Module 4 - Models for Nonstationary Time Series

MATH1318 Time Series Analysis

Prepared by Dr Haydar Demirhan based on the textbook by Cryer and Chan, Time Series Analysis with R, Springer.

Nomenclature

$\{X_t\}$: A zero-mean time series.
$\nabla^d Y_t$: The dth difference of successive Y-series.
d: Consistently with the above definition, the order of differencing.
M_t: A stochastic or deterministic series.
ARIMA(p,d,q): Integrated autoregressive moving average process of order p, q and d times of differencing.
IMA(d,q): Integrated moving average process of order q and d times of differencing.
ARI(p,d): Integrated autoregressive process of order p and d times of differencing.
$g(x)$: Box-Cox transformed rv.
λ: The parameter of Box-Cox transformation.
$log(\cdot)$: Natural logarithm of inner experssion.

Introduction

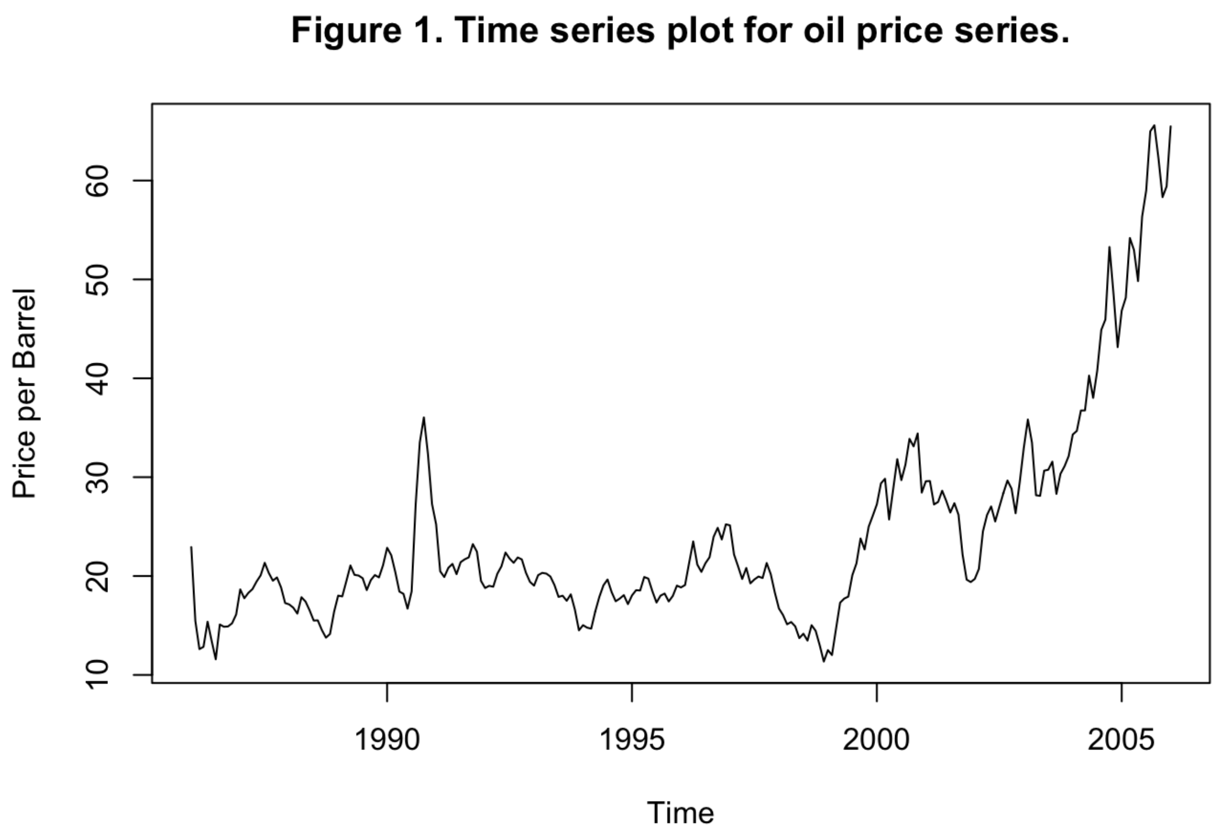
If the mean of time series is not constant over time then it is nonstationary. This means that in the model

$$Y_t = \mu_t + X_t,$$

the term μ_t changes with t and X_t is a zero-mean, stationary series. This type of models is more reasonable than the trend models of Module 2 in the sense that it is hardly possible to have a deterministic trend forever.

The series displayed in Figure 1 exhibits the monthly price of a barrel of crude oil from January 1986 through January 2006.

```
data(oil.price)
plot(oil.price, ylab='Price per Barrel',type='l', main = "Figure 1. Time
        series plot for oil price series.")
```



There is a highly variable trend in the series and a stationary model does not seem to be reasonable. However, one of the nonstationary models that have been described as containing stochastic trends seems to be reasonable.

In this module, we will discuss the models that incorporate stochastic trends. To analyse any time series, we first need to eliminate the effect of any kind of trend. So, we will first deal with the trend and then model the autocorrelation structure in the series.

Stationarity Through Differencing

We can transform a nonstationary time series into a stationary one by differencing. Let us consider the AR(1) process

$$Y_t = \phi Y_{t-1} + e_t.$$

When we assume e_t is uncorrelated with $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$, we must have $|\phi| < 1$. But what happens if $|\phi| > 1$?

For example, let us consider the case with $\phi = 3$. In this case, we have

$$Y_t = 3Y_{t-1} + e_t.$$

Iterating into the past we get

$$Y_t = e_t + 3e_{t-1} + 3^2 e_{t-2} + \cdots + 3^{t-1} e_{t-1} + 3^t Y_0.$$

The influence of past values of Y_t and e_t blow out as time passes as seen in the following table.

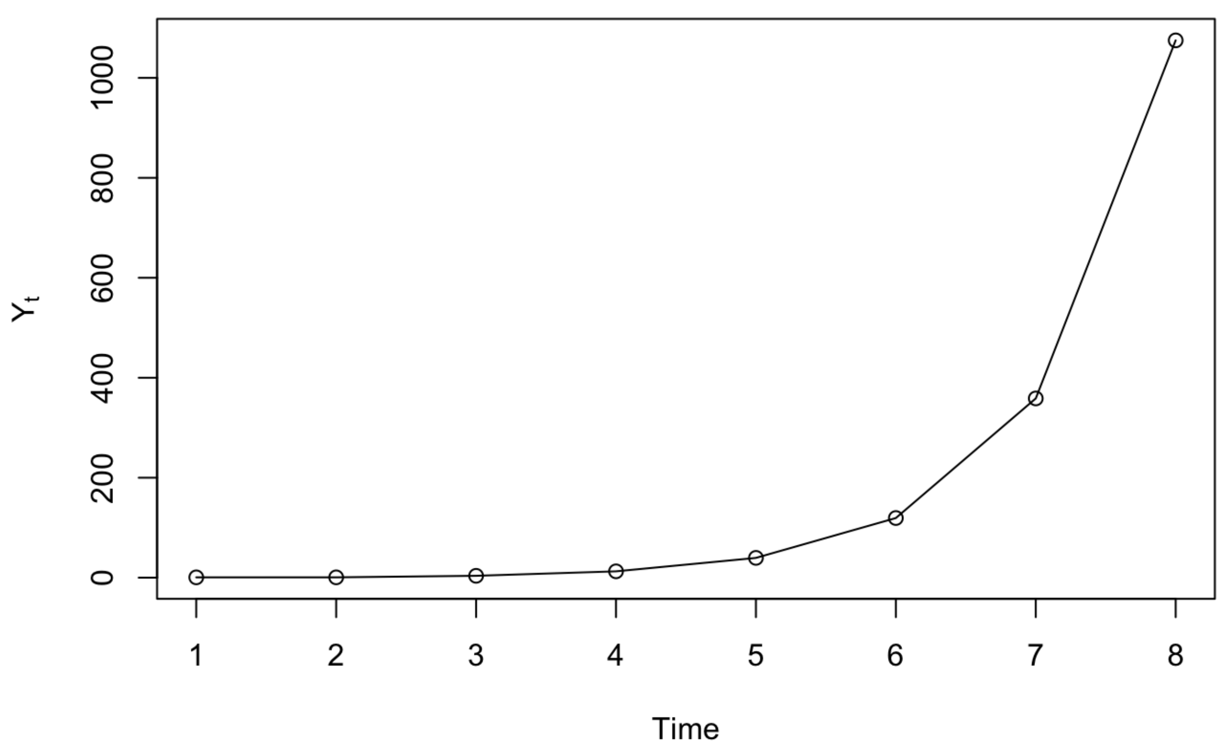
t	1	2	3	4	5	6	7	8
e_t	0.63	−1.25	1.80	1.51	1.56	0.62	0.64	–0.98
Y_t	0.63	0.64	3.72	12.67	39.57	119.33	358.63	1074.91

.

Figure 2 displays an example with initial conditions $Y_0 = 0$ and the white noise is generated from the standard normal distribution.

```
data(explode.s)
plot(explode.s,ylab=expression(Y[t]),type='o', main = "Figure 2. An examp
        le series for explosive behavior.")
```

Figure 2. An example series for explosive behavior.



It is also possible to observe the explosive behaviour of such a model in the model's variance and covariance functions.

$$Var(Y_t) = \frac{1}{8}(9^t - 1)\sigma_e^2$$

and

$$Cov(Y_t, Y_{t-1}) = \frac{3^k}{8}(9^{t-k} - 1)\sigma_e^2.$$

Consequently, for large t and moderate k we get

$$Corr(Y_t, Y_{t-1}) = 3^k\sqrt{\frac{9^{t-k} - 1}{p^t - 1}} \approx 1.$$

Instead, we can use a more reasonable model with $\phi = 1$. If $\phi = 1$, AR(1) model becomes

$$Y_t = Y_{t-1} + e_t.$$

This relationship is satisfied by the random walk process as well. Alternatively, we can write this model as follows:

$$\nabla Y_t = e_t.$$

where $\nabla Y_t = Y_t - Y_{t-1}$ is **the first difference** of Y_t. The random walk then is easily extended to a more general model whose first difference is some stationary process other than white noise.

For the second order differencing, we do the following calculations:

$$\begin{aligned}\nabla^2 Y_t &= \nabla Y_t - \nabla Y_{t-1} \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2}\end{aligned}$$

There will be $n - 1$ observations in the second diffferenced $\nabla^2 Y_t$ series.

Several somewhat different sets of assumptions can lead to models whose first difference is a stationary process. Consider

$$Y_t = M_t + X_t.$$

where M_t is a (either deterministic or stochastic) series that is changing only slowly over time. If we assume that M_t is approximately constant over every two consecutive time points, we might estimate M_t at t by choosing β_0 so that

$$\sum_{j=0}^{1}(Y_{t-j} - \beta_{0,t})^2$$

is minimised. This clearly leads to

$$\hat{M}_t = \frac{1}{2}(Y_t + Y_{t-1})$$

and the “detrended” series at time t is then

$$Y_t - \hat{M}_t = Y_t - \frac{1}{2}(Y_t + Y_{t-1}) = \frac{1}{2}(Y_t - Y_{t-1}) = \frac{1}{2}\nabla Y_t$$

Another set of assumptions might be that M_t is stochastic and changes slowly over time governed by a random walk model. Suppose, for example, that

$$Y_t = M_t + e_t$$

with

$$M_t = M_{t-1} + \epsilon_t$$

where e_t and ϵ_t are independent white noise series. Then we get

$$\nabla Y_t = \nabla M_t + \nabla e_t = \epsilon_t + e_t - e_{t-1}$$

which would have the autocorrelation function of an MA(1) series. In either of these situations, we are led to the study of ∇Y_t as a stationary process.

Let’s turn back to the oil prices example to see the effect of differencing in practice. Figure 3 and 4 display the time series plot and the difference of logarithms of oil price series.

```
par(mfrow=c(2,1))
par(cex=0.7)
plot(oil.price, ylab='Price per Barrel',type='l', main = "Figure 3. Time
         series plot for oil price series.")
first.diff = diff(log(oil.price))
plot(first.diff,ylab='Change in Log(Price)',type='l', main = "Figure 4. T
        ime series plot for the first difference of log-transformed oil p
        rice series.")
```

Figure 3. Time series plot for oil price series.

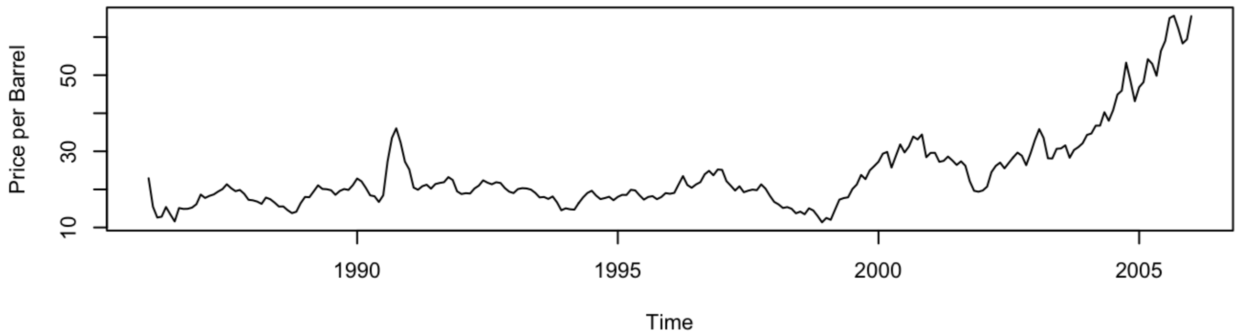
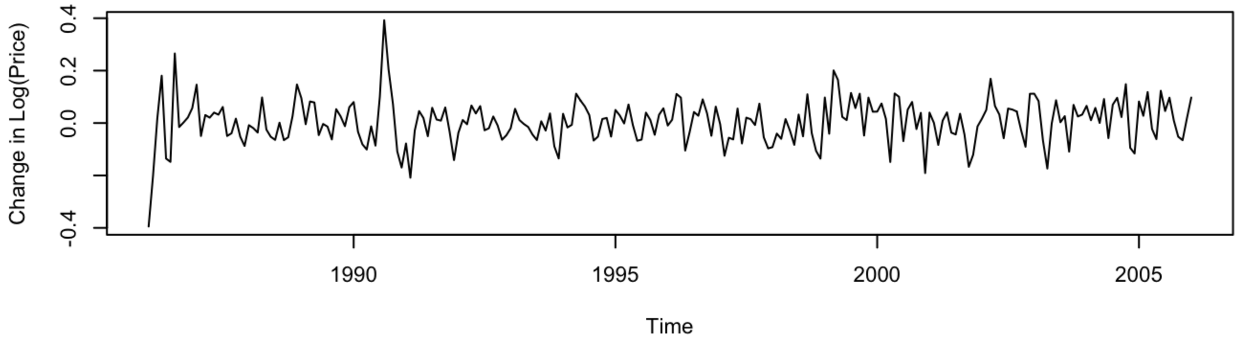


Figure 4. Time series plot for the first difference of log-transformed oil price series.



The differenced series looks much more stationary when compared with the original time series.

In a similar way, we can also make assumptions that lead to stationary second-difference models. Again, let's have the following model

$$Y_t = M_t + X_t.$$

But assume here that M_t is linear in time over *three* consecutive time points. We can now estimate M_t at the middle time point t by choosing and to minimize

$$\sum_{j=-1}^{1} (Y_{t-j} - (\beta_{0,t} + j\beta_{1,t}))^2.$$

The solution yields

$$\hat{M}_t = \frac{1}{3}(Y_{t+1} + Y_t + Y_{t-1})$$

and thus the detrended series is

$$
\begin{aligned}
Y_t - \hat{M}_t &= Y_t - \left(\frac{Y_{t+1} + Y_t + Y_{t-1}}{3}\right) \\
&= \left(-\frac{1}{3}\right)(Y_{t+1} - 2Y_t + Y_{t-1}) \\
&= \left(-\frac{1}{3}\right)\nabla(\nabla Y_{t+1}) \\
&= \left(-\frac{1}{3}\right)\nabla^2(Y_{t+1})
\end{aligned}
$$

a constant multiple of the centered **second difference** of Y_t. Notice that we have differenced twice, but both differences are at lag 1.

In another way, we would assume that

$$Y_t = M_t + e_t.$$

where

$$M_t = M_{t-1} + W_t$$

and

$$W_t = W_{t-1} + \epsilon_t$$

where e_t and ϵ_t are independent white noise series. Here the stochastic trend M_t is such that its “rate of change,” ∇M_t, is changing slowly over time. Then,

$$\nabla Y_t = \nabla M_t + \nabla e_t = W_t + \nabla e_t$$

and

$$
\begin{aligned}
\nabla^2 Y_t &= \nabla W_t + \nabla^2 e_t \\
&= \epsilon_t + (e_t - e_{t-1}) - (e_{t-1} - e_{t-2}) \\
&= \epsilon_t + e_t - 2e_{t-1} + e_{t-2}
\end{aligned}
$$

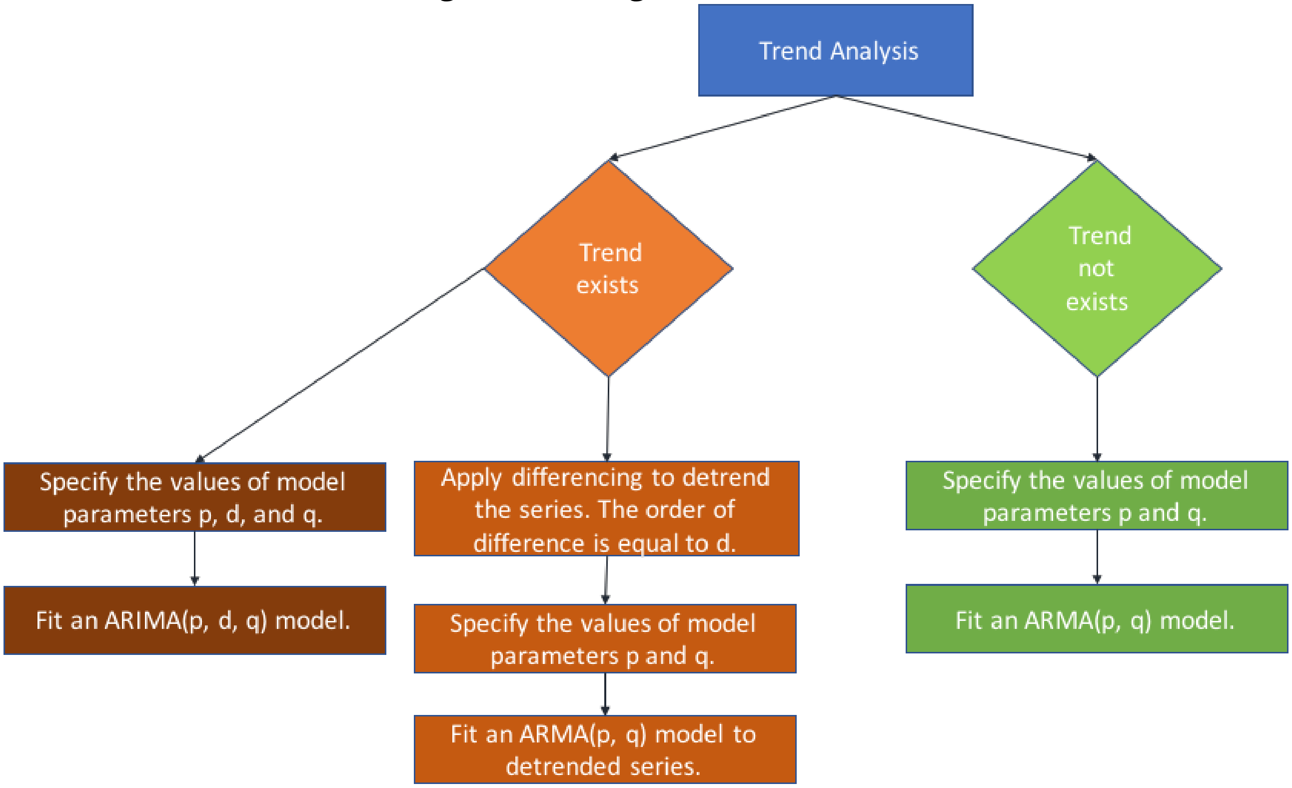
So we get a process with the autocorrelation function of the MA(2) process. The important point is that the second difference of the nonstationary process $\{Y_t\}$ is stationary. This leads us to the general definition of the important integrated autoregressive moving average time series models.

ARIMA Models

When we use differencing operation to make a nonstationary series stationary and then fit an ARMA model, actually we fit an integrated autoregressive moving average model, namely ARIMA(p,d,q). Here the parameter d is the order of differencing. If the dth difference $W_t = \nabla^d Y_t$ is a stationary ARMA process then $\{Y_t\}$ is an ARIMA(p,d,q) process. In practice, we usually set $d = 1$ or $d = 2$.

This equivalence between going on with ARIMA(p,d,q) and ARMA(p,q) after differencing is shown in the diagram in Figure 5.

Figure 5. Fitting ARIMA models.



For an ARIMA(q,1,q) process with $W_t = Y_t - Y_{t-1}$, we have in terms of transformed series,

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \\ - \cdots - \theta_q e_{t-q}$$

and in terms of observed series

$$Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + \cdots + \phi_p(Y_{t-p} - Y_{t-p-1}) \\ + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q},$$

Also, the *difference equation form* of the model is written as

$$Y_t = (1 + \phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + (\phi_3 - \phi_2)Y_{t-3} + \cdots \\ + (\phi_p - \phi_{p-1})Y_{t-p} - \phi_p Y_{t-p-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}.$$

IMA Models

If the process contains no autoregressive terms, we call it an integrated moving average model, namely IMA(d,q).

The model equation of the IMA(1,1) model is

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1}.$$

For this model, the weights on the white noise terms do not die out as we go into the past. $Var(Y_t)$ increases as t increases and it could be quite large. Also, the correlation between Y_t and Y_{t-k} will be strongly positive for many lags $k = 1, 2, \ldots$.

In difference-equation form, the model equation of the IMA(2,2) model is

$$\nabla^2 Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

or

$$Y_t = 2Y_{t-1} - Y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}.$$

The variance of Y_t increases rapidly with t and again is nearly 1 for all moderate k.

Exercise

Use the apps developed for the visualisation of

- Time plots (https://jbaglin.shinyapps.io/ts_app_4_1/)
- ACF plots (https://jbaglin.shinyapps.io/ts_app_4_2/)
- PACF plots (https://jbaglin.shinyapps.io/ts_app_4_3/)

for simulated IMA(d,q) processes and try to observe the following characteristics of the IMA(d,q) process.

Over the time series plot;

- Process values change smoothly due to the increasing variance and the strong, positive neighbouring correlations that dominate the appearance of the plot.
- We need to take dth difference to make an IMA(d,q) process stationary. So we can identify the value of d by the order of differencing that makes the series stationary.

Over the ACF and PACF plots;

- There is a smooth decay in the first lags of the ACF plot when the series is nonstationary.
- The duration of this decay is related to the degree of nonstationary (the true value of d).
- After the elimination of trend by differencing, ACF of the differenced series dies out to show an MA(q) characteristic.
- Accordingly, PACF has an exceptionally large value at the first lag.

ARI Models

If the process contains no moving average terms, we call it an integrated autoregressive model, namely ARI(q,d).

The model equation of the ARI(1,1) model is

$$Y_t = (1 + \phi)Y_{t-1} - \phi Y_{t-2} + e_t$$

where $|\phi| < 1$. ARI(1,1) model shows AR(1) characteristics after taking the first difference.

Exercise

Use the apps developed for the visualisation of

- Time plots (https://jbaglin.shinyapps.io/ts_app_4_4/)
- ACF plots (https://jbaglin.shinyapps.io/ts_app_4_5/)
- PACF plots (https://jbaglin.shinyapps.io/ts_app_4_6/)

for simulated ARI(p,d) processes and try to observe the following characteristics of the ARI(p,d) process.

Over the time series plot;

- Process values change smoothly due to the increasing variance and the strong correlations that dominate the appearance of the plot.
- We need to take dth difference to make an ARI(d,q) process stationary. So we can identify the value of d by the order of differencing that makes the series stationary.

Over the ACF and PACF plots;

- There is a smooth decay in the first lags of the ACF plot when the series is nonstationary.
- The duration of this decay is related to the degree of nonstationary (the true value of d).
- PACF does not exhibit such a smooth pattern.
- After the elimination of trend by differencing, ACF of the differenced series dies out to show an AR(p) characteristic.

Other Transformations

Natural logarithm

There are also some other transformations to ensure stationarity. The logarithm transformation is also a useful method in certain circumstances. We frequently encounter series where increased dispersion seems to be associated with higher levels of the series.

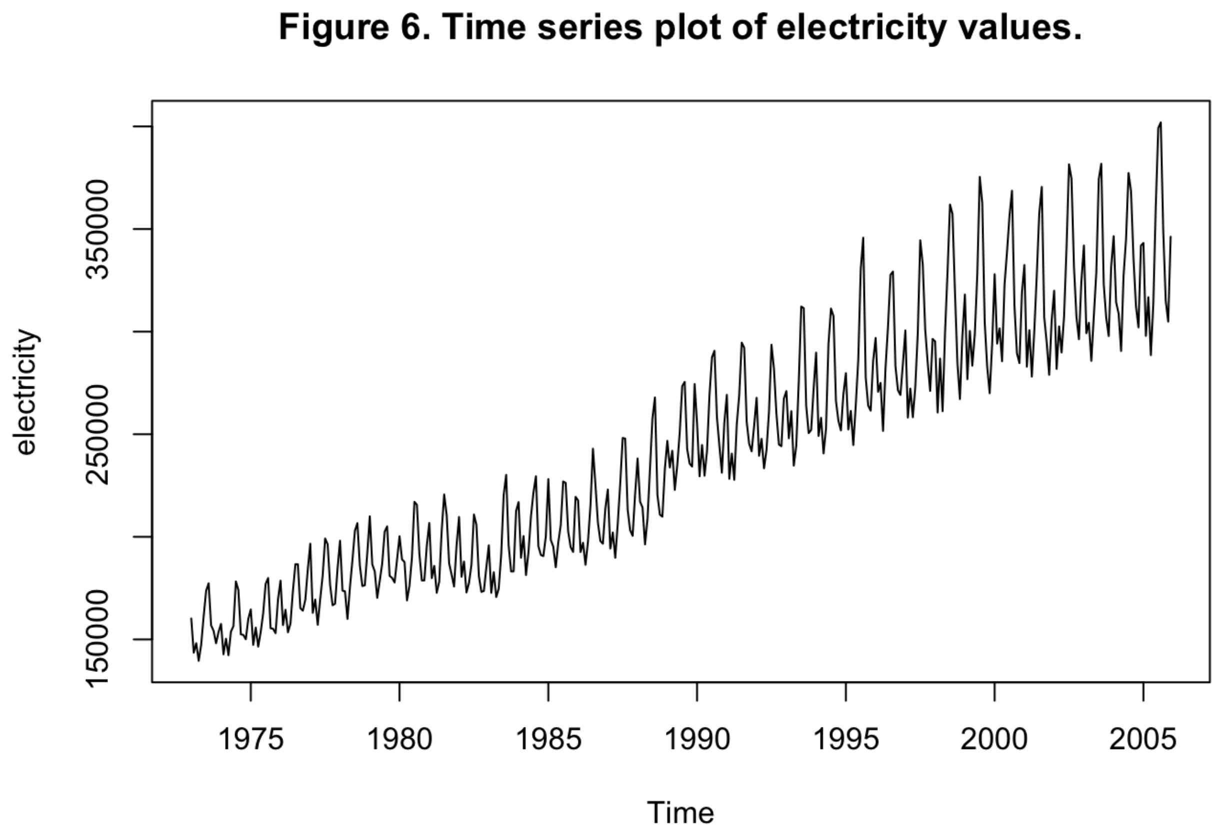
If the standard deviation of the series is proportional to the level of the series, then transforming to logarithms will produce a series with approximately constant variance over time. Also, if the level of the series is changing roughly exponentially, the log-transformed series will exhibit a linear time trend. Thus, we might then want to take the first differences.

> It is important to note that we take logs first and then compute first differences - the order does matter!

In financial literature, the differences of the (natural) logarithms are usually called *returns*.

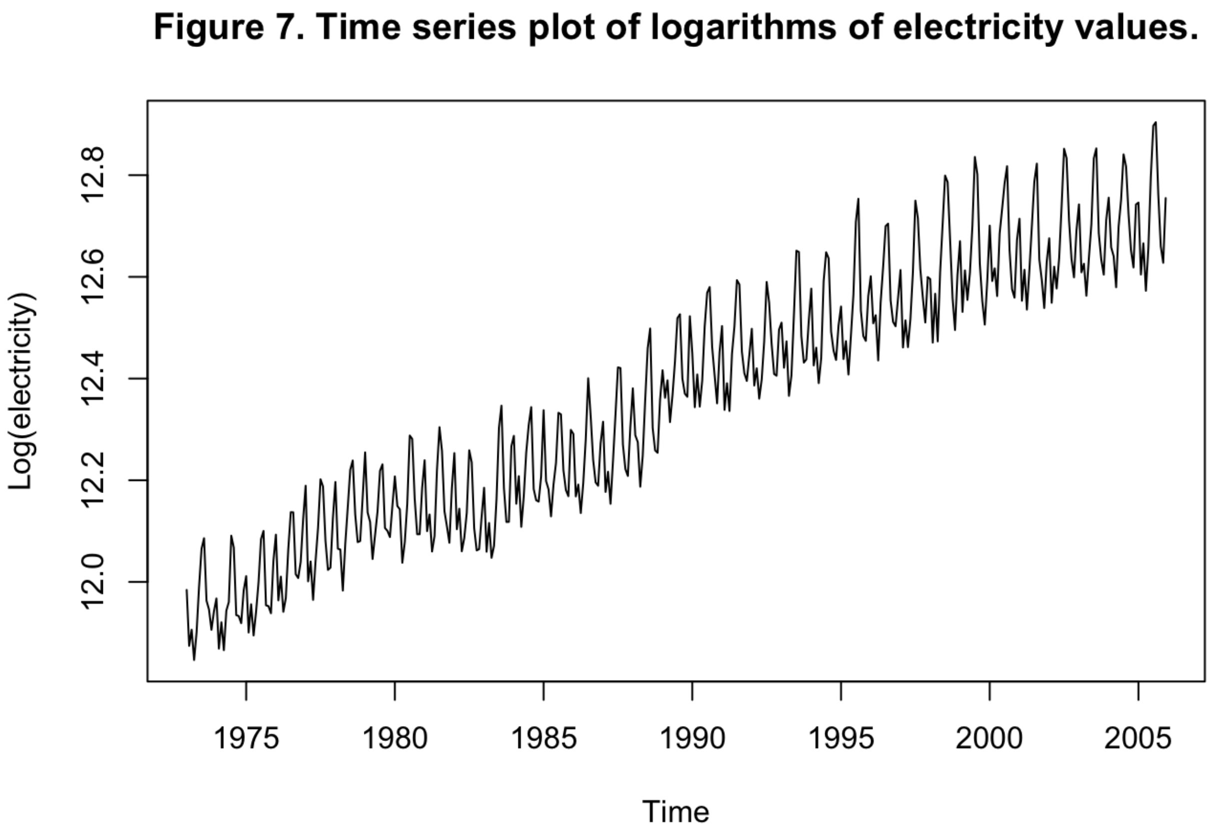
Figure 6 shows the total monthly electricity generated in the United States in millions of kilowatt-hours.

```
data(electricity)
plot(electricity,main = "Figure 6. Time series plot of electricity value
        s.")
```



In this series, the higher values display considerably more variation than the lower values. When we apply the natural log transformation, Figure 7 is obtained:

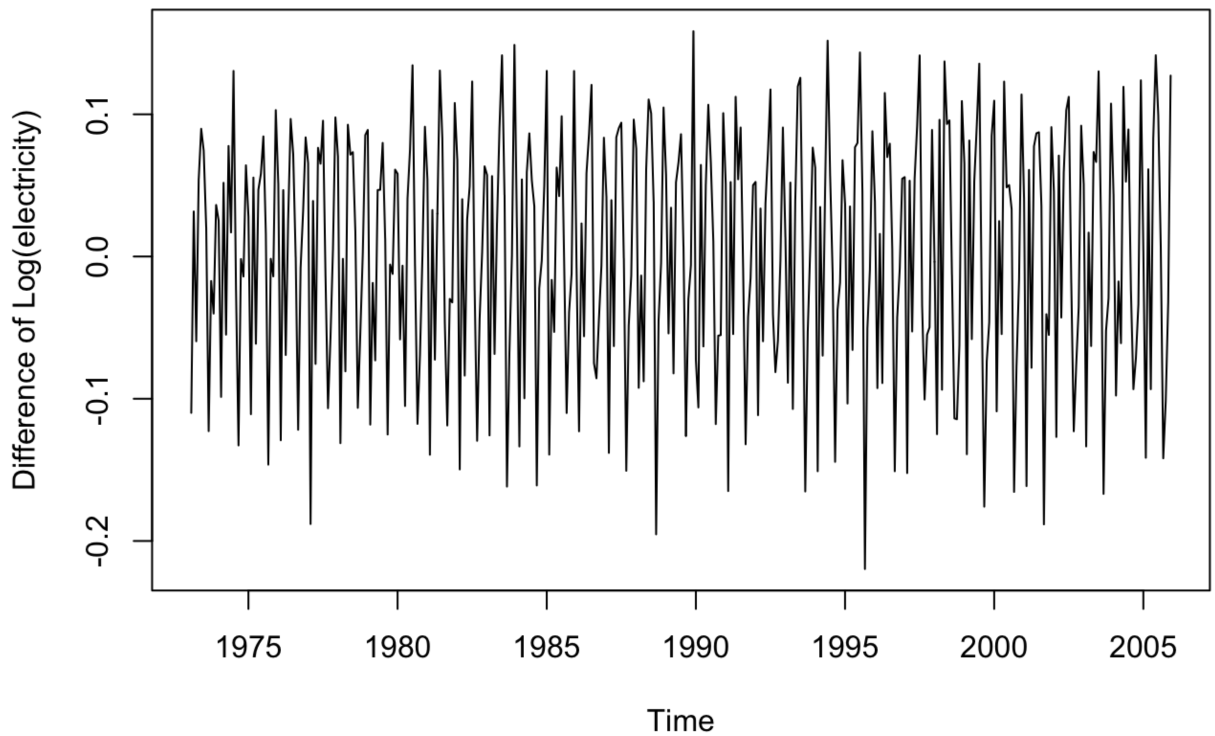
```
data(electricity)
plot(log(electricity),ylab='Log(electricity)', main = "Figure 7. Time ser
        ies plot of logarithms of electricity values.")
```



So, variance is stabilised across the time by the log transform. The differences of the logarithms of the electricity values are shown in Figure 8:

```
data(electricity)
plot(diff(log(electricity)), ylab='Difference of Log(electricity)',main =
        "Figure 8. Difference of logarithms for electricity series.")
```

Figure 8. Difference of logarithms for electricity series.



Power transformations

Box and Cox (1964) introduced a flexible and useful family of transformation called **power transformations** or **Box-Cox transformations**. For a given value of the parameter λ, the transformation is defined by

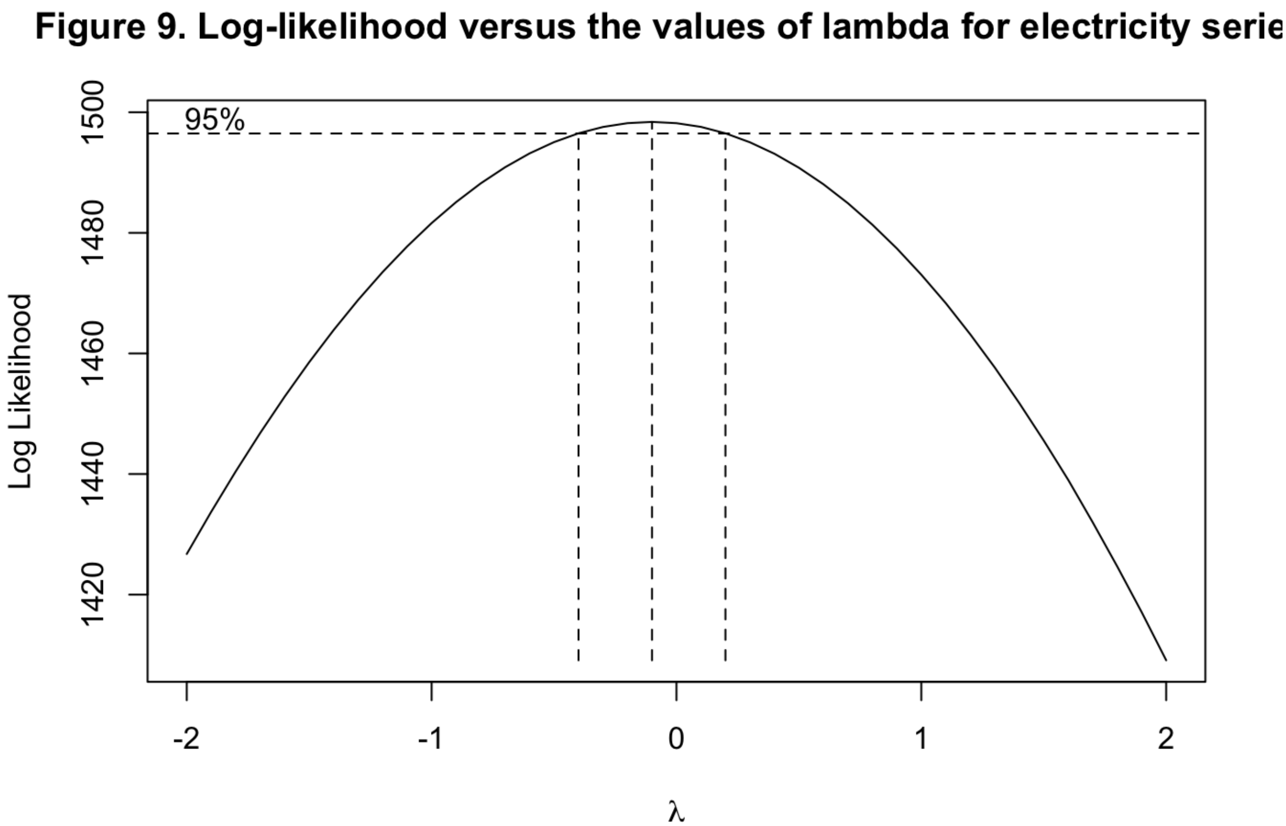
$$g(x) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log x & \text{for } \lambda = 0 \end{cases} .$$

When $\lambda = 0.5$, power transformation produces a square root transformation useful with Poisson-like data, and when $\lambda = -1$ it corresponds to a reciprocal transformation.

The power transformation applies only to positive data values. If some of the values are negative or zero, a positive constant may be added to all of the values to make them all positive before doing the power transformation. The shift is often determined subjectively.

We can consider λ as an additional parameter in the model to be estimated from the observed data. However, a precise estimation of λ is usually not warranted. Evaluation of a range of transformations based on a grid of λ values, say $\pm 1, \pm 1/2, \pm 1/3, \pm 1/4$, and 0, will usually suffice and may have some intuitive meaning. The software allows us to consider a range of lambda values and calculate a log-likelihood value for each lambda value based on a normal likelihood function. A plot of these values is generated by using the code below in Figure 9.

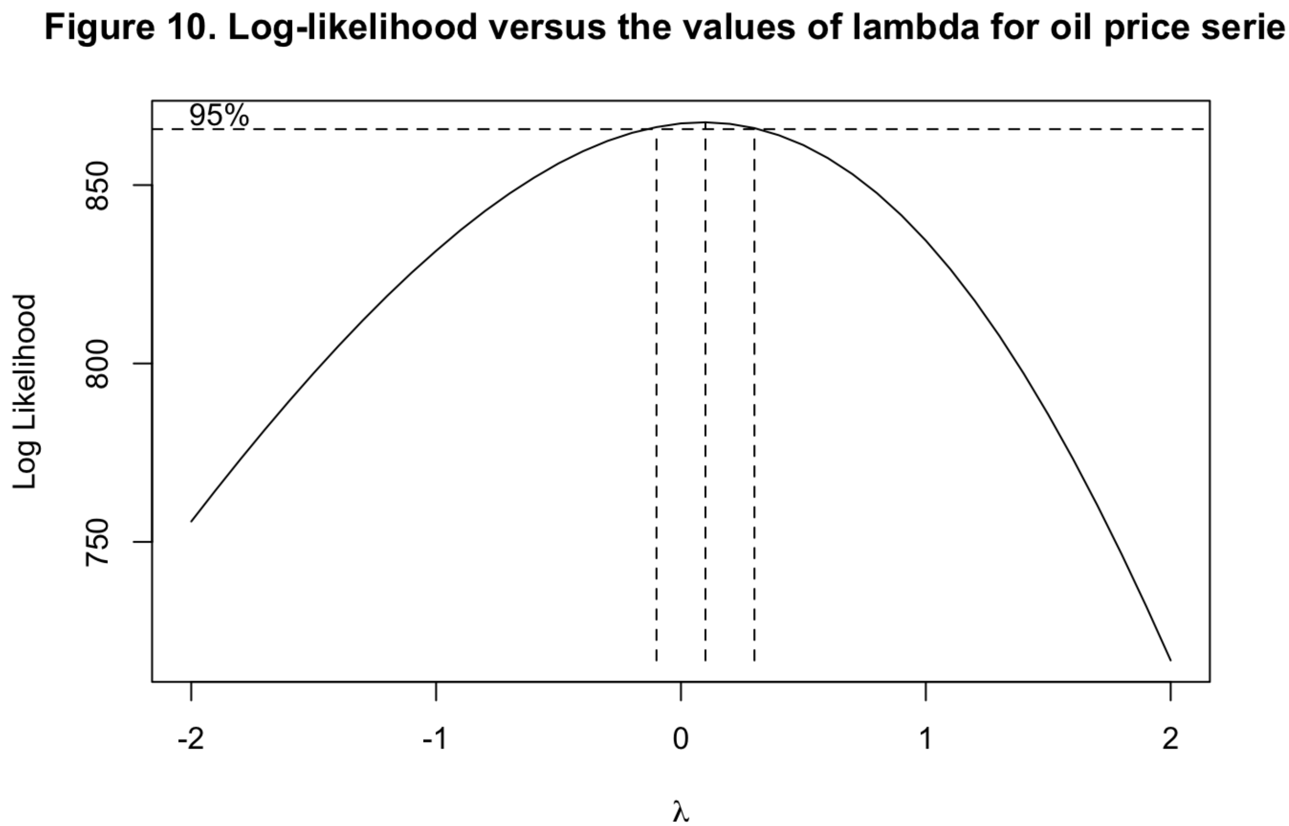
```
options(warn=-1)
BoxCox.ar(electricity)
title(main = "Figure 9. Log-likelihood versus the values of lambda for el
        ectricity series.")
```



The 95% confidence interval for λ contains the value of $\lambda = 0$ quite near its centre and strongly suggests a logarithmic transformation ($\lambda = 0$) for these data.

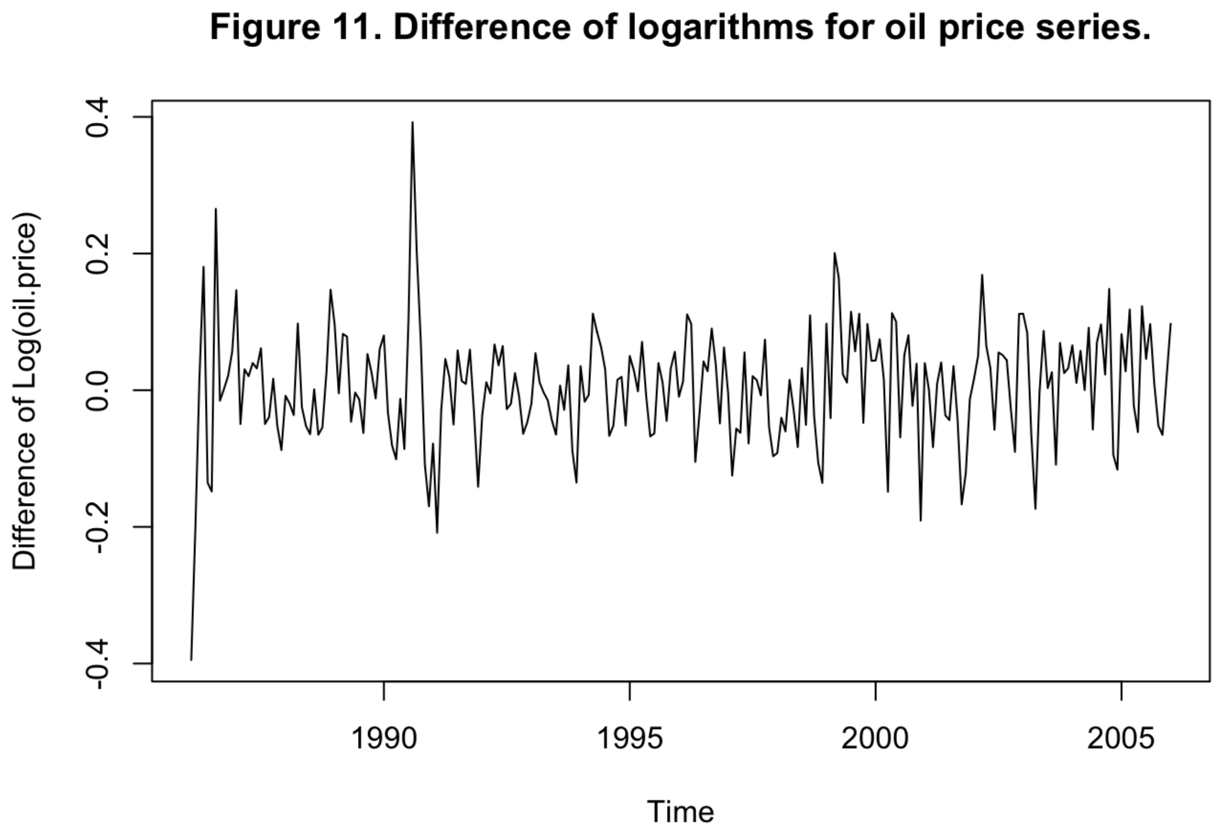
Consider the oil price series. We have the Box-Cox transformation plot in Figure 10 for this series.

```
options(warn=-1)
BoxCox.ar(oil.price)
title(main = "Figure 10. Log-likelihood versus the values of lambda for o
        il price series.")
```



The 95% confidence interval for λ again contains the value of $\lambda = 0$ quite near its center and suggests a logarithmic transformation ($\lambda = 0$).

```
plot(diff(log(oil.price)), ylab='Difference of Log(oil.price)',main = "Figure 11. Difference of logarithms for oil price series.")
```



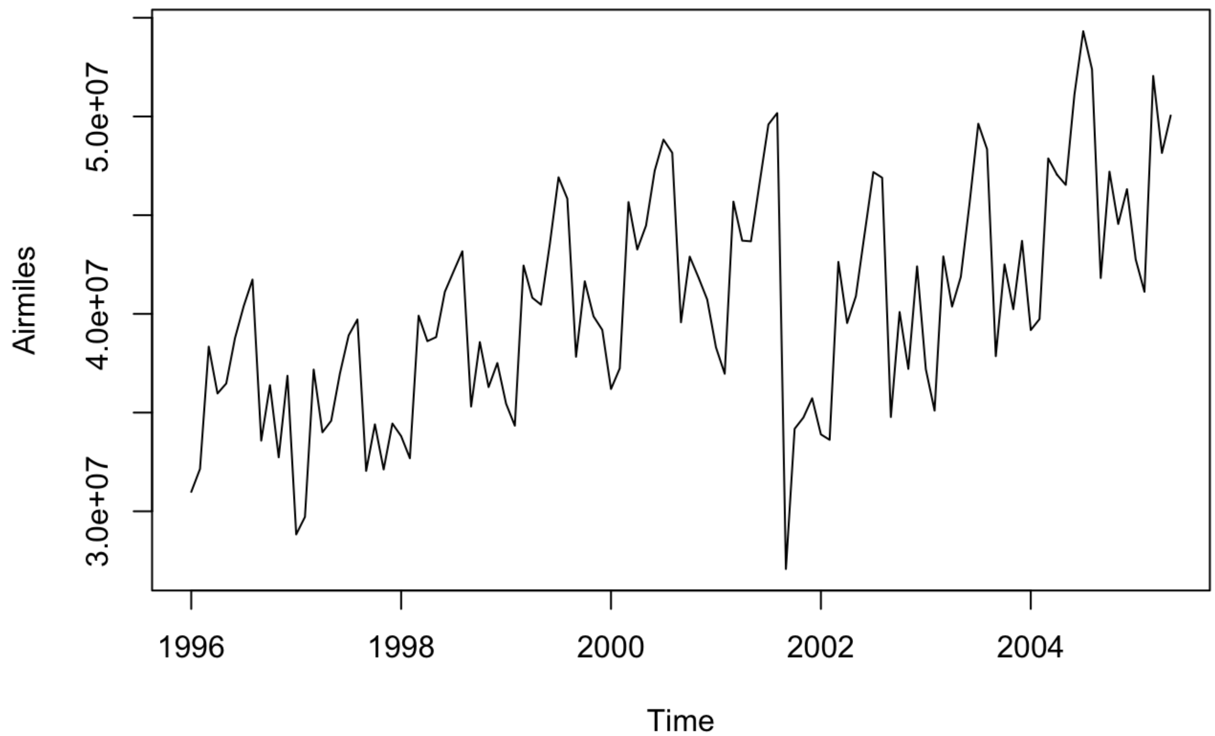
In Figure 11, we get a stationary series after taking the first difference of the log-transformed oil price series.

As another example, let us consider the Monthly U.S. airline passenger-miles series between 01/1996 - 05/2005.

We have Figure 12 for this series:

```
data(airmiles)
plot(airmiles, ylab='Airmiles',main = "Fgiure 12. Time series plot for Mo
        nthly U.S. airline passenger-miles.")
```

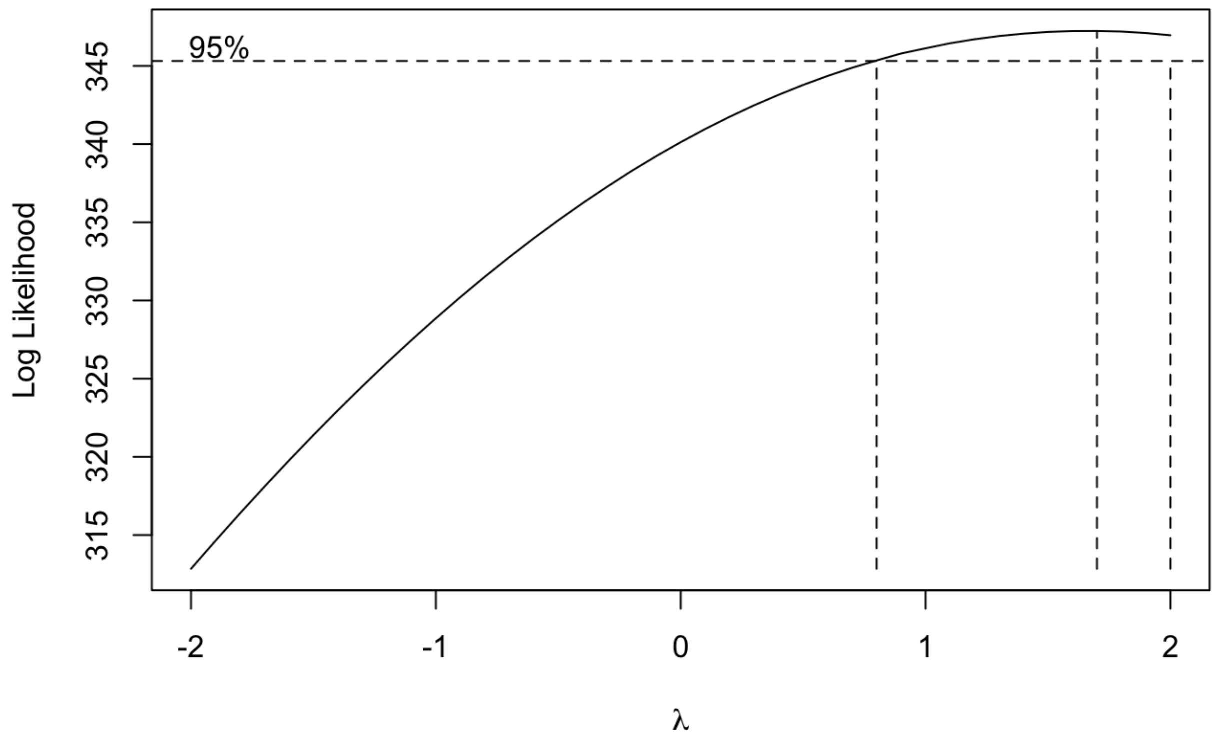
Fgiure 12. Time series plot for Monthly U.S. airline passenger-miles.



In this series, we observe both trend and a changing variance through time. So, first, we will overcome the changing variance by a transformation. For this aim, we display the Box-Cox transformation plot in Figure 13 for this series:

```
options(warn=-1)
airmiles.transf=BoxCox.ar(airmiles)
title(main = "Figure 13. Log-likelihood versus the values of lambda
      for airmiles series.")
```

Figure 13. Log-likelihood versus the values of lambda for airmiles series.




```
airmiles.transf$ci
```

```
## [1] 0.8 2.0
```

The 95% confidence interval for λ contains the values of λ between 0.8 and nearly 2. And the centre of the confidence interval corresponds to approximately 1.4, which will be used for the Box-Cox transformation. Then we will apply the transformation to the series with the following code chunk:

```
lambda = 1.4
airmiles.boxcox = (airmiles^lambda - 1) / lambda
```

Summary

In this module, we focused on the concept of differencing to induce stationarity in certain nonstationary processes. This generates important integrated autoregressive moving average models (ARIMA). The properties of these models were then thoroughly explored. Other transformations, namely percentage changes and logarithms, were then considered. More generally, power transformations or Box-Cox transformations were introduced as useful transformations to stationarity and often normality.