Time Series Analysis, Case 6

Andew, Carmen, Jimmy, Naresh

23 January 2016

We are going to analyze six simulated time series, and in all of them we have to decide, in a visual and graphical way, if they are or not Covariance-Stationary (CS), White Noise (WN), Strict White Noise (SWN) or Gaussian White Noise (GWN). We will also analyze the real time series for the daily Brent Dated spot prices and its corresponding returns.

Loading data

First, we load the two datasets

```
getwd()
```

```
## [1] "/Users/nareshshah/Downloads/ie_time_series_grp_e-master"
```

```
setwd("/Users/nareshshah/downloads")
library(ggplot2)
library(gridExtra)
real = read.csv("Session6real.csv", sep = "",dec = ",")
names(real) = "price"
sim = read.csv("Session6sim.csv", sep = ";",dec = ",")
```

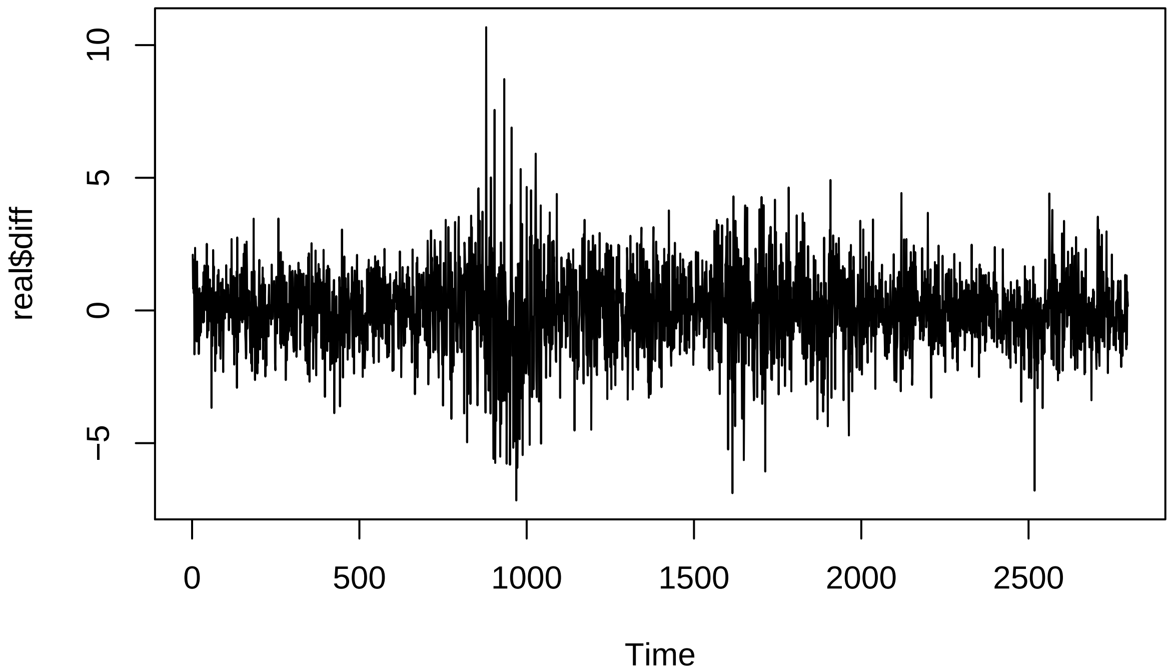
Analysis of the Real Data

We need to add a difference value that look at the difference between the price and the previous day. To do this we create a lag of t-1 and then calculate the different.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
real$lag = lag(real$price,1)
real$diff = real$price - real$lag
ts.plot(real$diff)
```



Step 1. Plot the series and compute basic statistics

First we compute the mean, standard deviation and the variance of the time series. We then plot the series and see the data is not stationary. The variance is different across the time series so can assume it is not stationary (is this right guys?).

```r
summary(real$price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   33.66   60.74   77.88   82.89  108.30  144.20
```

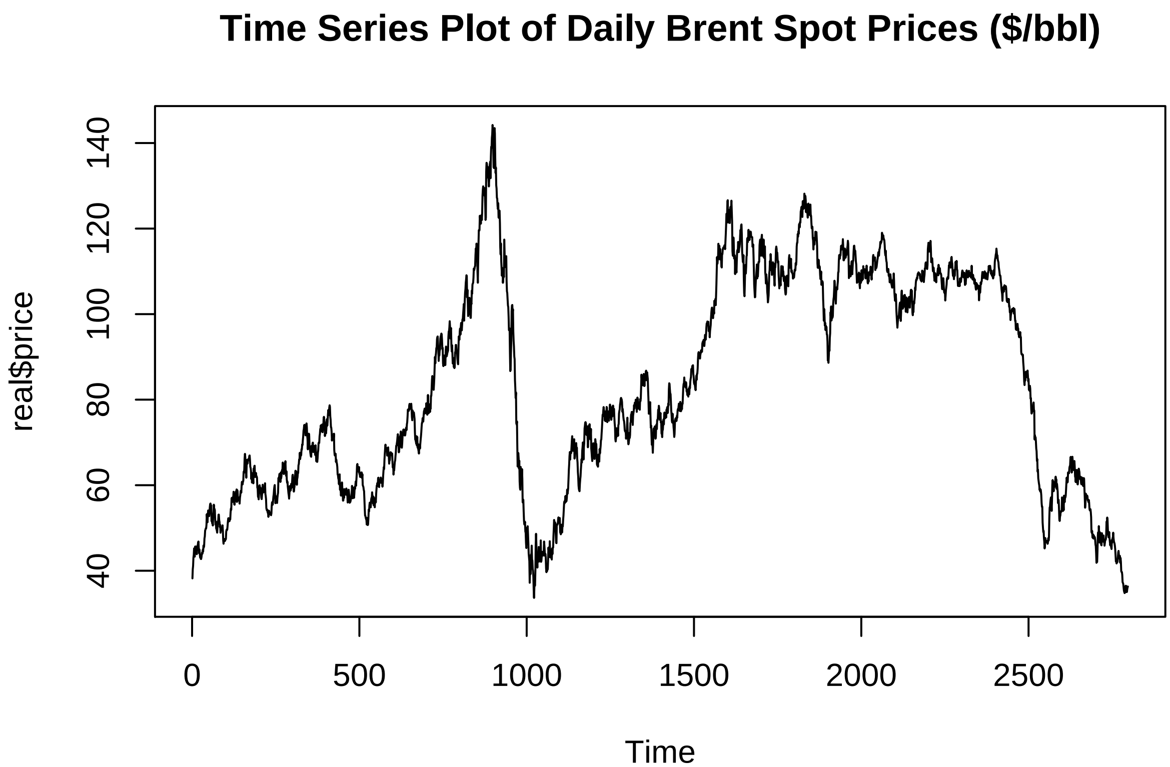
```r
sd(real$price)
```

```
## [1] 25.30475
```

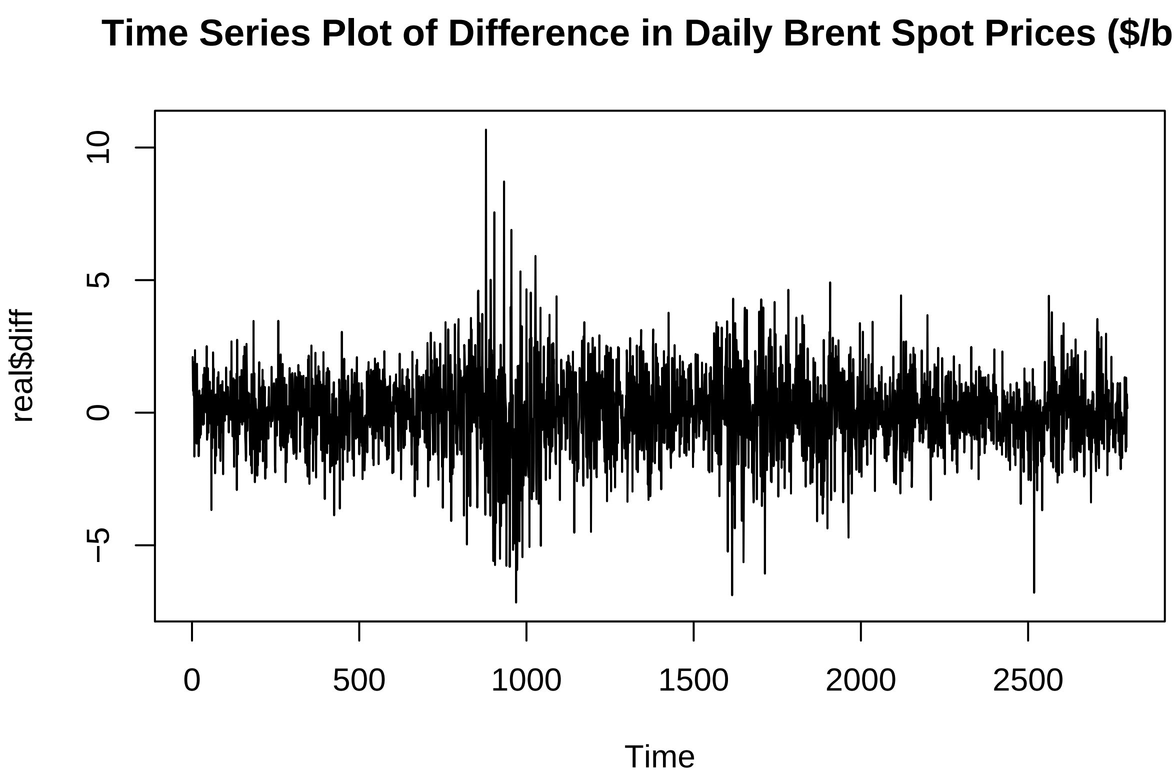
```r
var(real$price)
```

```
## [1] 640.3306
```

```r
ts.plot(real$price, main = "Time Series Plot of Daily Brent Spot Prices ($/bbl)")
```



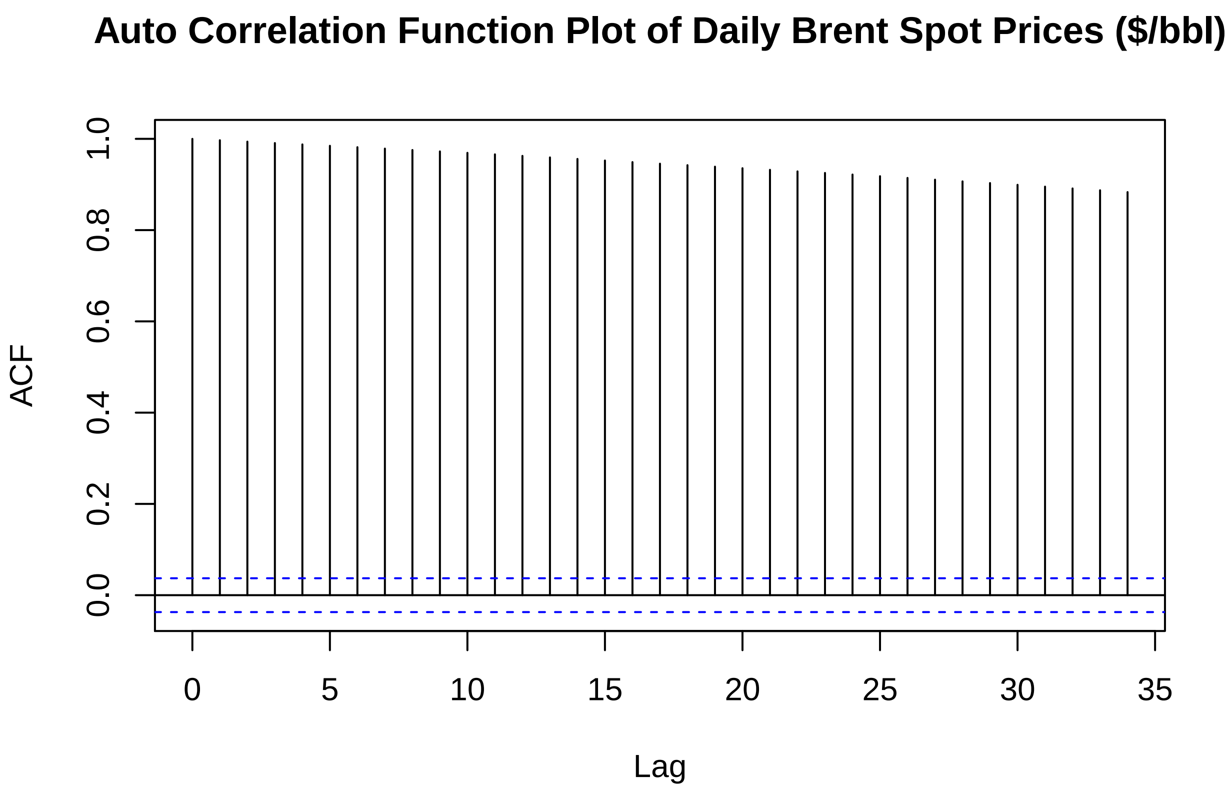

```
ts.plot(real$diff, main = "Time Series Plot of Difference in Daily Brent Spot Prices ($/bbl)")
```



Step 2. Plot the acf and pacf for the series

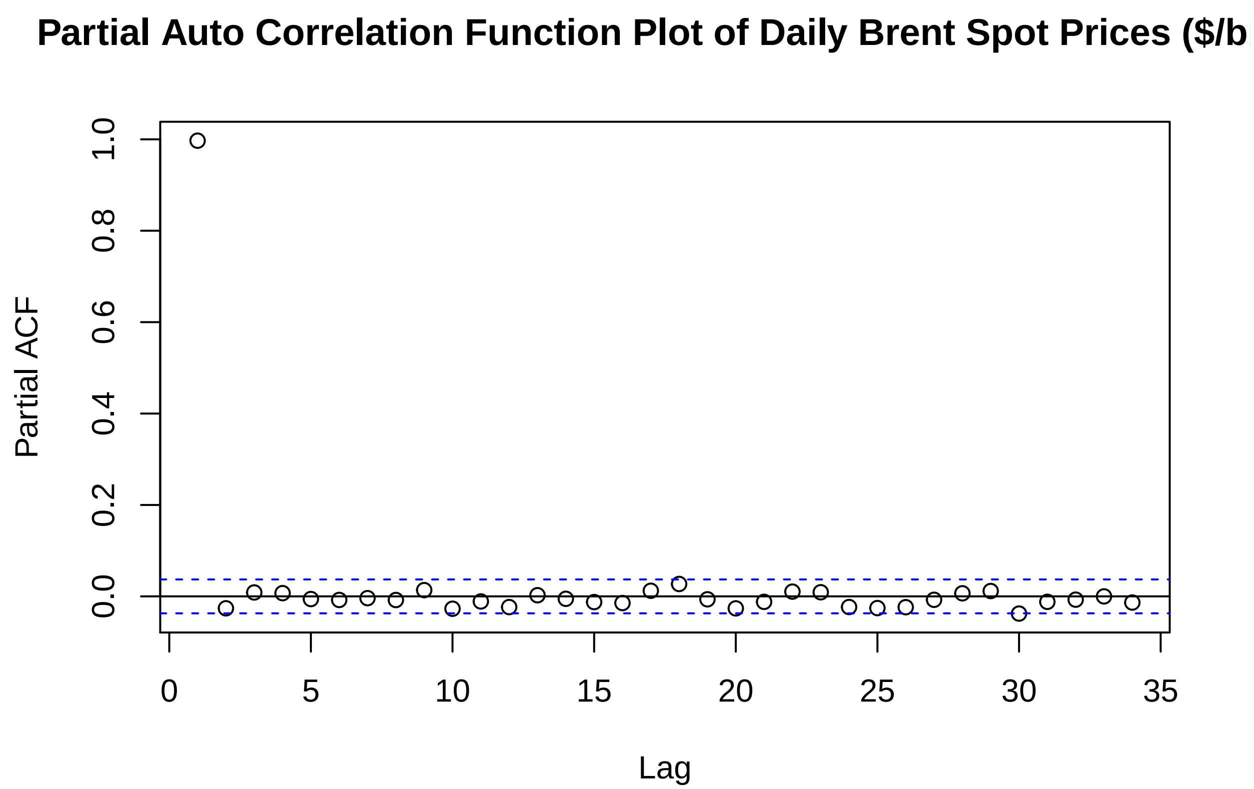
We now plot the Auto Correlation Function and the Partial AutoCorrelation Function

```
acf(real$price, main = "Auto Correlation Function Plot of Daily Brent Spot Prices ($/bbl)")
```




```
pacf(real$price,type = "partial", main = "Partial Auto Correlation Function Plot of Daily Brent Spot Pri
```

```
## Warning in plot.xy(xy, type, ...): plot type 'partial' will be truncated to
## first character
```




```
#Since the auto-correlation is relatively low, it is unlikely that this is white noise.
```

Step 3. Test Normality

1. A stochastic process is covariance stationary (or weak stationary) if

a. E(Yt) = µ, for all t
b. Var(Yt) = y2, for all t
c. Cov(Yt,Yt-h) = (h), for all t and h

2. A process is white noise if it is

a. Covariance Stationary
b. Uncorrelated
c. Zero mean

3. A process is strict white noise if it is

a. Covariance Stationary
b. Zero mean
c. Independent and identically distributed (iid)

ts.plot(log(real), main = "Time Series Plot of Daily Brent Spot Prices ($/bbl)")

Step 4: Checking simulated Data statistics

```r
summary(sim)
```

```
##      Serie1           Serie2            Serie3            Serie4
##  Min.   :-0.944   Min.   :-2.4650   Min.   :-5.7290   Min.   :-2.7120
##  1st Qu.: 1.210   1st Qu.:-0.6010   1st Qu.:-0.7345   1st Qu.:-0.7915
##  Median : 1.975   Median : 0.1430   Median :-0.1540   Median :-0.0960
##  Mean   : 2.025   Mean   : 0.1315   Mean   :-0.1049   Mean   :-0.0807
##  3rd Qu.: 2.723   3rd Qu.: 0.7715   3rd Qu.: 0.5833   3rd Qu.: 0.7075
##  Max.   : 5.578   Max.   : 2.8090   Max.   : 2.8210   Max.   : 3.1430
##  NA's   :1800     NA's   :1800      NA's   :1800      NA's   :1800
##      Serie5            Serie6
##  Min.   :-12.690   Min.   :-3.095000
##  1st Qu.: -8.922   1st Qu.:-0.459000
##  Median : -5.301   Median : 0.002500
##  Mean   : -5.227   Mean   :-0.000293
##  3rd Qu.: -2.017   3rd Qu.: 0.477250
##  Max.   :  3.367   Max.   : 2.611000
##  NA's   :1800
```

```r
mean(sim)
```

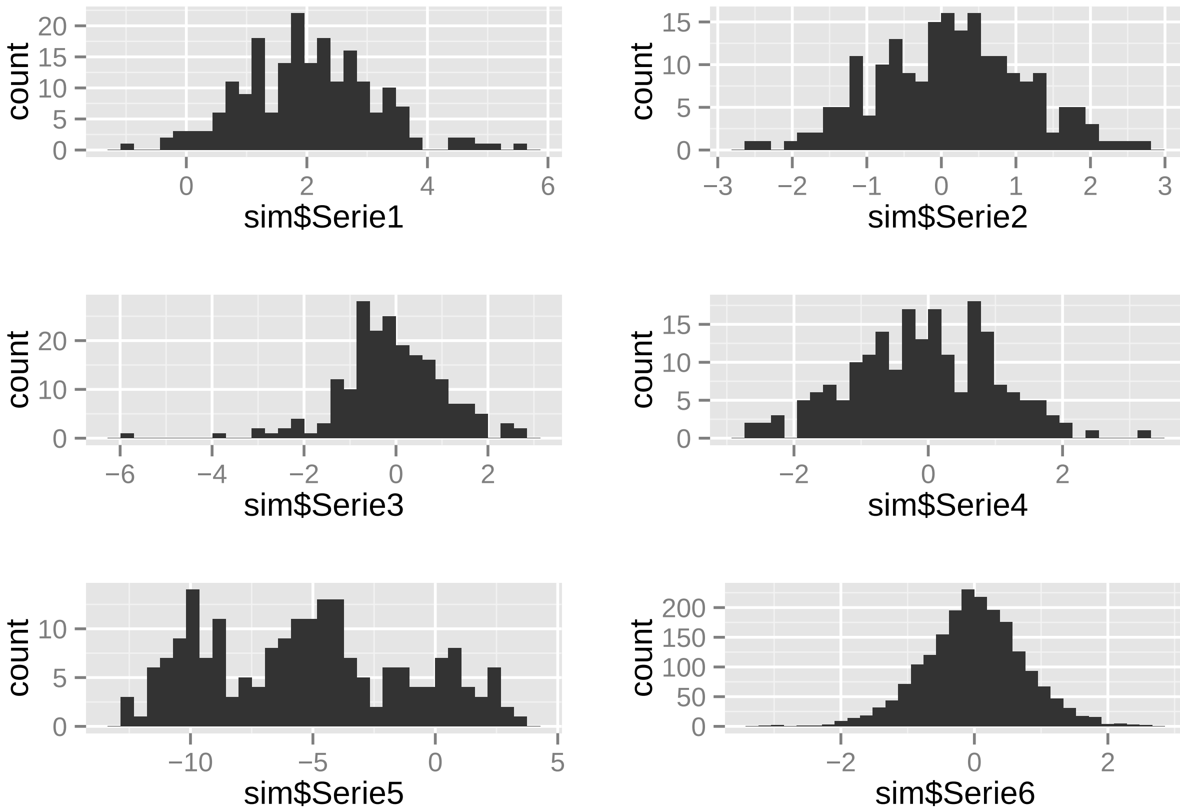
```
## Warning in mean.default(sim): argument is not numeric or logical: returning
## NA
```

```
## [1] NA
```

```r
p1 = qplot(sim$Serie1)
p2 = qplot(sim$Serie2)
p3 = qplot(sim$Serie3)
p4 = qplot(sim$Serie4)
p5 = qplot(sim$Serie5)
p6 = qplot(sim$Serie6)

grid.arrange(p1,p2,p3,p4,p5,p6, nrow = 3)
```

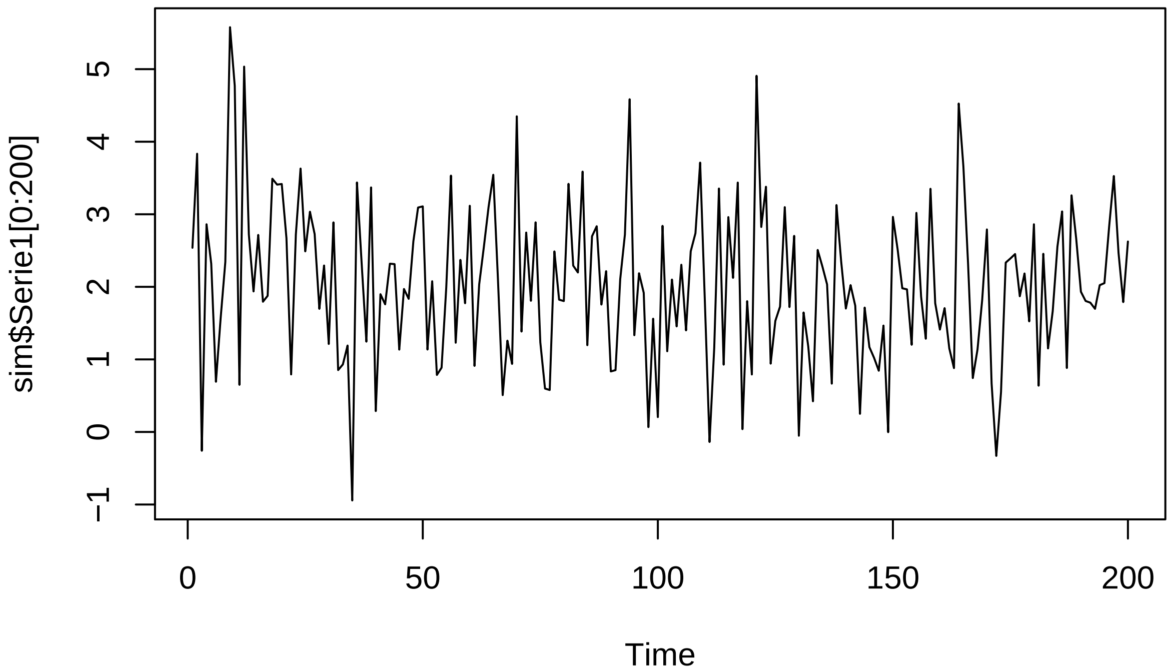
```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



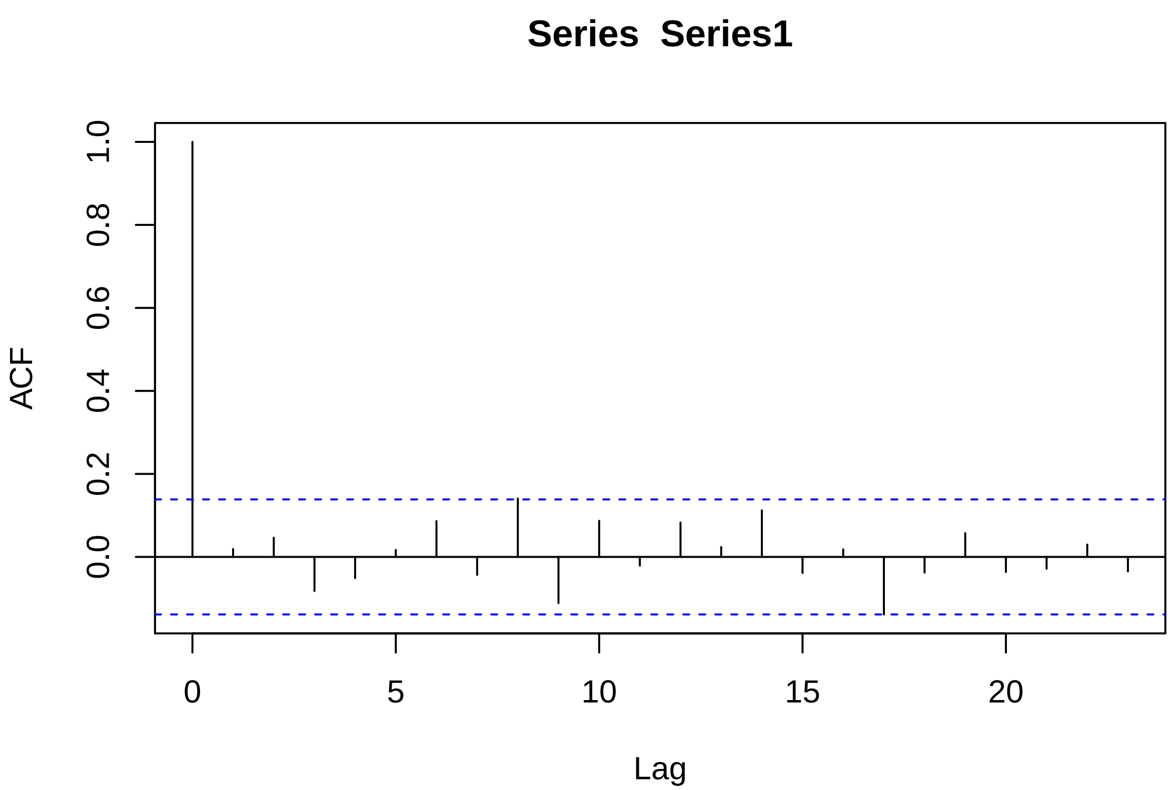

```
M = cor(sim)
#corrplot(M,method = 'number')
#corrplot(M,method = 'square')
#mean = summarize(sim,mean(sim,na.rm="TRUE"))
```

Step 5: Checking Series 1

```
mean1 = mean(sim$Serie1,na.rm = "TRUE")
sd1 = sd(sim$Serie1,na.rm="TRUE")
var1 = var(sim$Serie1,na.rm="TRUE")
ts.plot(sim$Serie1[0:200])
```



```r
Series1 = na.omit(sim$Serie1)
acf(Series1)
```



```r
#Since most auto-correlation values are close to 0, this might indicate white noise.
#Mean is not close to 0, not white noise.
# Total of 3 auto-correlation points above critical values according to the chart. 3/200 does not cross
```

```r
# Mean remains somewhat constant across time
mean(Series1[0:10])
```

```
## [1] 2.6242
```

```r
mean(Series1[0:100])
```

```
## [1] 2.12307
```

```r
mean(Series1[100:200])
```

```
## [1] 1.910228
```

```r
#Here we can use a hypothesis test to determine if mean is somewhat centered around 2, but it is quite
sd(Series1[0:10])
```

```
## [1] 1.769838
```

```r
sd(Series1[0:100])
```

```
## [1] 1.162364
```

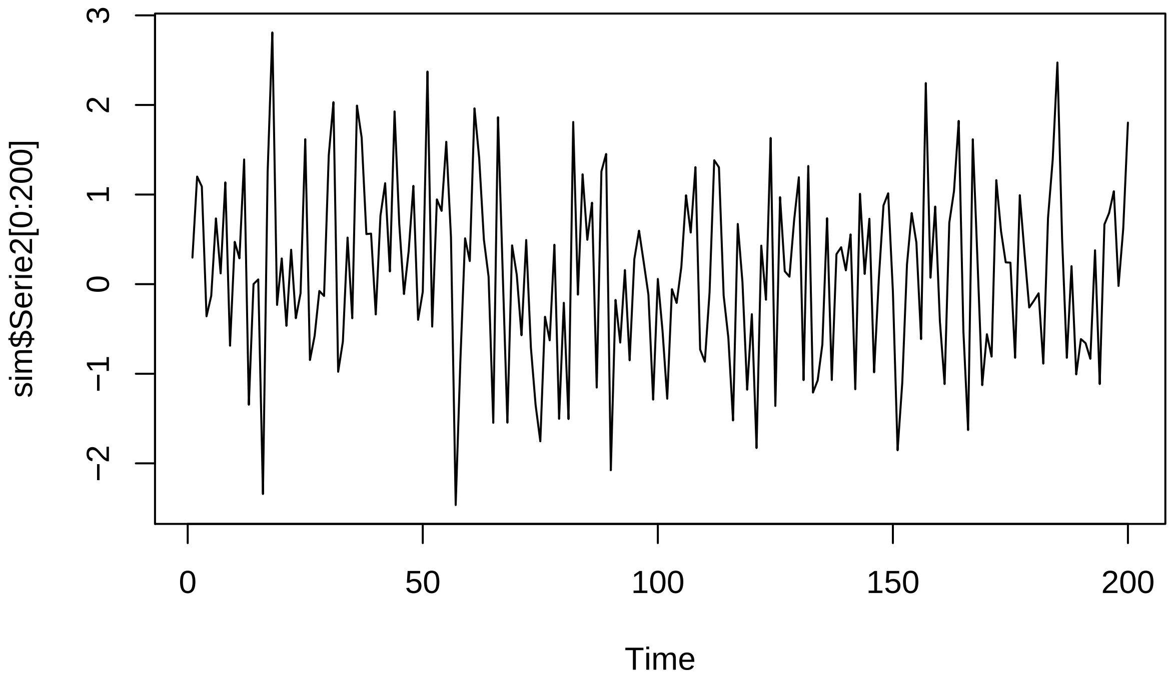
```r
sd(Series1[100:200])
```

```
## [1] 1.014593
```

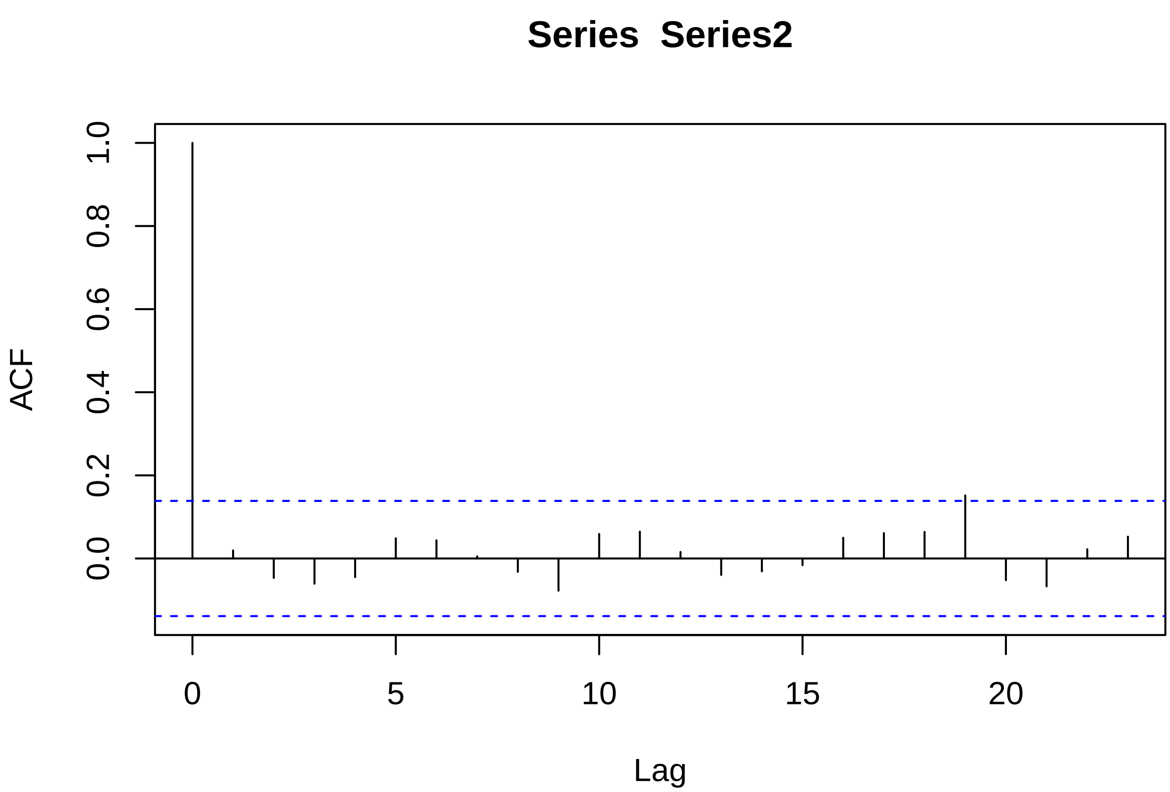
```r
#Here standard Deviation and variance seem to decrease with time. Therefore, this series is not station
```

Step 6: Checking Series 2

```r
mean2 = mean(sim$Serie2,na.rm = "TRUE")
sd2 = sd(sim$Serie2,na.rm="TRUE")
var2 = var(sim$Serie2,na.rm="TRUE")
ts.plot(sim$Serie2[0:200])
```




```
Series2 = na.omit(sim$Serie2)
acf(Series2)
```




```
#Since most auto-correlation values are close to 0, this might indicate white noise.
#Mean is close to 0, Mean = 0.13
# Total of 2 auto-correlation points above critical values according to the chart. 2/200 does not cross
```

```r
# Mean decreases across time
mean(Series2[0:10])
```

```
## [1] 0.3873
```

```r
mean(Series2[0:100])
```

```
## [1] 0.18383
```

```r
mean(Series2[100:200])
```

```
## [1] 0.07887129
```

```r
#Here we can use a hypothesis test to determine if mean is somewhat centered around 0, but it is quite
sd(Series2[0:10])
```

```
## [1] 0.6584738
```

```r
sd(Series2[0:100])
```

```
## [1] 1.053157
```

```r
sd(Series2[100:200])
```

```
## [1] 0.9406056
```

```r
#Here standard Deviation and variance seem to be roughly constant. Therefore, this series is not covari
lag2 = lag(Series2)
cov(Series2,lag2,use = 'complete')
```

```
## [1] 0.01963358
```

```r
cov(Series2[0:10],lag2[0:10],use = 'complete')
```

```
## [1] -0.1259348
```

```r
cov(Series2[0:100],lag2[0:100],use = 'complete')
```

```
## [1] 0.004840011
```

```r
cov(Series2[100:200],lag2[100:200],use = 'complete')
```

```
## [1] 0.02804846
```

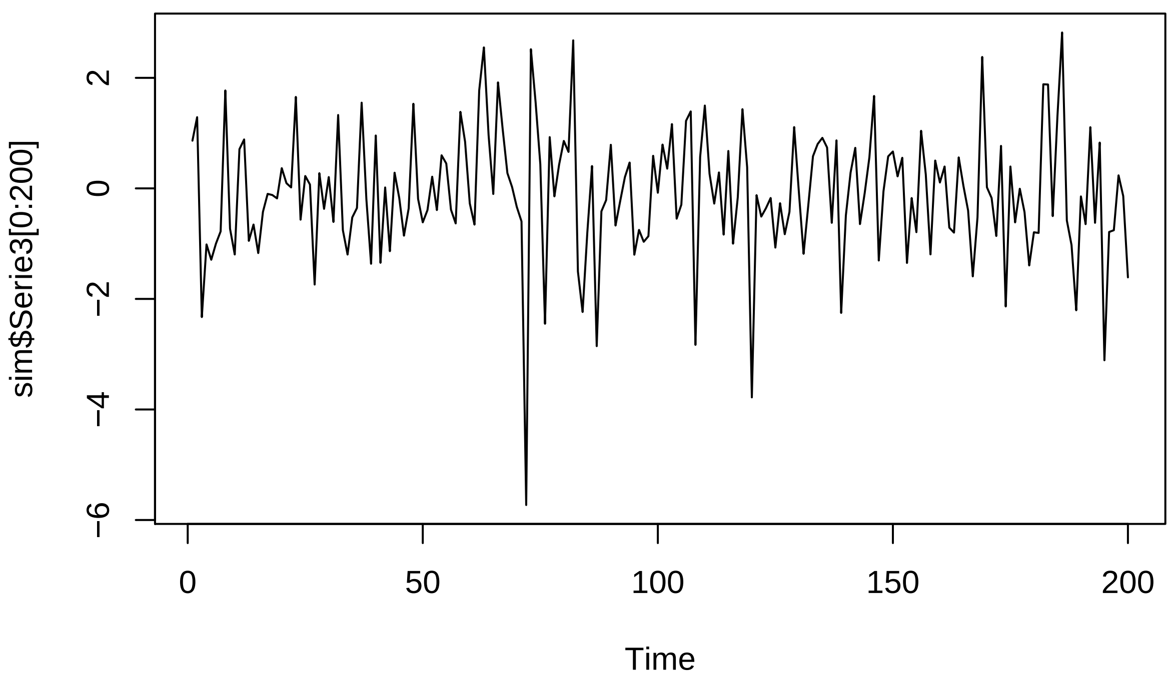
```r
#This tells us that covariance is roughly constant.
shapiro.test(Series2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Series2
## W = 0.99744, p-value = 0.9866
```

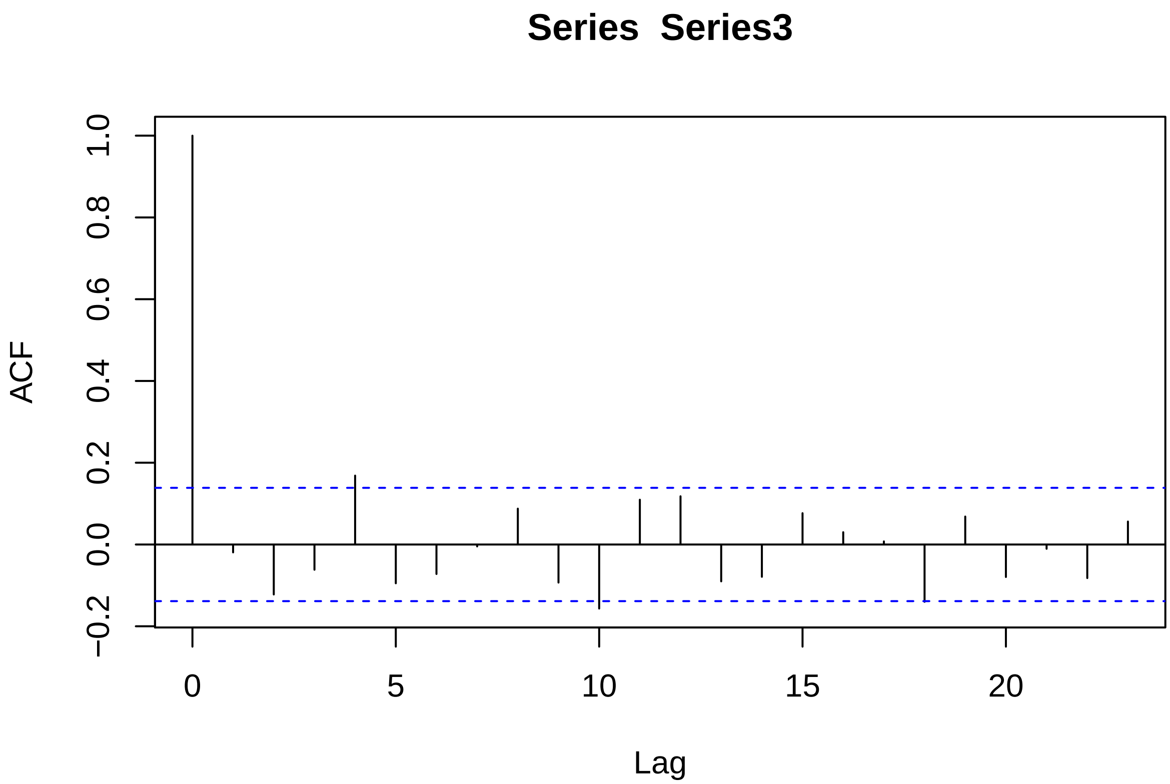
```r
#High p-value means the white noise is normal and therefore gaussian. (SWN,GWN)
```

Step 7: Checking Series 3

```r
mean3 = mean(sim$Serie3,na.rm = "TRUE")
sd3 = sd(sim$Serie3,na.rm="TRUE")
var3 = var(sim$Serie3,na.rm="TRUE")
ts.plot(sim$Serie3[0:200])
```




```r
Series3 = na.omit(sim$Serie3)
acf(Series3)
```




```
#Since most auto-correlation values are close to 0, this might indicate white noise.
#Mean is close to 0, Mean = -0.10494
# Total of 2 auto-correlation points above critical values according to the chart. 2/200 does not cross
# Mean increases across time
mean(Series3[0:10])
```

```
## [1] -0.4412
```

```
mean(Series3[0:100])
```

```
## [1] -0.10464
```

```
mean(Series3[100:200])
```

```
## [1] -0.1049802
```

```
#Here we can use a hypothesis test to determine if mean is somewhat centered around 0, but it is quite
sd(Series3[0:10])
```

```
## [1] 1.303044
```

```
sd(Series3[0:100])
```

```
## [1] 1.202217
```

```r
sd(Series3[100:200])
```

```
## [1] 1.096938
```

```r
#Here standard Deviation and variance seem to decrease with time. Therefore, this series is not covaria
lag3 = lag(Series3)
cov(Series3,lag3,use = 'complete')
```

```
## [1] -0.02569859
```

```r
cov(Series3[0:10],lag3[0:10],use = 'complete')
```

```
## [1] 0.02085346
```

```r
cov(Series3[0:100],lag3[0:100],use = 'complete')
```

```
## [1] -0.0114154
```

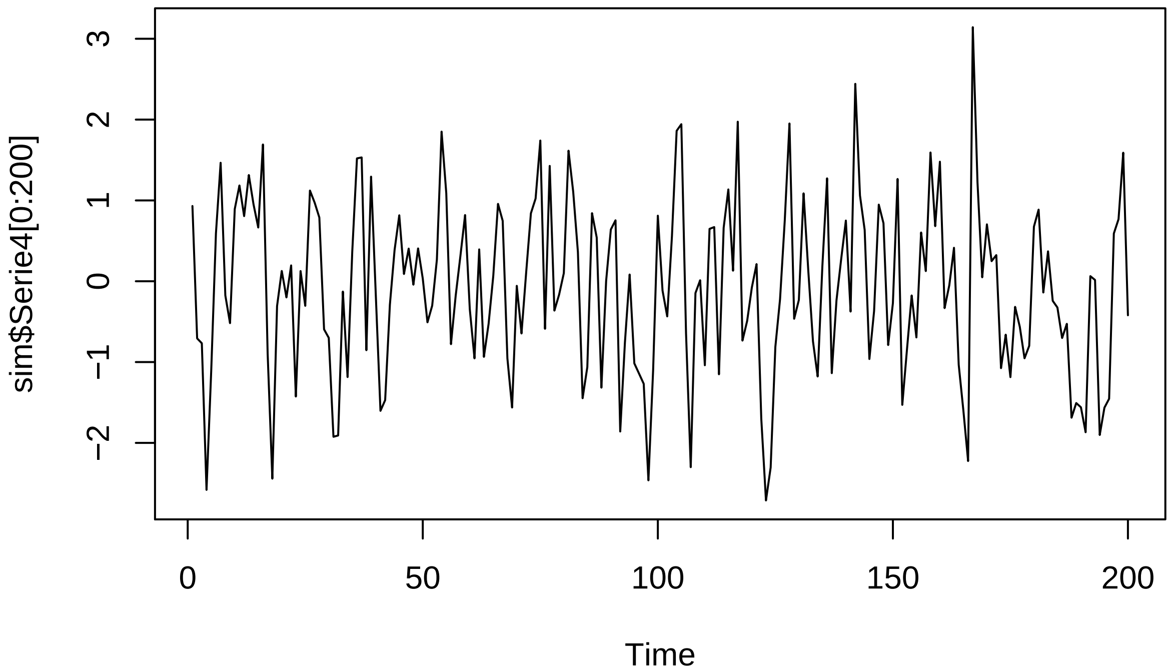
```r
cov(Series3[100:200],lag3[100:200],use = 'complete')
```

```
## [1] -0.0395878
```

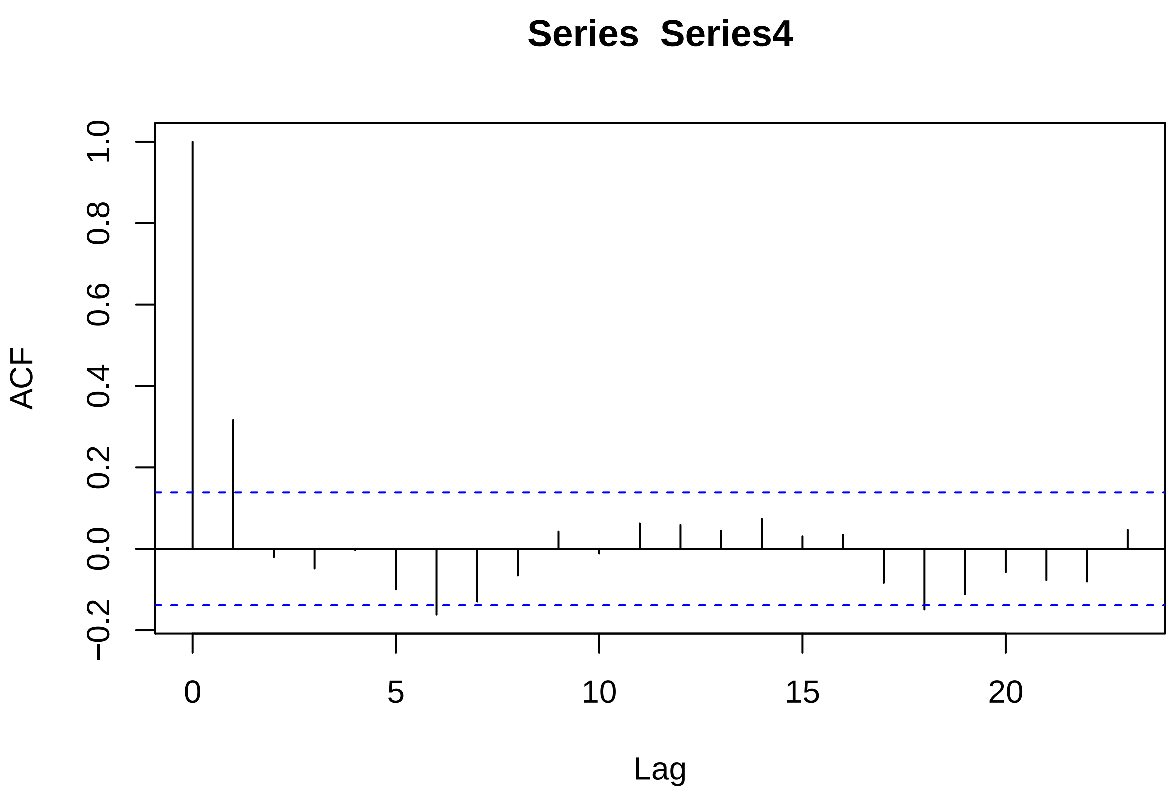
```r
#This tells us that covariance is dependent on time and constantly decreasing based on time. Which make
```

Step 8: Checking Series 4

```r
mean4 = mean(sim$Serie4,na.rm = "TRUE")
sd4 = sd(sim$Serie4,na.rm="TRUE")
var4 = var(sim$Serie4,na.rm="TRUE")
ts.plot(sim$Serie4[0:200])
```




```
Series4 = na.omit(sim$Serie4)
acf(Series4)
```




```
#Since most auto-correlation values are close to 0, this might indicate white noise.
#Mean is close to 0, Mean is -0.080695
# Total of 2 auto-correlation points above critical values according to the chart. 2/200 does not cross
```

```r
# Mean is roughly constant
mean(Series4[0:10])
```

```
## [1] -0.1967
```

```r
mean(Series4[0:100])
```

```
## [1] -0.05393
```

```r
mean(Series4[100:200])
```

```
## [1] -0.09836634
```

```r
#Here we can use a hypothesis test to determine if mean is somewhat centered around 0, but it is quite
sd(Series4[0:10])
```

```
## [1] 1.203595
```

```r
sd(Series4[0:100])
```

```
## [1] 1.025043
```

```r
sd(Series4[100:200])
```

```
## [1] 1.10992
```

```r
#Here standard Deviation and variance seem to roughly constant. Therefore, this series is covariance st
lag4 = lag(Series4)
cov(Series4,lag4,use = 'complete')
```

```
## [1] 0.3621047
```

```r
cov(Series4[0:10],lag4[0:10],use = 'complete')
```

```
## [1] 0.4215058
```

```r
cov(Series4[0:100],lag4[0:100],use = 'complete')
```

```
## [1] 0.3547752
```

```r
cov(Series4[100:200],lag4[100:200],use = 'complete')
```

```
## [1] 0.3594821
```

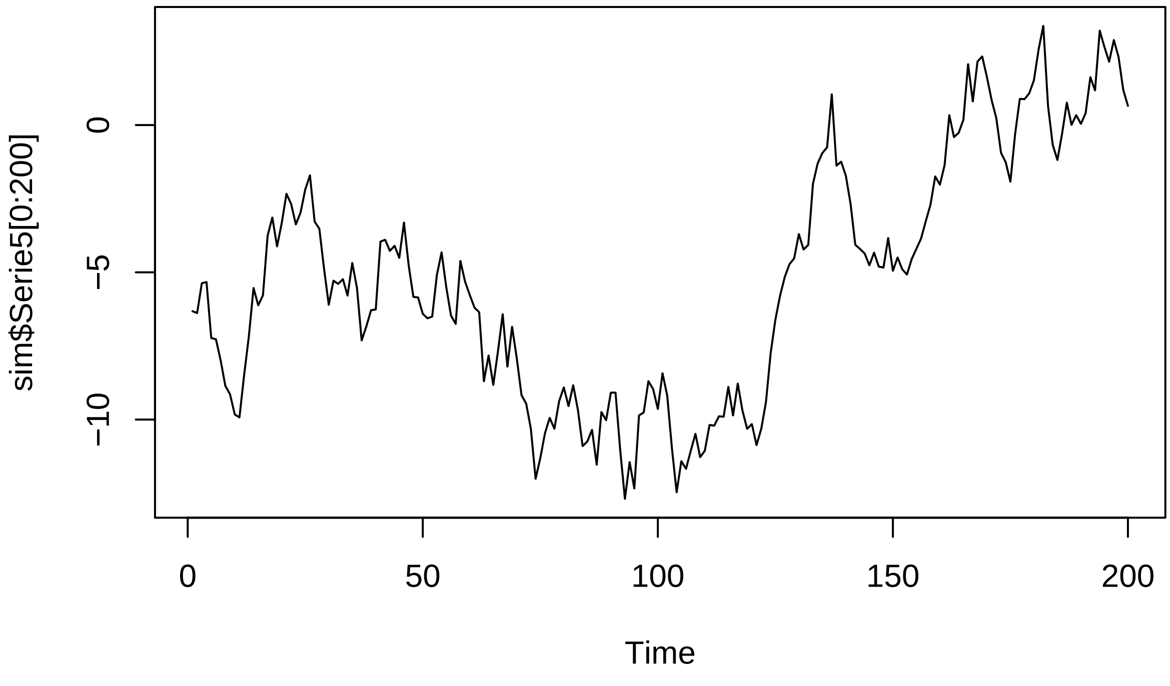
```r
#This tells us that covariance is independent of time and almost constant.
#Testing for normality
shapiro.test(Series4)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Series4
## W = 0.99602, p-value = 0.8851
```

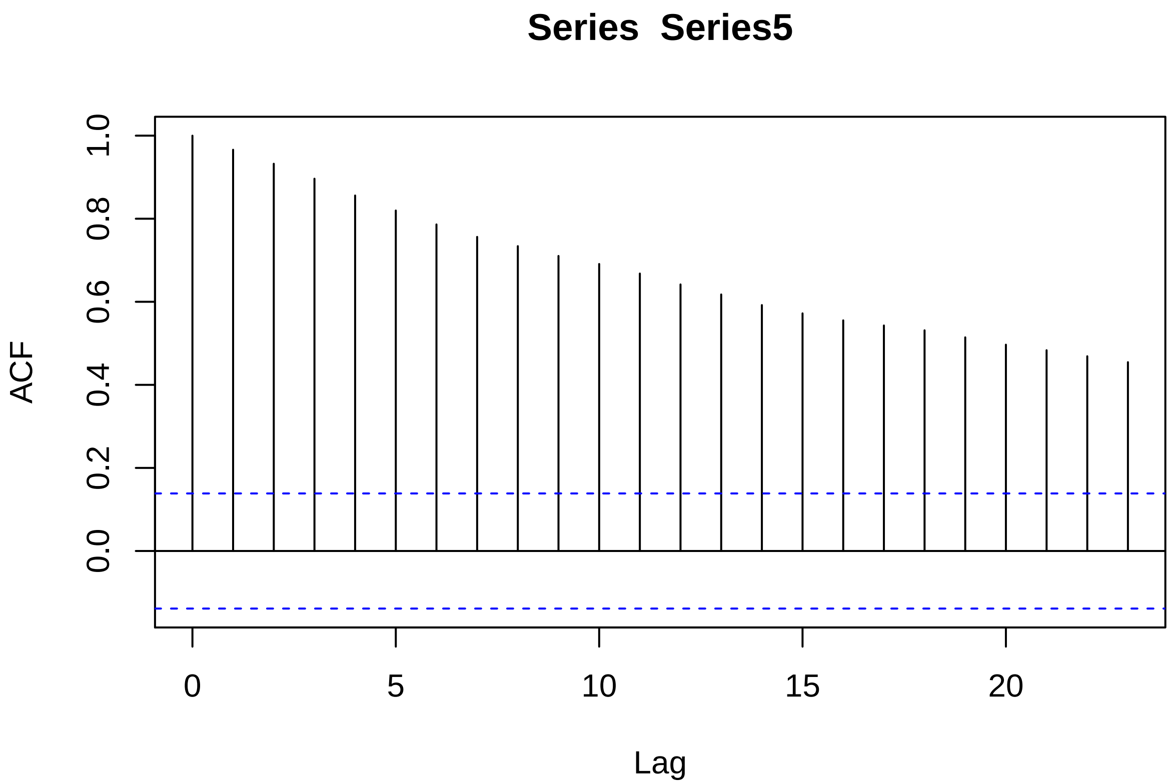
```r
#It seems like Series 4 is a normally distributed white noise (Gaussian White noise and strictly white
```

Step 9: Checking Series 5

```r
mean5 = mean(sim$Serie5,na.rm = "TRUE")
sd5 = sd(sim$Serie5,na.rm="TRUE")
var5 = var(sim$Serie5,na.rm="TRUE")
ts.plot(sim$Serie5[0:200])
```




```r
Series5 = na.omit(sim$Serie5)
acf(Series5)
```




```r
#Since most auto-correlation values are not close to 0, this indicates a definite time series pattern.
# Mean is far from 0
# Mean decreases across time
mean(Series5[0:10])
```

```
## [1] -7.3717
```

```r
mean(Series5[0:100])
```

```
## [1] -7.06722
```

```r
mean(Series5[100:200])
```

```
## [1] -3.448475
```

```r
#Here we can use a hypothesis test to determine if mean is somewhat centered around 0, but it is quite
sd(Series5[0:10])
```

```
## [1] 1.563711
```

```r
sd(Series5[0:100])
```

```
## [1] 2.640267
```

```r
sd(Series5[100:200])
```

```
## [1] 4.522227
```

```r
#Here standard Deviation and variance seem to increase with time. Therefore, this series is not covaria
lag2 = lag(Series2)
cov(Series2,lag2,use = 'complete')
```

```
## [1] 0.01963358
```

```r
cov(Series2[0:10],lag2[0:10],use = 'complete')
```

```
## [1] -0.1259348
```

```r
cov(Series2[0:100],lag2[0:100],use = 'complete')
```

```
## [1] 0.004840011
```

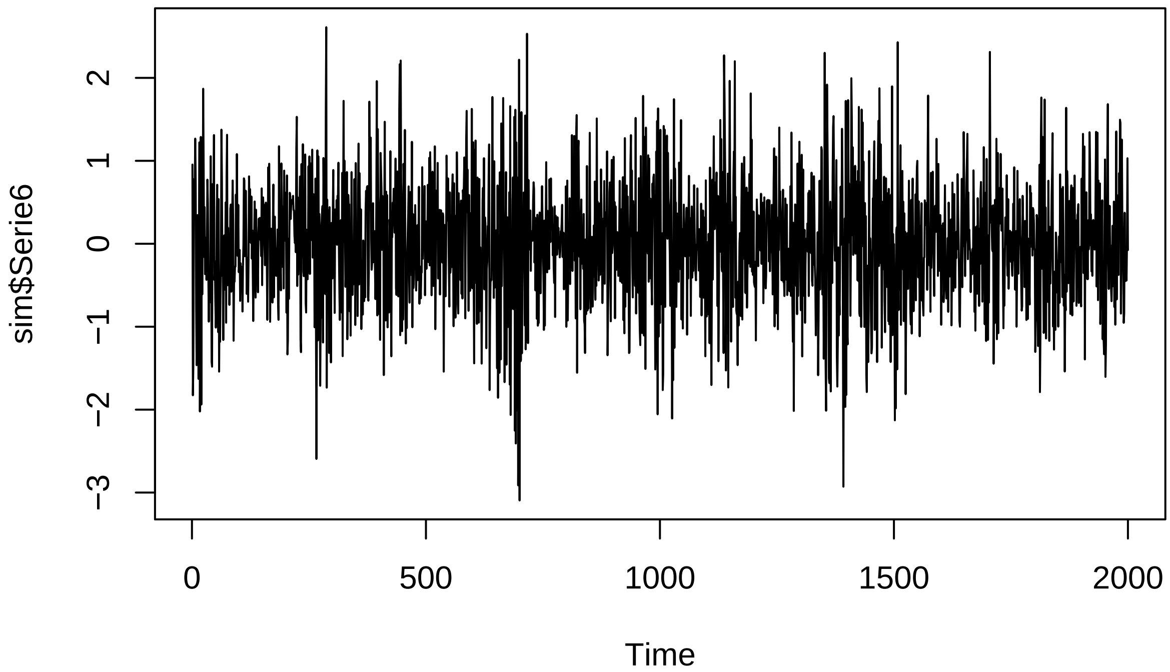
```r
cov(Series2[100:200],lag2[100:200],use = 'complete')
```

```
## [1] 0.02804846
```

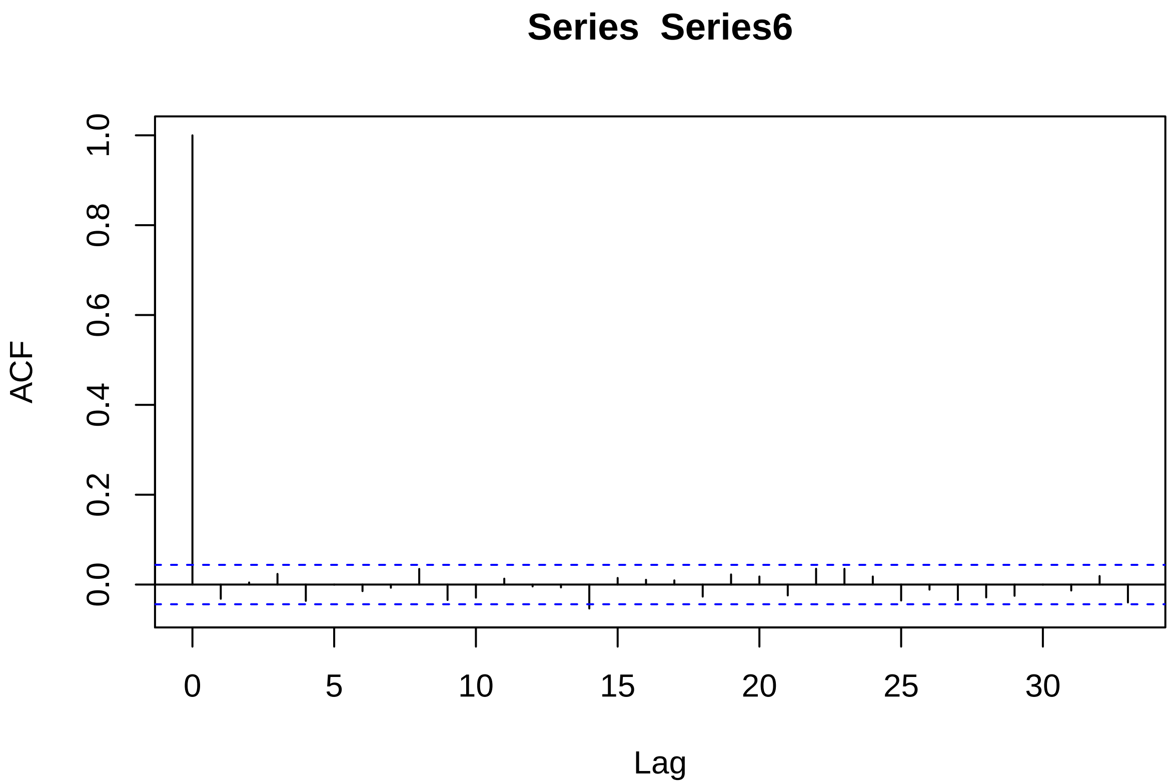
```r
#This tells us that covariance is independent of time and roughly constant.
```

Step 10: Checking Series 6

```r
mean6 = mean(sim$Serie6,na.rm = "TRUE")
sd6 = sd(sim$Serie6,na.rm="TRUE")
var6 = var(sim$Serie6,na.rm="TRUE")
ts.plot(sim$Serie6)
```




```
Series6 = na.omit(sim$Serie6)
acf(Series6)
```




```
#Since most auto-correlation values are close to 0, this might indicate white noise.
#Mean is close to 0, Mean is -0.0002929
# Mean roughly constant
mean(Series6[0:10])
```

```
## [1] -0.0819
```

```r
mean(Series6[0:1000])
```

```
## [1] 0.009737
```

```r
mean(Series6[1000:200])
```

```
## [1] 0.02179775
```

```r
#Here we can use a hypothesis test to determine if mean is somewhat centered around 0, but it is quite
sd(Series6[0:10])
```

```
## [1] 1.168371
```

```r
sd(Series6[0:1000])
```

```
## [1] 0.7526359
```

```r
sd(Series6[1000:2000])
```

```
## [1] 0.7437998
```

```r
#Here standard Deviation and variance seem to be roughly constant/decrease with time. Therefore, this s
lag6 = lag(Series6)
cov(Series6,lag6,use = 'complete')
```

```
## [1] -0.01774753
```

```r
cov(Series6[0:10],lag6[0:10],use = 'complete')
```

```
## [1] 0.4638805
```

```r
cov(Series6[0:1000],lag6[0:1000],use = 'complete')
```

```
## [1] -0.06485198
```

```r
cov(Series6[1000:2000],lag6[1000:2000],use = 'complete')
```

```
## [1] 0.02963085
```

```r
#This tells us that covariance is independent of time and roughly constant.
shapiro.test(Series6)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Series6
## W = 0.9968, p-value = 0.0003511
```

```
# Not a normally distributed but definitely white noise.(WN)
```