Tutorial Sheet 6 Solutions

Q6.1

A time series of 100 observations produced sample auto-correlation of $r_1 = -0.49$, $r_2 = 0.31$, $r_3 = -0.21$, $r_4 = 0.11$ and $|r_k| < 0.09$ for $k > 4$.

On this basis alone, what ARIMA model would we tentatively specify for the series?

Using $2/\sqrt{100} = 0.2$ as a guide, we might consider MA(2) or MA(3) as possibilities, because an MA(q) process enjoys the acf cut-off property at lag q.

r_1 and r_2 are definitely significant, but r_3 is borderline.

If we choose an MA model, then we calculate $Var(r_k)$ using the formula found on TSLecture6 Page 7:

$$Var(r_k) = \frac{1}{n}\left(1 + 2\sum_{j=1}^{q}\rho_j^2\right) \qquad \text{for } k > q.$$

For an observed time series, we can replace ρ's by r's.

If we tentatively chose an MA($q = 2$) model, then, for $k = 3$, i.e., $k > q$,

$$\begin{aligned} Var(r_3) &= \frac{1}{100}\left[1 + 2\{(-0.49)^2 + (0.31)^2\}\right] \\ &= 0.016724. \\ se(r_3) &= \sqrt{0.016724} \\ &\approx 0.129. \end{aligned}$$

Whether the r_3 coefficient is significant or not is tested via:

$$\begin{aligned} t_{r_3} &= \frac{r_3}{se(r_3)} \\ &= \frac{-0.21}{0.129} \\ &= -1.62. \end{aligned}$$

As $|t| < 2$, the value of r_3 is likely to have occurred by chance, i.e., it is not significant.

Hence we cannot reject MA(2).

Alternatively, if the interval of $r_3 \pm 2\text{se}(r_3)$ contains 0, then r_3 is not significant.

Q6.2

A stationary time series of 121 observations produced sample partial auto-correlation of $\hat{\phi}_{11} = 0.8$, $\hat{\phi}_{22} = -0.6$, $\hat{\phi}_{33} = 0.08$, $\hat{\phi}_{44} = 0.00$

On this basis alone, what ARIMA model would we tentatively specify for the series?

Using

$$\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{121}} = 0.181,$$

only $\hat{\phi}_{11} = 0.8$ and $\hat{\phi}_{22} = -0.6$ exceed this limit, so an AR(2) model should be specified.

Q6.3

A time series of 169 observations produced sample auto-correlation of $r_1 = 0.41$, $r_2 = 0.32$, $r_3 = 0.26$, $r_4 = 0.21$ and $r_5 = 0.16$

On this basis alone, what ARIMA model would we tentatively specify for the series?

We do not have $r_k \to r_1^k$ and hence this outrules an AR(1) model.

Note that $r_2/r_1 = 0.78$, $r_3/r_2 = 0.81$, $r_4/r_3 = 0.81$ and $r_5/r_4 = 0.76$.

This supports an ARMA(1,1) model with $\phi \approx 0.8$.

See the formula for theoretical ρ_k for an ARMA(1,1) model in TSLecture4c Page 16.

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

At lag 1, $\frac{\rho_{k+1}}{\rho_k} = \frac{\phi^{k+1}}{\phi^k} = \phi$.

Q6.4

The sample acf for a series and its first difference are given in the following table. Here $n = 100$.

lag	1	2	3	4	5	6
acf for Y_t	0.97	0.97	0.93	0.85	0.80	0.71
acf for ∇Y_t	-0.42	0.18	-0.02	0.07	-0.10	-0.09

Based on this information alone, which ARIMA model(s) would we consider for the series?

The lack of decay in the sample acf suggests non-stationarity.

After differencing, the correlations seem much more like an MA(1).

In particular, using the same formula as in Q6.1, and likewise, for an observed time series, replacing ρ's by r's, we have;

$$Var(r_2) = \frac{1}{100}\{1 + 2(-0.42)^2\} = 0.0135.$$

$$\begin{aligned} t_{r_2} &= \frac{r_2}{se(r_2)} \\ &= \frac{0.18}{\sqrt{0.0135}} \\ &= 1.55. \end{aligned}$$

r_2 is not significant because its t value is < 2.

Thus an IMA(1,1) model seems reasonable.

Q6.5

The sample pacf for a series of length 64 are given in the following table.

lag	1	2	3	4	5
pacf	0.47	-0.34	0.20	0.02	-0.06

Based on this information alone, which ARIMA model(s) would we consider for the series?

Notice that $\frac{2}{\sqrt{64}} = 0.25$.

All the partial auto-correlations from lag 3 onwards are smaller in amplitude than 0.25.

Thus an AR(2) model seems reasonable.

Q6.6

Simulate an AR(1) series with $n = 48$ and $\phi = 0.7$.

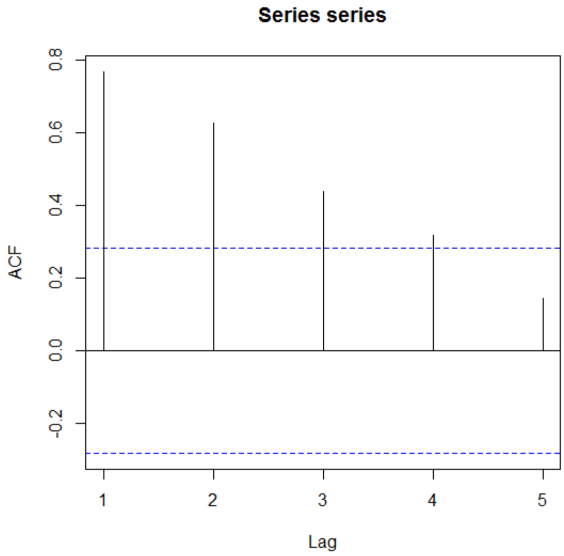
```
set.seed(241357)
series=arima.sim(n=48,list(ar=0.7))
```

(a) Calculate the theoretical auto-correlations at lag 1 and lag 5 for this model.

$$
\begin{aligned}
\rho_k &= \phi^k. \\
\rho_1 &= \phi^1 \\
&= 0.7. \\
\rho_5 &= \phi^5 \\
&= 0.7^5 \\
&= 0.16807.
\end{aligned}
$$

(b) Calculate the sample auto-correlations at lag 1 and lag 5 and compare the values with their theoretical values. Quantify the comparisons using standard error (se) formulae.

```
acf(series,lag.max=5)[1:5]

#Autocorrelations of series series, by lag
#    1     2     3     4     5
#0.768 0.626 0.436 0.318 0.143
```



See pages 4 and 5 of TSLecture6.

$$
\begin{aligned}
se(r_1) &= \sqrt{\frac{1-\phi^2}{n}} \\
&= \sqrt{\frac{1-0.7^2}{48}} \\
&\approx 0.10.
\end{aligned}
$$

$$
\begin{aligned}
se(r_5) &= \sqrt{\frac{1}{n}\frac{1+\phi^2}{1-\phi^2}} \\
&= \sqrt{\frac{1}{48}\frac{1+0.7^2}{1-0.7^2}} \\
&\approx 0.25.
\end{aligned}
$$

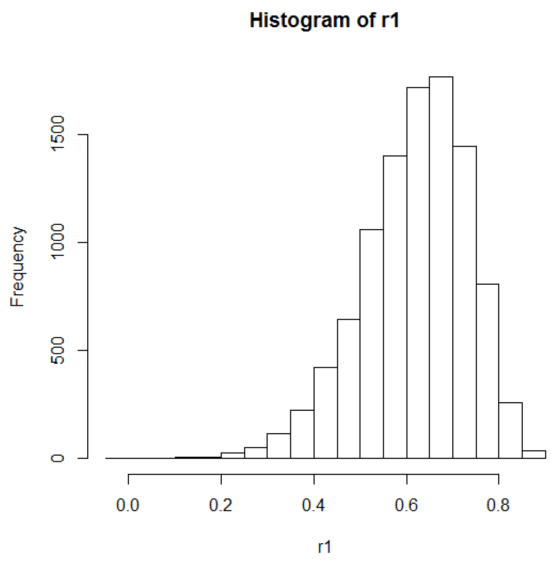
The values of r_1 and r_5 are well within 2 standard errors of ρ_1 and of ρ_5, and so r_1 and r_5 are good estimates of their theoretical counterparts, ρ_1 and ρ_5.

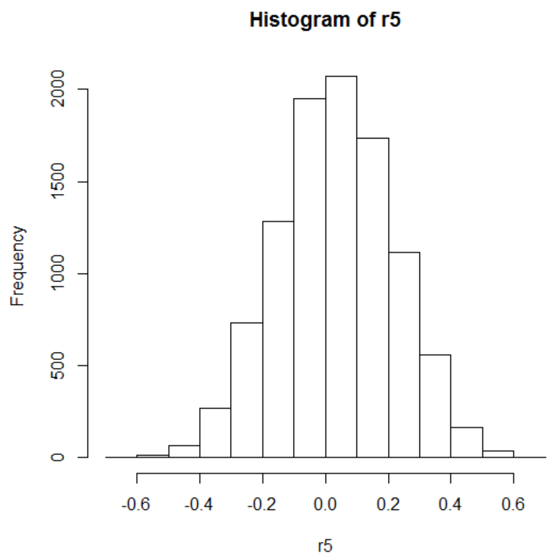
(c) Repeat the simulation of the series and calculation of r_1 and r_5 many times and form the sampling distributions of r_1 and r_5.

Describe how the precision of the estimate varies with different samples selected under identical conditions.

How well does the large sample variance approximate the variance in your sampling distribution?

```
set.seed(132435)
r1=rep(NA,10000)
r5=r1 # We are doing 10,000 replications.

for (k in 1:10000)
 {
  series=arima.sim(n=48, list(ar=0.7))
  r1[k]=acf(series,lag.max=1,plot=F)$acf[1]
  r5[k]=acf(series,lag.max=5,plot=F)$acf[5]
 }
hist(r1)
mean(r1)
#[1] 0.6184299
sd(r1)
#[1] 0.1145287
median(r1)
#[1] 0.6313287

hist(r5)
mean(r5)
#[1] 0.03277664
sd(r5)
#[1] 0.1848059
median(r5)
#[1] 0.03243878
```





For the sampling distribution of r_1, the mean is 0.618 ($\rho_1 = 0.7$) and the median is 0.631, and hence, the histogram shows a slight left skew. The standard deviation is 0.11 which is very similar to the asymptotic theory value of 0.10.

For the sampling distribution of r_5, the mean is 0.033 ($\rho_5 = 0.168$) and the median is 0.032, and hence, the histogram shows a symmetrical distribution.

The standard deviation is 0.18 which is not too different from the asymptotic theory value of 0.25.

Q6.7

Simulate an MA(1) time series with $n = 60$ and $\theta = 0.5$

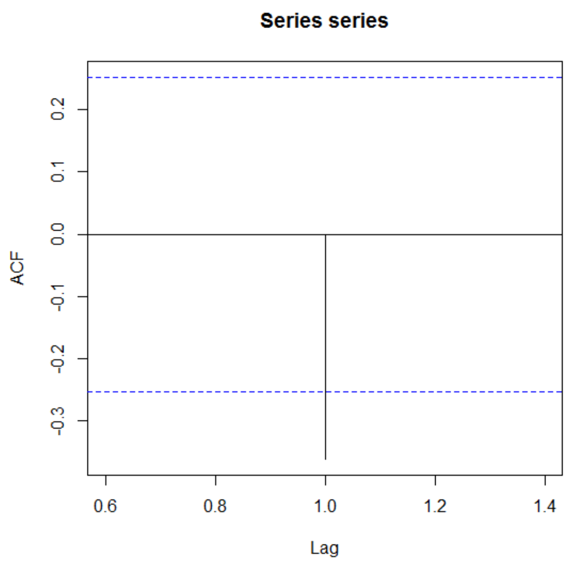
```
set.seed(6453421)
series=arima.sim(n=60,list(ma=-0.5))
```

(a) Calculate the theoretical auto-correlation at lag1 for this model.

$$\begin{aligned}\rho_1 &= -\frac{\theta}{1+\theta^2} \\ &= -\frac{0.5}{1+0.5^2} \\ &= -0.4.\end{aligned}$$

(b) Calculate the auto-correlation at lag 1 and compare the value with its theoretical value.

```
acf(series,lag.max=1)[1]

#Autocorrelations of series series, by lag
#     1
#-0.362
```



Quantify the comparisons using the standard error formula.

For an MA(1) process, the formula for $Var(r_k)$ is given in TSLecture6 Page6.

$$\begin{aligned} se(r_1) &= \sqrt{\frac{c_{11}}{n}} \\ &= \sqrt{\frac{1-3\rho_1^2+4\rho_1^4}{n}}. \end{aligned}$$

For an observed time series, we can replace ρ's by r's.

$$\begin{aligned} se(r_1) &\approx \sqrt{\frac{1-3r_1^2+4r_1^4}{n}} \\ &= \sqrt{\frac{1-3(-0.362)^2+4(-0.362)^4}{60}} \\ &\approx 0.106. \end{aligned}$$

The estimate, r_1, of -0.362 is well within 2 standard errors of the true value, ρ_1, of -0.4.

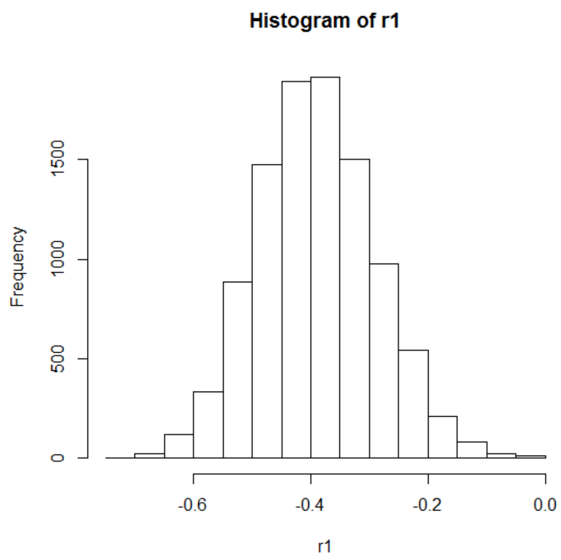
(d) Repeat the simulation of the series and calculation of r_1 many times and form the sampling distributions of r_1.

Describe how the precision of the estimate varies with different samples selected under identical conditions.

How well does the large sample variance approximate the variance in your sampling distribution?

```
set.seed(534261)
r1=rep(NA,10000)
# We are doing 10,000 replications.

for (k in 1:10000)
 {
  series=arima.sim(n=60, list(ma=-0.5))
  r1[k]=acf(series,lag.max=1,plot=F)$acf[1]
 }
```

```
hist(r1)
mean(r1)
#[1] -0.3901954
sd(r1)
#[1] 0.1002125
median(r1)
#[1] -0.3931737
```



The theoretical auto-correlation at lag 1 is $\rho_1 = -0.4$.

The mean of the sampling distribution is -0.390, the median is -0.393 and the standard deviation is 0.1.

The large sample standard deviation is 0.1.

Q6.8

Simulate an AR(1) time series with $n = 48$ and calculate theoretical ρ_1 and ρ_5.

(a) with $\phi = 0.9$

```
set.seed(5342310)
series1=arima.sim(n=48,list(ar=0.9))
```

$$\rho_1 = 0.9; \qquad \rho_5 = 0.9^5 = 0.59.$$

(b) with $\phi = 0.6$

```
set.seed(5342310)
series2=arima.sim(n=48,list(ar=0.6))
```

$$\rho_1 = 0.6; \qquad \rho_5 = 0.6^5 = 0.07776.$$

(c) with $\phi = 0.3$

```
set.seed(5342310)
series3=arima.sim(n=48,list(ar=0.3))
```

$$\rho_1 = 0.3; \qquad \rho_5 = 0.3^5 = 0.00243.$$

(d) Calculate r_1 and r_5 for (a), (b) and (c) above and the precision of the estimates.

Describe, in general, how the precision of the estimates varies with the value of ϕ.

```
acf(series1)[1:5]

#    1     2     3     4     5
#0.862 0.739 0.569 0.420 0.232
```

$$\begin{aligned} se(r_1) &= \sqrt{\frac{1-\phi^2}{n}} \\ &= \sqrt{\frac{1-0.9^2}{48}} \\ &\approx 0.06. \end{aligned}$$

$$
\begin{aligned}
se(r_5) &= \sqrt{\frac{1}{n}\frac{1+\phi^2}{1-\phi^2}} \\
&= \sqrt{\frac{1}{48}\frac{1+0.9^2}{1-0.9^2}} \\
&\approx 0.40.
\end{aligned}
$$

0.862 is within 2 se(r_1) of 0.9.

0.232 is within 2 se(r_5) of 0.59.

The two estimates are good estimates of their theoretical counterparts.

```
acf(series2)[1:5]

#    1     2     3     4     5
#0.617 0.388 0.392 0.228 0.191
```

$$
\begin{aligned}
se(r_1) &= \sqrt{\frac{1-\phi^2}{n}} \\
&= \sqrt{\frac{1-0.6^2}{48}} \\
&\approx 0.12.
\end{aligned}
$$

$$
\begin{aligned}
se(r_5) &= \sqrt{\frac{1}{n}\frac{1+\phi^2}{1-\phi^2}} \\
&= \sqrt{\frac{1}{48}\frac{1+0.6^2}{1-0.6^2}} \\
&\approx 0.21.
\end{aligned}
$$

0.617 is within 2 se(r_1) of 0.6.

0.191 is within 2 se(r_5) of 0.07776.

The two estimates are good estimates of their theoretical counterparts.

```
acf(series3)[1:5]
#     1      2      3      4      5
# 0.188  0.032  0.294 -0.081  0.048
```

$$\begin{aligned} se(r_1) &= \sqrt{\frac{1-\phi^2}{n}} \\ &= \sqrt{\frac{1-0.3^2}{48}} \\ &\approx 0.14. \end{aligned}$$

$$\begin{aligned} se(r_5) &= \sqrt{\frac{1}{n}\frac{1+\phi^2}{1-\phi^2}} \\ &= \sqrt{\frac{1}{48}\frac{1+0.3^2}{1-0.3^2}} \\ &\approx 0.16. \end{aligned}$$

0.188 is within 2 se(r_1) of 0.3.

0.048 is within 2 se(r_5) of 0.00243.

The two estimates are good estimates of their theoretical counterparts.

Refer to TSLecture6 Pages 4 and 5.

As $\phi \downarrow$, $se(r_1) \uparrow$, and $se(r_5) \downarrow$.

Q6.9

Simulate 3 AR(1) time series with $\phi = 0.6$ and $n = 24$, 60 and 120.

With each one, estimate ρ_1 with r_1 and quantify the error.

For an AR(1) process, $\rho_k = \phi^k$, and thus $\rho_1 = \phi^1 = 0.6$.

(a)

```
set.seed(162534)
series1=arima.sim(model=list(order=c(1,0,0),ar=0.6),n=24)
acf(series1)[1]
#0.459
```

(b)

```
series2=arima.sim(model=list(order=c(1,0,0),ar=0.6),n=60)
acf(series2)[1]
#0.385
```

(c)

```
series3=arima.sim(model=list(order=c(1,0,0),ar=0.6),n=120)
acf(series3)[1]
#0.627
```

The best estimate is the series 3 with the largest $n = 120$.

The three standard error are given by:

$$se(r_1) = \sqrt{\frac{1-\phi^2}{n}} = \sqrt{\frac{1-0.6^2}{24}} = 0.163,$$

and thus the estimate of 0.459 is well within 2 standard errors of ρ_1.

$$se(r_1) = \sqrt{\frac{1-\phi^2}{n}} = \sqrt{\frac{1-0.6^2}{60}} = 0.103,$$

and thus the estimate of 0.385 is more than 2 se from $\rho_1 = 0.6$.

$$se(r_1) = \sqrt{\frac{1-\phi^2}{n}} = \sqrt{\frac{1-0.6^2}{120}} = 0.073,$$

and thus the estimate of 0.627 is well within 2 standard errors of ρ_1.

Q6.10

Simulate 3 MA(1) time series with $\theta = 0.7$ and $n = 24$, 60 and 120.

With each one, estimate ρ_1 with r_1 and quantify the error.

$$\begin{aligned}\rho_1 &= -\frac{\theta}{1+\theta^2} \\ &= -0.4.\end{aligned}$$

(a)

```
set.seed(172534)
series1=arima.sim(n=24,list(ma=-0.7))
acf(series1)[1]
#-0.595
```

(b)

```
set.seed(172534)
series2=arima.sim(n=60,list(ma=-0.7))
acf(series2)[1]
#-0.527
```

(c)

```
set.seed(172534)
series3=arima.sim(n=120,list(ma=-0.7))
acf(series3)[1]
#-0.458
```

The best estimate is the series 3 with the largest $n = 120$.

The standard errors for an MA(1) series are given by:

$$\sqrt{\frac{1-3\rho_1^2+4\rho_1^4}{n}}.$$

For the observed series, we replace ρ_1 with r_1.

The three standard errors are given by:

$$se(r_1) = \sqrt{\frac{1-3(-0.595)^2+4(-0.595)^4}{24}} = 0.163,$$

and thus the estimate of -0.595 is within 2 standard errors of $\rho_1 = -0.4$.

$$se(r_1) = \sqrt{\frac{1 - 3(-0.527)^2 + 4(-0.527)^4}{60}} = 0.08904185,$$

and thus the estimate of -0.527 is within 2 standard errors of $\rho_1 = -0.4$.

$$se(r_1) = \sqrt{\frac{1 - 3(-0.458)^2 + 4(-0.458)^4}{120}} = 0.06749494,$$

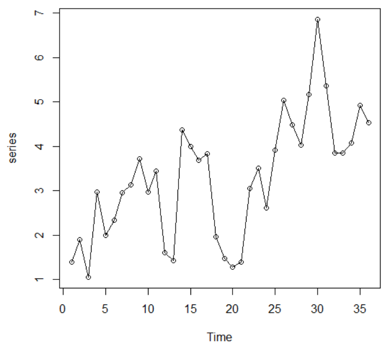
and thus the estimate of -0.458 is within 2 standard errors of $\rho_1 = -0.4$.

Q6.11

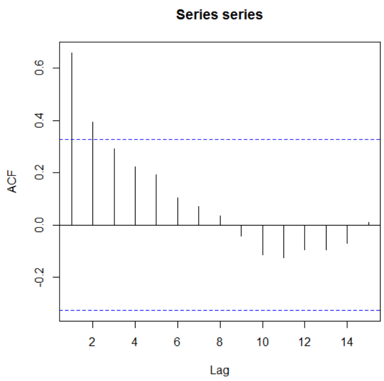
Simulate a stationary time series of length $n = 36$ according to an AR(1) model with $\phi = 0.95$.

(a) Plot the series and calculate the sample acf and pacf and describe them.

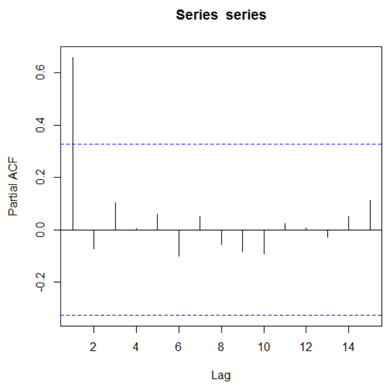
```
set.seed(274135)
series=arima.sim(n=36,list(ar=0.95))

plot(series,type='o')
```




```
acf(series)
```



```
pacf(series)
```



The acf and pacf would tend to suggest an AR(1) model, but such a model is stationary, and, from the time plot, there appears to be an upward trend which would indicate non-stationarity.

(b) Perform the augmented Dicky-Fuller test with k chosen by the software.

```
ar(diff(series))

#Call:
#ar(x = diff(series))

#Coefficients:
#      1        2
#-0.1573  -0.3090

#Order selected 2  sigma^2 estimated as  1.115

adf.test(series, k=2)

#data:  series
#Dickey-Fuller = -2.6607, Lag order = 2, p-value = 0.3161
#alternative hypothesis: stationary
```

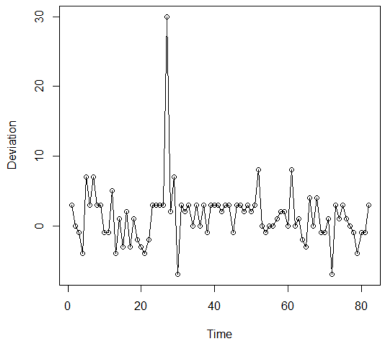
This suggests not rejecting H_0, i.e., the data are not stationary, and have a unit root.

Q6.12

The data file named "deere1" contains 82 consecutive values for the amount of deviation (in 0.000025 inch units) from a specified target value that an industrial machining process at Deere & Co. produced under specified operating conditions.

(a) Display the time series plot of this series and comment on any unusual points.

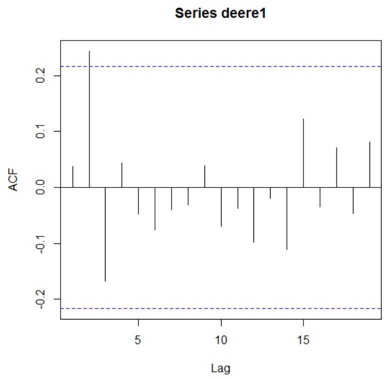
```
data(deere1)
plot(deere1,type='o',ylab='Deviation')
```



Except for one value of 30 at $t = 27$, the process seems relatively stable and stationary.

(b) Calculate the sample acf for this series and comment on the results.

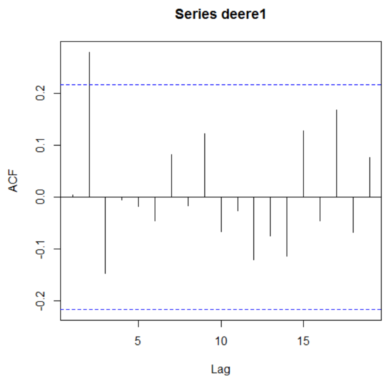
```
acf(deere1)
```



The graph reveals a statistically significant auto-correlation at lag 2.

(c) Now replace the unusual value by a much more typical value and recalculate the sample acf. Comment on the change from part(b).

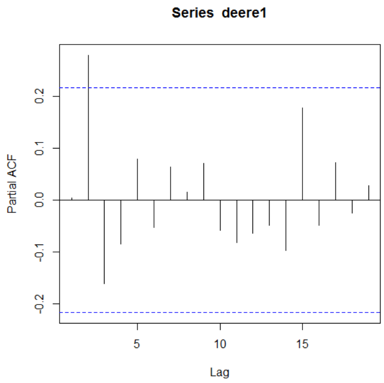
```
deere1[27]=8
acf(deere1)
```



The value of 30 at time 27 was replaced with the next largest value, i.e., 8. This only had a small effect on the sample acf.

(d) Calculate the sample pacf on the revised series and comment on the results.

```
deere1
pacf(deere1)
```



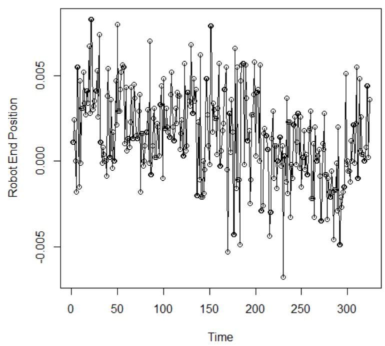
The pacf suggests an AR(2) model for the series.

Q6.13

The datafile named "robot" contains a time series obtained from an industrial robot. The robot was put through a sequence of maneuvers, and the distance from a desired ending point was recorded in inches. This was repeated 324 times to form the time series.

(a) Display the time series plot of the data. Based on this information, do these data appear to come from a stationary or non-stationary process?

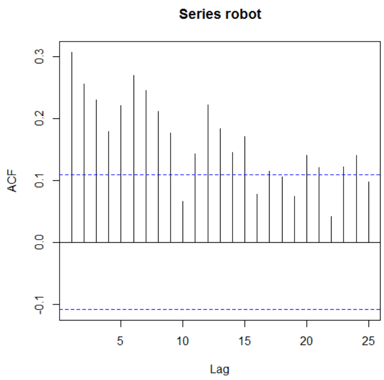
```
data(robot)
plot(robot,type='o',ylab='Robot End Position')
```

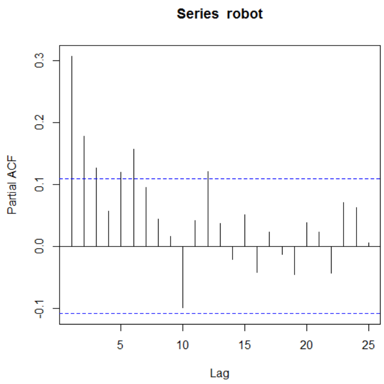


From this plot, we might try a stationary model, but there is also enough drift that we might try a non-stationary model.

(b) Calculate and plot the sample acf and pacf for these data. Based on this additional information, do these data appear to come from a stationary or non-stationary process?

```
acf(robot)
pacf(robot)
```





These plots are not especially definitive, but the pacf suggests possibly a (stationary) AR(3) model.

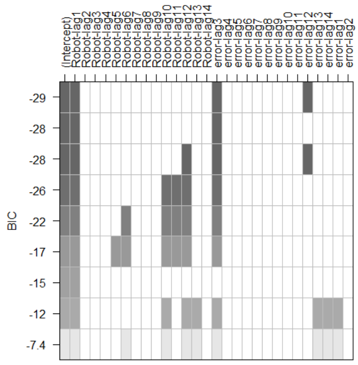
(c) Calculate and interpret the sample eacf.

```
eacf(robot)

AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x x o x  x  x  x
1 x o o o o o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x x o o o o o o o o o  o  o  o
4 x x x x o o o o o o o  o  x  o
5 x x x o o o o o o o o  o  x  o
6 x o o o o x o o o o o  o  o  o
7 x o o x o x x o o o o  o  o  o
```

This suggests an ARMA(1,1) model.

(d) Use the best subsets ARMA approach to specify a model for these data. Compare these results with what you discovered in parts (a), (b) and (c).

```
plot(armasubsets(y=robot,nar=14,nma=14,y.name='Robot',ar.method='ols'))
```



The best model here includes a lag 1 AR term, but lags 3 and 12 terms in the MA part of the model.

Q6.14

Calculate and interpret the sample eacf for the logarithms of the LA rainfall series.

Do the results confirm that the logs are white noise?

```
data(larain)
eacf(log(larain))
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o o o o o  o  o  o
1 o o o o o o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x o o o o o o o o o o  o  o  o
4 x o o o o o o o o o o  o  o  o
5 x x x x x o o o o o o  o  o  o
6 x x o o x o o o o o o  o  o  o
7 x o x o o o o o o o o  o  o  o
```

The eacf results confirm that the logs are white noise

Q6.15

Calculate and interpret the sample eacf for the colour property time series.

Do the results suggest the same model specified by the sample pacf.

```
data(color)
eacf(color,ar.max=7,ma.max=9)
AR/MA
  0 1 2 3 4 5 6 7 8 9
0 x o o o o o o o o o
1 o o o o o o o o o o
2 o o o o o o o o o o
3 x o o o o o o o o o
4 o o o o o o o o o o
5 x o o o o o o o o o
6 x o o o o o o o o o
7 x o o o o o o o o o
```

The sample eacf supports an AR(1) model for this series as suggested by the pacf.

See TSLecture6 Pages 42 and 43.