Homework7

Megan Jasek, Rohan Thakur, Charles Kekeh

Sunday, March 27, 2016

Exercise 1

1.1 Load hw07_series1.csv

```
hw07.series1 = ts(read.csv("hw07_series1.csv", header = TRUE))
```

1.2 Describe the basic structure of the data and provide summary statistics of the series

The data are a time series of 74 values.

```
str(hw07.series1)
```

```
##  ts [1:74, 1] 10.07 10.32 9.75 10.33 10.13 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : NULL
##   ..$ : chr "X10.01"
##  - attr(*, "tsp")= num [1:3] 1 74 1
```

```
describe(hw07.series1)
```

```
##        vars  n  mean   sd median trimmed  mad  min   max range skew
## X10.01    1 74 10.82 0.44  10.82    10.8 0.48 9.75 11.94  2.19  0.3
##        kurtosis   se
## X10.01    -0.11 0.05
```

```
summary(hw07.series1)
```

```
##      X10.01
##  Min.   : 9.75
##  1st Qu.:10.48
##  Median :10.82
##  Mean   :10.82
##  3rd Qu.:11.06
##  Max.   :11.94
```

```
quantile(hw07.series1, probs = c(0.25, 0.5, 0.75))
```

```
##     25%     50%     75%
## 10.4825 10.8250 11.0650
```

1.3 Plot histogram and time-series plot of the series. Describe the patterns exhibited in histograrm and time-series plot.

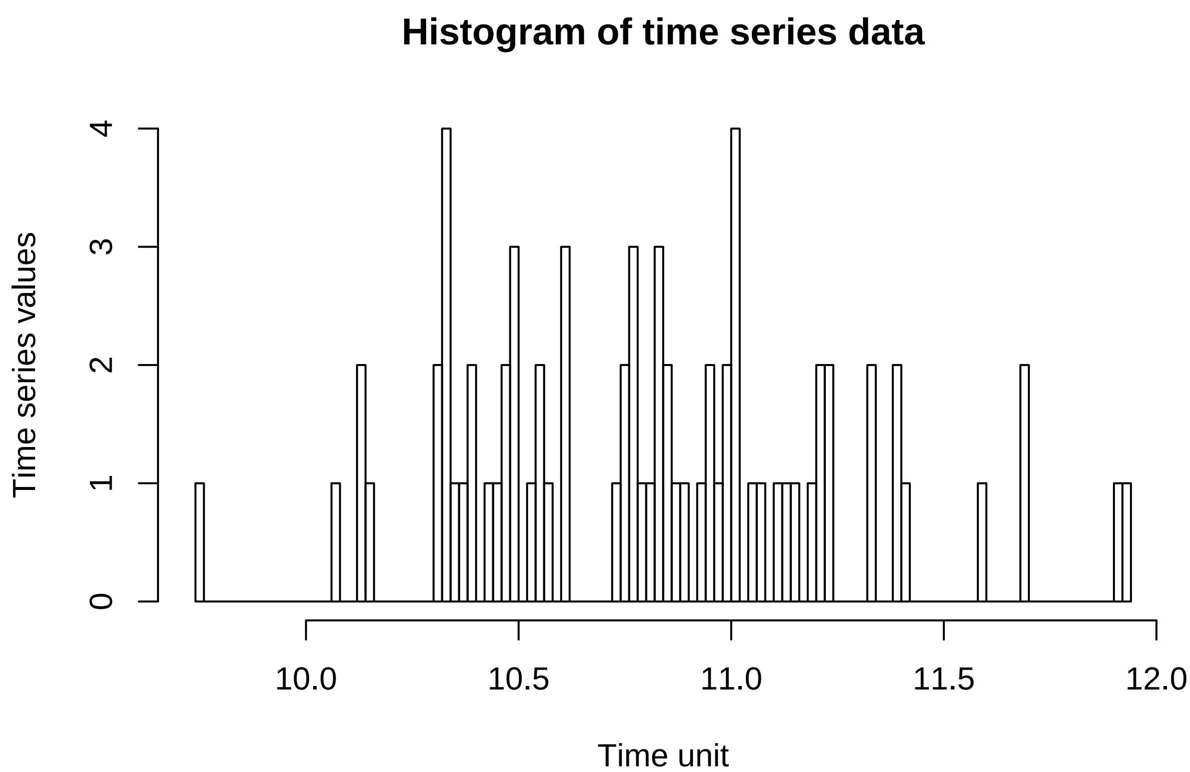
The histogram has values from 9.75 to 11.94 with a mean of 10.82. It has multiple modes. There are spikes around 10.3 and 11. The histogram generally rises on the left-hand side of the graph and falls on the right-hand side of the graph.

The plot generally trends up from time=0 to about time=22. From time=22 to about time=72, the plot trends down except for a spike at about time=42. At aboue time=72, it appears that the data is starting to trend up again. The data has some persistence in that when it goes up it tends to stay up and when it goes down it tends to stay down. There doesn't appear to be consistent seasonality.

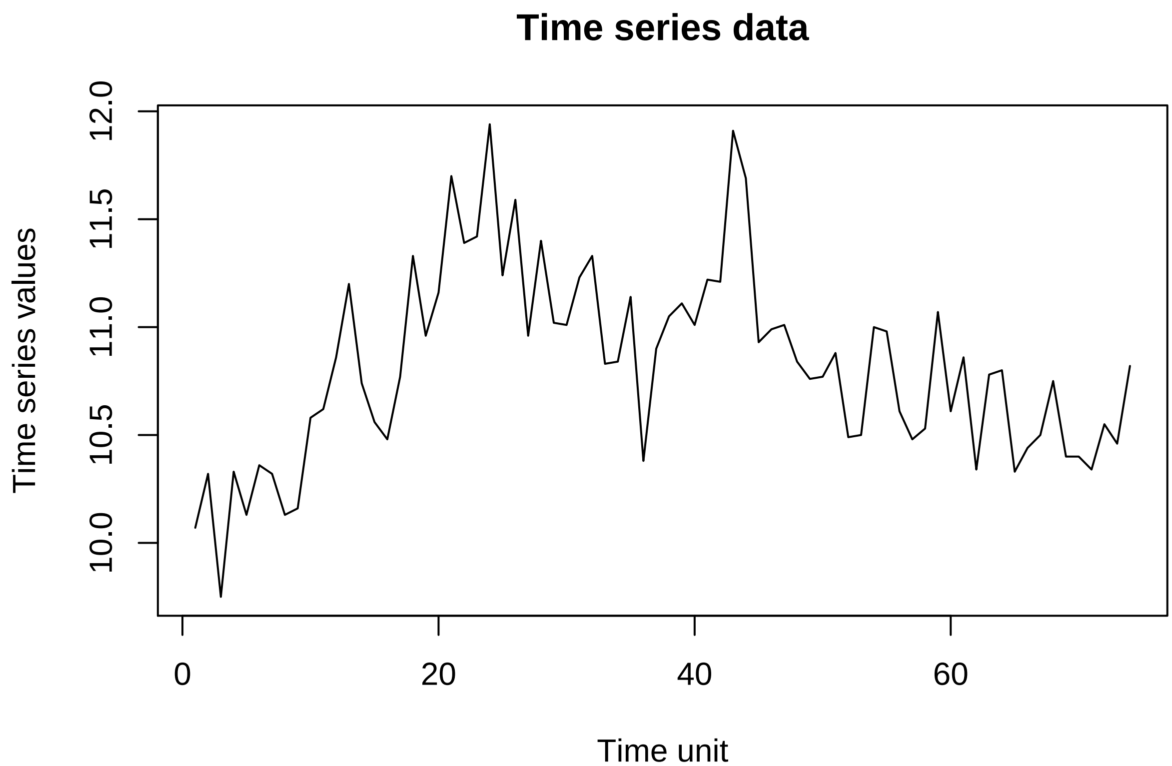
For time series analysis, is it sufficient to use only histogram to describe a series?

It is not sufficient to use only histogram for a time series because the histogram does not show the dependencies among the data over time. It does not show the time component. It only shows the values of the series and not how they relate to each other.

```r
hist(hw07.series1, 100, , xlab = "Time unit", ylab = "Time series values",
    main = "Histogram of time series data")
```




```r
plot(hw07.series1, type = "l", main = "Time series data", ylab = "Time series values",
    xlab = "Time unit")
```

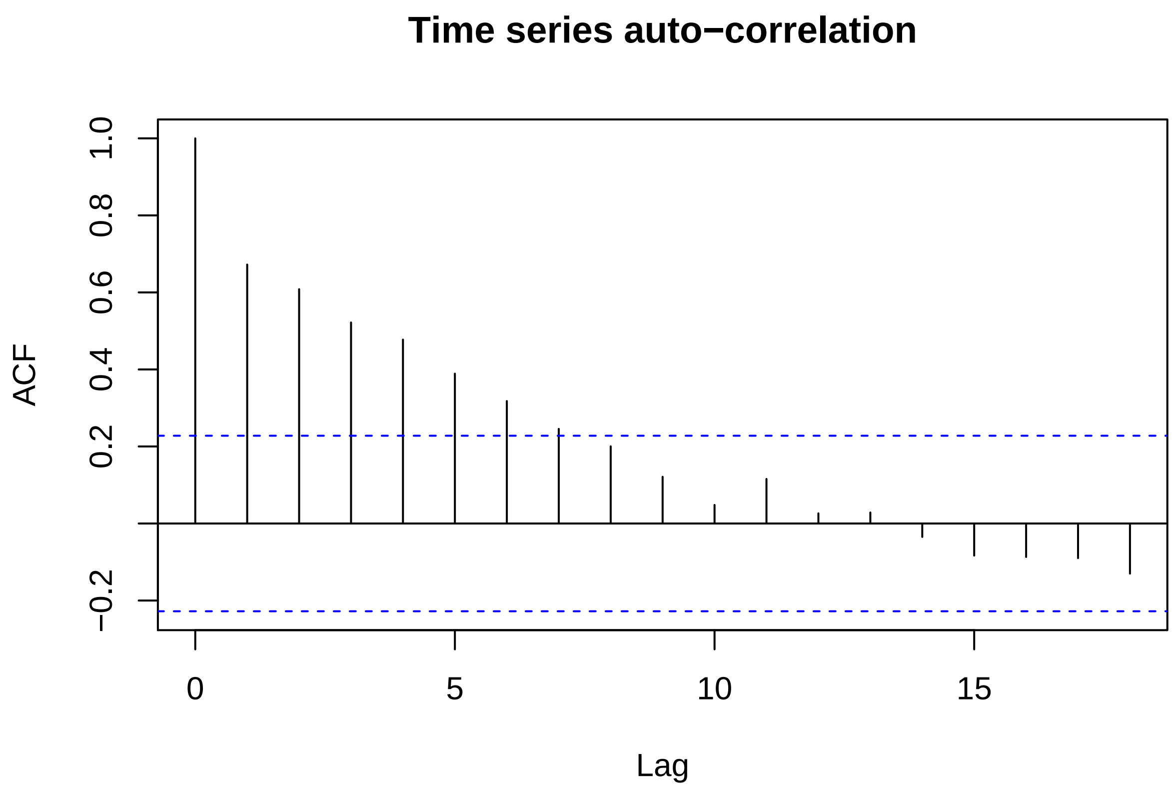


1.4 Plot the ACF and PACF of the series. Describe the patterns exhibited in the ACF and PACF.

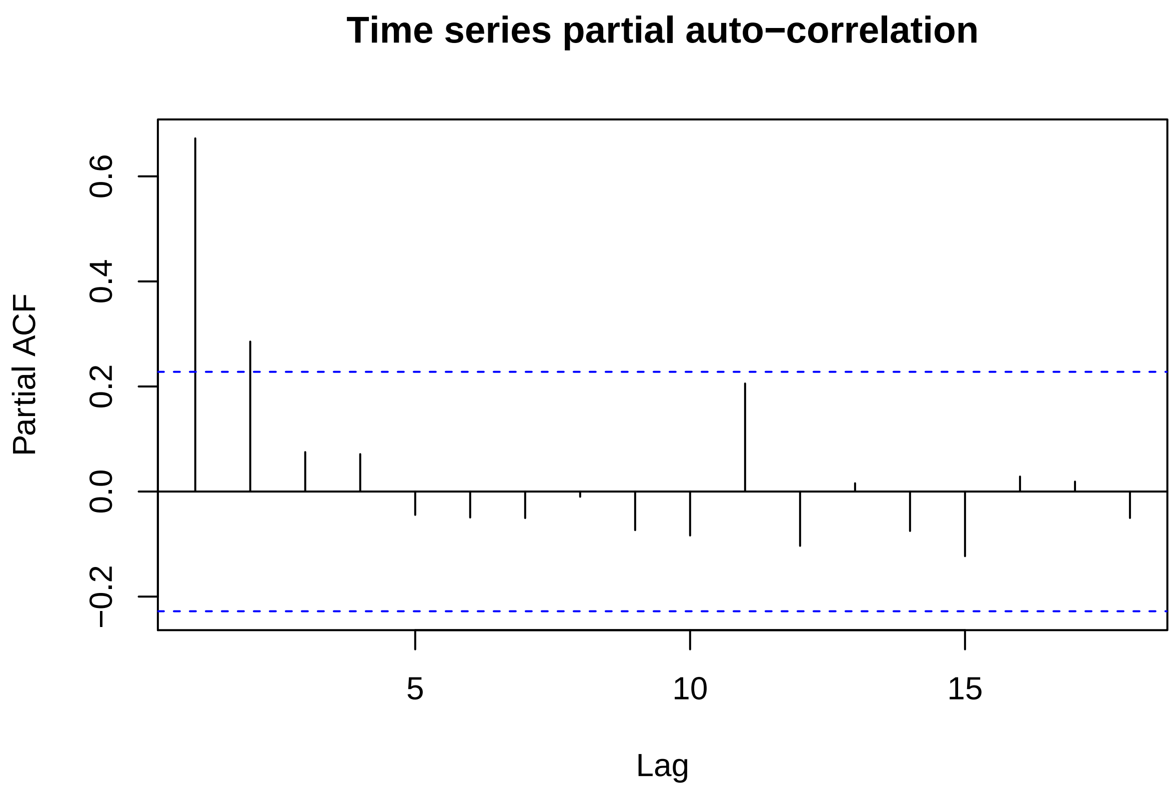
The ACF is positive from lag=0 to lag=13. At lag=14, the acf is negative. It has steadily decreasing amplitude. At lag=14, the amplitude of the lag is about -0.5 and then continues to decrease with negative amplitude. The ACF displays damped oscillations, indicating a time series of order greater than one with negative coefficients.

The PACF drops off to non-significant values after lag 2 indicating that it is an autoregressive model with order 2.

```
par(mfrow = c(1, 1))
acf(hw07.series1, main = "Time series auto-correlation")
```




```r
pacf(hw07.series1, main = "Time series partial auto-correlation")
```



1.5 Estimate the series using the ar() function.

```
hw07.series1.ar = ar(hw07.series1)
```

1.6 Report the estimated AR parameters, the order of the model, and standard errors.

- Estimated AR parameters - 0.4803726 and 0.2854828
- Order of the model - 2
- Standard error for 0.4803726 - 0.1137392
- Standard error for 0.2854828 - 0.1137392

```
# parameter estimates
hw07.series1.ar$ar
```

```
## [1] 0.4803726 0.2854828
```

```
# order of the model with lowest AIC
hw07.series1.ar$order
```

```
## [1] 2
```

```r
# standard deviation
sqrt(hw07.series1.ar$asy.var)
```

```
## Warning in sqrt(hw07.series1.ar$asy.var): NaNs produced
```

```
##           [,1]      [,2]
## [1,] 0.1137392       NaN
## [2,]       NaN 0.1137392
```

Exercise 2

2.1 Simulate a time series of length 100 for the following model. Name the series x.

$$x_t = \frac{5}{6}x_{t-1} - \frac{1}{6}x_{t-2} + \omega_t$$

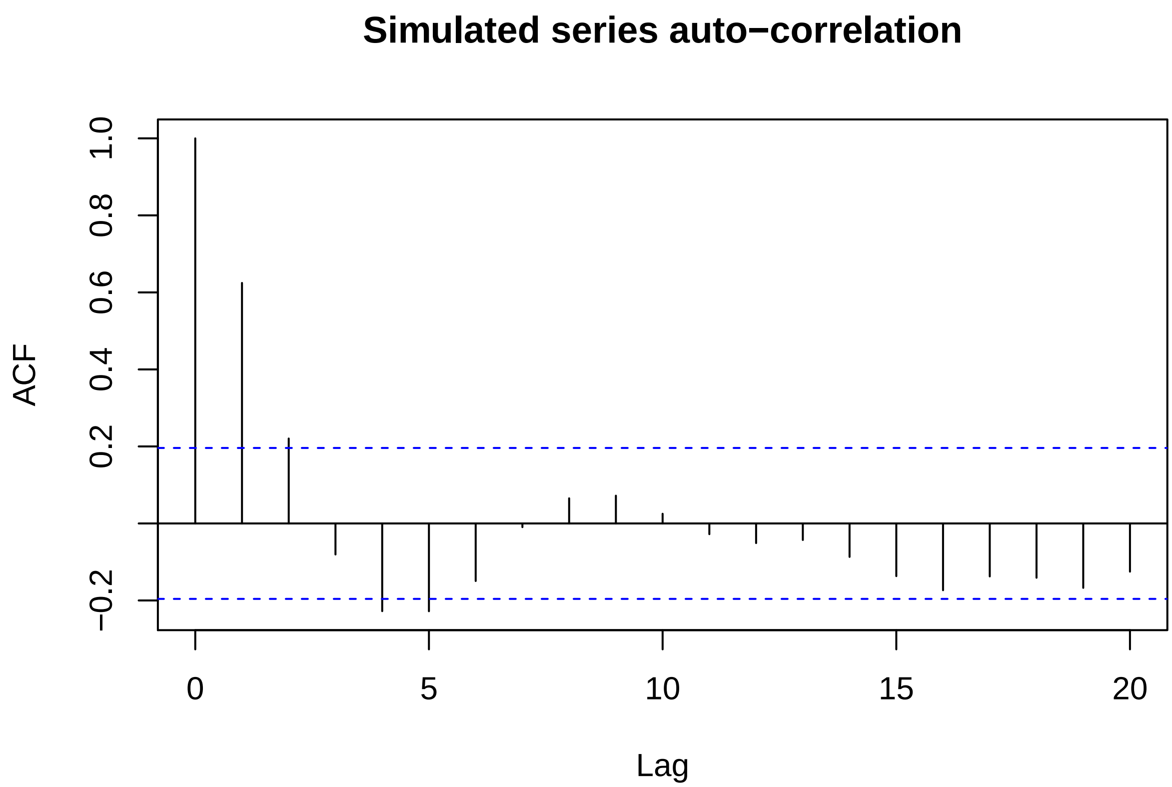
```r
white.noise = rnorm(100, 0, 1)
x = white.noise
for (t in 3:length(white.noise)) {
    x[t] = 5/6 * x[t - 1] - 1/6 * x[t - 2] + white.noise[t]
}
```

2.2 Plot the correlogram and partial correlogram for the simulated series. Comment on the plots.

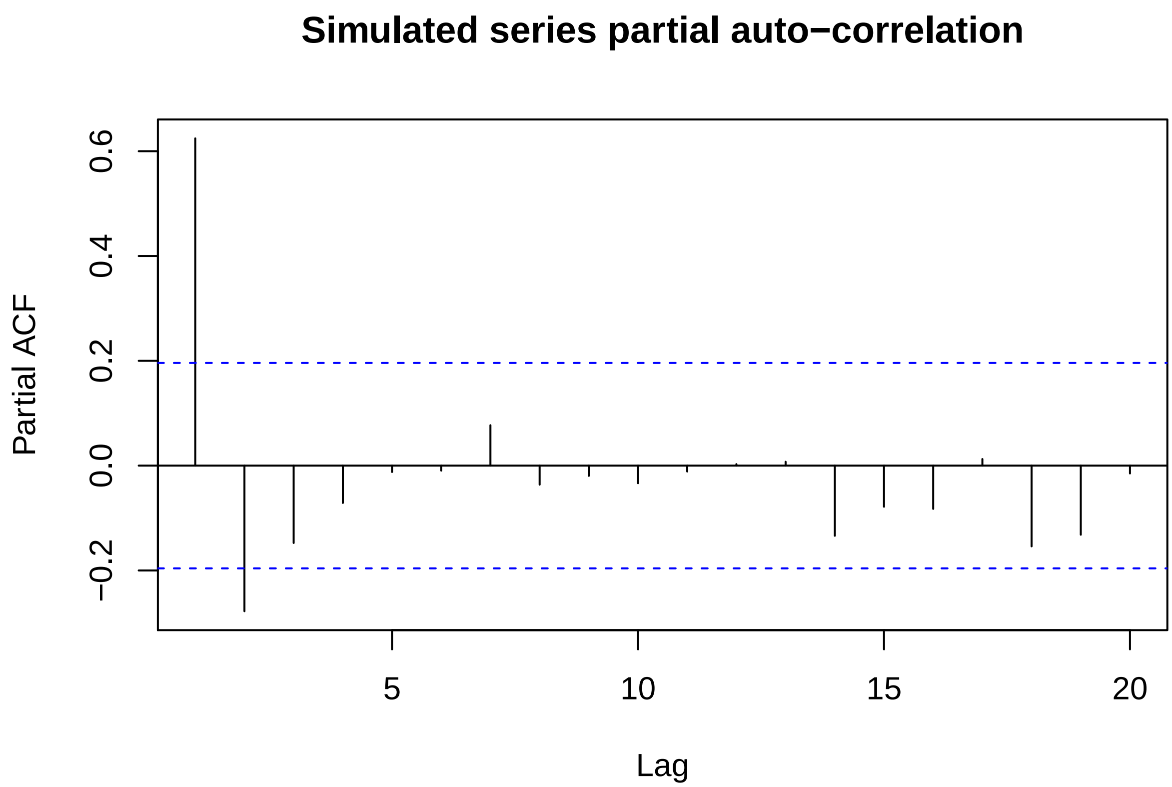
The ACF is slowly decaying toward zero. It oscillates from positive numbers to negative numbers. It has a damped oscillation that we would expect with an AR(p) model (P>1) with negative coefficients.

The PACF drops off to non-significant values after lag 2 which indicates that it could be an autoregressive series of order 2.

```r
par(mfrow = c(1, 1))
acf(x, main = "Simulated series auto-correlation")
```




```r
pacf(x, main = "Simulated series partial auto-correlation")
```



2.3 Estimate an AR model for this simulated series. Report the estimated AR parameters, standard erros, and the order of the AR model.

- Estimated AR parameters - 0.7789692 and -0.1630600 and -0.1502864
- Standard error for 0.7789692 - 0.09729566
- Standard error for -0.1630600 - 0.1215105
- Standard error for -0.1502864 - 0.09729566
- Order of the model - 3

The best models fitted is an AR(3) model. However we can see from the relative AICs that the AR(2) model that we know is teh true model, has the closest AIC value to the AR(3) model. Print the parameter estimates

```
# Estimate the model
x.ar = ar(x, method = "mle")
# Parameter estimates
x.ar$ar
```

```
## [1]  0.7789692 -0.1630600 -0.1502864
```

```
# Standard deviation
sqrt(x.ar$asy.var)
```

```
## Warning in sqrt(x.ar$asy.var): NaNs produced
```

```
##             [,1]      [,2]       [,3]
## [1,] 0.09729566       NaN 0.05128590
## [2,]        NaN 0.1215105        NaN
## [3,] 0.05128590       NaN 0.09729566
```

```
# Order
x.ar$order
```

```
## [1] 3
```

```
# This is the difference of each AIc with the lowest AIC
x.ar$aic
```

```
##          0          1          2          3          4          5
## 56.1893133  6.3682863  0.1956891  0.0000000  1.3848816  3.2127821
##          6          7          8          9         10         11
##  5.2088938  6.0804126  7.9627751  9.8821476 11.9017490 13.7317087
##         12
## 15.7059834
```

2.4 Construct a 95% confidence interval for the parameter estimates of the estimated model. Do the "true" model parameters fall within the confidence intervals? Explain the 95% confidence intervals in this context.

The computed confidence intervals for the model paramaters are as follows:

$$\varphi1 = [0.5882697, 0.9696687]$$

$$\varphi2 = [-0.40122049, 0.07510058]$$

$$\varphi3 = [-0.3409859, 0.0404131]$$

All true parameters fall inside the 95% confidence intervals of the estimated parameters. The 95% confidence interval is the interval where we expect 95% of the parameters to fall given our estimation of the parameters of the distribution of the time series parameters.

```
print((x.ar$ar[1]) + c(-1.96, 1.96) * sqrt(x.ar$asy.var[1, 1]))
```

```
## [1] 0.5882697 0.9696687
```

```
print((x.ar$ar[2]) + c(-1.96, 1.96) * sqrt(x.ar$asy.var[2, 2]))
```

```
## [1] -0.40122049  0.07510058
```

```
print((x.ar$ar[3]) + c(-1.96, 1.96) * sqrt(x.ar$asy.var[3, 3]))
```

```
## [1] -0.3409859  0.0404131
```

2.5 Is the estimated model stationary or non-stationary?

A model is stationary if the absolute value of all of the roots of its characteristic equation are greater than 1.

$$Characteristic\ Equation = \theta_p(B) = (1 - 0.7789692B + 0.1630600B^2 + 0.1502864B^3)$$

$$roots : 1.098508 + 0.905915i, 1.098508 - 0.905915i, -3.282012 + 0.000000i$$

The absolute value of a complex number is:

$$|a + bi| = \sqrt{a^2 + b^2}$$

The absolute value of the roots are:

$$\sqrt{1.098508^2 + 0.905915^2}, \sqrt{1.098508^2 + (-0.905915)^2}, |-3.282012|$$

$$1.423869, 1.423869, 3.282012$$

The estimated model is stationary since the absolute value of all of the roots of its characteristic equation are greater than 1.

```
# Calculate the roots of the characteristic equation
polyroot(c(1, -x.ar$ar))
```

```
## [1]  1.098508+0.905915i  1.098508-0.905915i -3.282012+0.000000i
```

```
# Calculate the absolute values of the roots which are
# complex numbers
sqrt(1.098508^2 + 0.905915^2)
```

```
## [1] 1.423869
```

```
sqrt(1.098508^2 + (-0.905915)^2)
```

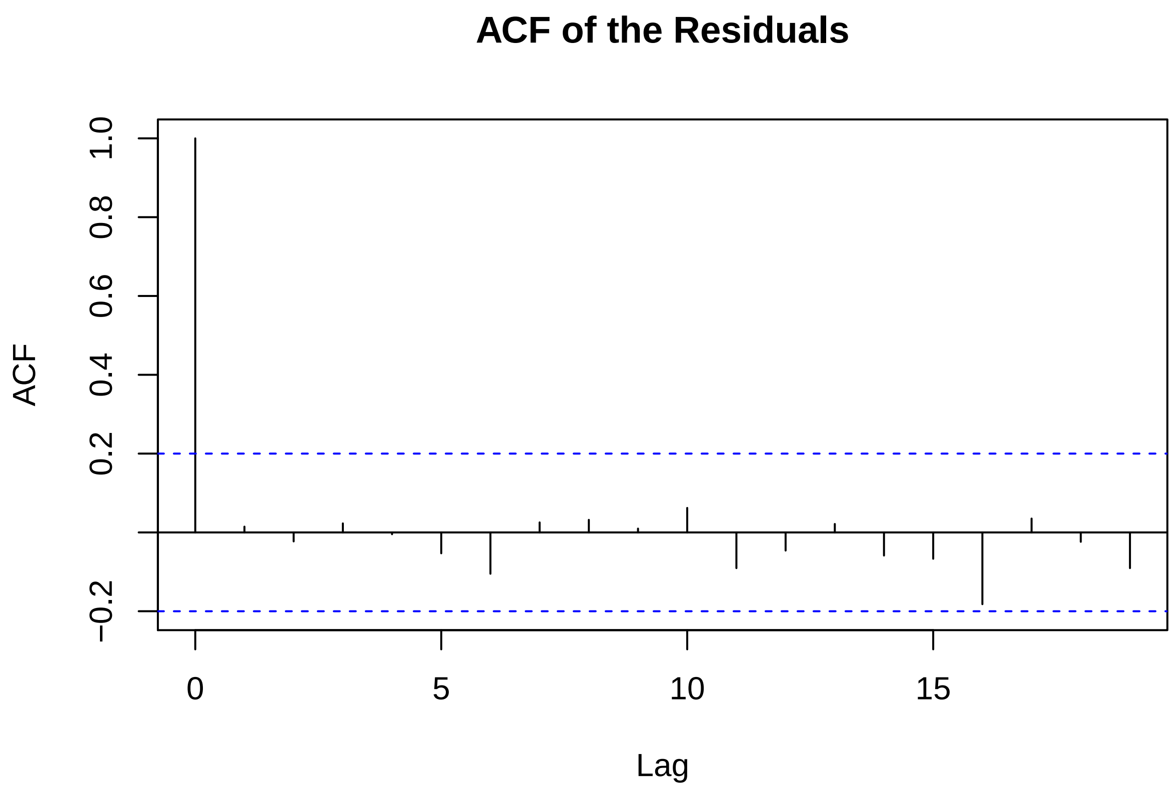
```
## [1] 1.423869
```

2.6 Plot the correlogram of the residuals of the estimated model. Comment on the plot.

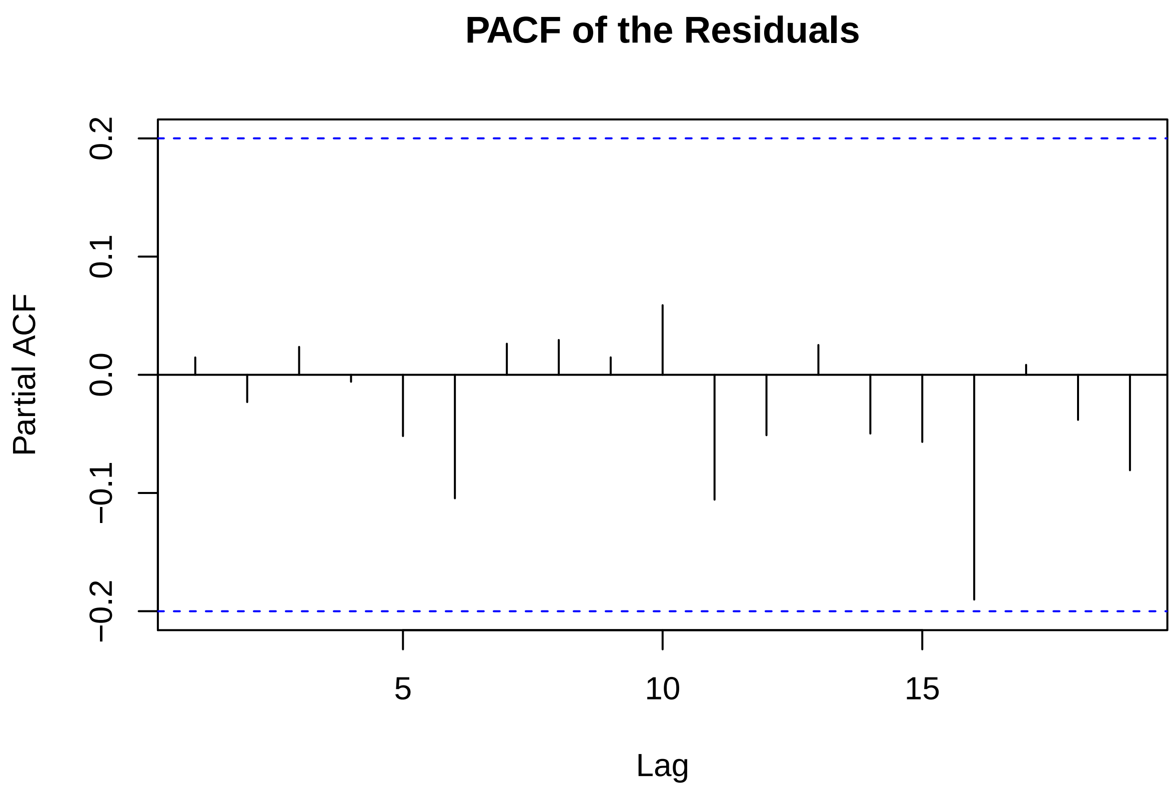
The autocorrelation has the pattern of a white noise series with significant correlation for leg 0 and non-significant for all other legs. The partial autp-correlation confirms the pattern of the acf with non-significant pacf for all legs greater than 0. The pattern is consistent with a white noise time series.

To confim the observed pattern of a white noise series for residuals, we plot a qqplot and qqline for the residuals. Both the qqplot and qqline indicate close simularity with a normal distribution, confirming the white noise pattern

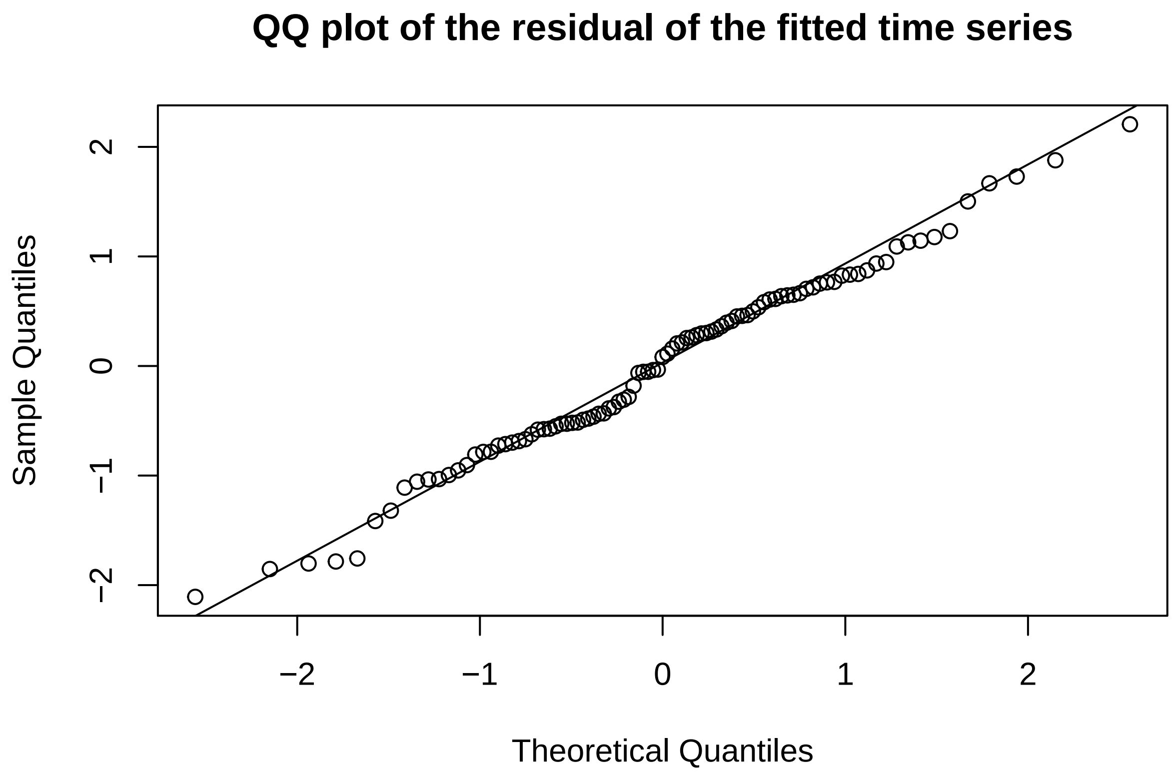
```
par(mfrow = c(1, 1))
acf(x.ar$resid[-c(1:4)], main = "ACF of the Residuals")
```




```r
pacf(x.ar$resid[-c(1:4)], main = "PACF of the Residuals")
```




```r
par(mfrow = c(1, 1))
qqnorm(x.ar$resid[-c(1:5)], main = "QQ plot of the residual of the fitted time series")
qqline(x.ar$resid[-c(1:5)], main = "QQ line of the residual of the fitted time series")
```


QQ plot of the residual of the fitted time series
Sample Quantiles
2
1
0
−1
−2
−2
−1
0
1
2
Theoretical Quantiles