Time Series Analysis – Lecture 5

Models for Non-Stationary Time Series

Dr. Kevin Burke

1 Introduction

In Lecture 3 we considered models which can be expressed as

$$Y_t = \mu_t + X_t$$

where μ_t is trend component (i.e., mean function which depends on time) and X_t is a stationary series.

One approach to time series analysis is to model the trend component using the methods of Lecture 3 and then modelling the residuals (i.e., the *detrended* series) using an $ARMA(p,q)$ model. If developing a model for μ_t is of interest then this approach is reasonable. However, we must be mindful of the inherent assumption that such trend continues beyond the observed data - which may or may not be valid (see example 3.1 in Lecture 3).

An alternative approach is to model the trend and stationary component together using ***integrated* autoregressive moving average (ARIMA) models**. This approach applies the use of differencing (which we have considered in Lecture 2 and 3) to eliminate trend *without assuming a trend model.*

2 Differenced / Integrated Series

In the context of time series analysis, the operation of differening is also called **integration**. Thus, a differenced series $W_t = \nabla Y_t$ may be referred to as an **integrated series**.

The operation of differencing is very useful in practice for eliminating trend where, *typically, differencing once or twice is enough to produce a stationary series.*

2.1 Deterministic Polynomial Trend

We have seen previously that differencing can eliminate deterministic trend, for example, consider the series:

$$Y_t = \beta_0 + \beta_1\, t + X_t$$

where X_t is stationary. Thus:

$$\begin{aligned}\nabla Y_t &= \beta_0 + \beta_1\, t + X_t - \beta_0 - \beta_1\,(t-1) - X_{t-1}\\ &= \beta_1 + X_t - X_{t-1}\\ &= \beta_1 + \nabla X_t.\end{aligned}$$

Which is stationary since it is a constant plus a stationary series (Note: if X_t is stationary, ∇X_t is also stationary - see Tutorial 2 Q7).

The above operation is known as integration of order one. It can be shown that second order integration removes quadratic deterministic trend and, more generally, order d polynomial trend ($\mu_t = \sum_{j=0}^{d} \beta_j t^j$) can be removed by integration of order d, for example,

$$Y_t = \beta_0 + \beta_1\, t + \beta_2\, t^2 + \beta_3\, t^3 + \beta_4\, t^4 + X_t$$

is non-stationary but $W_t = \nabla^4\, Y_t$ is stationary.

2.2 Stochastic Trend

Often stochastic processes can display apparent trend which is not due to any deterministic trend component but, rather, high correlation between neighbouring values. This is sometimes referred to as "stochastic trend".

Consider the random walk

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

which, from Lecture 2, we know is a non-stationary process. We now see that this is an AR(1) model with $\phi = 1$ and this is clearly non-stationary since we require $|\phi| < 1$ for stationarity. However, it is easy to see that

$$\begin{aligned}Y_t - Y_{t-1} &= e_t\\ \Rightarrow \nabla Y_t &= e_t\end{aligned}$$

is stationary since this is white noise.

Another similar non-stationary process is given by

$$Y_t = Y_{t-1} + X_t$$

where $X_t = X_{t-1} + e_t$, i.e., X_t is a random walk. In this case we have

$$\begin{aligned}Y_t - Y_{t-1} &= X_t\\ \Rightarrow \nabla Y_t &= X_t\end{aligned}$$

which is still non-stationary since X_t is a random walk. Differencing again gives

$$\nabla^2 Y_t = \nabla X_t = e_t$$

which is white noise and, hence, stationary.

2.3 The Backshift Operator

Note that differencing can be expressed in terms of the backshift operator $B^m Y_t = Y_{t-m}$ as follows:

$$\begin{aligned}\nabla Y_t &= (1-B)\,Y_t \\ \nabla^2 Y_t &= (1-B)^2\,Y_t \\ &\vdots \\ \nabla^d Y_t &= (1-B)^d\,Y_t\end{aligned}$$

3 ARIMA(p,d,q) Models

Recall that an ARMA(p,q) is a stationary time series model with autoregressive and moving average components which can be written in the general form:

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)\,Y_t = (1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q)\,e_t$$

or, more compactly,

$$\phi(B)\,Y_t = \theta(B)\,e_t$$

where $\phi(B)$ and $\theta(B)$ are the AR and MA characteristic polynomials.

Since the operation of differencing (or integration) is a powerful method for removing (deterministic and stochastic) trend we can **model non-stationary series as follows:**

1. Difference d times to produce a stationary series $W_t = (1-B)^d\,Y_t$ (often $d \in \{0,1,2\}$ in practice).
2. Model W_t as $\phi(B)\,W_t = \theta(B)\,e_t$, i.e., an ARMA process.

Therefore, the **ARIMA(p,d,q) (integrated autoregressive moving average)** model for Y_t is:

$$\phi(B)\,(1-B)^d\,Y_t = \theta(B)\,e_t$$

where $(1-B)^d\,Y_t = W_t$ is the differenced series. Thus,

- p is the order of the AR component.
- d is the order of integration.
- q is the order of the MA component.

It is worth noting that the AR characteristic polynomial implied by the ARIMA model is:

$$\phi^*(B) = \phi(B)\,(1-B)^d$$

which has d roots which are equal to 1. This shows that the ARIMA model is non-stationary (for stationarity, we require all AR roots to be greater than 1). If we multiplied out

$$\phi(B)\,(1-B)^d = (1-\phi_1 B-\cdots-\phi_p B^p)\,(1-B)^d,$$

it is clear that the highest order term is B^{p+d} which shows that the ARIMA(p,d,q) is just a *non-stationary* ARMA$(p+d,q)$ model.

Note:

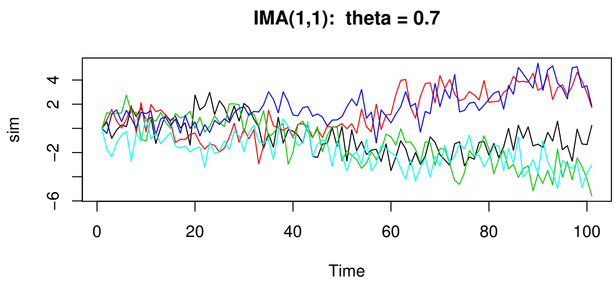
- ARIMA$(0,d,q)$ models are referred to as IMA(d,q).
- ARIMA$(p,d,0)$ models are referred to as ARI(p,d).

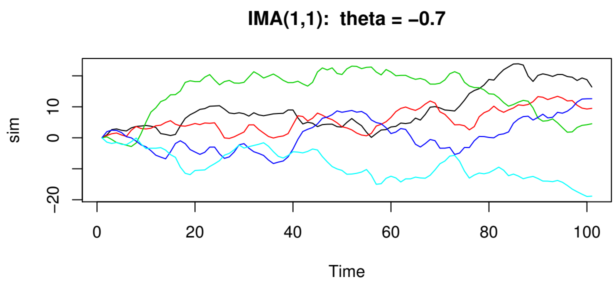
Example 3.1. *Generating ARIMA Series*
We can generate ARIMA processes using variations of the following R code (which generates 5 series from the IMA(1,1) model):

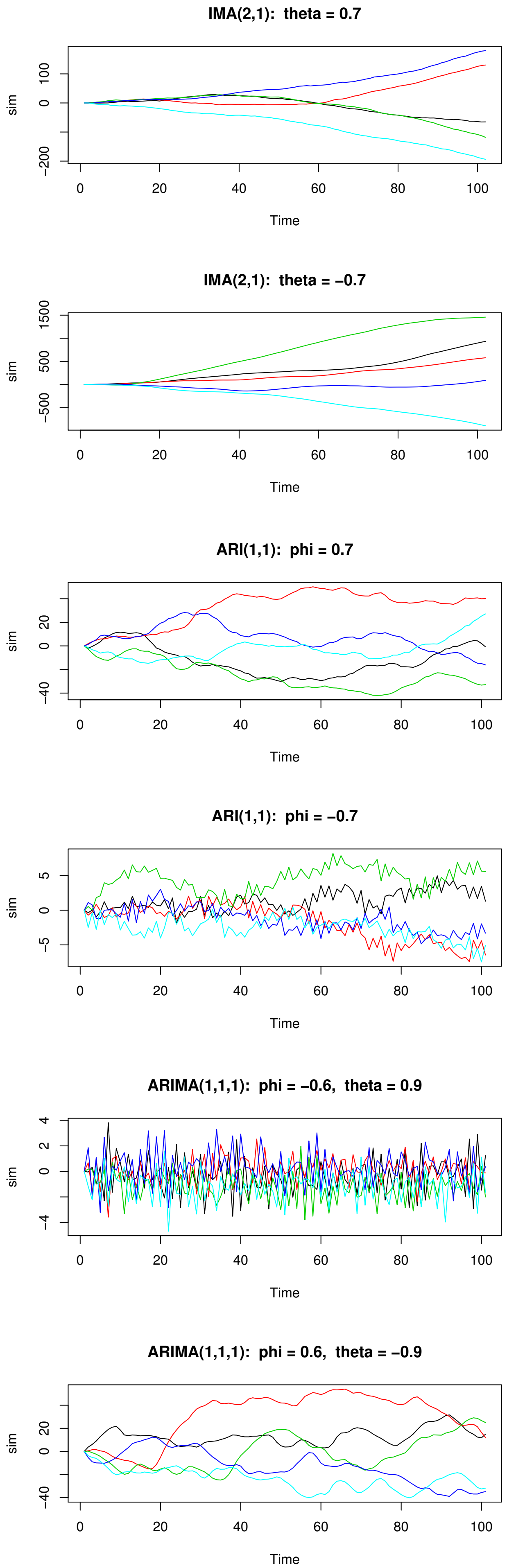
```
par(mfrow=c(2,1))
set.seed(171211)

sim <- ts( replicate(5, arima.sim(n=100,
                model=list(order=c(0,1,1), ma=c(-0.7))) ) )
plot(sim, main="IMA(1,1):  theta = 0.7", type="l",
                plot.type="single", col=1:5)

sim <- ts( replicate(5, arima.sim(n=100,
                model=list(order=c(0,1,1), ma=c(0.7))) ) )
plot(sim, main="IMA(1,1):  theta = -0.7", type="l",
                plot.type="single", col=1:5)
```







It is clear from the above that the ARIMA model can capture a large variety of time series structures.

---

4 Other Transformations

Although differencing is an important transformation in the context of time series data, log and power transformations are also useful.

Note: we will make use of the following Taylor series approximation

$$\begin{aligned} & \log(1+x) = x - \tfrac{x^2}{2} + \tfrac{x^3}{3} - \tfrac{x^4}{4} + \cdots \quad |x| < 1 \\ \Rightarrow\ & \log(1+x) \approx x \end{aligned}$$

4.1 Log Transformation

It is often the case that higher levels of the series are associated with greater variance, i.e.,

$$\mathrm{Var}(Y_t) = \mu_t^2\, \sigma^2$$

where μ_t is the mean function (which may depend on time) and σ^2 is a dispersion parameter.

It is easy to show that

$$\mathrm{Var}(\log Y_t) \approx \sigma^2$$

i.e., the variance of $\log Y_t$ is constant. Thus, the log-transformation can be used to **stabilise the variance**.

The above is achieved by letting $x = \frac{Y_t}{\mu_t} - 1$ in the Taylor approximation for $\log(1+x)$:

$$\begin{aligned} \log(1 + \tfrac{Y_t}{\mu_t} - 1) &\approx \tfrac{Y_t}{\mu_t} - 1 \\ \log(\tfrac{Y_t}{\mu_t}) &\approx \tfrac{Y_t}{\mu_t} - 1 \\ \log(Y_t) - \log(\mu_t) &\approx \tfrac{Y_t}{\mu_t} - 1 \\ \log(Y_t) &\approx \log(\mu_t) + \tfrac{Y_t}{\mu_t} - 1 \end{aligned}$$

$$\begin{aligned} \Rightarrow \mathrm{Var}(\log Y_t) &\approx \tfrac{1}{\mu_t^2}\mathrm{Var}(Y_t) \\ &= \tfrac{1}{\mu_t^2}\mu_t^2\, \sigma^2 \\ &= \sigma^2. \end{aligned}$$

Note also that

$$\begin{aligned} E(\log Y_t) &\approx \log(\mu_t) + \tfrac{1}{\mu_t}E(Y_t) - 1 \\ &= \log(\mu_t) + \tfrac{1}{\mu_t}\mu_t - 1 \\ &= \log(\mu_t). \end{aligned}$$

Therefore, if, in addition to variance growing with μ_t, the series displays exponential trend, i.e.,

$$\mu_t = \beta_0 e^{\beta_1 t}$$

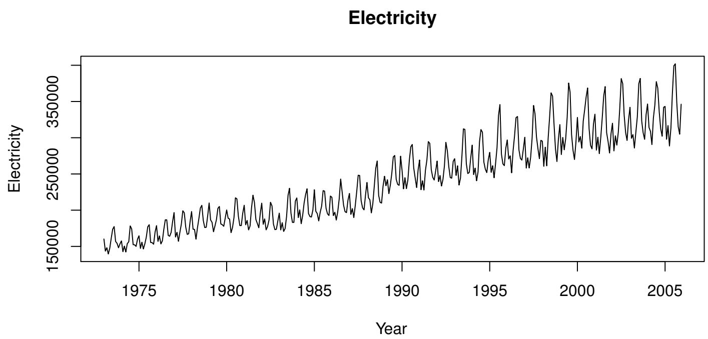
then the log-transformed series will display linear trend:

$$E(\log Y_t) \approx \log\left(\beta_0 e^{\beta_1 t}\right) \approx \log \beta_0 + \beta_1 t.$$

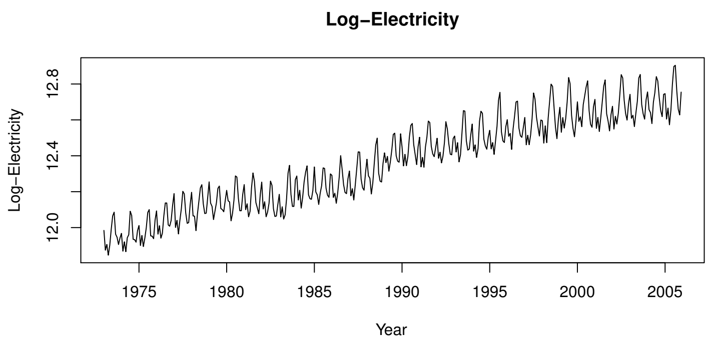
Since differencing eliminates linear trend the difference of the log transformation is suggested, i.e., we may call this the **difference-log transformation**:

$$\nabla \log Y_t = \log Y_t - \log Y_{t-1}$$

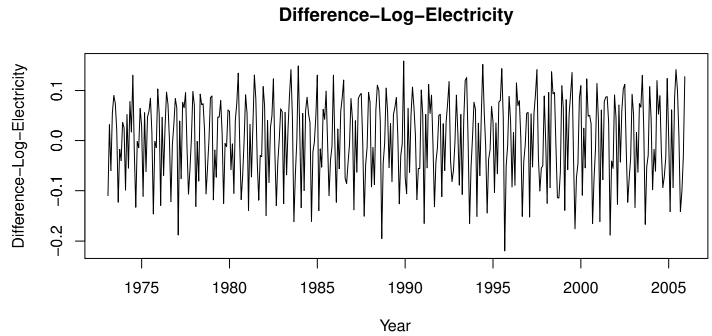
Example 4.1. *US Monthly Electricity Generated*
Consider the `electricity` series which gives the total monthly electricity generated in the United States in millions of kilowatt-hours. It is clear from the plot below that the variation in the series is increasing as the level of the series increases.



Applying the log-transformation leads to relatively constant (i.e., stable) variance:



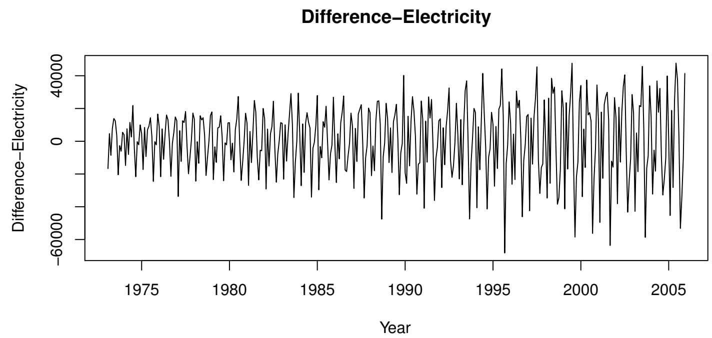
However, this log-transformed series displays linear trend which can further be removed by differencing:



This final difference-log series appears to be stationary (that is, the trend is eliminated and the variance is quite constant) and, hence, could be modelled using the ARMA framework.

Example 4.2. *US Monthly Electricity Generated*

Notice that above we **apply the log first** and then compute first differences - *the order matters*. **Never apply the log transformation to a differenced series**.



If we difference the `electricity` series we get the above series. Clearly the trend has been eliminated but the increasing variance is still present. We cannot apply the log at this stage however due to negative values in the differenced series (cannot take the log of a negative value).

4.2 Percentage Change

The difference log-transformation arose above due to exponential trend with increasing variance. It is worth noting that this transformation also arises if Y_t is non-stationary but percentage change, P_t, is stationary.

This often occurs in financial settings where a process evolves as a fairly small and stable percent-change.

Percentage change is defined as

$$P_t = \frac{Y_{t-1} - Y_t}{Y_{t-1}}$$

from which we get

$$
\begin{aligned}
Y_t &= (1+P_t)\,Y_{t-1} \\
\Rightarrow \log Y_t &= \log(1+P_t) + \log Y_{t-1} \\
&\qquad\qquad \text{(applying log to both sides)} \\
\log Y_t - \log Y_{t-1} &= \log(1+P_t) \\
\nabla \log Y_t &= \log(1+P_t) \\
&\approx P_t. \qquad \text{(taylor approximation)}
\end{aligned}
$$

Thus, the difference-log transformation is approximately equal to percentage change. Therefore, if percentage change is stationary (as it often is in financial settings), it makes sense to apply this transformation.

4.3 Box-Cox Power Transformations

Often **power transformations** are useful for stabilising variance or to make the data more normally distributed. Power transformations are those of the form

$$x^\lambda$$

where $\lambda \in (-\infty, \infty)$. However, $\lambda = 0$ does not yield a useful transformation.

Given the usefulness of the log-transformation, we can combine this with the above power transformation by defining the following transformation

$$g(x) = \begin{cases} x^\lambda & \lambda \neq 0 \\ \log x & \lambda = 0 \end{cases}$$

where we have set the vacant $\lambda = 0$ value to be the log-transformation. However, this is still not satisfactory as the transformation does not vary smoothly with λ, i.e., it jumps (in a discontinuous fashion) from x^λ to $\log x$ at $\lambda = 0$.

The **Box-Cox transformation** is given by

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

which now varies smoothly with λ, i.e., $\frac{x^\lambda-1}{\lambda} \to \log x$ as $\lambda \to 0$. Thus is easy to show:

$$\lim_{\lambda\to 0} \frac{x^\lambda - 1}{\lambda} = \frac{1-1}{0} = \frac{0}{0}$$

(need to apply l'Hopital's rule)

$$
\begin{aligned}
\lim_{\lambda\to 0} \frac{x^\lambda - 1}{\lambda} &= \lim_{\lambda\to 0} \frac{\frac{d}{d\lambda}(x^\lambda - 1)}{\frac{d}{d\lambda}\lambda} \\
&= \lim_{\lambda\to 0} \frac{x^\lambda \log x}{1} \\
&= \frac{(1)\log x}{1} \\
&= \log x
\end{aligned}
$$

Of course we have now changed the power transformation, x^λ, to a slightly different transformation, $\frac{x^\lambda-1}{\lambda}$, in favour of smoothly incorporating the log-transformation.

However, the operation of subtracting one and dividing by λ only alters the location and scale of the data (i.e., mean and variance) which is immaterial in terms of the transformation. The most important part of $\frac{x^\lambda-1}{\lambda}$ is x^λ.

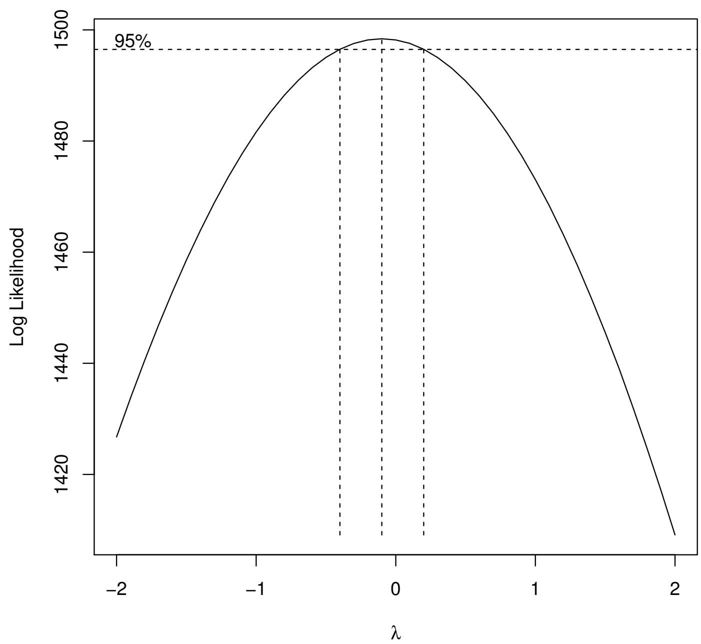
The `BoxCox.ar` function in `R` carries out the above transformation over a range of λ values and calculates a log-likelihood based on a normal likelihood function with AR covariance structure. Thus, it aims to find the λ value which transforms the data to a stationary series with normal white noise term.

Note: The `BoxCox.ar` function is quite computationally intensive and takes a little while to complete.

Example 4.3. *Box Cox Transformation*
The following code tests the Box-Cox transformation over a range of λ values from -2 to 2 (in increments of 0.1) to the `electricity` series:

```
BC <-BoxCox.ar(electricity, lambda=seq(-2,2,0.1) )
BC$mle
BC$ci
```

Over the range of values λ values explored, the maximum likelihood estimator is $\hat{\lambda} = -0.1$ with 95% confidence interval $[-0.4,\ 0.2]$. This is shown in the associated plot:



Thus, the log-transformation ($\lambda = 0$) is supported here.