Tutorial Sheet 5 Solutions

1

Identify as specific ARIMA models, i.e., what are p, d and q and what are the values of the parameters- the ϕ's and the θ's?

(a) $Y_t = Y_{t-1} - 0.25Y_{t-2} + e_t - 0.1e_{t-1}$.

This looks like an combination of AR(2) and MA(1), i.e., ARMA(2,1), with $\phi_1 = 1$, $\phi_2 = -0.25$ and with $\theta_1 = 0.1$.

If it is an ARMA model, it must be stationary, thus we need to check the stationarity conditions. These pertain to the AR component.

$\phi_1 + \phi_2 = 0.75 < 1$.

$\phi_2 - \phi_1 = -1.25 < 1$.

$|\phi_2| = 0.25 < 1$.

So the process is a stationary ARMA(2,1) model with $\phi_1 = 1$, $\phi_2 = -0.25$ and $\theta_1 = 0.1$

(b) $Y_t = 2Y_{t-1} - Y_{t-2} + e_t$.

Initially, it looks like an AR(2) model, but $\phi_1 + \phi_2 = 2 - 1 = 1$ which is strictly not < 1.

Rewrite it as $Y_t - Y_{t-1} = Y_{t-1} - Y_{t-2} + e_t$, i.e., $\nabla Y_t = \nabla Y_{t-1} + e_t$.

This gives an AR(1) process in the difference of the series.

However, here $\phi_1 = 1$ which is strictly not < 1.

$Y_t - Y_{t-1} = Y_{t-1} - Y_{t-2} + e_t$ can be re-written as a second difference. i.e.,

$\nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2} = e_t$, which can be seen to be white noise.

Thus this process is IMA(2,0).

(c) $Y_t = 0.5Y_{t-1} - 0.5Y_{t-2} + e_t - 0.5e_{t-1} + 0.25e_{t-2}$.

The AR part is stationary because

$\phi_1 + \phi_2 = 0 < 1$.

$\phi_2 - \phi_1 = -1 < 1$.

$|\phi_2| = 0.5 < 1$.

For the MA part, we have $-\theta_1 = -0.5$ and $-\theta_2 = 0.25$, and so $\theta_1 = 0.5$ and $\theta_2 = -0.25$.

Applying the same equations, we have

$\theta_1 + \theta_2 = 0.25 < 1$.

$\theta_2 - \theta_1 = -0.75 < 1$.

$|\theta_2| = 0.25 < 1$.

Thus the MA part is invertible.

Therefore the model is a stationary and invertible ARMA(2,2) model with $\phi_1 = 0.5$, $\phi_2 = -0.5$, $\theta_1 = 0.5$ and $\theta_2 = -0.25$.

2

For each of the ARIMA models below, give the values for $E(\nabla Y_t)$ and $Var(\nabla Y_t)$.

(a) $Y_t = 3 + Y_{t-1} + e_t - 0.75e_{t-1}$.

$$\begin{aligned}\nabla Y_t &= Y_t - Y_{t-1} \\ &= 3 + e_t - 0.75e_{t-1}.\end{aligned}$$

$$\begin{aligned}E(\nabla Y_t) &= E(3 + e_t - 0.75e_{t-1}) \\ &= 3.\end{aligned}$$

$$\begin{aligned}Var(\nabla Y_t) &= Var(3 + e_t - 0.75e_{t-1}) \\ &= Var(e_t) + (-0.75)^2 Var(e_{t-1}) \\ &= \frac{25}{16}\sigma_e^2.\end{aligned}$$

(b) $Y_t = 10 + 1.25Y_{t-1} - 0.25Y_{t-2} + e_t - 0.1e_{t-1}$.

$$\begin{aligned}
\nabla Y_t &= Y_t - Y_{t-1} \\
&= 10 + 0.25Y_{t-1} - 0.25Y_{t-2} + e_t - 0.1e_{t-1} \\
&= 10 + 0.25(Y_{t-1} - Y_{t-2}) + e_t - 0.1e_{t-1}.
\end{aligned}$$

So the model is a stationary and invertible ARIMA(1,1,1) model with $\phi = 0.25$, (i.e., < 1) and $\theta = 0.1$ and $\theta_0 = 10$.

Hence

$$\begin{aligned}
E(\nabla Y_t) &= \frac{\theta_0}{1-\phi} \qquad \text{TSLecture5 Page 22} \\
&= \frac{10}{1-0.25} \\
&= \frac{40}{3}. \\
Var(\nabla Y_t) &= \frac{1-2\phi\theta+\theta^2}{1-\phi^2}\sigma_e^2 \qquad \text{TSLecture4c Page 14} \\
&= \frac{1-2(0.25)(0.1)+(0.1)^2}{1-(0.25)^2}\sigma_e^2 \\
&= 1.024\sigma_e^2.
\end{aligned}$$

(c) $Y_t = 5 + 2Y_{t-1} - 1.7Y_{t-2} + 0.7Y_{t-3} + e_t - 0.5e_{t-1} + 0.25e_{t-2}$.

The AR characteristic equation is given by:

$$1 - 2x + 1.7x^2 - 0.7x^3 = 0.$$

We can get the roots of this as follows:

```
polyroot(c(1,-2,1.7,- 0.7))
#[1] 1.0000000-0.0000000i 0.7142857+0.9583148i 0.7142857-0.9583148i
```

$x = 1$ is a root.

We can divide this into the AR characteristic polynomial to get:

$$1 - 2x + 1.7x^2 - 0.7x^3 = (1 - x)(1 - x + 0.7x^2).$$

This shows that the first difference is needed, and after this, a stationary AR(2) with complex roots results. See page 8 of TSLecture5.

Thus the model may be re-written as:

$$\begin{aligned}
Y_t - Y_{t-1} &= 5 + Y_{t-1} - 1.7Y_{t-2} + 0.7Y_{t-3} + e_t - 0.5e_{t-1} + 0.25e_{t-2} \\
\nabla Y_t &= 5 + Y_{t-1} - Y_{t-2} - 0.7Y_{t-2} + 0.7Y_{t-3} + e_t - 0.5e_{t-1} + 0.25e_{t-2} \\
&= 5 + \nabla Y_{t-1} - 0.7\nabla Y_{t-2} + e_t - 0.5e_{t-1} + 0.25e_{t-2}.
\end{aligned}$$

So the model is an ARIMA(2,1,2) model with $\phi_1 = 1$, $\phi_2 = -0.7$, $\theta_1 = 0.5$, $\theta_2 = -0.25$ and $\theta_0 = 5$.

$$\begin{aligned}
E(\nabla Y_t) &= \frac{\theta_0}{1 - \phi_1 - \phi_2} \\
&= \frac{5}{1 - 1 - (-0.7)} \\
&\approx 7.14.
\end{aligned}$$

Variance for an $ARMA(2, 2)$ in the difference has not been covered on this course.

3

Consider two models:

$$
\begin{aligned}
A : Y_t &= 0.9Y_{t-1} + 0.09Y_{t-2} + e_t. \\
B : Y_t &= Y_{t-1} + e_t - 0.1e_{t-1}.
\end{aligned}
$$

(a) Identify each as a specific ARIMA model, i.e, what are p, d and q and what are the values of the parameters- the ϕ's and the θ's?

A: Since

$\phi_1 + \phi_2 = 0.99 < 1$,

$\phi_2 - \phi_1 = -0.81 < 1$,

and $|\phi_2| = 0.09 < 1$,

then the process is a stationary AR(2) model with $\phi_1 = 0.9$ and $\phi_2 = 0.09$.

B: Since

$Y_t - Y_{t-1} = e_t - 0.1e_{t-1}$,

then the process is an IMA(1,1) model with $\theta = 0.1$.

(b) In what way are the two models different?

One is stationary while the other is non-stationary.

(c) In what way are the two models similar? (Compare ψ weights and π weights.)

When the AR process is expressed as an MA(∞) process, the ψ weights are the e coefficients.

See page 30 of TSLecture4b.

```
psi=NULL
phi1=0.9
phi2=0.09
max.lag=20
psi[1]=1
psi[2]=phi1

for (k in 3:max.lag)
psi[k]=phi1*psi[k-1]+phi2*psi[k-2]
psi

# [1] 1.0000000 0.9000000 0.9000000 0.8910000 0.8829000 0.8748000 0.8667810
# [8] 0.8588349 0.8509617 0.8431607 0.8354312 0.8277725 0.8201841 0.8126652
#[15] 0.8052152 0.7978336 0.7905196 0.7832726 0.7760921 0.7689775
```

Alternatively, we can use the ARMAtoMA function as follows:

```
ARMAtoMA(ar=c(phi1,phi2),lag.max=20)

# [1] 0.9000000 0.9000000 0.8910000 0.8829000 0.8748000 0.8667810 0.8588349
# [8] 0.8509617 0.8431607 0.8354312 0.8277725 0.8201841 0.8126652 0.8052152
#[15] 0.7978336 0.7905196 0.7832726 0.7760921 0.7689775 0.7619280
```

For the IMA(1,1) model, we have

$$
\begin{aligned}
Y_t &= Y_{t-1} + e_t - 0.1e_{t-1}. \\
Y_{t-1} &= Y_{t-2} + e_{t-1} - 0.1e_{t-2}. \\
Y_t &= Y_{t-2} + e_{t-1} - 0.1e_{t-2} + e_t - 0.1e_{t-1} \\
&= Y_{t-2} + e_t + (1-0.1)e_{t-1} - 0.1e_{t-2} \\
&= Y_{t-3} + e_{t-2} - 0.1e_{t-3} + e_t + (1-0.1)e_{t-1} - 0.1e_{t-2}
\end{aligned}
$$

$$= \quad Y_{t-3} + e_t + (1-0.1)e_{t-1} + (1-0.1)e_{t-2} - 0.1e_{t-3}.$$

Thus the ψ weights for the IMA(1,1) model are 1, $1-0.1 = 0.9$, $1-0.1 = 0.9$, $1-0.1 = 0.9, \ldots$

Comparing these values with the computer-generated output, the ψ weights for the two models are quite similar for many lags.

When the MA process is expressed as an AR(∞) process, the π weights are the Y coefficients.

The π weights for the IMA(1,1) model are obtained via

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \ldots + \pi_k Y_{t-k} + \ldots + e_t.$$

We have

$$\begin{aligned}
Y_t &= Y_{t-1} + e_t - 0.1e_{t-1}. \\
e_t &= Y_t - Y_{t-1} + 0.1e_{t-1} \\
&= Y_t - Y_{t-1} + 0.1(Y_{t-1} - Y_{t-2} + 0.1e_{t-2}) \\
&= Y_t - 0.9Y_{t-1} - 0.1Y_{t-2} + 0.01e_{t-2} \\
&= Y_t - 0.9Y_{t-1} - 0.1Y_{t-2} + 0.01(Y_{t-2} - Y_{t-3} + 0.1e_{t-3}) \\
&= Y_t - 0.9Y_{t-1} - 0.09Y_{t-2} - 0.01Y_{t-3} + 0.001e_{t-3} \\
&= Y_t - 0.9Y_{t-1} - 0.09Y_{t-2} - 0.01Y_{t-3} + 0.001(Y_{t-3} - Y_{t-4} + 0.1e_{t-4}) \\
&= Y_t - 0.9Y_{t-1} - 0.09Y_{t-2} - 0.009Y_{t-3} - 0.001Y_{t-4} + 0.0001e_{t-4}.
\end{aligned}$$

So $\pi_1 = (1 - 0.1) = 0.9$, $\pi_2 = (1 - 0.1)(0.1) = 0.09$, $\pi_3 = (1 - 0.1)(0.1)^2 = 0.009$, etc..

Thus $\pi_k = (1 - \theta)\theta^{k-1}$ for $k = 1, 2 \ldots$.

The first two π weights for the two models are identical and the remaining π weights are close to zero. Thus it is very similar to the AR(2) model. These two models would be essentially impossible to distinguish in practice.

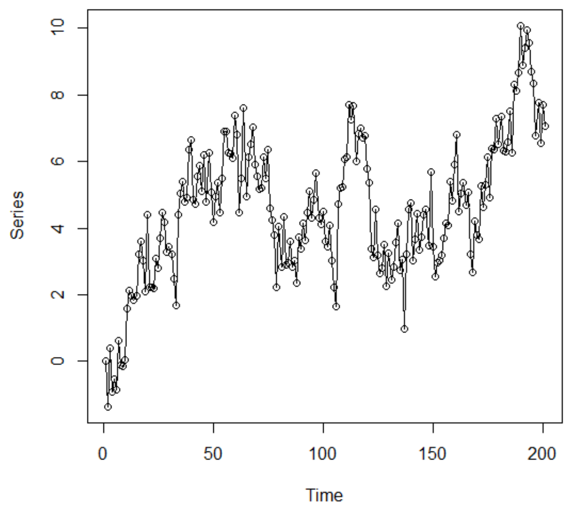
4

Non-stationary ARIMA series can be simulated by first simulating the corresponding stationary ARMA series and then "integrating" it (really partially summing it).

Use statistical software to simulate a variety of IMA(1,1) and IMA(2,2) series with a variety of parameter values.

Note any stochastic "trends" in the simulated series.

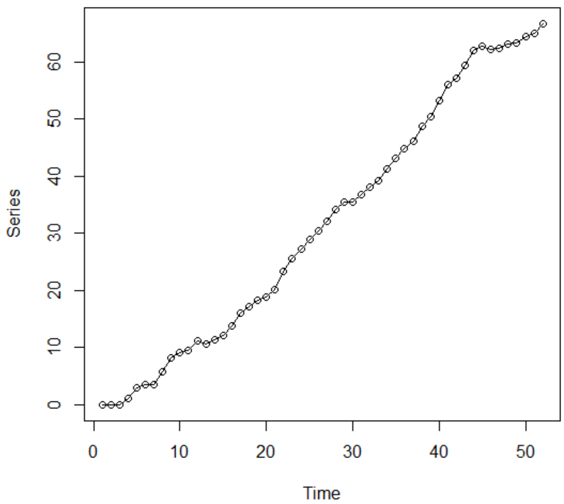
```
plot(arima.sim(model=list(order=c(0,1,1),ma=-0.5),n=200),type='o',ylab='Series')
```



Do this several times with the same parameters to see the possible variation, and then, change the various parameters, *ma* and *n*.

For the IMA(2,2) use, for example,

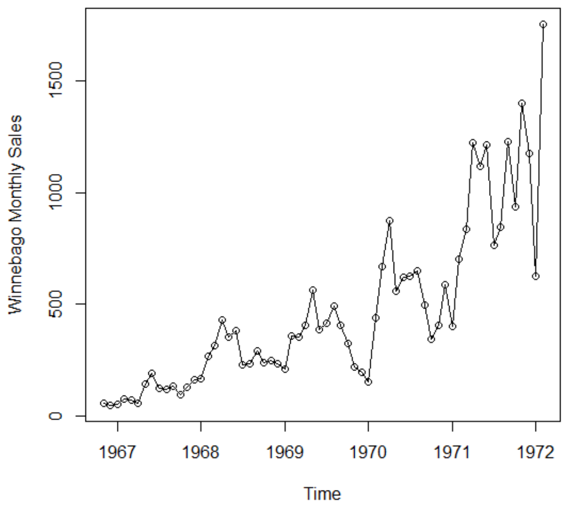
```
plot(arima.sim(model=list(order=c(0,2,2),ma=c(-0.7,-0.1)),n=50),
type='o',ylab='Series')
```



5
The datafile 'winnebago' contains monthly units sales of recreational vehicles from Winnebago Inc., from November 1966 through February 1972.

(a) Display and interpret the time series plot for these data.

```
data(winnebago)
plot(winnebago,type='o',ylab='Winnebago Monthly Sales')
```

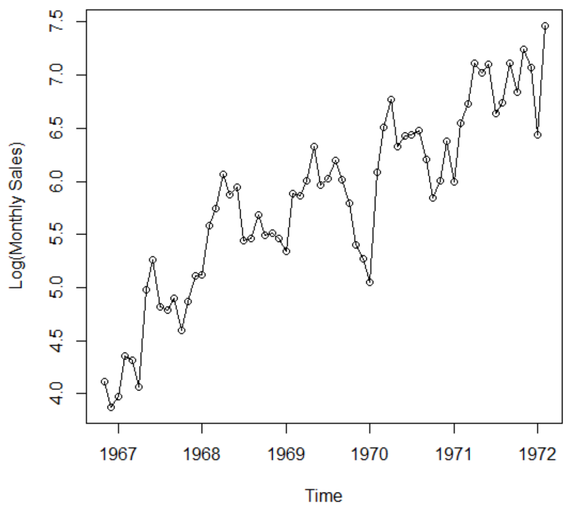


The series increases over time and the variation is larger as the series level gets higher.

This tells us we should look to take the log of the series.

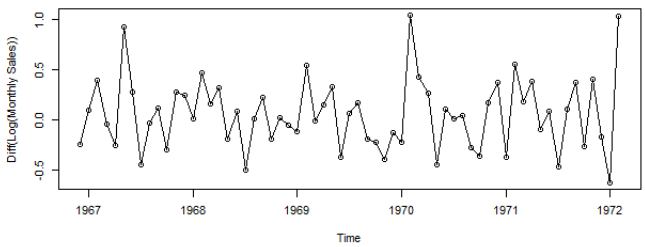
(b) Now take natural logarithms of the monthly sales figures and display the time series plot of the transformed values. Describe the effects of the logarithms on the behaviour of the series.

```
plot(log(winnebago),type='o',ylab='Log(Monthly Sales)')
```



The series still increases over time, but the variation around the general level is quite similar all levels of the series.

```
plot(diff(log(winnebago)),type='o',ylab='Diff(Log(Monthly Sales))')
```



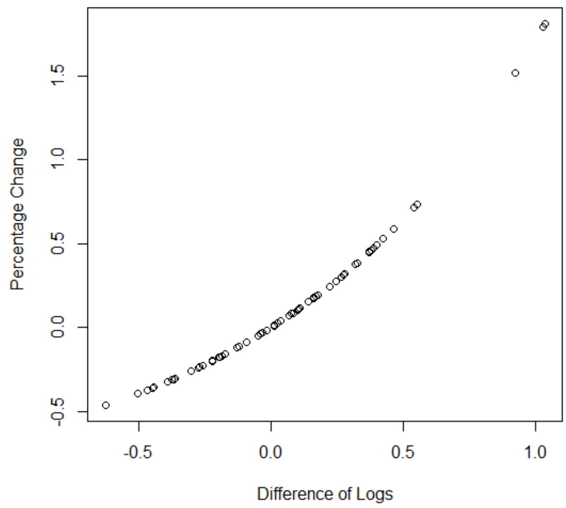
The difference of the log of the series is stationary.

(c) Calculate the fractional relative changes, $(Y_t - Y_{t-1})/Y_{t-1}$, and compare them with the differences of natural logarithms, $\nabla \log(Y_t) = \log(Y_t) - \log(Y_{t-1})$. How do they compare for smaller values and for larger values?

```
percentage=na.omit((winnebago-zlag(winnebago))/zlag(winnebago))

plot(x=diff(log(winnebago))[-1],y=percentage[-1],
ylab='Percentage Change',xlab='Difference of Logs')

cor(diff(log(winnebago))[-1],percentage[-1])
#[1] 0.9646886
```



If there were a perfect relationship, the above plot would be a straight line.

While the relationship is good, it is not perfect.

At lower values, the relationship is linear, but is more curvilinear at higher values.

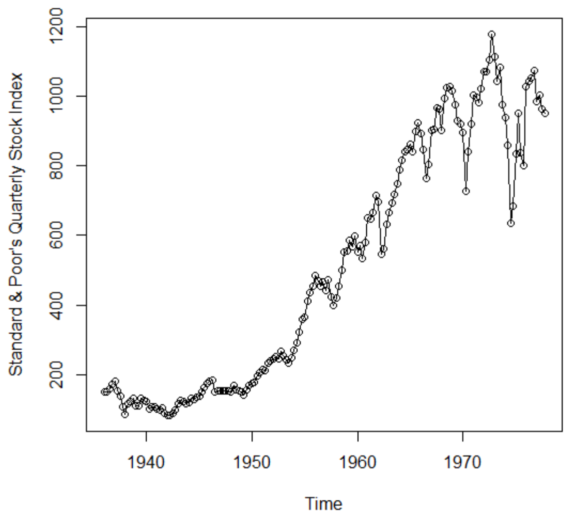
The correlation coefficient of 0.96 tells us that there is strong agreement between the two measures.

The seasonality has not been modelled.

6
The datafile 'SP' contains quarterly Standard & Poor's Composite Index stock price values from the first quarter of 1936 through the fourth quarter of 1977.

(a) Display and interpret the time series plot for these data.

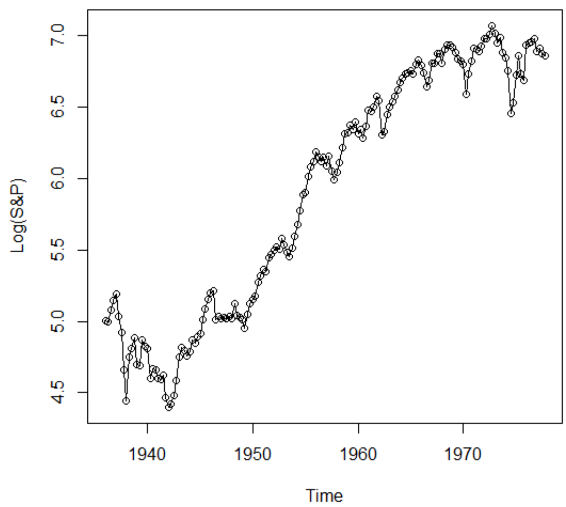
```
data(SP)
plot(SP,type='o',ylab='Standard & Poor's Quarterly Stock Index')
```



The series increases over time and the variation is larger as the series level gets higher.

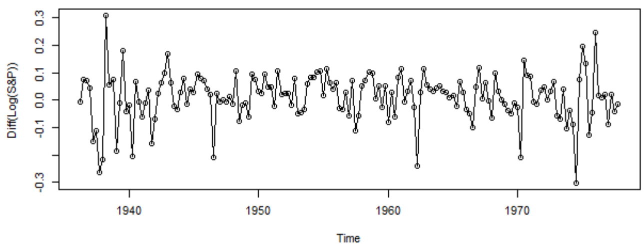
(b) Now take natural logarithms of the monthly sales figures and display the time series plot of the transformed values. Describe the effects of the logarithms on the behaviour of the series.

```
plot(log(SP),type='o',ylab='Log(S&P)')
```



The series still increases over time, but the variation around the general level is stabilised.

```
plot(diff(log(SP)),type='o',ylab='Diff(Log(S&P))')
```

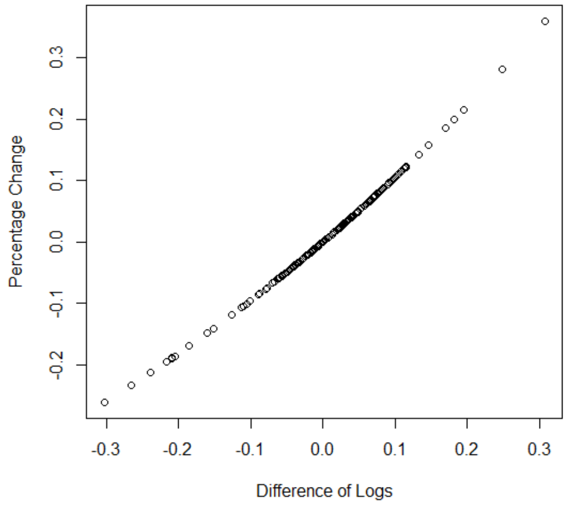


The trend has been removed.

(c) Calculate the fractional relative changes, $(Y_t - Y_{t-1})/Y_{t-1}$, and compare them with the differences of natural logarithms, $\nabla \log(Y_t) = \log(Y_t) - \log(Y_{t-1})$. How do they compare for smaller values and for larger values?

```
percentage=na.omit((SP-zlag(SP))/zlag(SP))
plot(x=diff(log(SP))[-1],y=percentage[-1],
ylab='Percentage Change',xlab='Difference of Logs')

cor(diff(log(SP))[-1],percentage[-1])
#[1] 0.9963475
```

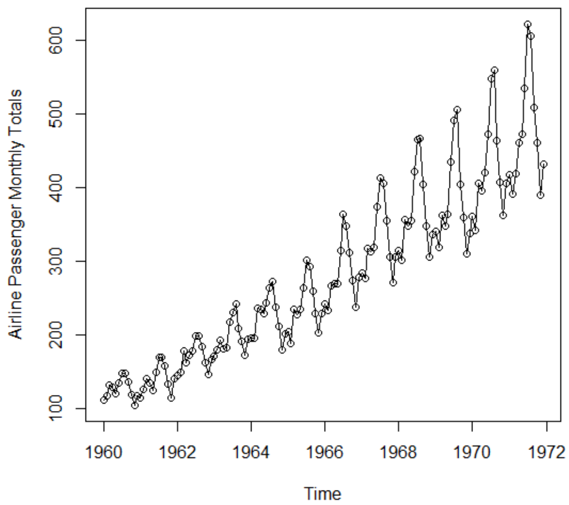


The agreement between the two measures is very good with correspondingly high correlation coefficient.

7
The datafile 'airpass' contains international airline passenger monthly totals in thousands flown from January 1960 through December 1971.

(a) Display and interpret the time series plot for these data.

```
data(airpass)
plot(airpass,type='o',ylab='Airline Passenger Monthly Totals')
```

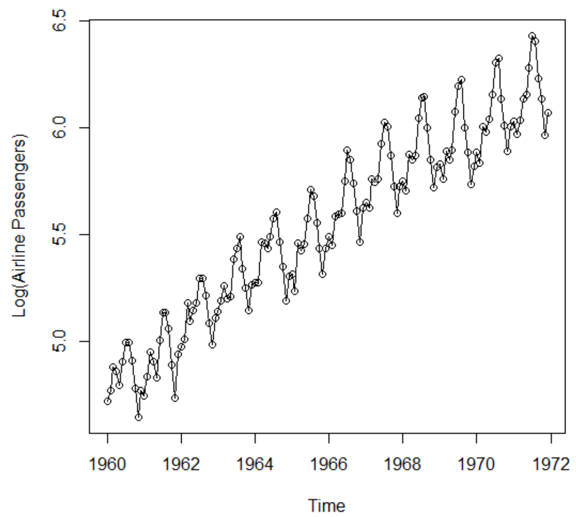


The series increases over time and the variation is larger as the series level gets higher.

There is also evidence of seasonality.

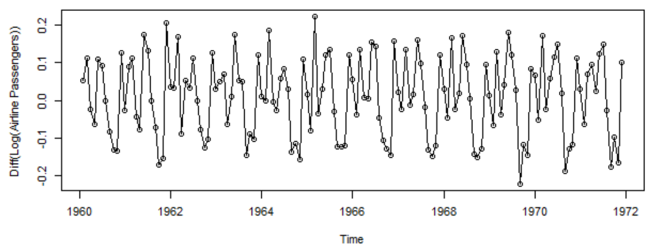
(b) Now take natural logarithms of the monthly sales figures and display the time series plot of the transformed values. Describe the effects of the logarithms on the behaviour of the series.

```
plot(log(airpass),type='o',ylab='Log(Airline Passengers)')
```



The variation is similar at all levels.

```
plot(diff(log(airpass)),type='o',ylab='Diff(Log(Airline Passengers))')
```



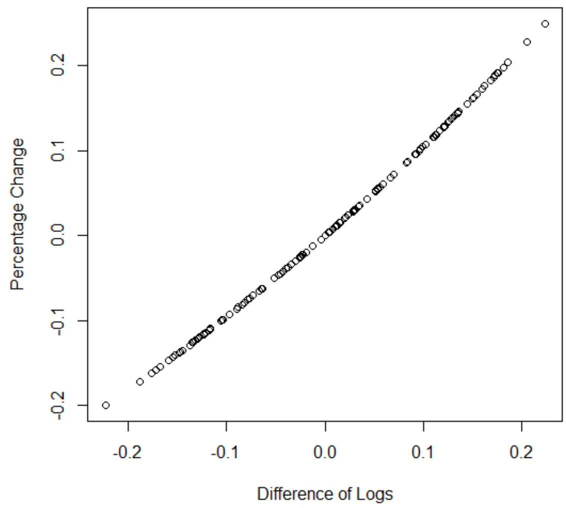
There is still evidence of seasonality in the transformed data, but the mean and variance are constant.

(c) Calculate the fractional relative changes, $(Y_t - Y_{t-1})/Y_{t-1}$, and compare them with the differences of natural logarithms, $\nabla \log(Y_t) = \log(Y_t) - \log(Y_{t-1})$. How do they compare for smaller values and for larger values?

```
percentage=na.omit((airpass-zlag(airpass))/zlag(airpass))

plot(x=diff(log(airpass))[-1],y=percentage[-1],
ylab='Percentage Change',xlab='Difference of Logs')

cor(diff(log(airpass))[-1],percentage[-1])
#[1] 0.9986814
```



There is excellent agreement between the two transformed series in this case.

The correlation coefficient is 0.999.

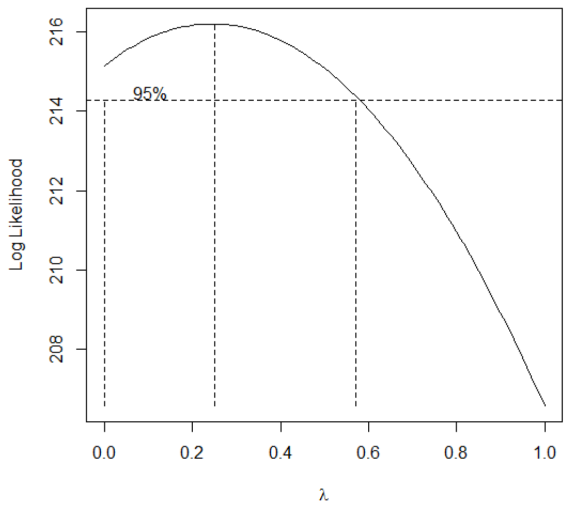
Either transformation would be extremely helpful in modelling this series further.

8
The datafile 'larain' contains yearly rainfall in Los Angeles as shown in Lecture 1.

The QQ plot of these data in Lecture 3 showed us that the data were not normal.

(a) Use software to produce a BoxCox plot and determine the best value of 'λ' for a power transformation of the series.

```
data(larain)
laraintransform<-BoxCox.ar(larain, lambda=seq(0,1,0.01))
laraintransform
```



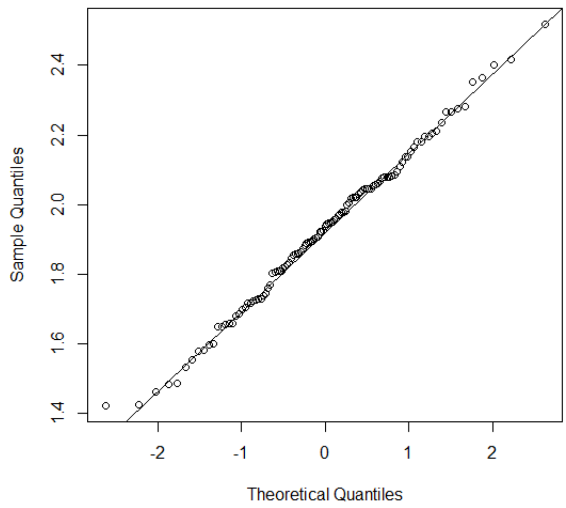
The mle is $\lambda = 0.25$ and the 95% confidence interval includes the log transformation of $\lambda = 0$ and the square root transformation of $\lambda = 0.5$.

(b) Display a QQ plot of the transformed data. Are they more normal?

```
qqnorm((larain)^(0.25),main='')
qqline((larain)^(0.25))
shapiro.test((larain)^(0.25))
#Shapiro-Wilk normality test
#
#data:  (larain)^0.25
#W = 0.9941, p-value = 0.9096
```

For the Shapiro-Wilk test, the Null hypothesis is that the data are Normally distributed.

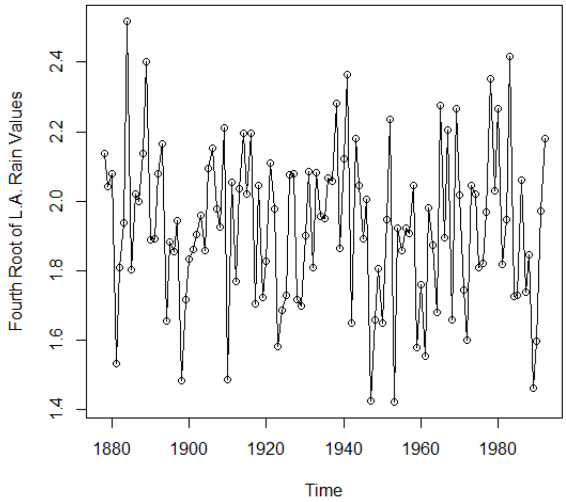
If this is the case, the test statistic, (W), will be small and the probability of seeing such a value by chance will be high. Thus the $p-$ value will be greater than 0.05.



The values transformed by the fourth root look normal.

(c) Produce the time series plot of the transformed values.

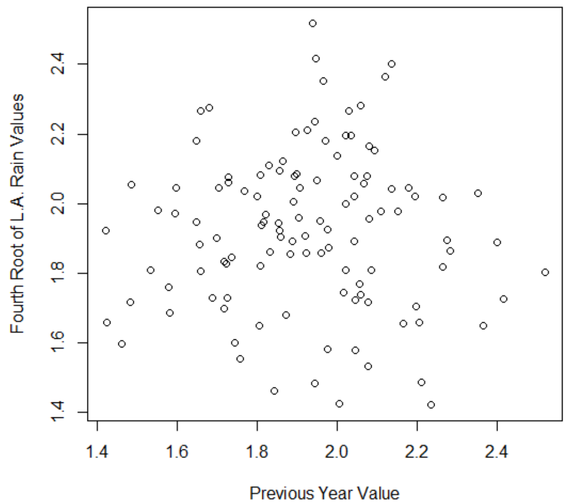
```
plot(larain^(0.25),type='o',ylab='Fourth Root of L.A. Rain Values')
```



The transformed valued could be considered to be white noise with a non-zero mean.

(d) Use the transformed values to display a plot of Y_t versus Y_{t-1}. Should we expect the transformation to change the dependence or lack of dependence in the series?

```
plot(y=(larain)^(0.25),x=zlag((larain)^(0.25)),
ylab='Fourth Root of L.A. Rain Values',
xlab='Previous Year Value')
```



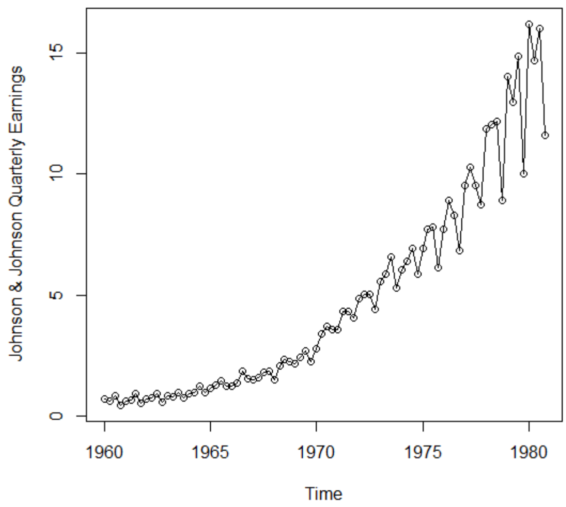
The lack of correlation of any kind between year values is clear.

Instantaneous transformations cannot induce correlation where none was present.

9
The datafile 'JJ' contains quarterly earnings per share for the Johnson & Johnson Company from 1960 through 1980.

(a) Display and interpret the time series plot for these data.

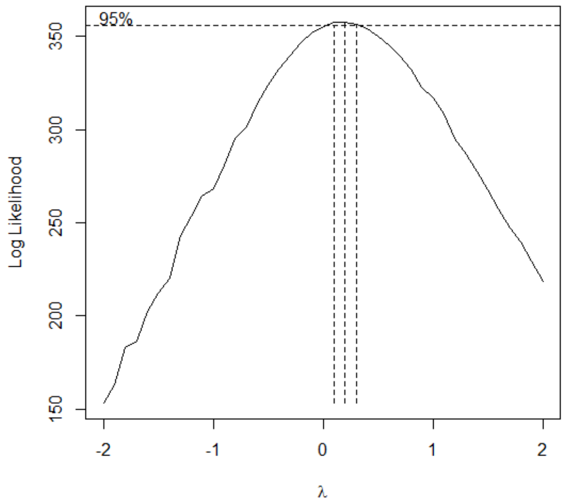
```
data(JJ)
plot(JJ,type='o',ylab='Johnson & Johnson Quarterly Earnings')
```



Once again an increasing trend with increasing variance over time.

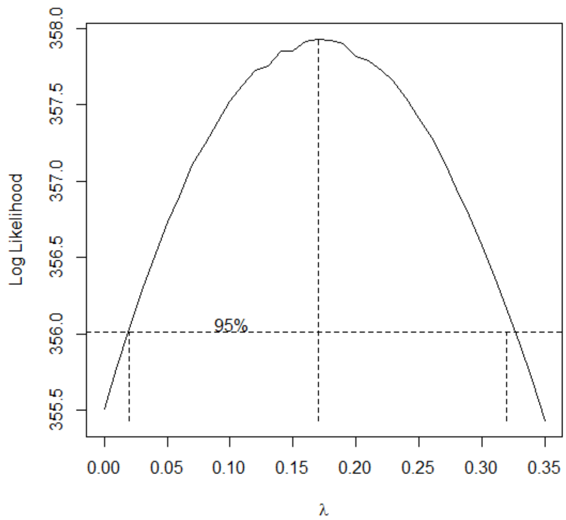
(b) Use software to produce a BoxCox plot and determine the best value of 'λ' for a power transformation of the series.

```
BC=BoxCox.ar(JJ)
BC
```



Restrict to a smaller range of λ based on initial BoxCox plot.

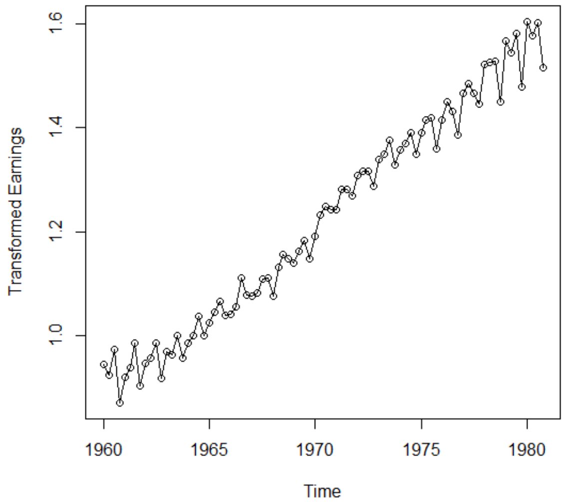
```
BC=BoxCox.ar(JJ,lambda=seq(0.0,0.35,0.01))
BC
```



The mle is $\lambda = 0.17$.

(c) Display the time series plot of the transformed values. Does this plot suggest that a stationary model might be appropriate?

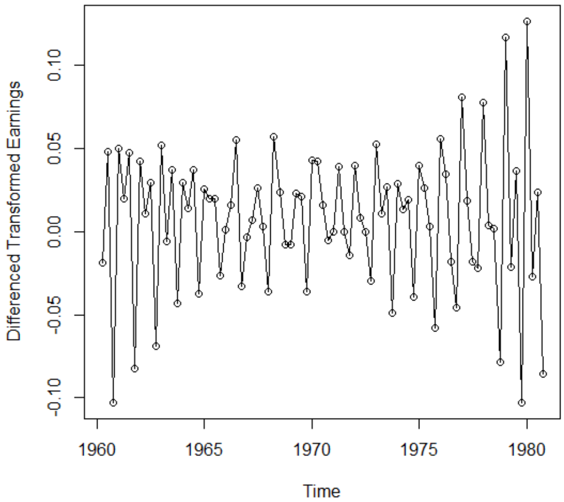
```
plot((JJ)^(0.17),type='o',ylab='Transformed Earnings')
```



The variance has been stabilised, but the strong trend must be accounted for before we can entertain a stationary model.

(d) Display the time series plot of the differences of the transformed values. Does this plot suggest that a stationary model might be appropriate?

```
plot(diff((JJ)^(0.17)),type='o',ylab='Differenced Transformed Earnings')
```



The trend is now gone, but the variation does not appear to be constant across time.

There may be quarterly seasonality to deal with.