Tutorial Sheet 8 Solutions

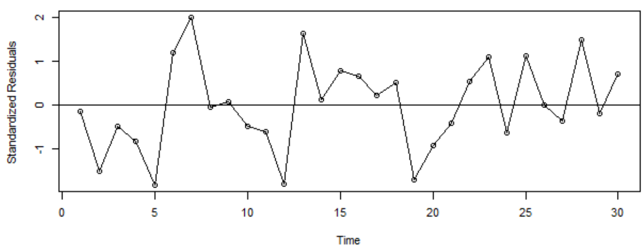
Q8.1

Simulate an AR(1) model with $n = 30$ and $\phi = 0.5$.

```
set.seed(12347)
series=arima.sim(n=30,list(ar=0.5))
```

(a) Fit the correctly specified AR(1) model and look at a time series plot of the residuals. Does the plot support the AR(1) specification?

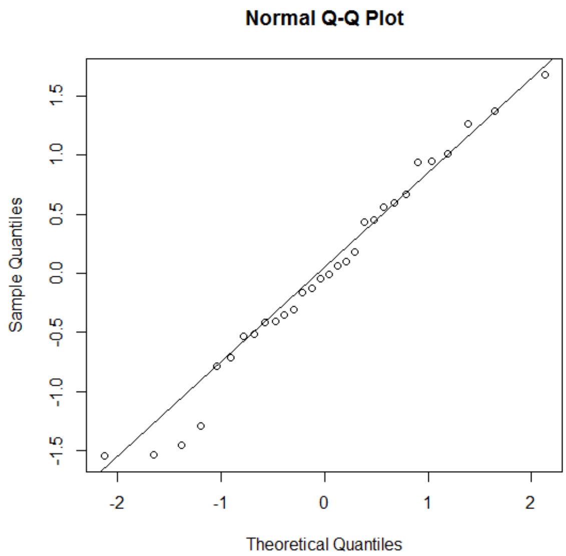
```
model=arima(series,order=c(1,0,0))
win.graph(width=6.5,height=3,pointsize=8)
plot(rstandard(model),ylab ='Standardized Residuals', type='o')
abline(h=0)
# alternatively we can plot with:
plot(resid(model),ylab ='Standardized Residuals', type='o')
abline(h=0)
```



These standardised residuals look fairly random with no particular patterns.

(b) Display a normal QQ plot of the standardised residuals. Does the plot support the AR(1) specification?

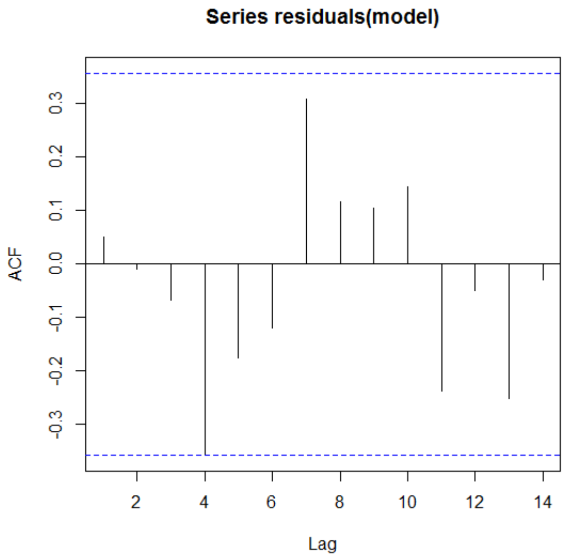
```
qqnorm(residuals(model))
qqline(residuals(model))
```



With a few minor exceptions in the lower tail, the QQ plot of the standardised residuals looks reasonably 'normal'.

(c) Display the sample acf of the standardised residuals. Does the plot support the AR(1) specification?

```
acf(residuals(model))
```



The sample acf at lag 4 is the only auto-correlation that comes close to being significant.

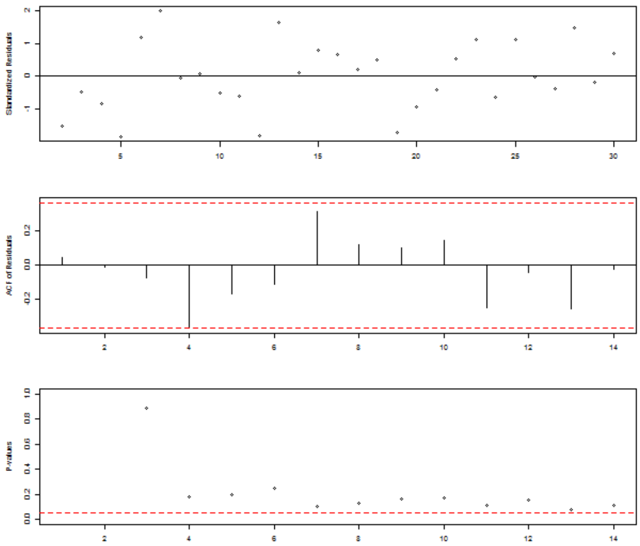
(d) Calculate the Ljung-Box statistic summing to $K = 8$. Does this statistic support the AR(1) specification?

```
LB.test(model,lag=8)

# Box-Ljung test

#data:  residuals from  model
#X-squared = 11.2399, df = 7, p-value = 0.1285
```

This test does not reject randomness of the error terms based on the first 8 auto-correlations of the residuals.

The 'tsdiag()' will produce a display that answer all 3 parts, (a), (c) and (d).

```
win.graph(width=6.5,height=6,pointsize=8)
tsdiag(model)
```



The bottom display shows the $p-$ values for a variety of values of the "K" parameter - the highest lag used in the sum.

The top display will flag, if any, potential outliers.

Q8.2

Simulate an MA(1) model with $n = 36$ and $\theta = -0.5$.

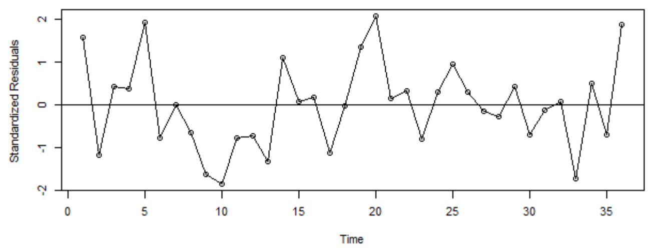
```
set.seed(64231)
series=arima.sim(n=36,list(ma=0.5))
```

Remember, when simulating an MA process using an inbuilt R function, we use the negative of the given θ value.

(a) Fit the correctly specified MA(1) model and look at a time series plot of the residuals.

Does the plot support the MA(1) specification.

```
model=arima(series,order=c(0,0,1))
win.graph(width=6.5,height=3,pointsize=8)
plot(rstandard(model),ylab ='Standardized Residuals', type='o')
abline(h=0)

#Equivalently we can use:
plot(residuals(model),ylab ='Standardized Residuals', type='o')
abline(h=0)
```

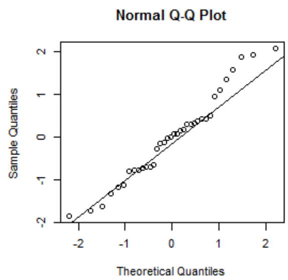


The standardised residuals look fairly random, implying the model fit is a good fit.

(b) Display a Normal QQ plot of the standardised residuals.

Does this plot support the MA(1) process?

```
win.graph(width=3,height=3,pointsize=8)
qqnorm(residuals(model))
qqline(residuals(model))
shapiro.test(residuals(model))
#        Shapiro-Wilk normality test

#data:  residuals(model)
#W = 0.968, p-value = 0.374
```



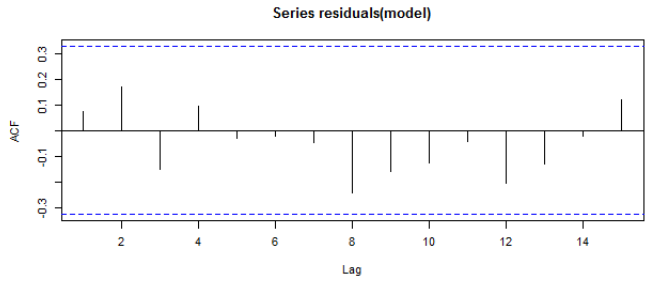
The QQ plot is quite straight, but there may be some problem with the upper tail.

The sample size is small, so that could be easily enlarged if required.

The high $p-$ value in the Shapiro-Wilk test means we fail to reject H_0, i.e., the residual data are indeed Normally distributed.

(c) Display the sample acf of the residuals. Does the plot support the MA(1) specification?

```
win.graph(width=6.5,height=3,pointsize=8)
acf(residuals(model))
```



There is no problem with significant auto-correlation in this simulation.

(d) Calculate the Ljung-Box statistic summing to $K = 6$.

Does this statistic support the MA(1) specification?

```
LB.test(model,lag=6)
        Box-Ljung test

data:  residuals from  model
X-squared = 2.7428, df = 5, p-value = 0.7396
```

There is no problem with large residual auto-correlations jointly out to lag 6.

Q8.3

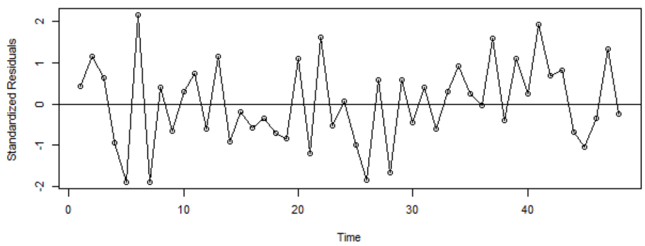
Simulate an AR(2) model with $n = 48$, $\phi_1 = 1.5$ and $\phi_2 = -0.75$.

```
set.seed(65423)
series=arima.sim(n=48,list(ar=c(1.5,-0.75)))
```

(a) Fit the correctly specified AR(2) model and look at a time series plot of the residuals.

Does the plot support the AR(2) specification?

```
model=arima(series,order=c(2,0,0))
win.graph(width=6.5,height=3,pointsize=8)
plot(resid(model),ylab ='Standardized Residuals', type='o')
abline(h=0)
```

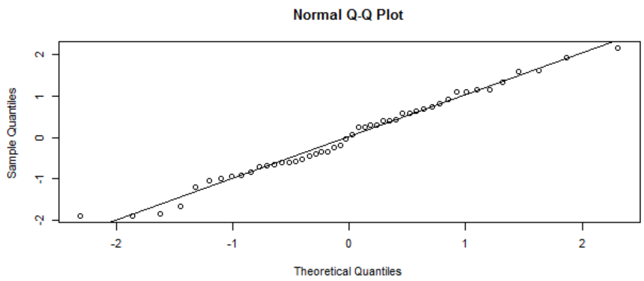


The residuals look random.

(b) Display a Normal QQ plot of the standardised residuals.

Does this plot support the AR(2) specification?

```
qqnorm(resid(model))
qqline(resid(model))
```

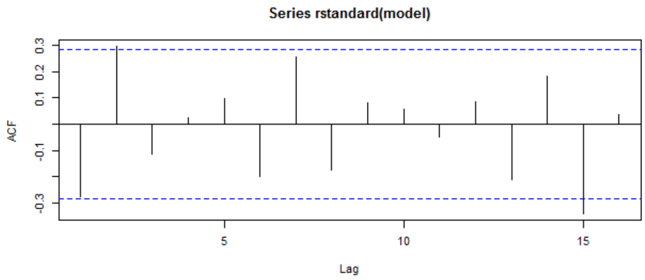


There is no problem with the Normality of the residuals.

(c) Display the sample acf of the residuals.

Does this plot support the AR(2) specification?

```
win.graph(width=6.5,height=3,pointsize=8)
acf(rstandard(model))
```



There are two significant auto-correlations, at lag 2 and at lag 15.

(d) Calculate the Ljung-Box statistic summing to $K = 12$.

Does this statistic support the AR(2) specification?

```
LB.test(model,lag=12)

#        Box-Ljung test

#data:  residuals from  model
#X-squared = 18.7997, df = 10, p-value = 0.04288
```

Because the $p-$ value is < 0.05, we reject H_0 that the error terms are independent at the 5% level.

Q8.4

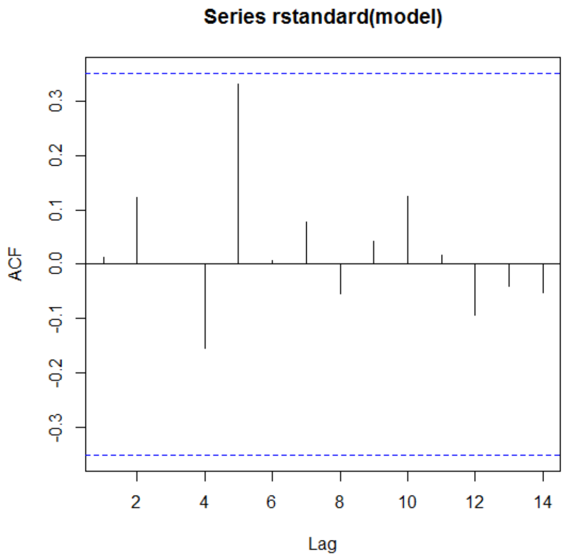
(a) Fit an AR(3) model by maximum likelihood to the square root of the 'hare abundance' data series.

```
data(hare)
model=arima(sqrt(hare),order=c(3,0,0))
model

#Call:
#arima(x = sqrt(hare), order = c(3, 0, 0))

#Coefficients:
#         ar1      ar2      ar3  intercept
#      1.0519  -0.2292  -0.3931     5.6923
#s.e.  0.1877   0.2942   0.1915     0.3371

#sigma^2 estimated as 1.066:  log likelihood = -46.54,  aic = 101.08
```

The ar1, ar3 and intercept terms are significant, but the ar2 term is not significant.

(b) Plot the sample acf of the residuals and comment on the size of the auto-correlations.

```
acf(rstandard(model))
```



The residual auto-correlations look very good.

(c) Calculate the Ljung-Box statistic summing to $K = 9$.

Does this statistic support the AR(3) specification?

```
LB.test(model,lag=9)

#        Box-Ljung test

#data:  residuals from  model
#X-squared = 6.2475, df = 6, p-value = 0.396
```

The high $p-$ value tells us not to reject the Null Hypothesis that the error terms are independent.

(d) Perform a runs test on the residuals and comment on the results.

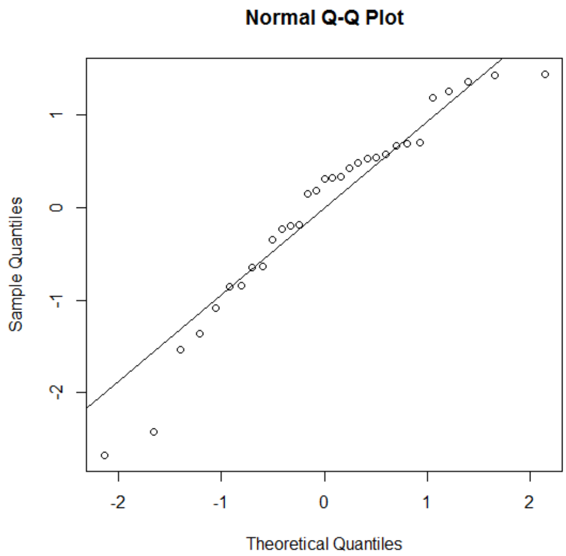
```
runs(resid(model))

$pvalue
[1] 0.602

$observed.runs
[1] 18

$expected.runs
[1] 16.09677

$n1
[1] 13

$n2
[1] 18

$k
[1] 0
```

The high $p-$ value tells us not to reject the Null Hypothesis that the error terms are independent.

The number of runs is not unusual.

(e) Display the QQ Normal plot of the residuals and comment on the plot.

```
qqnorm(residuals(model))
qqline(residuals(model))
```



There is some minor curvature to the plot with some possible outliers at both extremes.

(f) Perform the Shapiro-Wilk test of Normality on the residuals.

```
shapiro.test(residuals(model))

#        Shapiro-Wilk normality test

#data:  residuals(model)
#W = 0.9351, p-value = 0.06043
```

We do not reject Normality of residuals at the 5% signficance level.

Q8.5

We met the 'oilfiters' data series in TSLecture1.

(a) Fit an AR(1) model to this series.

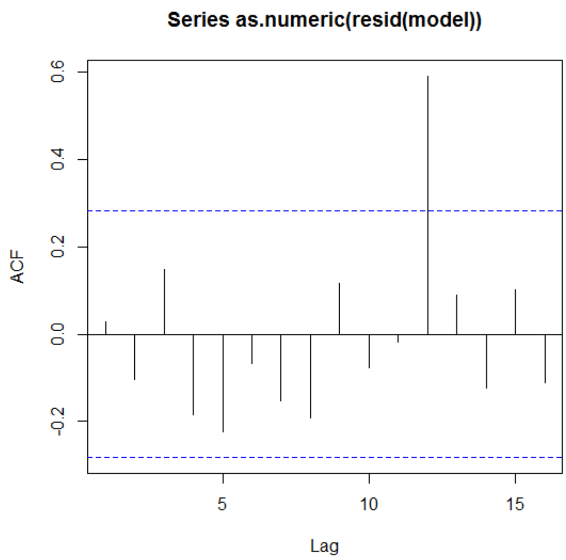
Is the estimate of the ϕ parameter significantly different from zero statistically?

```
data(oilfilters)
model=arima(oilfilters,order=c(1,0,0))
model

#Call:
#arima(x = oilfilters, order = c(1, 0, 0))

#Coefficients:
#         ar1  intercept
#      0.3115  3370.6744
#s.e.  0.1368   253.1499

#sigma^2 estimated as 1482802:  log likelihood = -409.19,  aic = 822.37
```

The estimate of ϕ is more than two standard errors away from zero and would thus be deemed to be significant at the usual significant levels.

(b) Display the sample acf of the residuals from the AR(1) fitted model and comment on the display.

```
acf(resid(model))
```



The sample auto-correlation of the residuals displays a highly significant correlation at lag 12.

This series contains substantial seasonality that this model does not capture.

Q8.6

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.

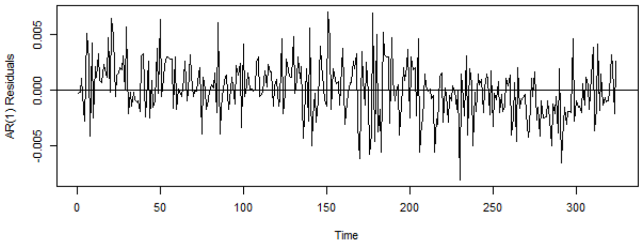
Compare the fits of an AR(1) model and an IMA(1,1) model for these data in terms of the diagnostic tests discussed in the lecture.

```
data(robot)
mod1=arima(robot,order=c(1,0,0))
res1=resid(mod1)
mod1

#Call:
#arima(x = robot, order = c(1, 0, 0))

#Coefficients:
#         ar1  intercept
#      0.3074     0.0015
#s.e.  0.0528     0.0002

#sigma^2 estimated as 6.482e-06:  log likelihood = 1475.54,  aic = -2947.08


mod2=arima(robot,order=c(0,1,1))
res2=resid(mod2)
mod2

#Call:
#arima(x = robot, order = c(0, 1, 1))

#Coefficients:
#          ma1
#      -0.8713
#s.e.   0.0389

#sigma^2 estimated as 6.069e-06:  log likelihood = 1480.95,  aic = -2959.9
```

Both models have statistically significant parameter estimates.

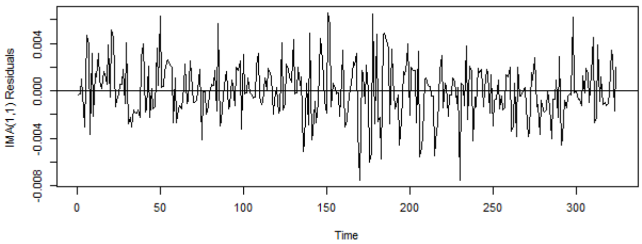
The log likelihood and AIC values are just a little better in the IMA(1,1) model.

```
win.graph(width=6.5,height=3,pointsize=8)
plot(res1,ylab='AR(1) Residuals')
abline(h=0)
```



There may be a slight drift in the residuals here with more positive values early on and more negative values later.

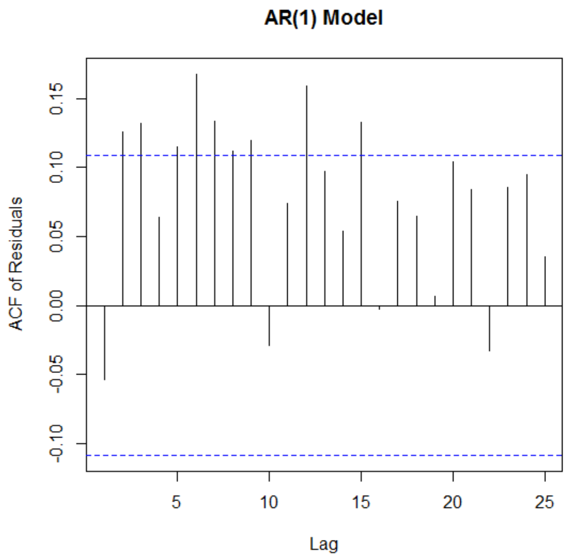
```
plot(res2,ylab='IMA(1,1) Residuals')
abline(h=0)
```



There does not appear to be any drift with the IMA(1,1) model.

We now look at the correlation of the residuals.

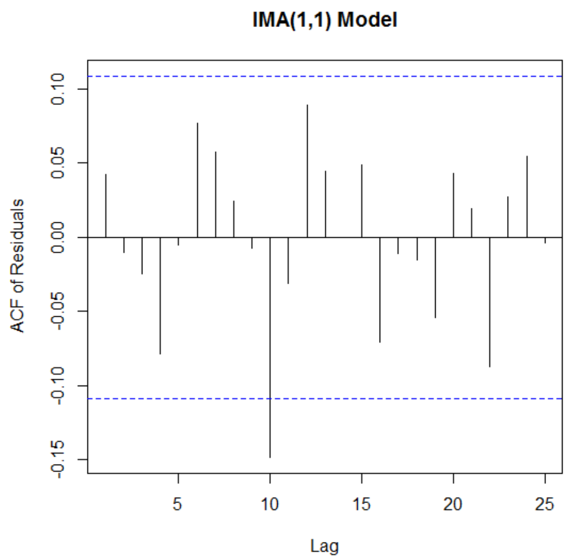
```
acf(residuals(mod1), main='AR(1) Model',ylab='ACF of Residuals')
LB.test(mod1)

#        Box-Ljung test

#data:  residuals from  mod1
#X-squared = 52.5123, df = 11, p-value = 2.201e-07
```



The residuals from the AR(1) model have too much auto-correlation.

The low $p-$ values show that these are unlikely to have arisen by chance.

```
acf(residuals(mod2), main='IMA(1,1) Model',ylab='ACF of Residuals')
LB.test(mod2)

#        Box-Ljung test

#data:  residuals from  mod2
#X-squared = 16.5435, df = 10, p-value = 0.0851
```

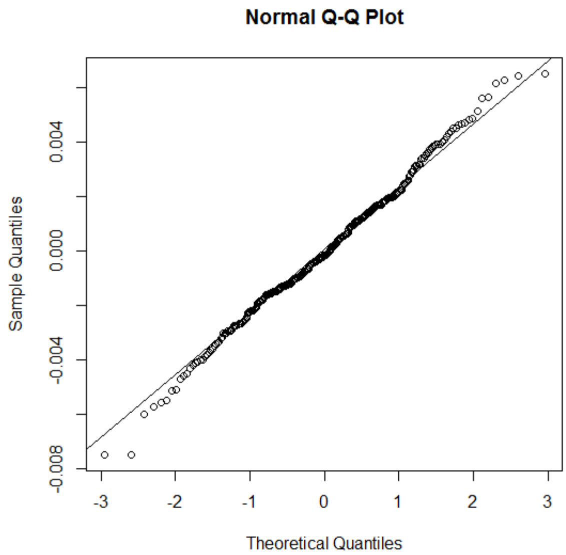


The residuals from the IMA(1,1) model have little auto-correlation.

The high $p-$ values show that these are likely to have arisen by chance.

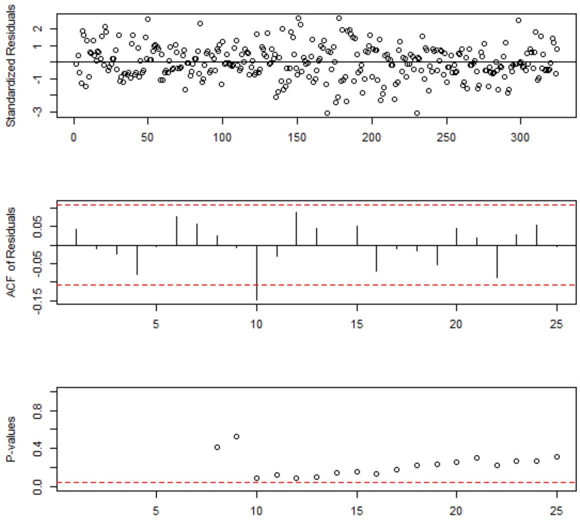
The normality of the error terms is checked by displaying a QQ plot of the residuals.

```
win.graph(width=3,height=3,pointsize=8)
qqnorm(residuals(mod2))
qqline(residuals(mod2))
shapiro.test(residuals(mod2))

#        Shapiro-Wilk normality test

#data:  residuals(mod2)
#W = 0.995, p-value = 0.3717
```



The normality in the QQ plot is confirmed by the high $p-$ value Shapiro-Wilk test, i.e., we fail to reject H_0 that the residuals are Normally distributed.

```
win.graph(height=6,width=6.5,pointsize=8)
tsdiag(mod2)
```



Overall, the robot time series is well-represented by an IMA(1,1) model.

Q8.7

The datafile named 'deere3' contains 57 consecutive values from a complex machine tool at Deere & Co.

The values given are deviations from a target value in units of ten millionths of an inch. The process employs a control mechanism that resets some of the parameters of the machine tool depending on the magnitude of deviation from target of the last item produced.

Diagnose the fit of an AR(1) model for these data.

```
data(deere3)
model=arima(deere3,order=c(1,0,0))
model

#Call:
#arima(x = deere3, order = c(1, 0, 0))

#Coefficients:
#         ar1  intercept
#      0.5255   124.3832
#s.e.  0.1108   394.2069

#sigma^2 estimated as 2069355:  log likelihood = -495.51,  aic = 995.02
```

The ar1 estimate, i.e., the maximum likeihood estimate of ϕ, is statistically significant, but the intercept could be removed from the model because it is not significant.

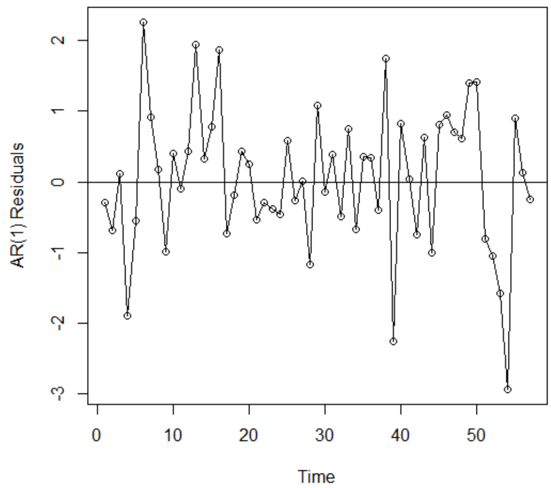
Fit the model excluding a mean or intercept term.

```
model=arima(deere3,order=c(1,0,0),include.mean=F)
model

#arima(x = deere3, order = c(1, 0, 0), include.mean = F)

#Coefficients:
#         ar1
#      0.5291
#s.e.  0.1103

#sigma^2 estimated as 2072748:  log likelihood = -495.56,  aic = 993.12
```

There is very little change in the estimate of ϕ.

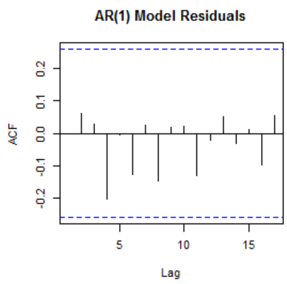
The maximum likelihood is a little bit worse but the AIC is a little bit better.

```
res=rstandard(model)
plot(res,ylab='AR(1) Residuals')
abline(h=0)
```



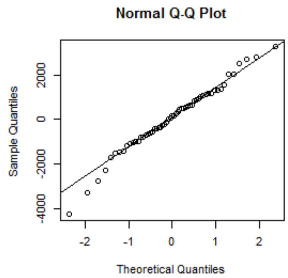
The residuals look reasonably random.

```
acf(res,main='AR(1) Model Residuals')
```



There is little evidence of auto-correlation in the error terms for this model.

```
qqnorm(res)
qqline(res)
shapiro.test(res)

#         Shapiro-Wilk normality test

#data:  res
#W = 0.9829, p-value = 0.5966
```



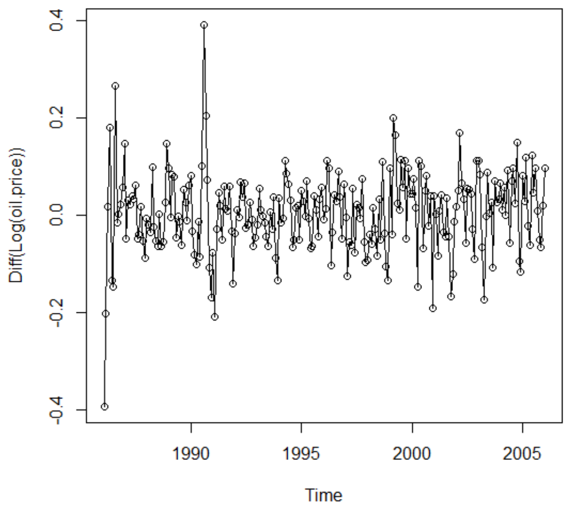
The QQ plot shows some deviation from the 45° line at the lower end, but the high $p-$value in the Shapiro-Wilk test tells us that we cannot reject the Null Hypothesis that the error terms are Normally distributed.

Q8.8

In TSLecture6 page 48, using best subset regression on the difference of the log of oil price, the pictorial output suggested either an AR(1) model or an AR(4) model.

Plot the first difference of the log of the oil price for reference.

```
data(oil.price)
plot(diff(log(oil.price)),type='o',ylab='Diff(Log(oil.price))')
```



The difference looks fairly stable apart from possible outliers at the beginning in February 1986 and in August 1990.

We can see these minimum and maximum values by using the commands:

```
diff(log(oil.price))
min(diff(log(oil.price)))
max(diff(log(oil.price)))
```

(a) Estimate both of these models using maximum likelihood and compare the results using the diagnostic tests considered in this chapter.

```
data(oil.price)
mod1=arima(log(oil.price),order=c(1,1,0))
mod1

#Call:
#arima(x = log(oil.price), order = c(1, 1, 0))

#Coefficients:
#         ar1
#      0.2364
#s.e.  0.0660

#sigma^2 estimated as 0.006787:  log likelihood = 258.55,  aic = -515.11


mod2=arima(log(oil.price),order=c(4,1,0))
mod2

#Call:
#arima(x = log(oil.price), order = c(4, 1, 0))

#Coefficients:
#         ar1      ar2     ar3      ar4
#      0.2673  -0.1550  0.0238  -0.0970
#s.e.  0.0669   0.0691  0.0691   0.0681

#sigma^2 estimated as 0.006603:  log likelihood = 261.82,  aic = -515.64
```

In the ARIMA(4,1,0) model, the ar3 and ar4 coefficients are not significantly different from 0.

This model also has a slightly better (lower) AIC value.

Given the standard errors of the two ar1 coefficients, there is no real difference between the estimates of the ar1 coefficients in the two models.

(b) Try an ARIMA(2,1,0) model for comparison.

```
mod3=arima(log(oil.price),order=c(2,1,0))
mod3

#Call:
#arima(x = log(oil.price), order = c(2, 1, 0))

#Coefficients:
#         ar1      ar2
#      0.2630  -0.1436
#s.e.  0.0666   0.0673

#sigma^2 estimated as 0.00666:  log likelihood = 260.81,  aic = -517.61
```

This model has the lowest AIC value of the three models considered thus far.

(c) TSLecture6 page 30 suggested specifying an MA(1) model for the difference of the logs.

Estimate this model by maximum likelihood and perform the diagnostic test.

```
mod4=arima(log(oil.price),order=c(0,1,1))
mod4

#Call:
#arima(x = log(oil.price), order = c(0, 1, 1))

#Coefficients:
#         ma1
#      0.2956
#s.e.  0.0693

#sigma^2 estimated as 0.006689:  log likelihood = 260.29,  aic = -518.58
```

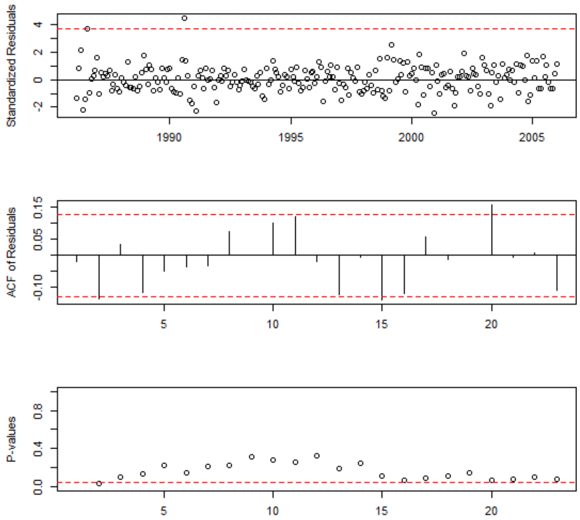
This model has a significant θ coefficient.

The log-likelihood and AIC values are quite similar to the ARIMA(1,1,0) and ARIMA(0,1,1) models.

This IMA(1,1) model does have the best AIC value.

(d) Which of the four models, AR(1), AR(4), AR(2) or MA(1) would you prefer, given the results of parts (a) and (b).

```
win.graph(width=6.5,height=6,pointsize=10)
tsdiag(mod1,main='Model 1')
```

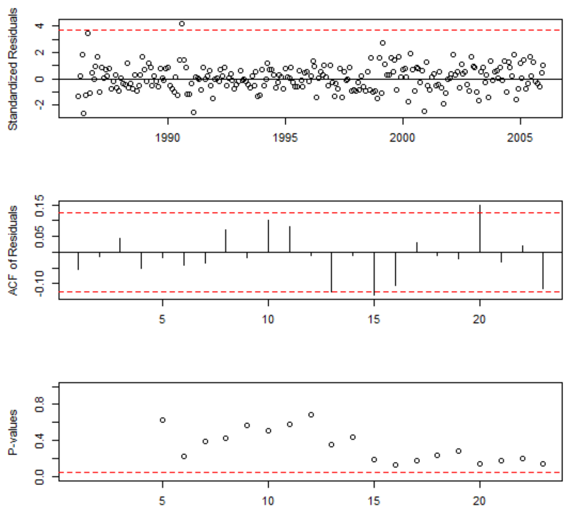


The possible outlier in August 1990 stands out in the plot of residuals.

There are also three residual acf values outside the critical limits.

(d) Which of the four models, AR(1), AR(4), AR(2) or MA(1) would you prefer, given the results of parts (a) and (b).

```
win.graph(width=6.5,height=6,pointsize=10)
tsdiag(mod2,main='Model 2')
```

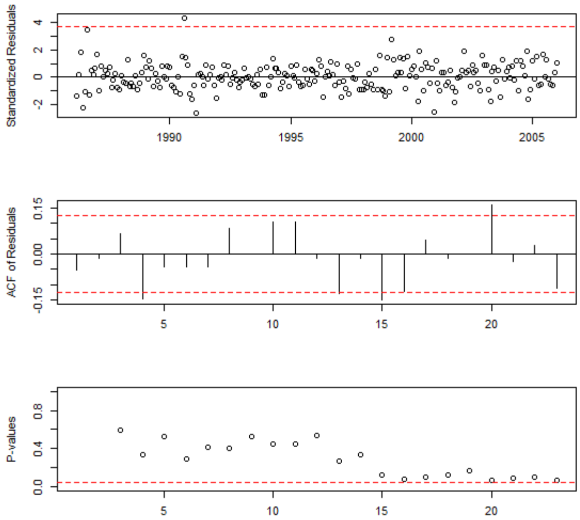


The possible outlier in August 1990 stands out in the plot of residuals.

There are also two residual acf values outside the critical limits.

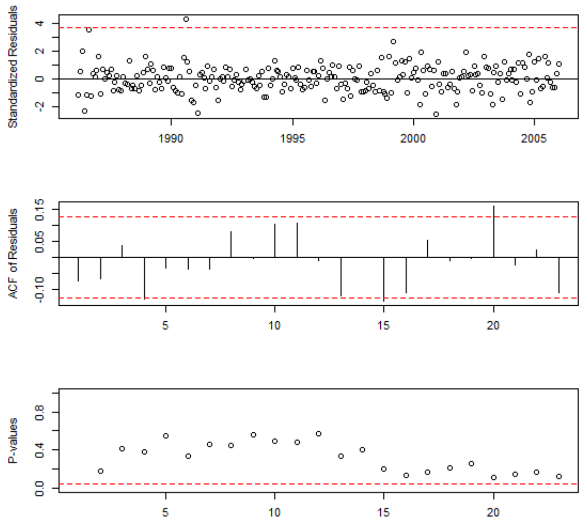
The Ljung-Box plot looks better than that for the AR(1) model.

```
tsdiag(mod3,main='Model 3')
```



Model 3 diagnostics are similar to those of Model 1 with the exception that the Ljung-Box statistics are better as shown in the bottom display.

```
tsdiag(mod4,main='Model 4')
```



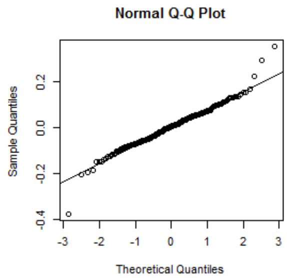
Model 4 diagnostics are similar to those for Models 1 and 3.

The Ljung-Box statistics are the best of the lot as shown in the bottom display.

Based on the AIC value and the Ljung-Box picture, Model 4 looks like it could be the best model.

(e) Look at the Normality of the error terms for Model 4.

```
win.graph(width=3,height=3,pointsize=8)
qqnorm(residuals(mod4))
qqline(residuals(mod4))
shapiro.test(residuals(mod4))

#        Shapiro-Wilk normality test

#data:  residuals(mod4)
#W = 0.9688, p-value = 3.937e-05
```



Both the QQ plot and the low $p-$ value of the Shapiro-Wilk test indicate that we should reject the Null Hypothesis that the error terms are Normally distributed. This could be caused by the outliers in the series.