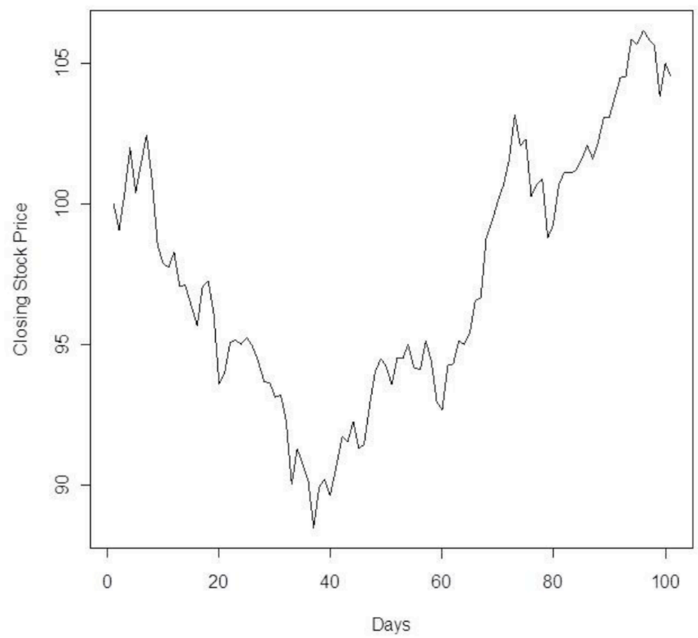
Outline

- The basics
- ARIMA(p,d,q)
- Heteroskedasticity

The basics

What is a time series?

- A time series is a collection of values X_t where $t = 1, 2, \ldots, n$
- The index t here refers to the time points (e.g. days, months, years) at which observations X_t are recorded.
- Examples of X_t include temperature and, of course, stock prices.



Distinctive feature of time series

- In classical setting (e.g. t-test, regression…), we assume X_t's are independent. On the contrary, in time series studies, we assume X_t's are dependent.
- The dependence structure in a time series is usually characterized by the auto-covariance. Essentially, for small lag k,

$$\mathrm{Cov}(X_t, X_{t+k}) \neq 0.$$

- By considering the auto-correlation structure of a time series, one can obtain a more accurate estimate of the parameter of interest.

Using correlation to find a "better" C.I of μ

- Assume that

$$X_t = \mu + a_t - \theta a_{t-1}$$

where $a_t \sim N(0,1)$. It can be shown easily that

$$E(X_t) = \mu + E(a_t) - \theta E(a_{t-1}) = \mu$$

- Observe that there exists time dependence because

$$Cov(X_t, X_{t-k}) = \begin{cases} -\theta & k = 1 \\ 1 + \theta^2 & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

Using correlation to find a "better" C.I of μ

- Recall that, the 95% Confidence interval for the mean μ is given by

$$\overline{X} \pm 2\sqrt{Var(\overline{X})}$$

- It can be shown that

$$Var(\overline{X}) = \frac{1}{n}[(1-\theta)^2 + \frac{2\theta}{n}]$$

Hence, the corresponding 95% CI is

$$\overline{X} \pm \frac{2}{\sqrt{n}}\left[(1-\theta)^2 + \frac{2\theta}{n}\right]^{1/2}$$

Using correlation to find a "better" C.I of μ

- If we ignore the dependence structure (hence $\theta = 0$), the corresponding CI is given by

$$\overline{X} \pm \frac{2}{\sqrt{n}}$$

- The ratio between the lengths of these two CIs is

$$L(\theta) = \left\{ \frac{2}{\sqrt{n}} \left[(1-\theta)^2 + \frac{2\theta}{n} \right]^{1/2} \right\} \Big/ \left\{ \frac{2}{\sqrt{n}} \right\} = \left[(1-\theta)^2 + \frac{2\theta}{n} \right]^{1/2}$$

Using correlation to find a "better" C.I of μ

- Implications:

1. If , we overestimate the standard error and thus our CI is "too wide".
2. If , we underestimate the standard error and hence the CI obtained is "too narrow".
3. As a result, we need to incorporate a correct correlation structure so as to produce a correct inference conclusion.

θ	$L(\theta), n = 50$
-1	2
-0.5	1.34
0	1
0.5	0.45
1	0.14

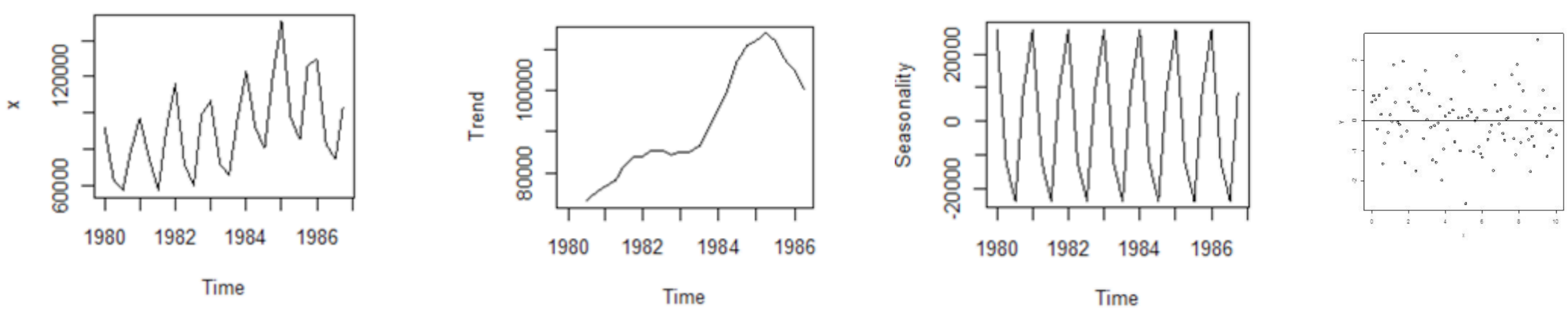
Some descriptive techniques

- Essentially, all time series can be decomposed into the following components, namely (i) trend, (ii) seasonality and (iii) idiosyncratic noise:

$$\begin{array}{cccccc} X_t = & T_t & + & S_t & + & N_t \\ & (Trend) & & (Seasonality) & & (Noise) \end{array}$$

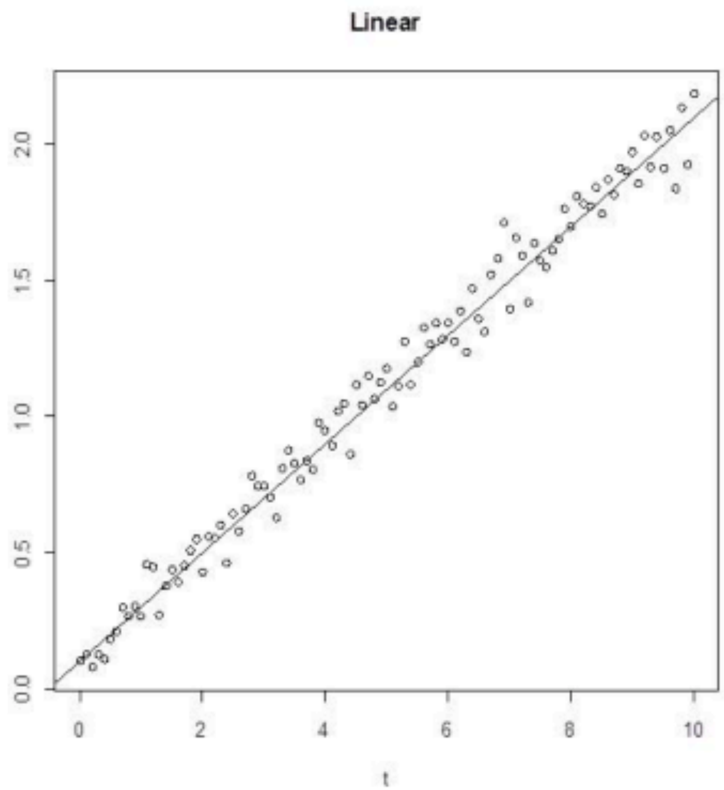
- The first two components can be considered as the macroscopic feature of the time series whereas the noise term captures momentary deviations of the series from the macroscopic structure.

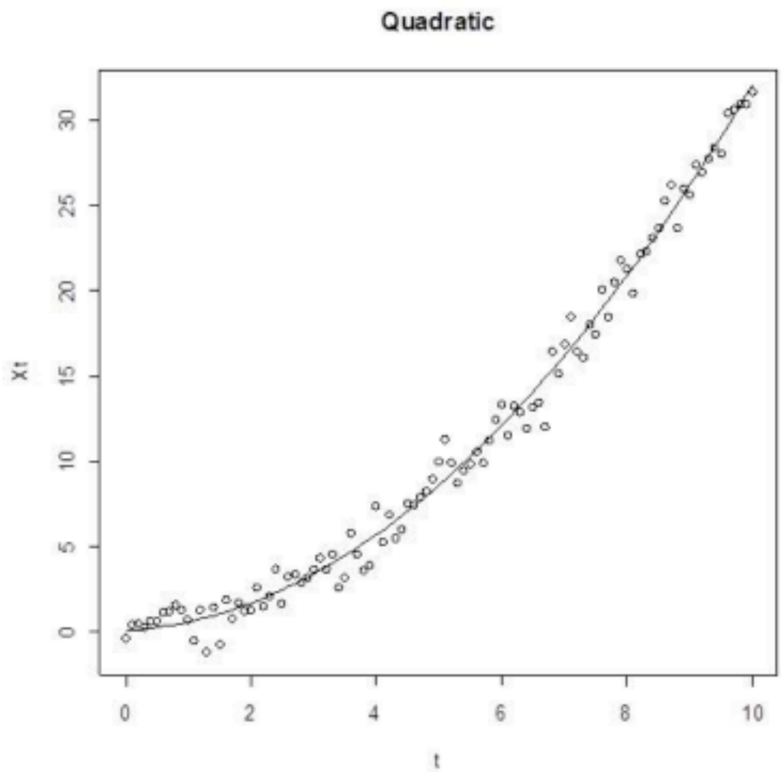
Series Trend Seasonality Noise



Trends

- Linear trend: $T_t = \alpha + \beta t$
- Quadratic trend: $T_t = \alpha + \beta t + \gamma t^2$





Estimation of Trends

- Least Squares Method

$$\text{min. } \sum (X_t - T_t)^2$$

- Filtering

$$Y_t = S_m(X_t)$$

- Differencing

$$\Delta X_t = X_t - X_{t-1}$$

Least Squares Method

- The least square method finds α and β in $T_t = \alpha + \beta t$ such that $\Sigma_t (X_t - T_t)^2$ is minimised.

- However, only simple form of T_t can be handled by this method as otherwise, it is hard to find the right form of T_t for this minimisation.

Filtering

- The idea of filtering is to smooth the series using local data so as to estimate the trend:

$$\underbrace{Y_t = S_m(X_t)}_{\text{"smoothed" series}} = \underbrace{\sum_{r=-q}^{s} a_r X_{t+r}}_{\text{Weighted average of } \{X_{t-q}, X_{t-q-1}, X_{t+s}\}}$$

- The weights $\{a_r\}$'s are usually assumed to be symmetric ($a_r = a_{-r}$) and normalised ($\Sigma_r a_r = 1$).
- Typical examples include (i) moving average filter, (ii) Spencer 15-point filter.

Filtering

- Moving Average (MA) Filter

$$Y_t = \frac{1}{2q+1}\sum_{r=-q}^{q} X_{t+r}$$

Observe that if $X_t = \alpha + \beta t$,

$$\begin{aligned} Y_t &= S_m(X_t) = \frac{1}{2q+1}\sum_{r=-q}^{q}\{[\alpha + \beta\,(t+r)] + N_{t+r}\} \\ &\approx \alpha + \beta t \end{aligned}$$

Filtering

- Spencer 15-point filter

$$(a_0, a_1, \ldots, a_7) = \frac{1}{320}(74, 67, 46, 21, 3, -5, -6, -3)$$

Property:

It does not distort a cubic trend: If $X_t = T_t + N_t$, where $T_t = a_t^3 + b_t^2 + ct + d$, then

$$\begin{aligned} S_m(X_t) &= \sum_{r=-7}^{7} a_r T_{t+r} + \sum_{r=-7}^{7} a_r N_{t+r} \\ &= at^3 + bt^2 + ct + d + \underline{\sum_{r=-7}^{7} a_r N_{t+r}} \end{aligned}$$

Since $\Sigma_r a_r N_{t+r} \approx 0$ (due to smoothing), we have

$$S_m(X_t) \approx at^3 + bt^2 + ct + d = T_t$$

Differencing

Define

$$\begin{aligned} \Delta X_t &= X_t - X_{t-1} \\ \Delta^2 X_t &= \Delta\left(\Delta X_t\right) \end{aligned}$$

and the *backshift operator (B)*

$$BX_t = X_{t-1}$$
$$B^k X_t = X_{t-k},\ k = 1, 2, \ldots$$

Hence, we can write

$$\Delta X_t = (1 - B)X_t$$
$$\Delta^k X_t = (1 - B)^k X_t,\ k = 1, 2, \ldots$$

Differencing

- If $X_t = \alpha+\beta t$, then

  $\Delta X_t = X_t - X_{t-1} = \alpha+\beta t-\{\alpha+\beta(t-1)\} = \beta$ (no trend)

- In general, if $X_t = T_t + N_t$ and $T_t = \Sigma_{j=\{1, \ldots, p\}} a_j t^j$, then

  $\Delta^p X_t = p! a_p + \Delta^p N_t$;

  in other words, if the trend is a j^{th} degree polynomial, the trend can be eliminated by differencing the series j times.

Seasonal Cycles

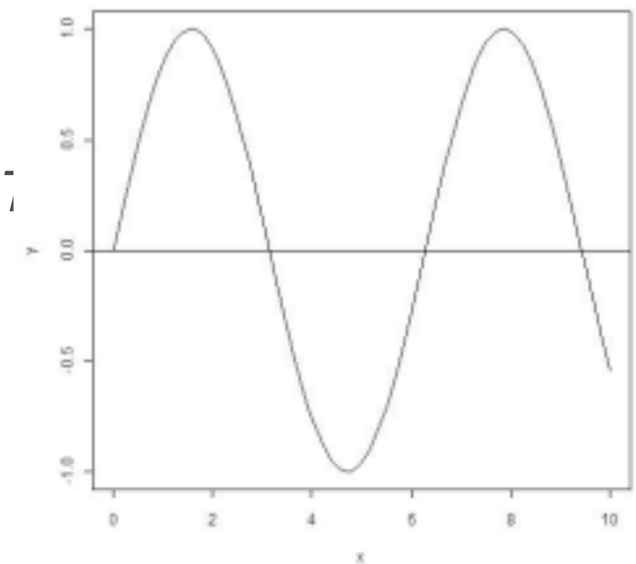
- Additive Seasonal Component

$$X_t = T_t + S_t + N_t$$

- The seasonal part is specified by a period denoted as *d*. The period has to satisfy the following requirements:

  1. $S_{t+d} = St$
  2. $S_1 + \ldots + S_d = 0$

  (common effect is explained by T



Estimating/Removing Seasonal Effects

- Moving average method

1. Estimate the trend T_t by a moving average filter

The moving average filter must be of length d. Note that the estimated trend is free of the seasonal effect because $S_1 + \ldots + S_d = 0$.

$$\hat{T}_t = \begin{cases} \frac{1}{d}\left(\frac{1}{2}X_{t-q} + X_{t-q+1} + \ldots + \frac{1}{2}X_{t+q}\right) & , \quad d = 2q \\ \frac{1}{d}\sum_{r=-q}^{q} X_{t+r} & , \quad d = 2q+1 \end{cases}$$

2. Estimate the seasonal component

3. Use any filter for the series Xt – St and obtain an improved Tt

Estimating/Removing Seasonal Effects

- Moving average method

2. Estimate the seasonal component

$$
\begin{aligned}
\widetilde{S}_j &= \sum_{t=j,d+j,\ldots,kd+j} (X_t - \widehat{T}_t) \\
S_j &= \frac{\widetilde{S}_j - \mu_s}{n_d}
\end{aligned}
$$

where

$\mu_s = \frac{1}{d} \sum_{j=1}^{d} \widetilde{S}_j$

n_d : number of complete cycle

3. Use any filter for the series $X_t - S_t$ and obtain an improved T_t. Finally, we obtain

$$X_t = T_t + S_t + N_t$$

Estimating/Removing Seasonal Effects

- Seasonal differencing

$$\begin{aligned}\Delta^d X_t &= \left(1 - B^d\right) X_t \\ &= X_t - X_{t-d}\end{aligned}$$

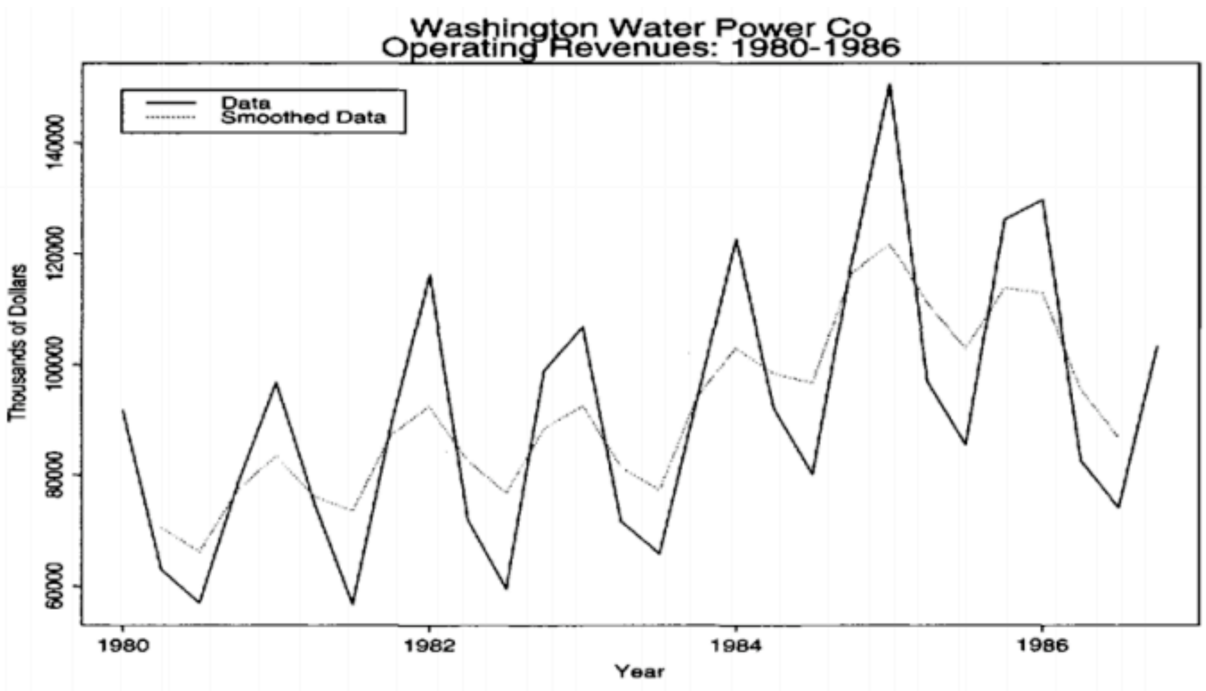
- Seasonal differencing removes the seasonal effects in the following way:

  If $X_t = S_t + N_t$ and period $=$ d, then
  $\Delta^d X_t = S_t - S_{t-d} + N_t - N_{t-d} = N_t - N_{t-d}.$

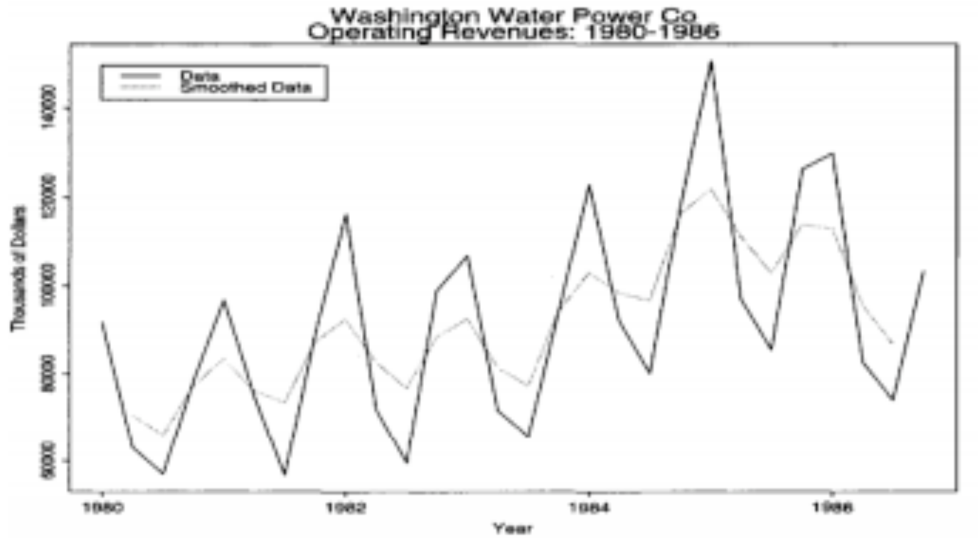
- But there is a drawback where *d* data points are lost due to *d*–differencing.

Implementation in R

- Step 0: Download the quarterly revenues of Washington power company 80-86 from http://www.sta.cuhk.edu.hk/NHCHAN/TSBook2nd/dataset.html



Implementation in R



- Observations: (i) there is a slightly increasing trend and (ii) it appears to be there is an annual cycle *(higher power consumption in winter for heating? [Domain knowledge])*

```
#R code:
x<-read.delim("C://washpower.dat",head=FALSE)
x<-ts(x, frequency = 4, start = c(1980, 1))
ts.plot(x,main="Wasington Water Power Co")
```

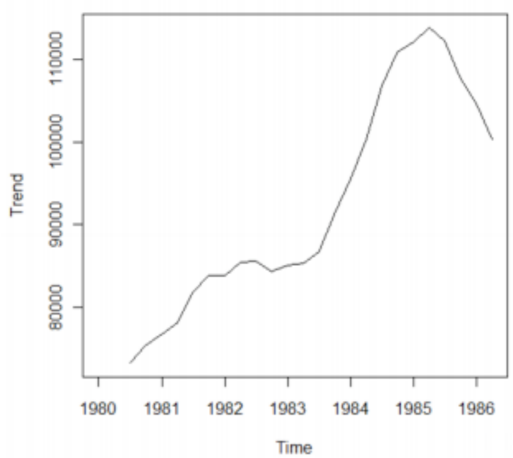
Implementation in R

- Step 1: Estimate the trend by a filter running over one complete season ($d = 4$)

$$
\begin{aligned}
\widehat{T}_3 &= \frac{\frac{1}{2}X_1 + X_2 + X_3 + X_4 + \frac{1}{2}X_5}{4} \\
\widehat{T}_4 &= \frac{\frac{1}{2}X_2 + X_3 + X_4 + X_5 + \frac{1}{2}X_6}{4} \\
\cdots & \quad \cdots \\
\widehat{T}_{26} &= \frac{\frac{1}{2}X_{24} + X_{25} + X_{26} + X_{27} + \frac{1}{2}X_{28}}{4}
\end{aligned}
$$

Implementation in R

```
#R code:

x <-
read.delim("C://washpower.dat",head=FALSE)

x <- ts(x, frequency = 4, start = c(1980, 1))

n <- length(x)

t <- rep(0,n-4)


for ( k in 1:(n-4))
{


t[k]=1/8*x[k]+1/4*x[k+1]+1/4*x[k+2]+1/4*x[k+3
]+1/8*x[k+4]

}



t <- c(NA,NA,t)

t <- ts(t,frequency=4,start=c(1980,1))

ts.plot(t,ylab="Trend")
```



Implementation in R

- Step 2: Estimate seasonal effect from the trend-removed series D_t. Estimate the seasonal part S_i as follows:

$$
\begin{aligned}
S_1 &= [(D_5 - \overline{D}) + (D_9 - \overline{D}) + \cdots + (D_{25} - \overline{D})]/6 \\
S_2 &= [(D_6 - \overline{D}) + (D_{10} - \overline{D}) + \cdots + (D_{26} - \overline{D})]/6 \\
S_3 &= [(D_3 - \overline{D}) + (D_7 - \overline{D}) + \cdots + (D_{23} - \overline{D})]/6 \\
S_4 &= [(D_4 - \overline{D}) + (D_8 - \overline{D}) + \cdots + (D_{24} - \overline{D})]/6
\end{aligned}
$$

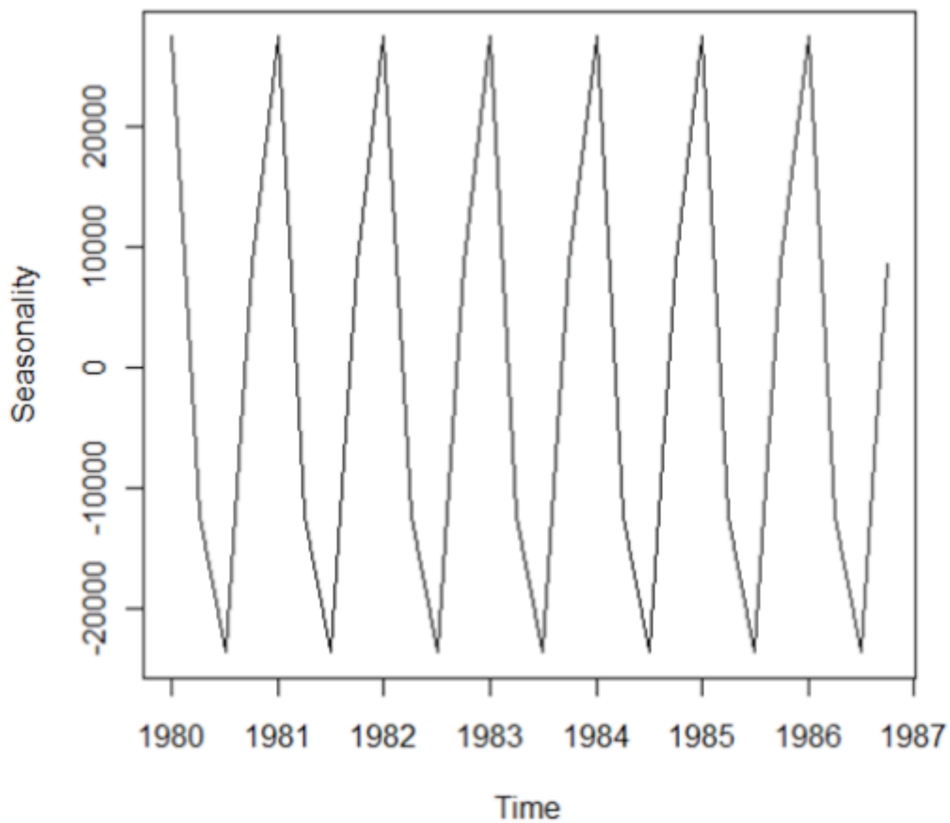
Hence $S_{i+4j} = S_i, \quad \text{all } i, j$

Implementation in R

```
#R code:
# run the previous code to load X etc …
t <- ts(t,frequency=4,start=c(1980,1))
d <- rep(0,n-4)
for (k in 1:(n-4)){d[k]= x[k+2]-t[k+2]}
m <- mean(d)
d < -c(NA,NA,d)

s<-rep(0,28)
for (k in 1:6){ s[1]= s[1]+1/6*(d[1+4*k]-m)}
for (k in 1:6){ s[2]= s[2]+1/6*(d[2+4*k]-m)}
for (k in 1:6){ s[3]= s[3]+1/6*(d[4*k-1]-m)}
for (k in 1:6){ s[4]= s[4]+1/6*(d[4*k]-m)}
for (k in 5:n){ s[k]= s[k-4]}
```

Implementation in R

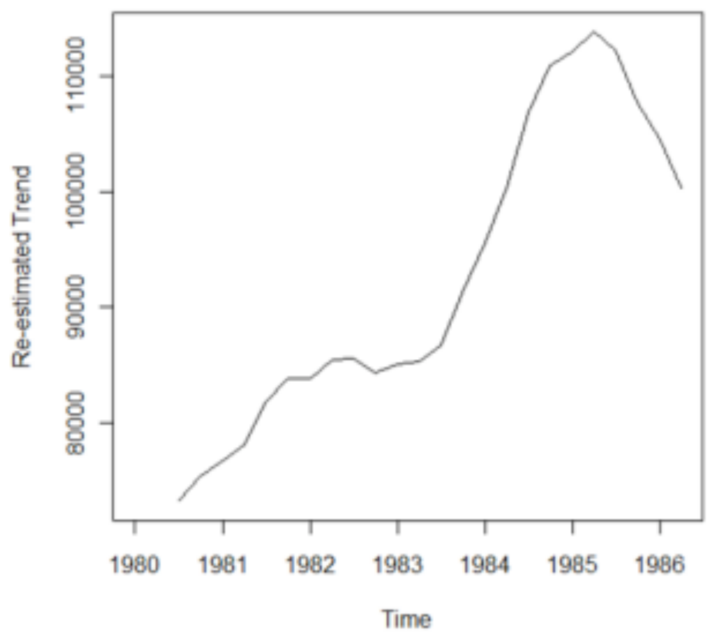
```
#R code:
s <-ts(s,frequency=4,start=c(1980,1))
ts.plot(s,ylab= "Seasonality")
```



Implementation in R

- Step 3: Re-estimate the trend from the deseasonalised data $Q_t = X_t - S_t$.

$$
\begin{aligned}
T_3 &= \frac{\frac{1}{2}Q_1 + Q_2 + Q_3 + Q_4 + \frac{1}{2}Q_5}{4} \\
T_4 &= \frac{\frac{1}{2}Q_2 + Q_3 + Q_4 + Q_5 + \frac{1}{2}Q_6}{4} \\
\cdots T_{26} &= \cdots
\end{aligned}
$$



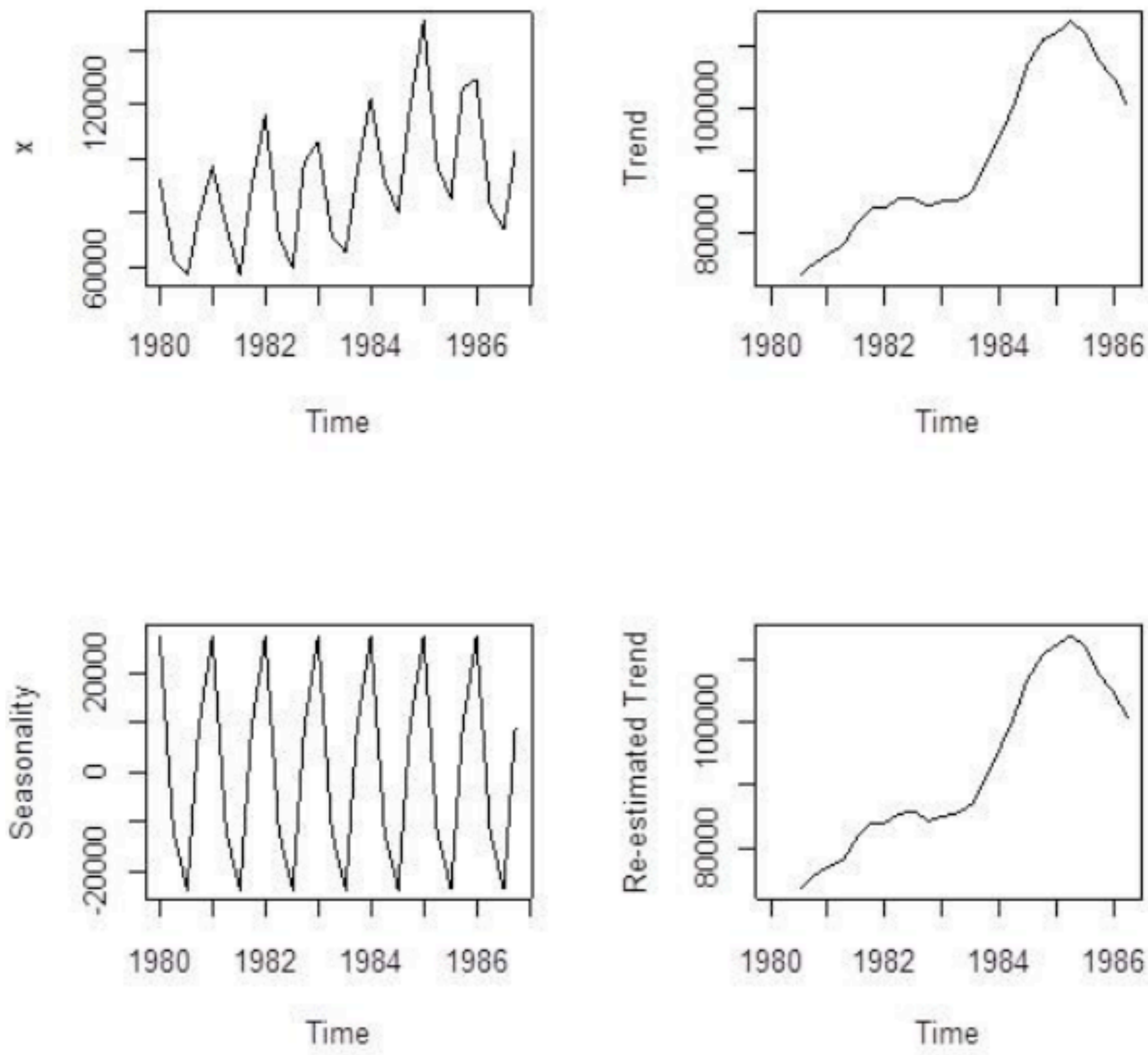
Implementation in R

```
#R code:

# run the previous code to load X etc …

s <- ts(s,frequency=4,start=c(1980,1))

q <- rep(0,n)

for (k in 1:n)

{
        q[k]=x[k]-s[k]
}

t1 <- c(NA,NA,rep(0,n-4))

for (k in 3: (n-2))
{

        t1[k]=1/8*q[k-2]+1/4*q[k-
1]+1/4*q[k]+1/4*q[k+1]+1/8*q[k+2]
}
```

Implementation in R

```
#R code:
# run the previous code to load X etc …
t1 <- ts(t1,frequency=4,start=c(1980,1))
ts.plot(t1,ylab="Re-estimated Trend")

par(mfrow=c(2,2))
ts.plot(x)
ts.plot(t,ylab="Trend")
ts.plot(s,ylab= "Seasonality")
ts.plot(t1,ylab="Re-estimated Trend")
```

Implementation in R



Implementation in R

- Are we done yet after decomposition?

  No, unfortunately. You should also:

  1. Check the residual N_t to detect any further structure.
  2. Sophisticated models fro the microscopic structure in N_t will be discussed in the immediately next two sections.

ARIMA(p,d,q)

For further reference, pleaser refer to Shumway & Stoffer (2005): Chapter 3

Introduction: AR(p), MA(q)

- A moving average MA(q) time series model:

$$Y_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \qquad Z_t \sim WN(0, \sigma^2)$$

- An autoregression AR(p) time series model:

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + Z_t.$$

- Backshift operator (B): $(1 - B)Y_t = Y_t - Y_{t\quad 1}$

- So, we can write an AR(1) model as

$$\begin{aligned} Y_t &= \phi Y_{t\quad 1} + Z_t = \phi(\phi Y_{t\quad 2} + Z_{t\quad 1}) + Z_t \\ &= \phi^2 Y_{t-2} + \phi Y_{t-1} + Y_t \\ &= \ldots \\ &= \phi^k Y_{t-k} + \sum_{j=0}^{k\quad 1} \phi^j Z_{t-j} \end{aligned}$$

AR and Causality

- What if $|\phi| > 1$?

  Although when $|\phi| > 1$, the process $\{Y_t\}$ is no longer convergent, we can rewrite the model as follows:

$$Y_t = \phi^{-1} Y_{t+1} - \phi^{-1} Z_{t+1}.$$

- By substituting the above expression into the AR(1) iteratively, we yield

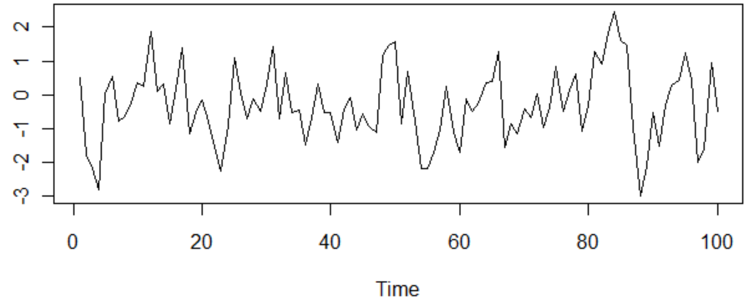
$$\begin{aligned} Y_t &= -\phi^{-1} Z_{t+1} + \phi^{-1} Y_{t+1} \\ &= -\phi\ ^{1} Z_{t+1} + \phi\ ^{1}(\phi\ ^{1} Y_{t+2} - \phi\ ^{1} Z_{t+2}) \\ &= \dots \\ &= \phi^{-1} Z_{t+1} - \phi^{-2} Z_{t+2} - \dots + \phi^{-(k+1)} Y_{t+k+1}. \end{aligned}$$

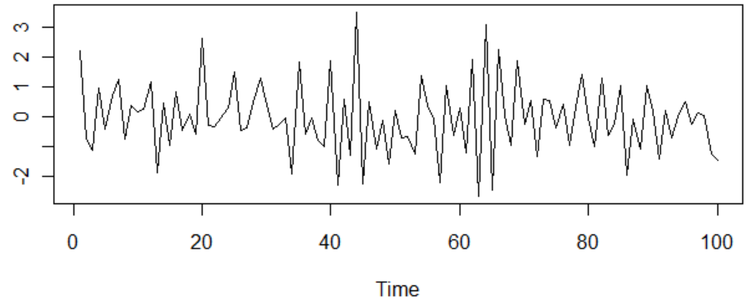
- But Y_t now depends on the *future* values ➔ Not causal!

Simulations of AR(p) model

- Use **arima.sim**

```
#R code:
>par(mfrow=c(2,1))

>plot(arima.sim(list(order=
+c(1,0,0), ar=0.5), n=100))

>plot(arima.sim(list(order=
+c(1,0,0), ar=-0.5), n=100))
```





MA, Invertibility and Simulations

- If $|\theta| < 1$, ...

$$
\begin{aligned}
Z_t - \theta Z_{t-1} &= Y_t \\
(1 - B\theta)Z_t &= Y_t \\
Z_t &= (1 - B\theta)^{-1} Y_t \\
&= (1 + \theta B + \theta^2 B^2 + \ldots) Y_t \\
&= Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \ldots .
\end{aligned}
$$

- When we want to interpret the residuals Z_t, it is more desirable to deal with a convergent expression. In this case, the MA(1) model $\{Y_t\}$ is said to be *invertible*.

```
#R code:

plot(arima.sim(list(order=c(0,0,1), ma=0.5), n=100))
```

ARMA(p,q) = AR(p)+MA(q)

- Autoregressive Moving Average Models(p,q) is defined as follows:

$$\phi(B)Y_t = \theta(B)Z_t,$$

where

$$\begin{aligned} \phi(x) &= 1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p \\ \theta(x) &= 1 - \theta_1 x - \theta_2 x^2 - \ldots - \theta_q x^q \end{aligned}$$

- For a causal ARMA(p,q) model , where the zeros of (z) are outside the unit circle, recall that we may write

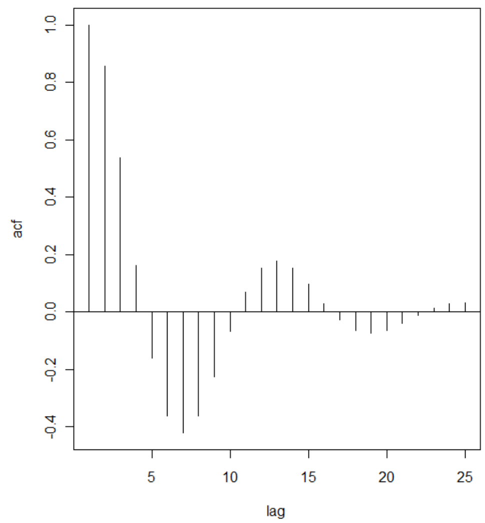
$$Y_t = \phi^{-1}(B)\theta(B)Z_t = \sum_{j=0}^{\infty} \psi_j Z_{t\ \ j}$$

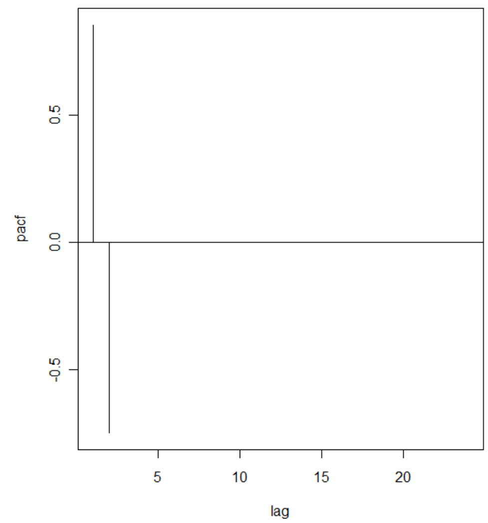
```
#R code:

ARMAtoMA(ar=.9, ma=.5, 50)        #for an array of values
plot(ARMAtoMA(ar=.9, ma=.5, 50))  #for a graph
```

Partial autocorrelation function (PACF)

- Recall that we can identify the order of an MA model by inspecting its ACF. We now introduce a similar device to identify the order of an AR, which is called *partial autocorrelation function* (PACF).

```
#R code:

Acf<-ARMAacf(ar=c(1.5, -0.75), ma=0, 24)
pacf<-ARMAacf(ar=c(1.5, -0.75), ma=0, 24, pacf=T)
par(mfrow=c(1,2))
plot(acf, type="h", xlab="lag")
abline(h=0)
plot(pacf, type="h", xlab="lag")
abline(h=0)
```

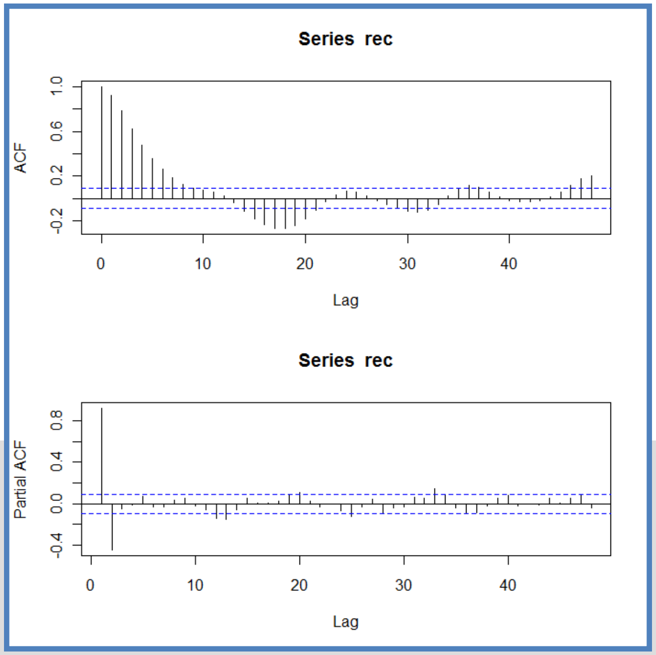




Behavior of the ACF and PACF for Causal and Invertible ARMA Models

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

```
#R code:

Rec<-scan("recruit.dat")
par(mfrow=c(2,1))
acf(rec, 48)
pacf(rec, 48)
fit<-
ar.ols(rec,aic=F,order.max=2,demean=F,intercept=T)
fit
fit$asy.se
```




```
#R printout:
…
Coefficients:
      1         2
 1.3541   -0.4632

Intercept: 6.737 (1.111)

Order selected 2  sigma^2 estimated as  89.72
```

ARIMA(p,d,q)

- We now generalise ARMA models into the ARIMA model.
- Let $W_t = (1-B)^d Y_t$ and suppose that W_t is an ARMA(p, q), $\phi(B)W_t = \theta(B)Z_t$.

  Then $$\phi(B)(1-B)^d Y_t = \theta(B)Z_t.$$

  The process $\{Y_t\}$ is said to be an ARIMA(p, d, q), autoregressive integrated moving average model. Usually, d is a small integer (≤ 3).
- By differencing appropriately, we can obtain a stationary time series model. E.g. random walk / ARIMA(0,1,0) model. In general,

$$Y_t = \beta_0 + \beta_1 t + X_t := \mu_t + X_t \Longrightarrow \nabla Y_t = Y_t Y_{t-1} = \beta_1 + \nabla X_t.$$

ARIMA(p,d,q): Fitting Strategy

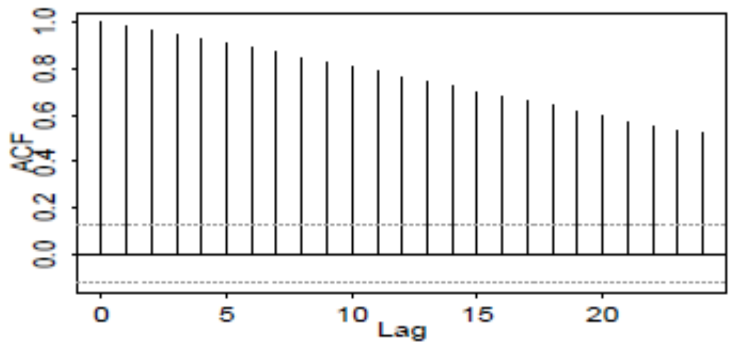
- In practice, to model possibly nonstationary time series data, we may apply the following steps:

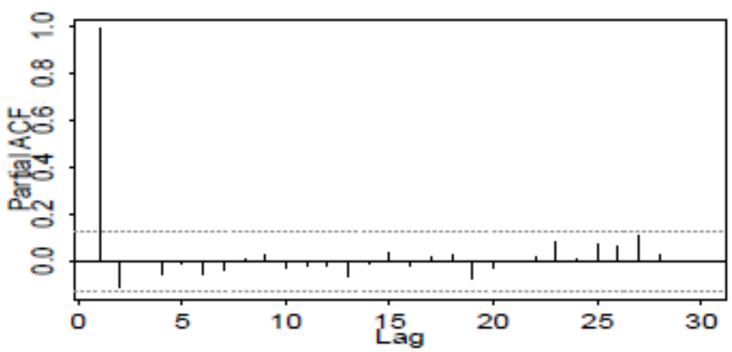
1. Look at the ACF to determine if the data are stationary.
2. If not, process the data, probably by means of differencing.
3. After differencing, fit an ARMA(p, q) model to the differenced data.

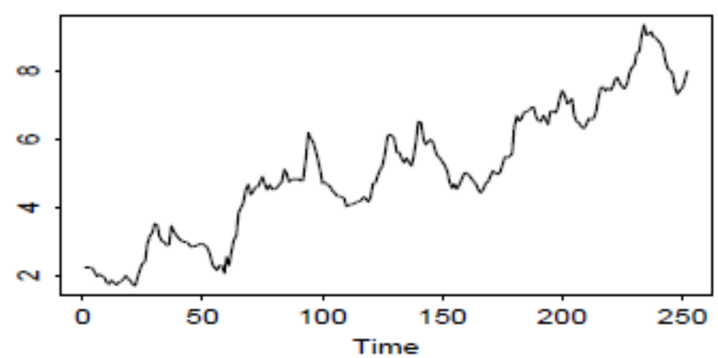
- Example: The yield of short-term government securities for 21years for a country in Europe in a period in the 1950s and 1960s are given. Several observations can be made about this series.

Example: Yields

```
#R code:

yield.ts<-scan("yields.dat")
plot.ts(yield.ts)
acf(yield.ts)
acf(yield.ts,30,type=
"partial")
```

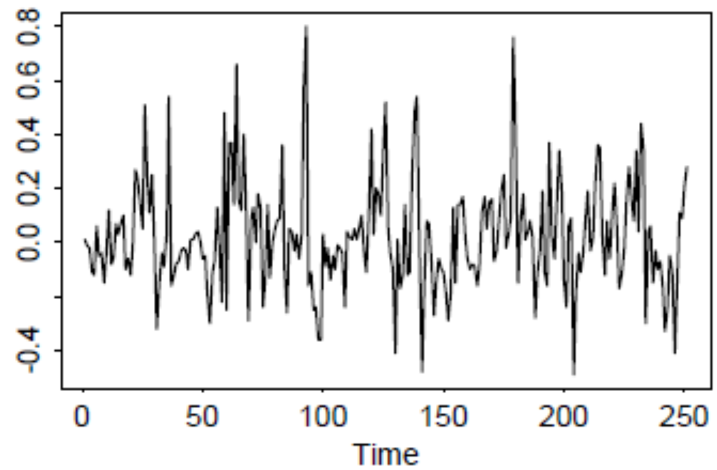


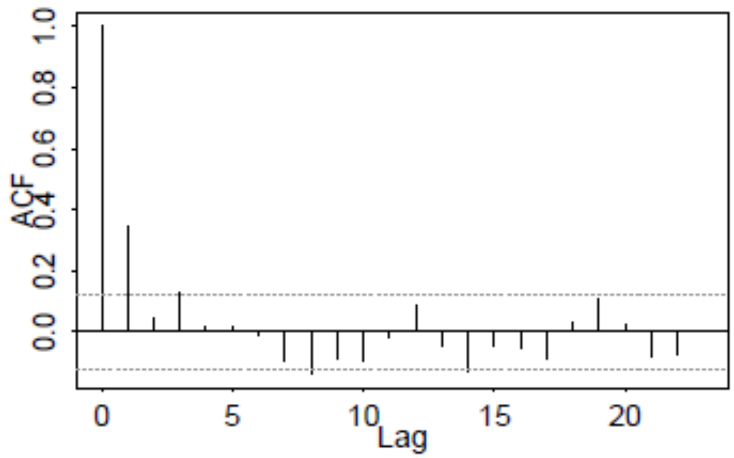


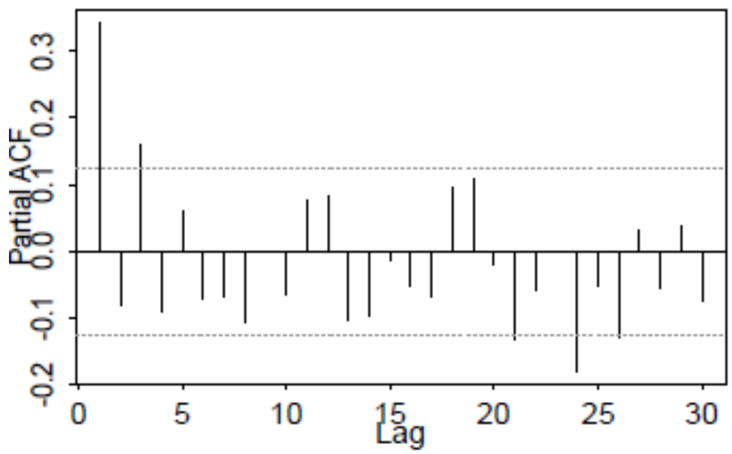


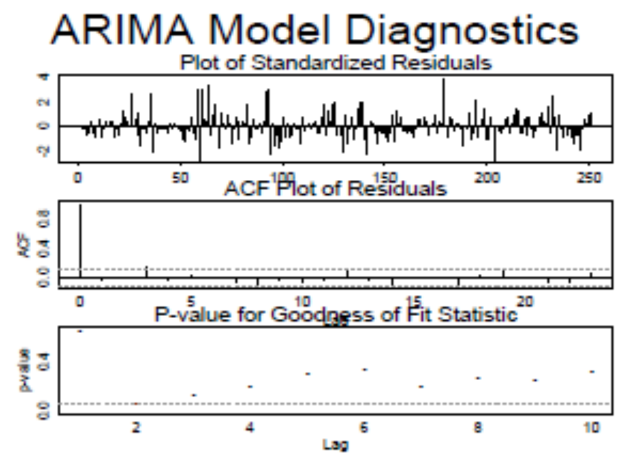
- From the acf and pacf plots, we see that the data are clearly nonstationary. This phenomenon is quite common for financial series. Therefore, we should first difference the observations using `diff(yield.ts)`.

Plots & Diagnostics









Example: Yields (cont'd)

- Since there is a lag-one correlation in the ACF that is different from zero, we may attempt an ARIMA(0,1,1) model for the original data, that is, fit an MA(1) for the differenced yield series.

```
#R code:

w<-diff(yield.ts)
w.1<-arima(w,order=c(0,0,1),method="ML")
tsdiag(w.1)
```

- This is also equivalent to fitting an ARIMA(0,1,1) to the demeaned series.
- A simple ARIMA(0,1,1) model works reasonably well. This is a typical example for financial data, as most of them fit quite well with the random walk hypothesis

Using AIC(C) as a criterion for choosing orders

- Although we can use the ACF and the PACF to determine the tentative orders p and q, it seems more desirable to have a systematic order selection criterion for a general ARMA model. One commonly used method is comparing the AIC (Akaike's information criterion) that a model gives.

```
#R code:

aic<- matrix(data = NA, nrow = 4, ncol = 4)
for (i in 1:4){
   for (j in 1:4){
     aic[i,j]<- arima(<data>,order = c(i-1,0,
                 j-1),method="ML")$aic
   }
}
```

Some Remarks regarding the use of AIC(C)

- To use these criteria, fit models with orders p and q such that AIC is minimised.
- Since we never know the true order for sure, in using AIC we should not choose a model based on the minimal value of AIC solely.
- Within a class of competing models, a minimum value of $p \pm C$ may still be legitimate when other factors, such as parsimony and whiteness of residuals, are taken into consideration.