STAT 221 Project

John Rapp Farnes | 405461225

3/5 2020

```r
library(astsa)

data <- read.csv("beer.csv")

data <- data.frame(week = data[-1,0], volume = as.numeric(as.character(data[-1,1])))

nrow(data)
```

```
## [1] 261
```

```r
data
```

```
##            volume
## 2015-03-01     64
## 2015-03-08     68
## 2015-03-15     76
## 2015-03-22     66
## 2015-03-29     68
## 2015-04-05     76
## 2015-04-12     67
## 2015-04-19     69
## 2015-04-26     68
## 2015-05-03     69
## 2015-05-10     69
## 2015-05-17     71
## 2015-05-24     78
## 2015-05-31     74
## 2015-06-07     78
## 2015-06-14     79
## 2015-06-21     81
## 2015-06-28     87
## 2015-07-05     78
## 2015-07-12     81
## 2015-07-19     86
## 2015-07-26     83
## 2015-08-02     86
## 2015-08-09     91
## 2015-08-16     83
## 2015-08-23     78
## 2015-08-30     76
## 2015-09-06     79
## 2015-09-13     75
## 2015-09-20     75
## 2015-09-27     72
## 2015-10-04     71
## 2015-10-11     71
```

```
## 2015-10-18       69
## 2015-10-25       66
## 2015-11-01       63
## 2015-11-08       64
## 2015-11-15       63
## 2015-11-22       72
## 2015-11-29       63
## 2015-12-06       66
## 2015-12-13       69
## 2015-12-20       80
## 2015-12-27       78
## 2016-01-03       62
## 2016-01-10       63
## 2016-01-17       69
## 2016-01-24       68
## 2016-01-31       67
## 2016-02-07       72
## 2016-02-14       70
## 2016-02-21       69
## 2016-02-28       67
## 2016-03-06       69
## 2016-03-13       81
## 2016-03-20       75
## 2016-03-27       73
## 2016-04-03       85
## 2016-04-10       74
## 2016-04-17       72
## 2016-04-24       72
## 2016-05-01       73
## 2016-05-08       75
## 2016-05-15       80
## 2016-05-22       82
## 2016-05-29       88
## 2016-06-05       84
## 2016-06-12       83
## 2016-06-19       84
## 2016-06-26       85
## 2016-07-03       88
## 2016-07-10       80
## 2016-07-17       81
## 2016-07-24       81
## 2016-07-31       82
## 2016-08-07       74
## 2016-08-14       74
## 2016-08-21       75
## 2016-08-28       75
## 2016-09-04       77
## 2016-09-11       75
## 2016-09-18       74
## 2016-09-25       74
## 2016-10-02       74
## 2016-10-09       71
## 2016-10-16       69
## 2016-10-23       70
```

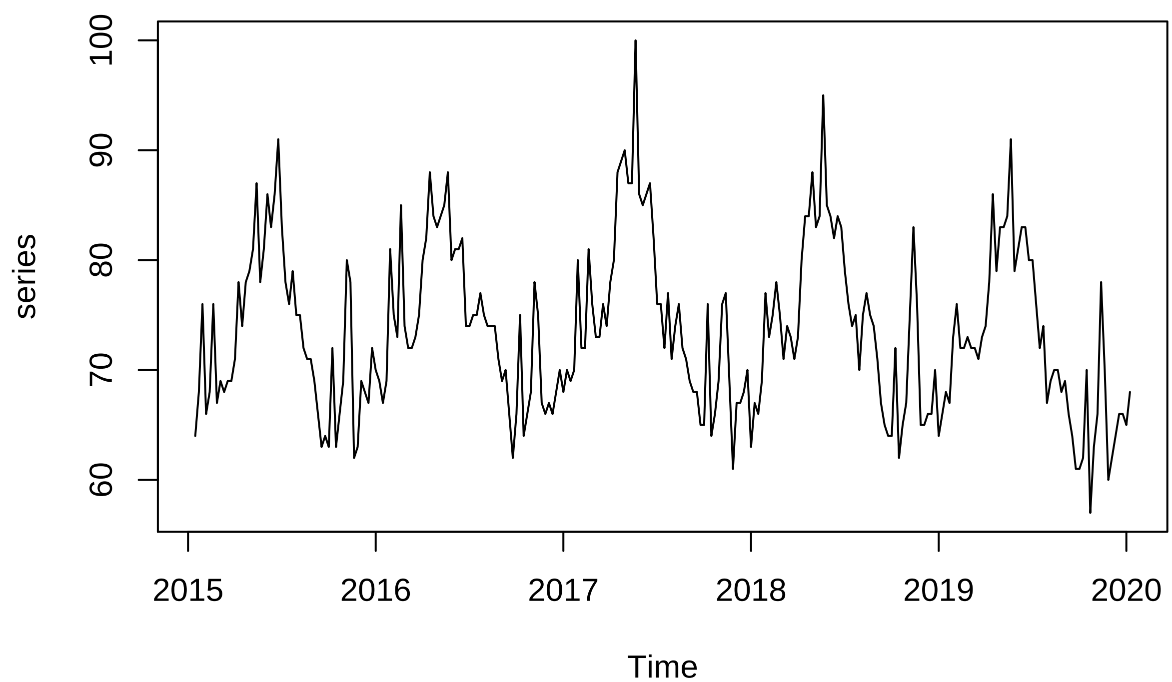
```
## 2016-10-30      66
## 2016-11-06      62
## 2016-11-13      66
## 2016-11-20      75
## 2016-11-27      64
## 2016-12-04      66
## 2016-12-11      68
## 2016-12-18      78
## 2016-12-25      75
## 2017-01-01      67
## 2017-01-08      66
## 2017-01-15      67
## 2017-01-22      66
## 2017-01-29      68
## 2017-02-05      70
## 2017-02-12      68
## 2017-02-19      70
## 2017-02-26      69
## 2017-03-05      70
## 2017-03-12      80
## 2017-03-19      72
## 2017-03-26      72
## 2017-04-02      81
## 2017-04-09      76
## 2017-04-16      73
## 2017-04-23      73
## 2017-04-30      76
## 2017-05-07      74
## 2017-05-14      78
## 2017-05-21      80
## 2017-05-28      88
## 2017-06-04      89
## 2017-06-11      90
## 2017-06-18      87
## 2017-06-25      87
## 2017-07-02     100
## 2017-07-09      86
## 2017-07-16      85
## 2017-07-23      86
## 2017-07-30      87
## 2017-08-06      82
## 2017-08-13      76
## 2017-08-20      76
## 2017-08-27      72
## 2017-09-03      77
## 2017-09-10      71
## 2017-09-17      74
## 2017-09-24      76
## 2017-10-01      72
## 2017-10-08      71
## 2017-10-15      69
## 2017-10-22      68
## 2017-10-29      68
## 2017-11-05      65
```

```
## 2017-11-12      65
## 2017-11-19      76
## 2017-11-26      64
## 2017-12-03      66
## 2017-12-10      69
## 2017-12-17      76
## 2017-12-24      77
## 2017-12-31      69
## 2018-01-07      61
## 2018-01-14      67
## 2018-01-21      67
## 2018-01-28      68
## 2018-02-04      70
## 2018-02-11      63
## 2018-02-18      67
## 2018-02-25      66
## 2018-03-04      69
## 2018-03-11      77
## 2018-03-18      73
## 2018-03-25      75
## 2018-04-01      78
## 2018-04-08      75
## 2018-04-15      71
## 2018-04-22      74
## 2018-04-29      73
## 2018-05-06      71
## 2018-05-13      73
## 2018-05-20      80
## 2018-05-27      84
## 2018-06-03      84
## 2018-06-10      88
## 2018-06-17      83
## 2018-06-24      84
## 2018-07-01      95
## 2018-07-08      85
## 2018-07-15      84
## 2018-07-22      82
## 2018-07-29      84
## 2018-08-05      83
## 2018-08-12      79
## 2018-08-19      76
## 2018-08-26      74
## 2018-09-02      75
## 2018-09-09      70
## 2018-09-16      75
## 2018-09-23      77
## 2018-09-30      75
## 2018-10-07      74
## 2018-10-14      71
## 2018-10-21      67
## 2018-10-28      65
## 2018-11-04      64
## 2018-11-11      64
## 2018-11-18      72
```

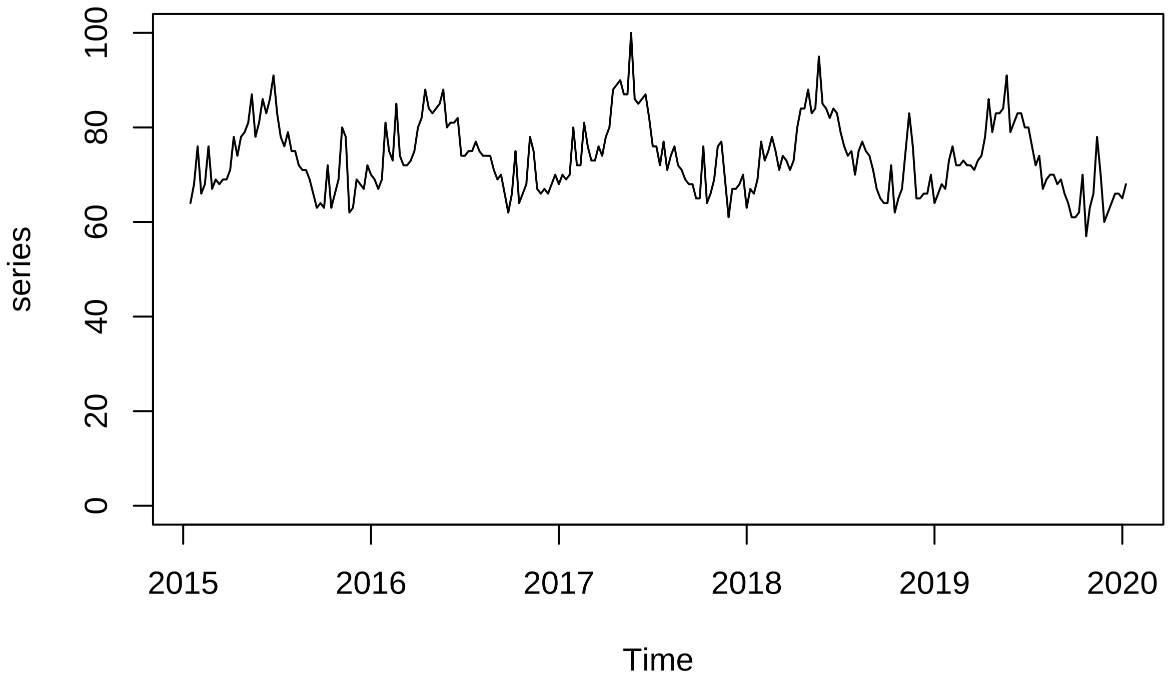
```
## 2018-11-25      62
## 2018-12-02      65
## 2018-12-09      67
## 2018-12-16      75
## 2018-12-23      83
## 2018-12-30      76
## 2019-01-06      65
## 2019-01-13      65
## 2019-01-20      66
## 2019-01-27      66
## 2019-02-03      70
## 2019-02-10      64
## 2019-02-17      66
## 2019-02-24      68
## 2019-03-03      67
## 2019-03-10      73
## 2019-03-17      76
## 2019-03-24      72
## 2019-03-31      72
## 2019-04-07      73
## 2019-04-14      72
## 2019-04-21      72
## 2019-04-28      71
## 2019-05-05      73
## 2019-05-12      74
## 2019-05-19      78
## 2019-05-26      86
## 2019-06-02      79
## 2019-06-09      83
## 2019-06-16      83
## 2019-06-23      84
## 2019-06-30      91
## 2019-07-07      79
## 2019-07-14      81
## 2019-07-21      83
## 2019-07-28      83
## 2019-08-04      80
## 2019-08-11      80
## 2019-08-18      76
## 2019-08-25      72
## 2019-09-01      74
## 2019-09-08      67
## 2019-09-15      69
## 2019-09-22      70
## 2019-09-29      70
## 2019-10-06      68
## 2019-10-13      69
## 2019-10-20      66
## 2019-10-27      64
## 2019-11-03      61
## 2019-11-10      61
## 2019-11-17      62
## 2019-11-24      70
## 2019-12-01      57
```

```
## 2019-12-08     63
## 2019-12-15     66
## 2019-12-22     78
## 2019-12-29     70
## 2020-01-05     60
## 2020-01-12     62
## 2020-01-19     64
## 2020-01-26     66
## 2020-02-02     66
## 2020-02-09     65
## 2020-02-16     68
## 2020-02-23     87
```

```r
series <- ts(data$volume, start=c(2015, 3, 1), end=c(2020, 2, 23), frequency=52)

plot(series)
```

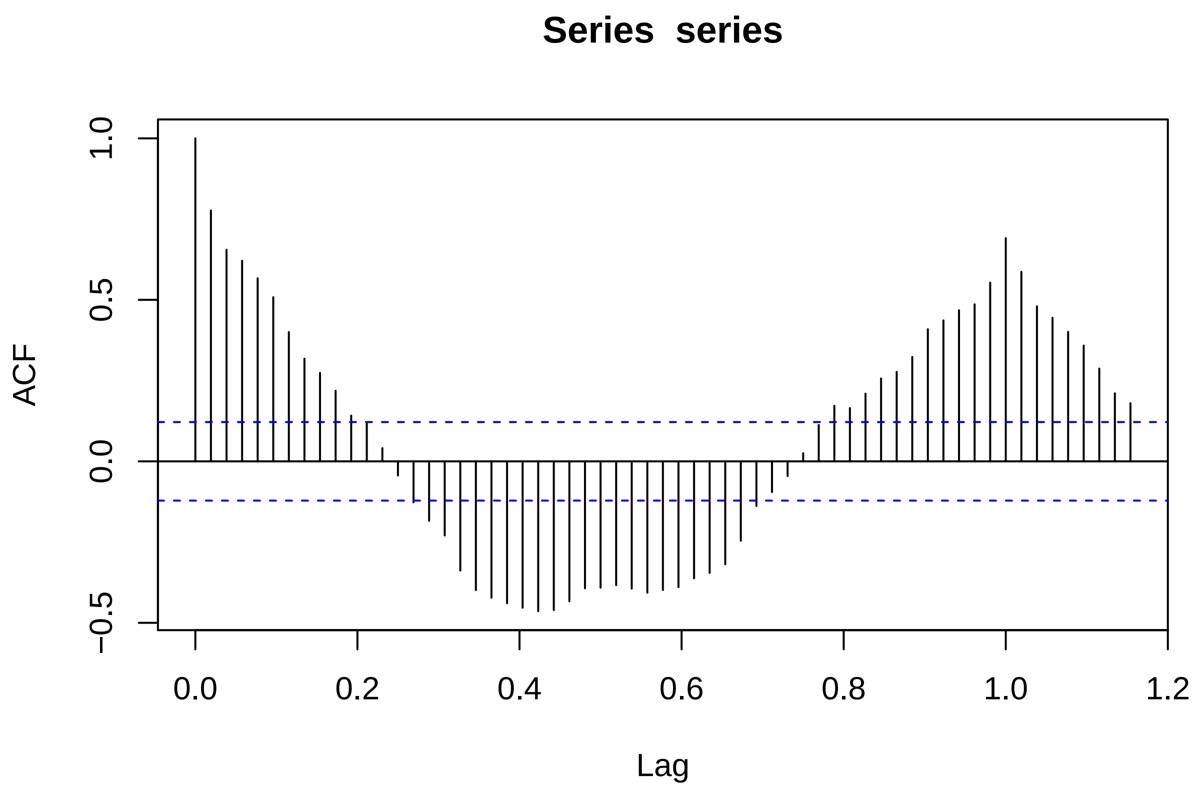



```r
plot(series, ylim=c(0, 100))
```

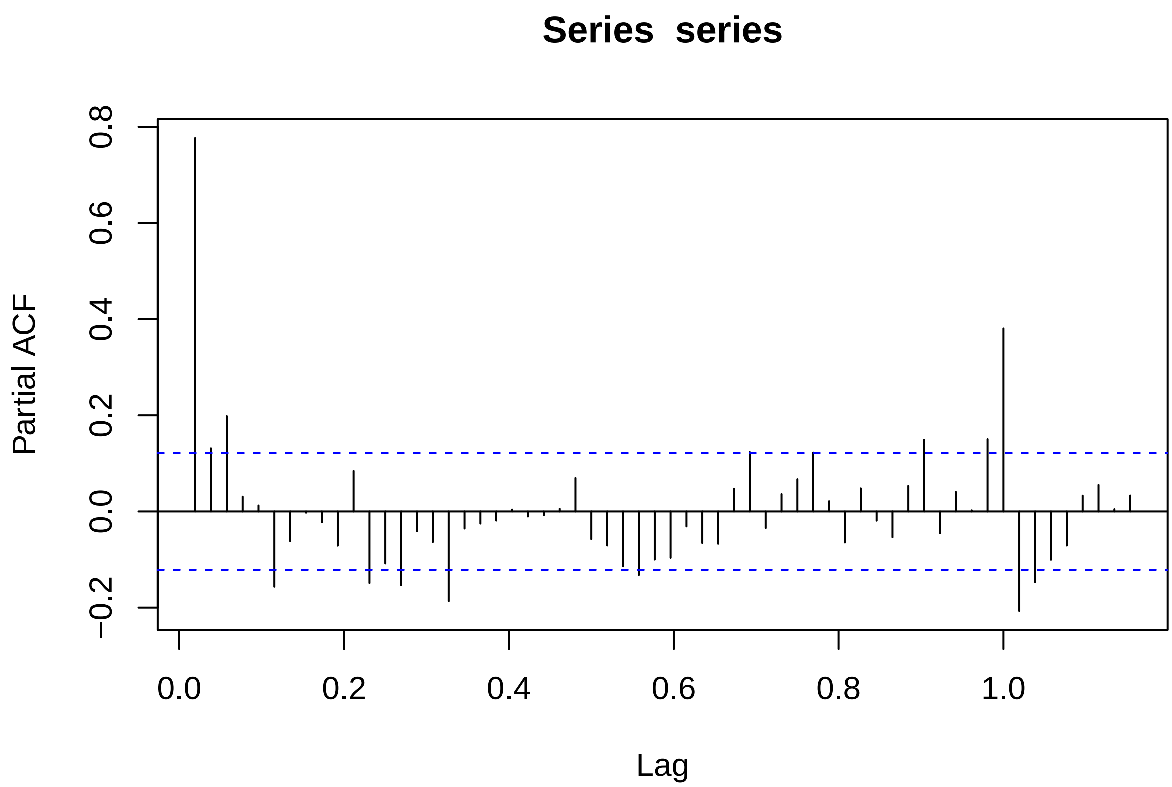


ACF

```
acf(series, lag.max = 60)
```




```
pacf(series, lag.max = 60)
```

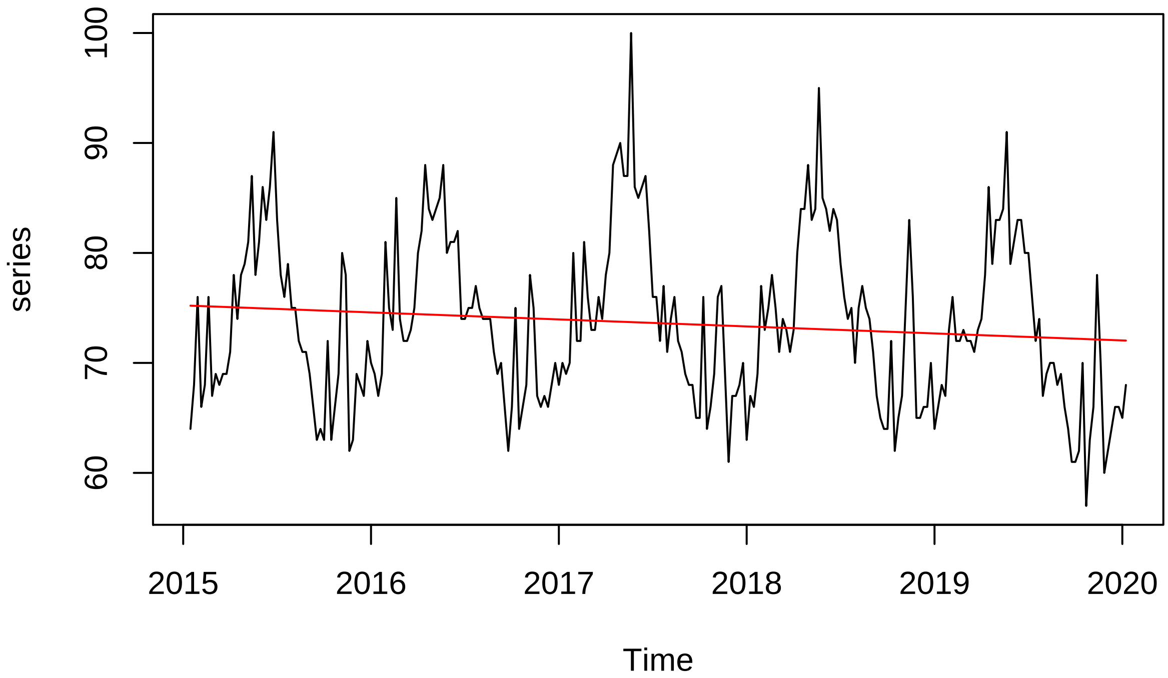


Detrend

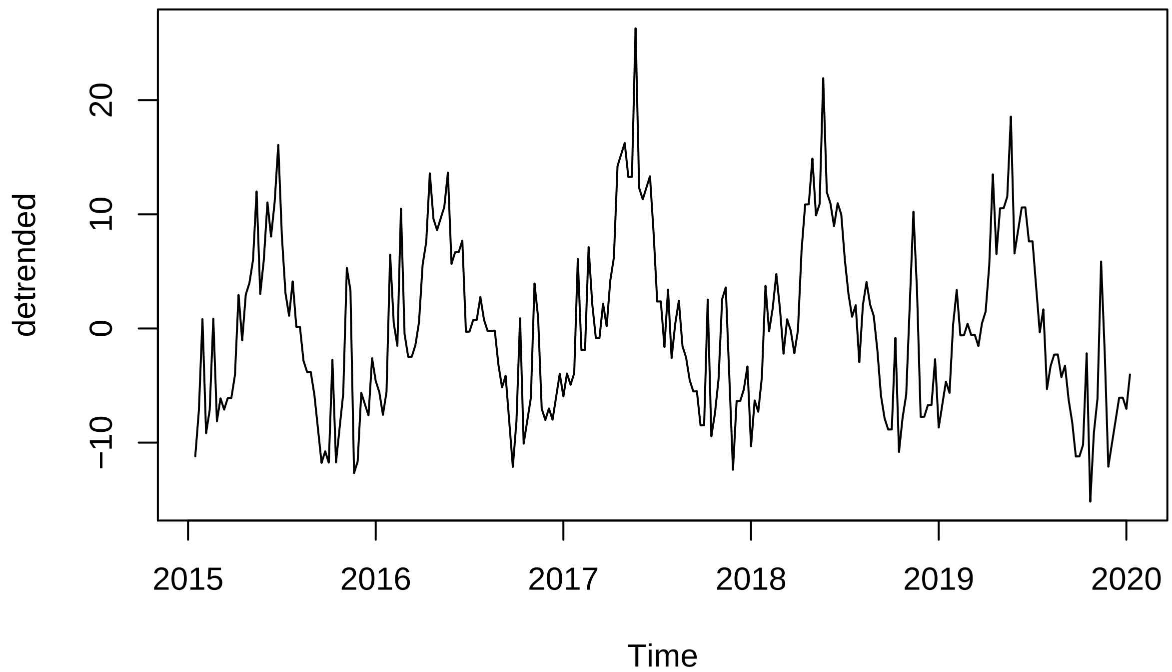
```
model <- lm(series ~ time(series))
summary(model)
```

```
##
## Call:
## lm(formula = series ~ time(series))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.1675  -5.9998  -0.5656   4.9022  26.2889
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1358.8996   644.9614   2.107   0.0361 *
## time(series)   -0.6371     0.3197  -1.993   0.0473 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.44 on 258 degrees of freedom
## Multiple R-squared:  0.01516,    Adjusted R-squared:  0.01134
## F-statistic: 3.971 on 1 and 258 DF,  p-value: 0.04734
```

```r
plot(series)
lines(series*0 + predict(model, time(series)), type="l", col="red")
```




```r
detrended <- series - predict(model, time(series))

plot(detrended)
```

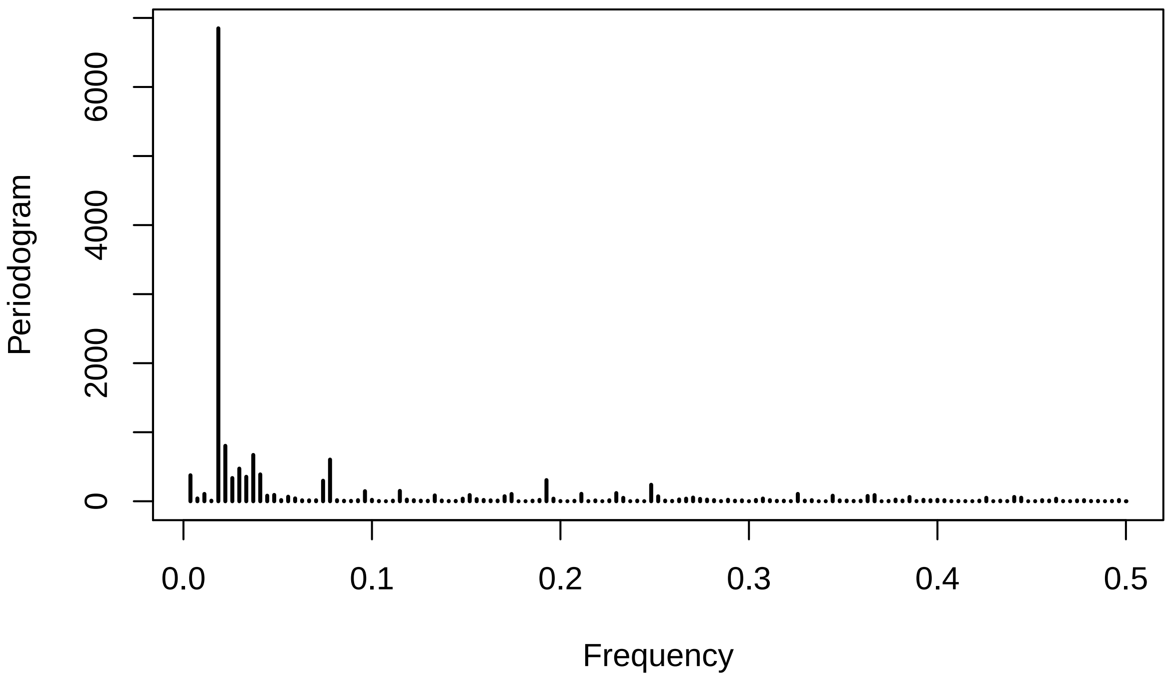


Periodogram

```r
library(TSA)
```

```
##
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':
##
##     acf, arima

## The following object is masked from 'package:utils':
##
##     tar
```

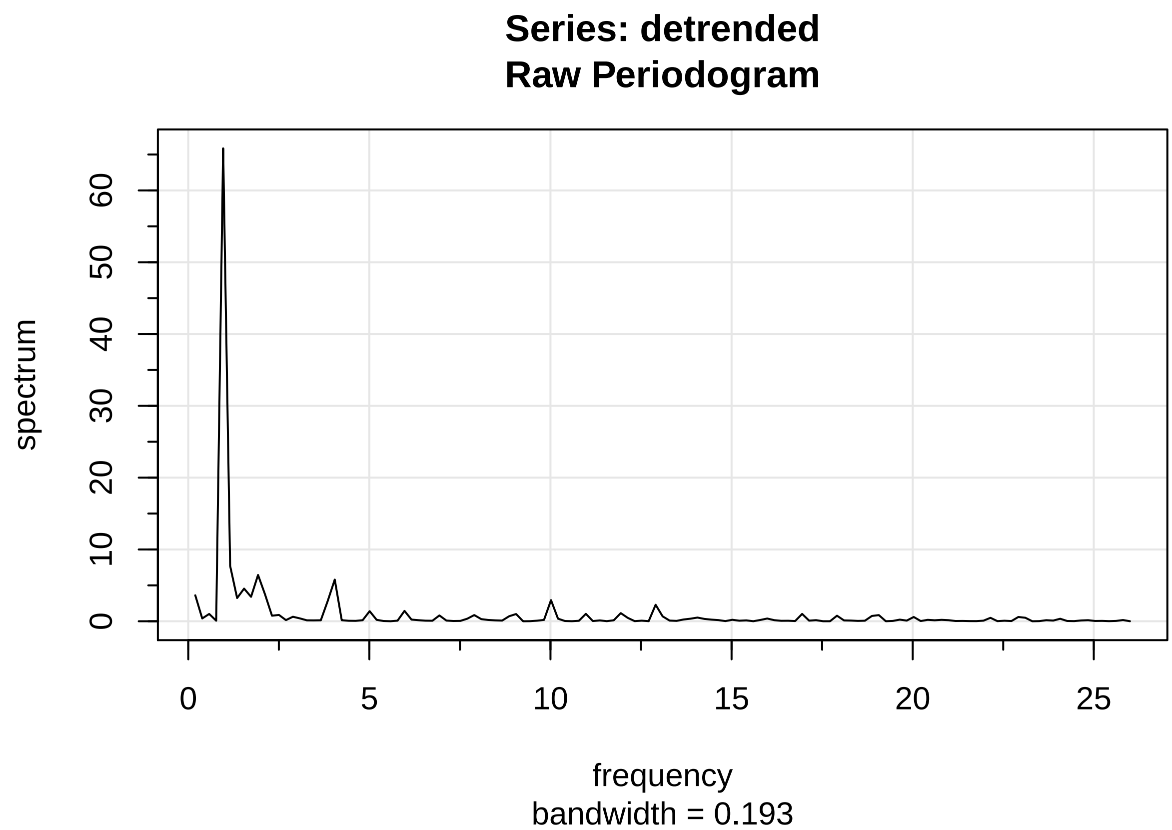
```r
pg <- periodogram(detrended)
```




```
max_index <- which(pg$spec == max(pg$spec))
freq <- pg$freq[max_index]

1/freq
```

```
## [1] 54
```

```
pg <- mvspec(detrended, log="no")
```

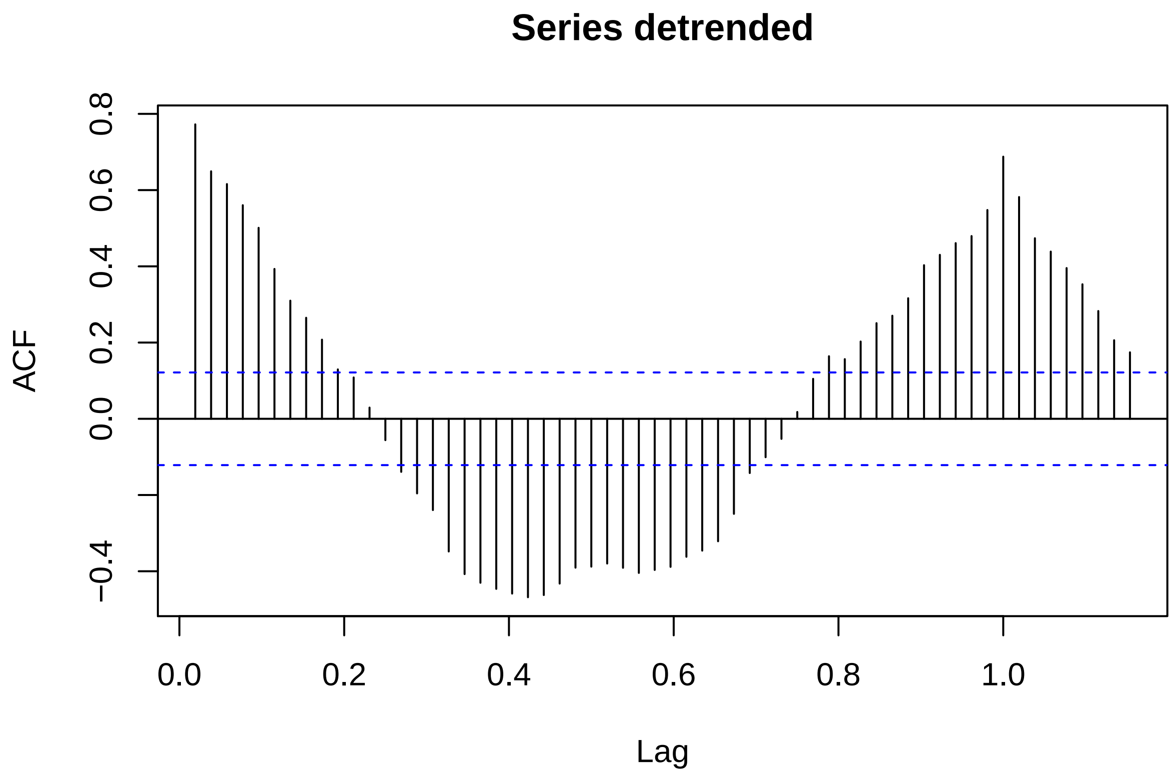



```
max_index <- which(pg$spec == max(pg$spec))
freq <- pg$freq[max_index]

1/freq
```

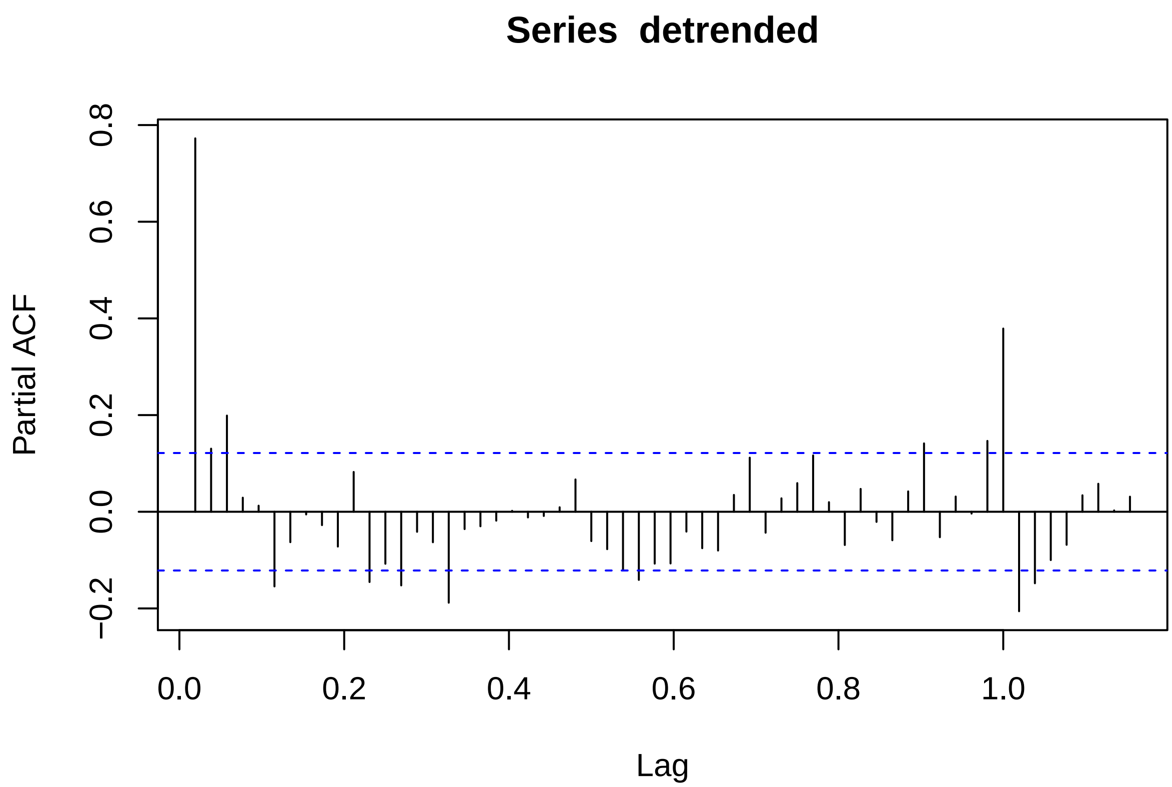
```
## [1] 1.038462
```

ACF

```
acf(detrended, lag.max = 60)
```




```
pacf(detrended, lag.max = 60)
```



Fit ARIMA

Show residuals and measure of how well fit

Comment on fitted model

Estimate the spectral density using the periodogram and smoothed periodogram.

Comment on any clear cycles and the overall distribution of the variance by frequency, and its relationship to the smoothness of your time series.

2.9

a)

```r
library(astsa)

model <- lm(soi ~ time(soi))

(sum <- summary(model))
```

```
##
## Call:
## lm(formula = soi ~ time(soi))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04140 -0.24183  0.01935  0.27727  0.83866
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.70367    3.18873   4.298 2.12e-05 ***
## time(soi)   -0.00692    0.00162  -4.272 2.36e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3756 on 451 degrees of freedom
## Multiple R-squared:  0.0389, Adjusted R-squared:  0.03677
## F-statistic: 18.25 on 1 and 451 DF,  p-value: 2.359e-05
```

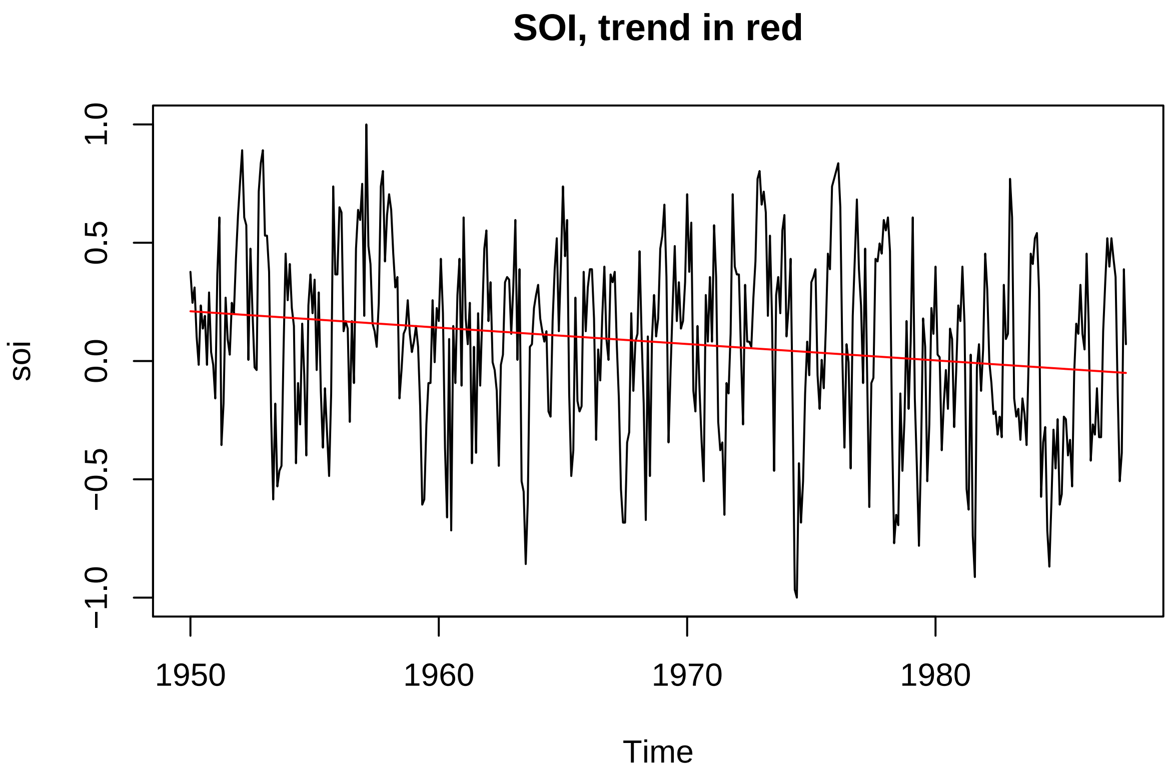
```r
sum$coefficients[2]
```

```
## [1] -0.006919645
```

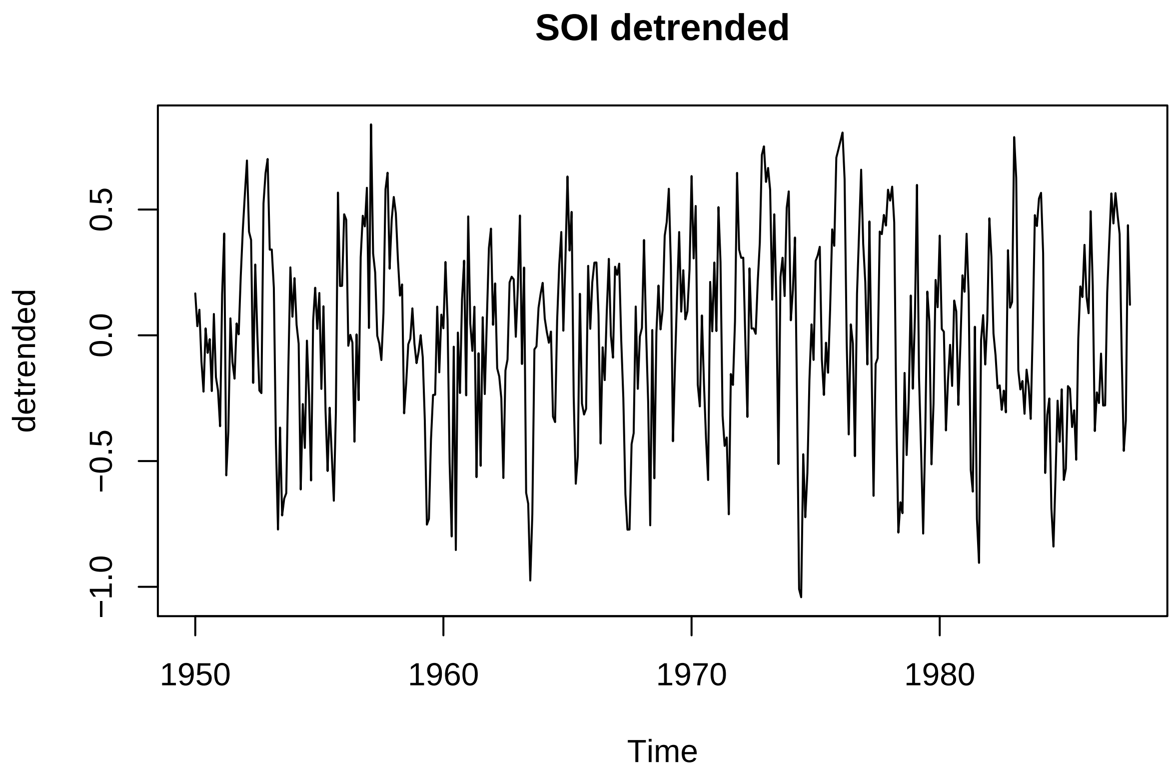
There is a significant trend in SOI in relation to time, where the SOI decreases ~0.007 every year.

b)

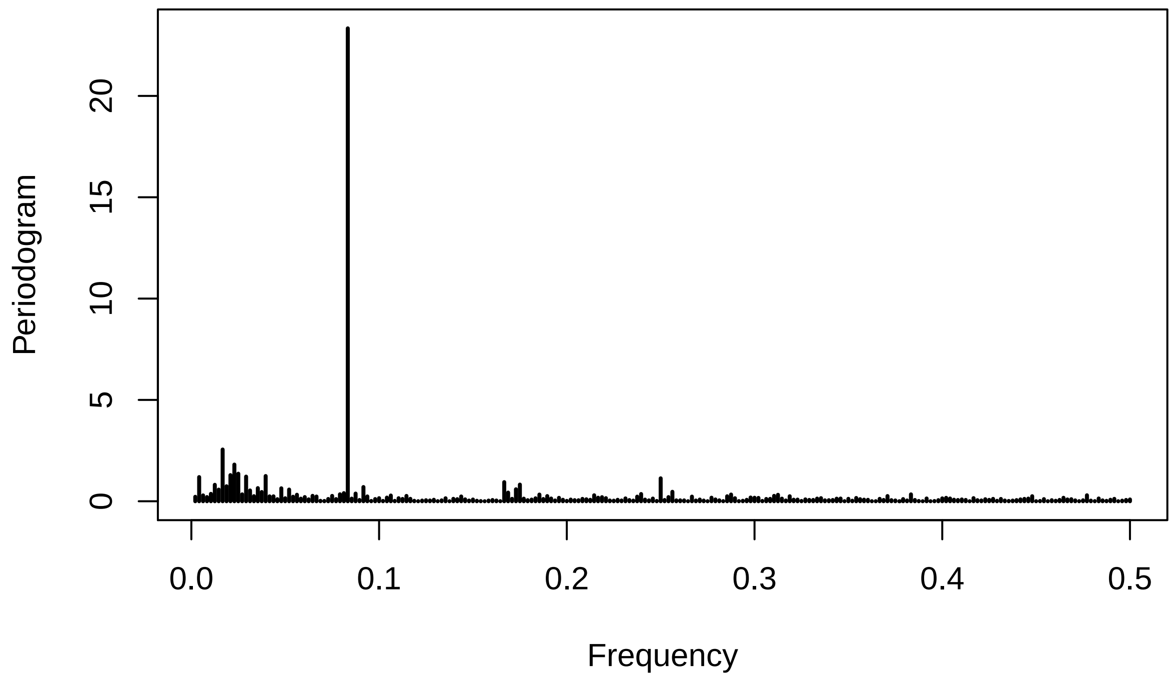
```r
plot(soi, main="SOI, trend in red")
lines(soi*0 + predict(model, time(soi)), type="l", col="red")
```




```r
detrended <- soi - predict(model, time(soi))

plot(detrended, main="SOI detrended")
```




```
library(TSA)
pg <- periodogram(detrended)
```



Two peaks, first at period:

```
max_index <- which(pg$spec == max(pg$spec))
freq <- pg$freq[max_index]

1/freq
```

```
## [1] 12
```

Or every year. Second at period:

```
freq <- pg$freq[which(pg$spec[1:(max_index - 1)] == max(pg$spec[1:(max_index - 1)]))]

1/freq
```

```
## [1] 60
```

Or every five years, which is the probable El Nino cycle.

3.2

a)

By expanding “backwards” in time, we have

$$
\begin{aligned}
x_t &= \phi x_{t-1} + w_t \\
&= \phi(\phi x_{t-2} + w_{t-1}) + w_t \\
&= \phi^2 x_{t-2} + \phi w_{t-1} + w_t \\
&\vdots \\
&= \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}
\end{aligned}
$$

When $k = t$ we get

$$
\begin{aligned}
x_t &= \phi^t x_0 + \sum_{j=0}^{t-1} \phi^j w_{t-j} \\
&= \phi^t w_0 + \sum_{j=0}^{t-1} \phi^j w_{t-j} \\
&= \sum_{j=0}^{t} \phi^j w_{t-j}
\end{aligned}
\tag{1}
$$

b)

We have

$$
\begin{aligned}
E(x_t) &= E(\sum_{j=0}^{t} \phi^j w_{t-j}) \\
&= \sum_{j=0}^{t} \phi^j E(w_{t-j}) = 0
\end{aligned}
\tag{2}
$$

as $E(w_i) = 0 \, \forall i$

c)

As $Var(w_i) = \sigma_w^2 \, \forall i$, we have

$$
\begin{aligned}
Var(x_t) &= Var(\sum_{j=0}^{t} \phi^j w_{t-j}) \\
&= \sum_{j=0}^{t} (\phi^j)^2 Var(w_{t-j}) \\
&= \sigma_w^2 \sum_{j=0}^{t} \phi^{2j} \\
&= \sigma_w^2 \frac{1-\phi^{2(t+1)}}{1-\phi^2}
\end{aligned} \tag{3}
$$

by geometric series formula, as $|\phi| < 1 \implies |\phi^2| < 1$.

d)

$$
\begin{aligned}
cov(x_{t+h}, x_t) &= E[(\sum_{j=0}^{t+h} \phi^j w_{t+h-j})(\sum_{j=0}^{t} \phi^j w_{t-j})] \\
&= E[(w_{t+h} + \cdots + \phi^h w_t + \phi^{h+1} w_{t-1} + \cdots + \phi^{h+t} w_0)(w_t + \phi w_{t-1} + \cdots + \phi^t w_0]) \\
&= \sigma_w^2 \sum_{j=0}^{t} \phi^{h+2j} \\
&= \sigma_w^2 \phi^h \sum_{j=0}^{t} \phi^{2j} \\
&= \phi^h \sigma_w^2 \frac{1-\phi^{2(t+1)}}{1-\phi^2} \\
&= \phi^h Var(x_t)
\end{aligned} \tag{4}
$$

e) We have $E(x_t) = 0$ constant, however $\gamma(h) = \phi^h Var(x_t) = \phi^h \sigma_w^2 \frac{1-\phi^{2(t+1)}}{1-\phi^2}$ depends on time. Hence, the series is not stationary.

f) We have $t \to \infty \implies \gamma(h) \to \phi^h \sigma_w^2 \frac{1}{1-\phi^2}$ depends only on h and not on t, hence it is "asymptotically stationary."

g) As we have just proved that this process estimates an AR(1) process, a AR(1) process could be simulated by simulating n i.i.d $N(0,1)$ noise values as $w_t, t = 1..n$ and then calculating x_t by

$$
\begin{cases}
x_t = \phi x_{t-1} + w_t, t = 1..n \ldots \\
x_0 = w_0
\end{cases} \tag{5}
$$

h) We have, with $k = t$,

$$
\begin{aligned}
x_t &= \phi^t x_0 + \sum_{j=0}^{t-1} \phi^j w_{t-j} \\
&= \phi^t w_0/\sqrt{1-\phi^2} + \sum_{j=0}^{t-1} \phi^j w_{t-j} \implies \\
Var(x_t) &= \frac{(\phi^t)^2}{1-\phi^2}\sigma_w^2 + \sigma_w^2 \frac{1-\phi^{2t}}{1-\phi^2} \\
&= \frac{\sigma_w^2}{1-\phi^2}(\phi^{2t} + 1 - \phi^{2t}) \\
&= \frac{\sigma_w^2}{1-\phi^2}
\end{aligned} \tag{6}
$$

So $Var(x_t)$ is constant, and $E(x_t)$ is still constant, hence the series is stationary.

3.6

We have

$$
\begin{aligned}
x_t = -.9x_{t-2} + w_t \implies \\
x_t + .9x_{t-2} = w_t \implies \\
(1 + 0.9B^2)x_t = w_t
\end{aligned} \tag{7}
$$

So $\phi(z) = 1 + 0.9z^2$, which has the roots z_i:

```
(z <- polyroot(c(1,0,0.9)))
```

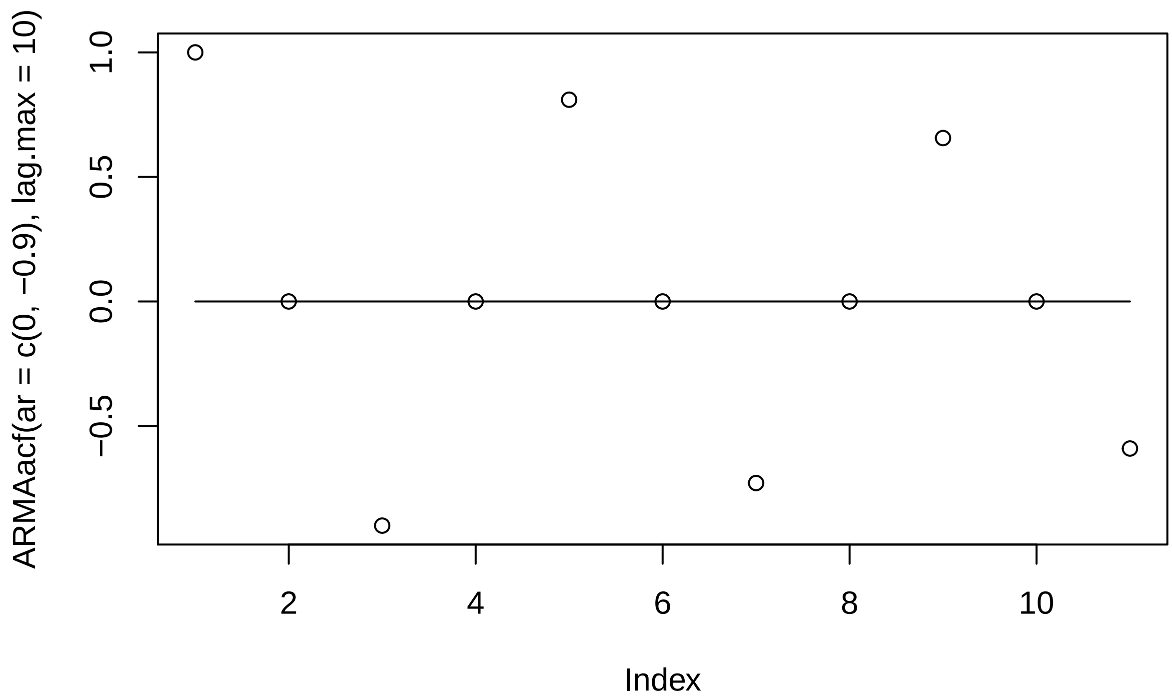
```
## [1] 0+1.054093i 0-1.054093i
```

```
abs(polyroot(c(1,0,0.9)))
```

```
## [1] 1.054093 1.054093
```

Hence $|z_i| > 1 \, \forall i$ and z_i complex conjugate, as such the acf will have periodic behavior, see plot.

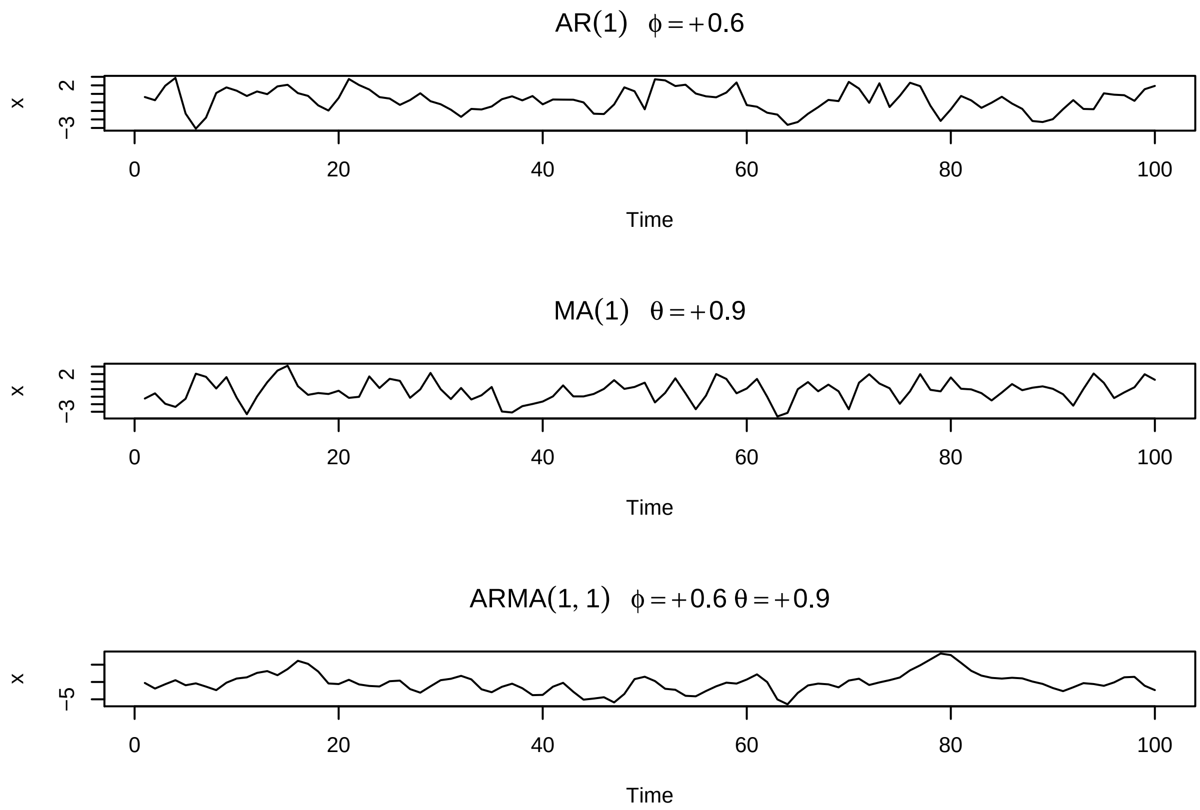
```
plot(ARMAacf(ar = c(0, -0.9), lag.max = 10))
lines(rep(0, 11))
```



3.9

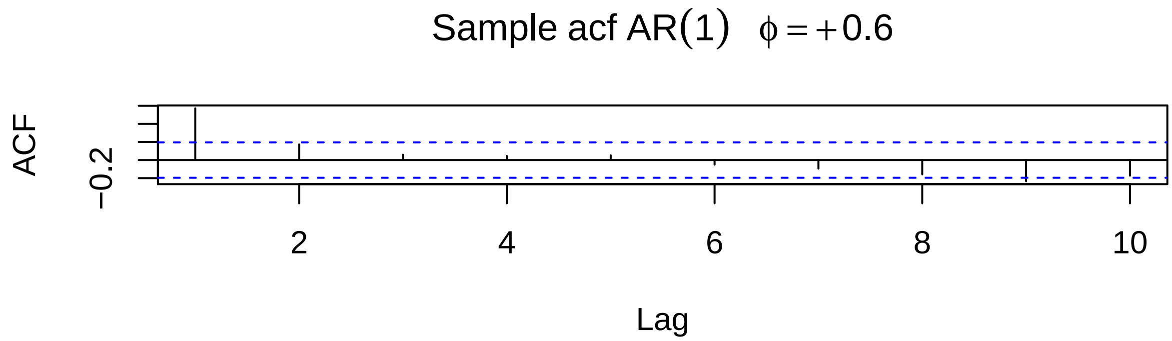
First we'll plot the simulated series

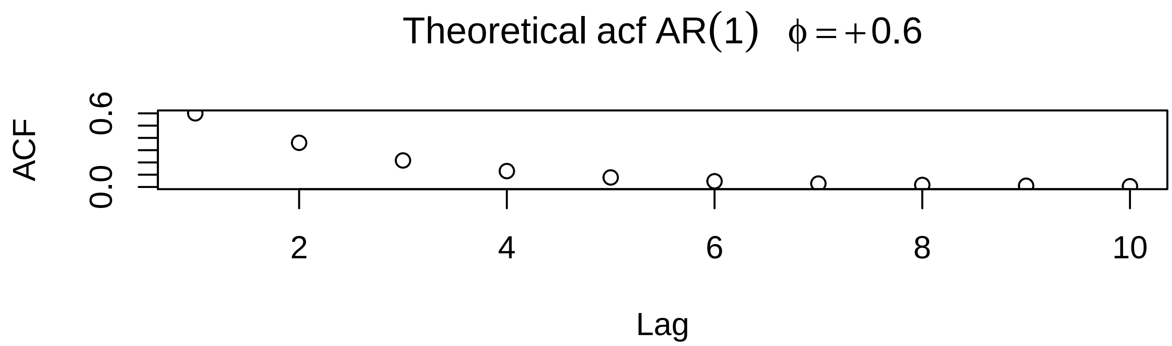
```
set.seed(2020)
ar <- arima.sim(list(order=c(1,0,0), ar=.6), n=100)
ma <- arima.sim(list(order=c(0,0,1), ma=.9), n=100)
arma <- arima.sim(list(order=c(1,0,1), ar=.6, ma=.9), n=100)

par(mfrow=c(3, 1))
plot(ar, ylab="x", main=(expression(AR(1)~~~phi==+.6)))
plot(ma, ylab="x", main=(expression(MA(1)~~~theta==+.9)))
plot(arma, ylab="x", main=(expression(ARMA(1, 1)~~~phi==+.6~theta==+.9)))
```



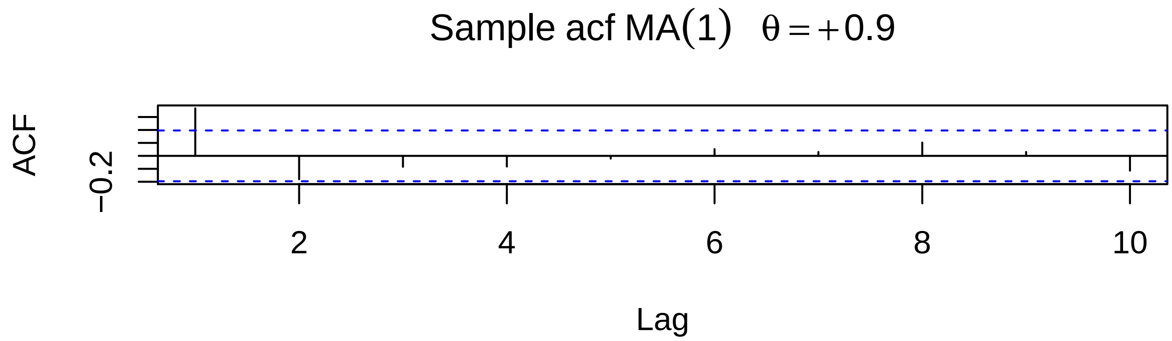
Next, we'll plot and compare the sample acf with the theoretical acf.

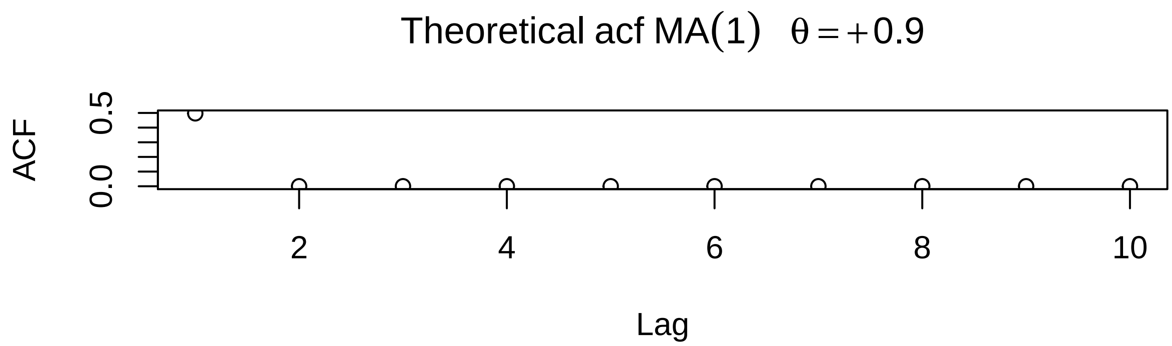
```
par(mfrow=c(2, 1))
plot(acf(ar, lag.max = 10, plot=FALSE), main=(expression(Sample~acf~AR(1)~~~phi==+.6)))
plot(ARMAacf(ar = 0.6, lag.max = 10)[-1], main=(expression(Theoretical~acf~AR(1)~~~phi==+.6)), ylab = "
```



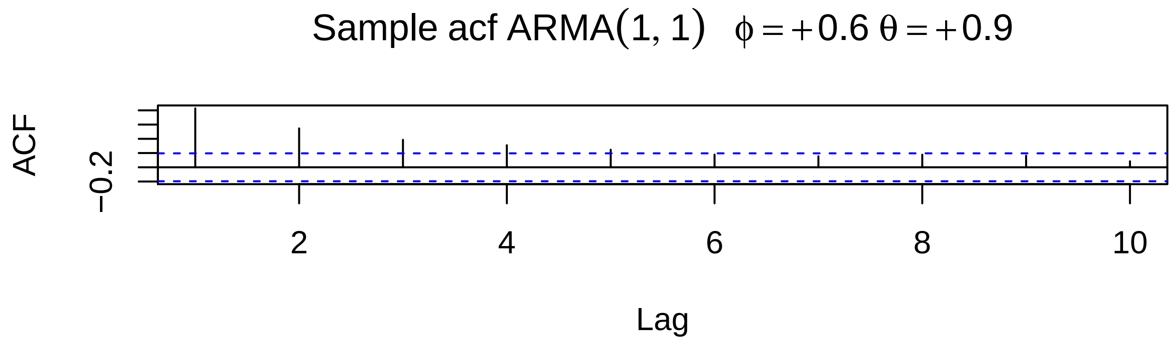


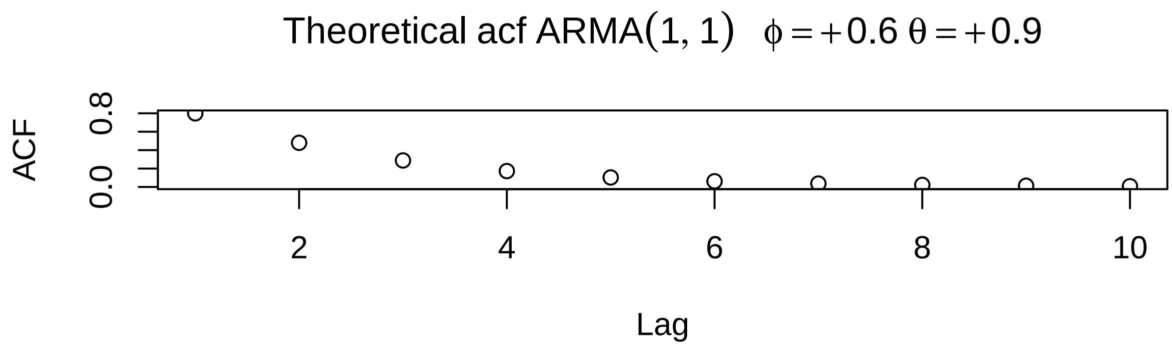

```
par(mfrow=c(2, 1))
plot(acf(ma, lag.max = 10, plot=FALSE), main=(expression(Sample~acf~MA(1)~~~theta==+.9)))
plot(ARMAacf(ma = 0.9, lag.max = 10)[-1], main=(expression(Theoretical~acf~MA(1)~~~theta==+.9)), ylab =
```





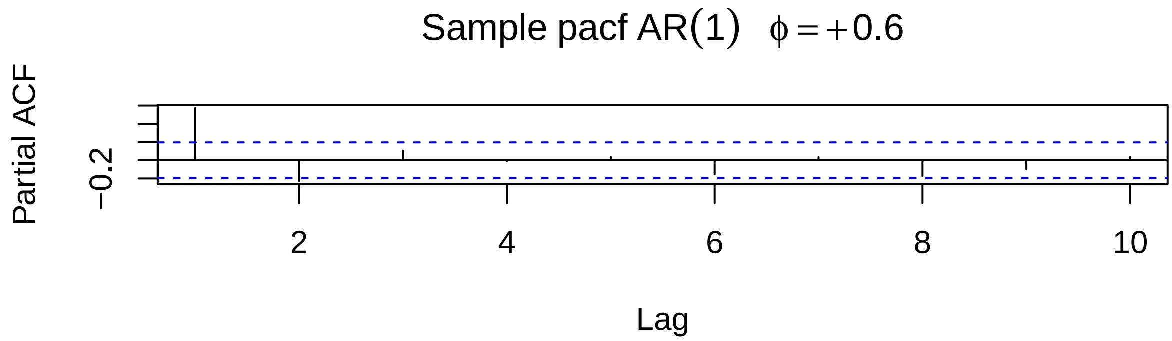

```
par(mfrow=c(2, 1))
plot(acf(arma, lag.max = 10, plot=FALSE), main=(expression(Sample~acf~ARMA(1, 1)~~~phi==+.6~theta==+.9)
plot(ARMAacf(ar = 0.6, ma = 0.9, lag.max = 10)[-1], main=(expression(Theoretical~acf~ARMA(1, 1)~~~phi==
```

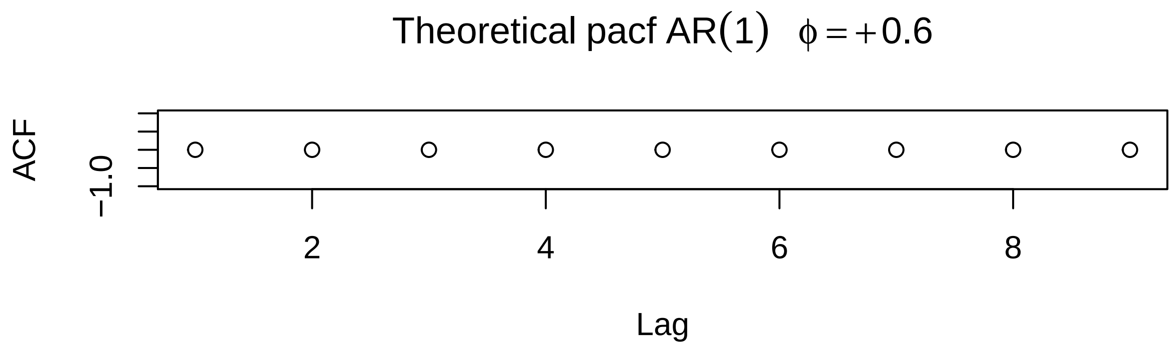




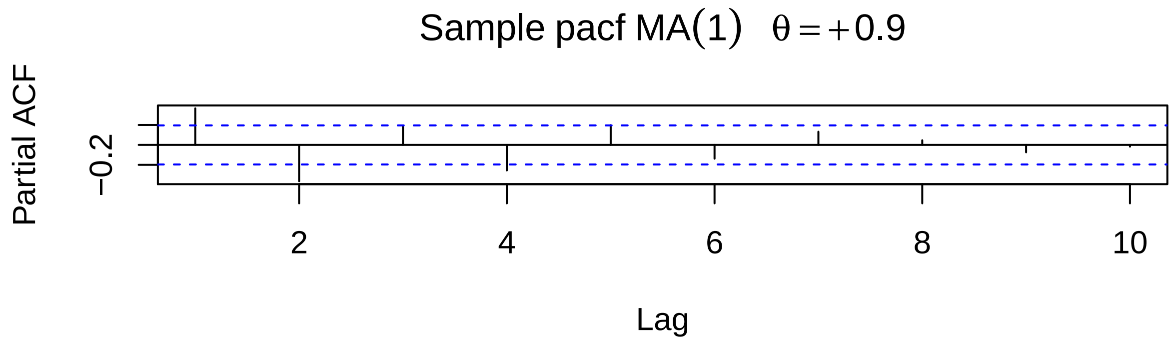
The plots are consistent with the general results of Table 3.1, as the AR(1) as well as ARMA(1,1) process tails off while the MA(1) cutts of after lag 1= q.

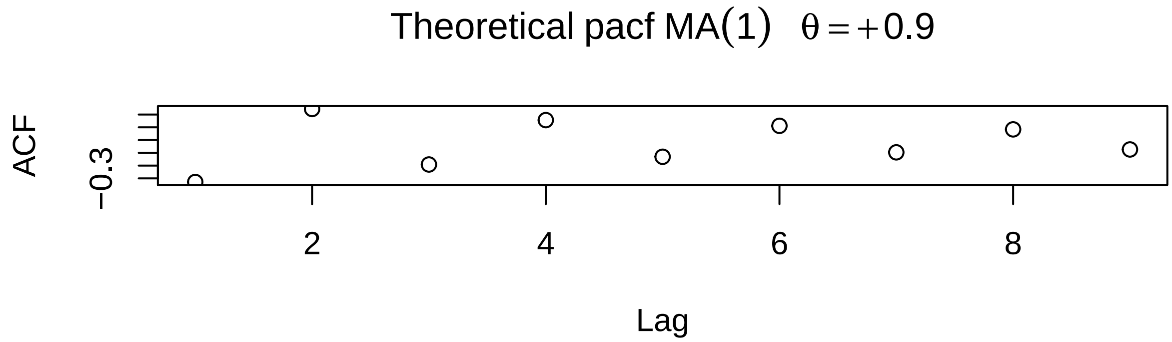
```
par(mfrow=c(2, 1))
plot(pacf(ar, lag.max = 10, plot=FALSE), main=(expression(Sample~pacf~AR(1)~~~phi==+.6)))
plot(ARMAacf(ar = 0.6, lag.max = 10, pacf = TRUE)[-1], main=(expression(Theoretical~pacf~AR(1)~~~phi==+
```



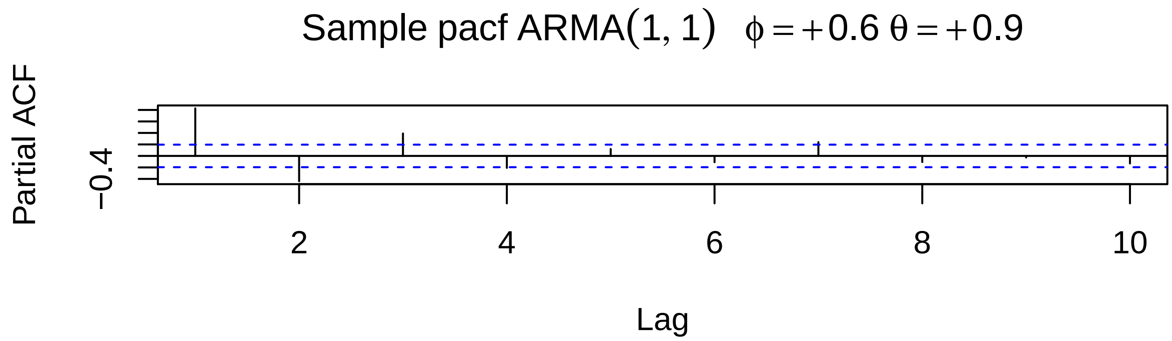


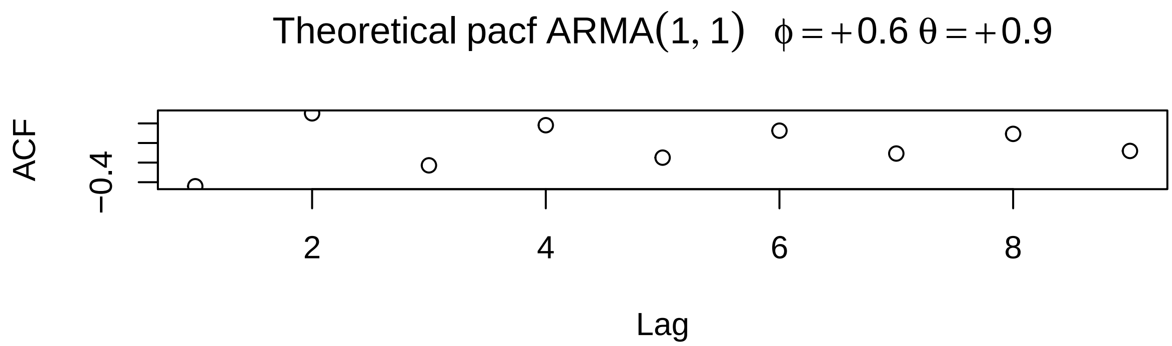

```
par(mfrow=c(2, 1))
plot(pacf(ma, lag.max = 10, plot=FALSE), main=(expression(Sample~pacf~MA(1)~~~theta==+.9)))
plot(ARMAacf(ma = 0.9, lag.max = 10, pacf = TRUE)[-1], main=(expression(Theoretical~pacf~MA(1)~~~theta=
```






```
par(mfrow=c(2, 1))
plot(pacf(arma, lag.max = 10, plot=FALSE), main=(expression(Sample~pacf~ARMA(1, 1)~~~phi==+.6~theta==+.
plot(ARMAacf(ar = 0.6, ma = 0.9, lag.max = 10, pacf = TRUE)[-1], main=(expression(Theoretical~pacf~ARMA
```





The plots are consistent with the general results of Table 3.1, as the MA(1) as well as ARMA(1,1) process tails off while the AR(1) cutts of after lag 1= q.