Time Series Analysis with R   Code ▾

R. Nicholls / D.-L. Couturier / M. Fernandes

Last modified: 04 Mar 2019

We will consider a dataset corresponding to the Monthly Southern Oscillation Index, measured as the difference in sea-surface air pressure between Darwin and Tahiti.
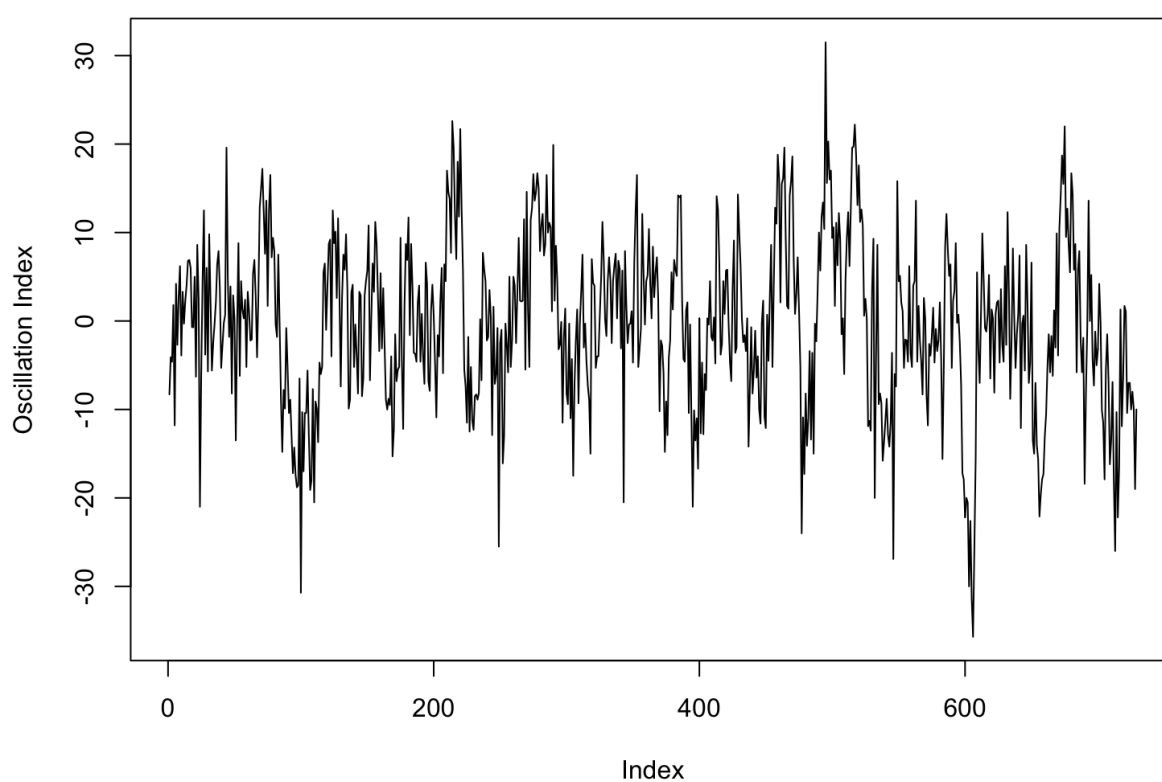
Hide

```r
x=read.table("data/OscillationIndex.txt",header=TRUE)
x$Index
```

```
##   [1]  -8.3  -4.1  -4.6   1.8 -11.8   4.2  -2.7   2.8   6.2  -3.9   3.3
##  [12]  -0.3   2.3   4.1   6.8   6.9   6.0  -0.7  -0.7   5.0  -6.3   8.6
##  [23]   2.7 -21.0  -5.9   4.7  12.5  -3.8   6.0  -5.7   9.8   2.1  -5.6
##  [34]  -2.5  -0.4   2.1   6.4   7.9   3.6  -5.3  -2.7  -0.2   0.7  19.6
##  [45]   4.7  -1.8   3.9  -8.2   2.9   0.3 -13.5  -0.7   8.8  -6.2   4.5
##  [56]   1.3   0.3   2.4  -5.2   3.3   1.1  -2.2  -2.1   5.4   6.9   2.7
##  [67]  -4.1   2.8  12.8  14.9  17.2  12.4   7.6  13.6   1.7  12.5  16.5
##  [78]   7.2   9.4   7.9  -0.4  -1.8   7.5  -0.3  -8.8 -14.8  -7.8  -9.9
##  [89]  -0.8  -5.2 -10.4  -8.9 -13.0 -17.2 -14.3 -17.4 -18.8 -18.6  -6.5
## [100] -30.7 -10.3 -17.0 -10.4 -10.4  -5.6 -13.0 -19.1 -18.0  -7.7 -20.5
## [111]  -9.1  -9.9 -13.7  -4.7  -6.0  -5.2   5.5   6.5  -1.0   3.9   8.7
## [122]   9.2  -4.0  12.5   8.8  10.1   2.6  11.6   3.3  -7.4   2.7   7.5
## [133]   5.8   9.8   3.6  -9.9  -8.9   3.2   4.1  -5.2  -0.4  -3.9  -8.2
## [144]   3.3   2.9  -8.5  -6.5   2.9   4.5   5.7  10.8  -6.7   0.3   6.5
## [155]   3.3  11.2   8.7   2.9  -3.4   5.4  -3.1   3.7  -2.7  -8.9 -10.0
## [166]  -8.8  -9.5  -4.0 -15.3 -12.3  -1.5  -6.8  -5.5  -5.2   9.4  -4.5
## [177] -12.2   1.7   8.7   6.9  11.7  -1.6   8.7   3.9  -3.6  -3.7  -4.6
## [188]   2.1   4.0  -4.6   0.8  -4.0  -7.1   6.6   4.2  -6.8  -7.9   1.2
## [199]   4.1   0.6  -4.8 -10.9  -1.6  -4.0   2.3   6.0  -5.9   6.4   4.5
## [210]  17.0  14.6  13.8   7.7  22.6  19.6  11.8   7.0  18.0  11.8  21.7
## [221]  12.7   5.7  -5.5  -7.4 -11.5  -1.8 -12.5  -5.2 -11.2 -12.3  -8.5
## [232]  -8.3  -8.9  -8.1   0.2  -6.7   7.7   5.8   4.5  -2.2  -1.8   3.5
## [243]   0.4 -12.9   1.6  -7.1  -6.0  -0.8 -25.5  -2.5  -1.0 -16.1 -13.0
## [254]  -0.3  -2.7  -5.8   5.0  -5.2  -2.2   5.0   4.0  -2.5   3.3   9.4
## [265]   2.3   2.2   2.3  11.5  -5.5  14.6   1.2  -5.2  11.4  12.8  16.6
## [276]  13.6  14.6  16.7  15.0   7.9  10.8  12.1   7.4   8.7  16.5  10.0
## [287]  11.1  10.6   1.1  19.9   2.3   8.5   4.5  -3.2  -2.7  -0.1 -11.5
## [298]  -1.8   1.4  -8.2  -9.4  -0.3 -11.0  -4.3 -17.5  -7.1  -2.2   1.3
## [309]  -9.3  -0.4   3.3   7.5  -3.0  -0.3  -4.6  -7.3  -8.9 -15.0   7.0
## [320]   4.3   4.0  -5.3  -4.0  -4.0   0.5   4.7  11.2   6.9   0.2  -1.7
## [331]   4.5   7.2   4.7  -2.5   4.5   6.3   7.6   0.3   6.8   5.9  -3.1
## [342]   5.7 -20.5   7.9   1.8  -2.5  -0.4  -0.3   1.1  -4.7   6.8  12.5
## [353]  16.5  -5.2  -3.1  -0.8  12.1   5.1  -0.4   4.5   5.2  10.4   4.2
## [364]   0.3   8.4   2.7   5.5   7.2   2.5 -10.2  -2.2  -2.8  -5.9 -14.8
## [375]  -9.1 -12.9  -4.1  -2.2   5.5   1.3   6.9   5.8   5.1  14.2  14.0
## [386]  14.2   2.3  -4.3  -4.6   1.2   2.1 -10.4  -0.4 -10.9 -21.0 -10.1
## [397] -13.5 -11.0 -16.7   0.3 -12.7  -4.7 -12.8  -6.0  -7.8   0.3  -0.4
## [408]   4.5  -1.8  -2.2   0.4  -4.8  14.1  12.6   6.5  -3.8  -2.6   4.5
## [419]   0.8   5.7   5.8  -0.3  -4.6  -6.8   3.6   9.1  -3.6  -3.0  14.3
## [430]  10.0   6.3   0.3  -2.4  -1.6  -3.4   0.3 -14.2  -7.6  -0.7  -8.2
## [441]  -5.6  -1.1  -6.4  -4.0 -10.0 -11.6  -0.2   2.3 -10.8 -12.1   0.7
## [452]  -4.5   2.5   8.6  -5.2   3.9  12.8  11.0  18.8  16.1   2.1  15.5
## [463]  16.1  19.6   9.2   1.7   1.4  14.2  15.8  18.6   6.8   0.8   3.1
## [474]   7.2   1.2  -5.2 -24.0 -10.9 -17.3  -8.2 -14.1 -11.0  -3.4 -13.4
## [485]  -3.6 -15.0  -0.3  -2.3   3.3  10.0   5.7  11.8  13.4  10.4  31.5
## [496]  15.6  20.3  16.0  17.0   9.4  10.6   1.7  11.1   6.3  12.2   9.2
## [507]  -1.5   0.3  -6.0   4.7   9.4  12.3   6.2  12.8  19.6  19.7  22.2
## [518]  18.6  13.1  17.6  11.2  12.6  10.8   0.6   2.5   0.3 -11.9 -11.3
## [529] -12.4   3.5   9.3 -20.0  -4.1   8.6  -9.4  -8.2  -9.3 -15.8 -13.7
## [540] -11.3  -8.8 -12.9 -14.2 -11.4  -3.6 -26.9  -6.0  -7.4  15.8   4.5
## [551]   5.1   2.1   1.1  -5.3  -2.1  -2.2  -4.6   6.2  -3.6  -5.2   4.0
## [562]   4.5  13.6  -4.6   1.7  -2.2  -4.6  -8.3   2.6   0.3  -8.4 -11.8
## [573]  -2.6  -3.9  -1.6   1.5  -4.7  -0.9  -3.4  -2.2   2.1  -4.2 -15.6
## [584]  -5.2   8.4  12.1   8.1   5.1   6.4  -5.3   2.3   3.4   8.8  -0.2
## [595]   0.7  -2.3  -7.1 -17.2 -17.9 -22.2 -20.0 -20.5 -30.0 -22.6 -31.4
## [606] -35.7 -25.7 -15.5   5.5  -3.2  -7.0   0.9   9.9   4.7  -0.8  -1.2
## [617]   0.7   5.2  -6.5   1.3   0.3  -8.1   0.8   2.1   2.3  -4.7   3.6
```

```
## [628]  -2.7  -4.6   6.2  -2.7  12.3   3.3  -8.8  -2.2   8.2   0.5  -5.3
## [639]  -1.5   0.8   7.4 -12.1  -0.3   0.6  -5.6   8.6   2.0  -7.0  -4.7
## [650]   6.6 -13.5 -15.0  -7.0 -14.0 -15.6 -22.1 -19.6 -17.9 -17.3 -13.1
## [661] -10.6  -5.3  -1.5  -5.8  -1.7  -6.2   1.2  -3.0   9.9  -3.9  10.5
## [672]  14.2  18.7  15.5  22.0   9.5  12.7   8.6   5.5  16.7  14.3   5.8
## [683]   8.7  -5.8   5.8   7.9  -2.1  -5.8  -1.9 -18.4  -8.2  -0.7  13.6
## [694]   0.0   5.2  -4.4  -7.3  -1.2  -5.0  -3.2   4.2  -0.2 -10.1 -11.5
## [705] -17.9  -5.5  -1.5  -6.8 -16.2 -13.5  -6.9 -18.3 -26.0 -10.3 -22.2
## [716] -16.5   1.3 -11.9  -6.5   1.7   1.1 -10.4  -7.0  -7.0 -10.0  -8.0
## [727] -10.0 -19.0 -10.0
```
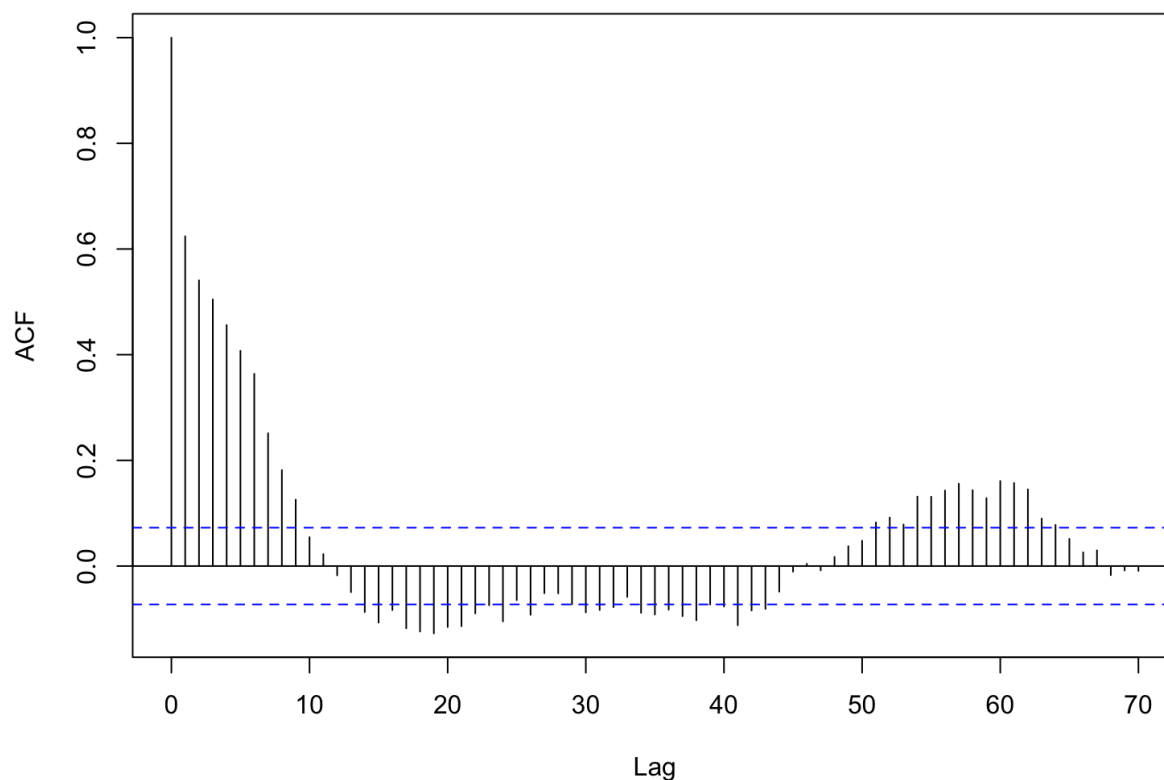
Hide

```r
plot(x$Index,type="l",ylab="Oscillation Index")
```


Now let's look at the autocorrelation function corresponding to this dataset:

Hide
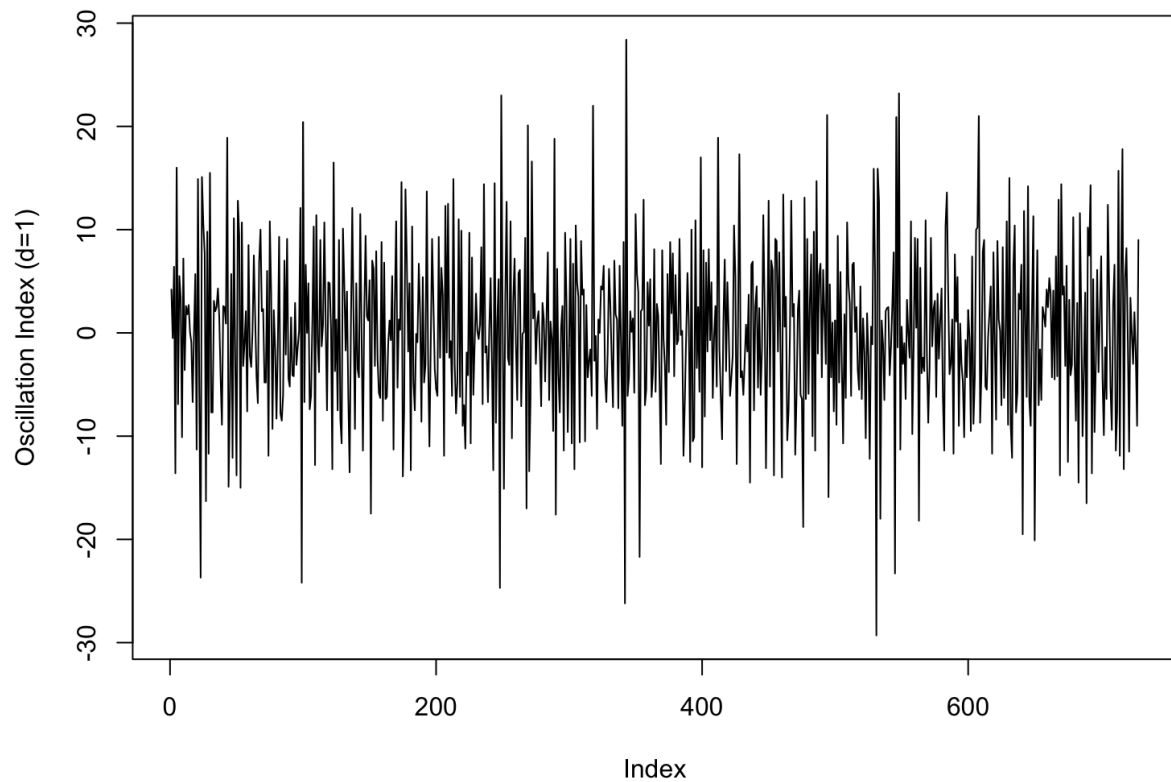
```r
acf(x$Index,lag.max=70,main="")
```

There is clear long-range oscillatory behaviour in the autocorrelation function, indicating that the process is not stationary.
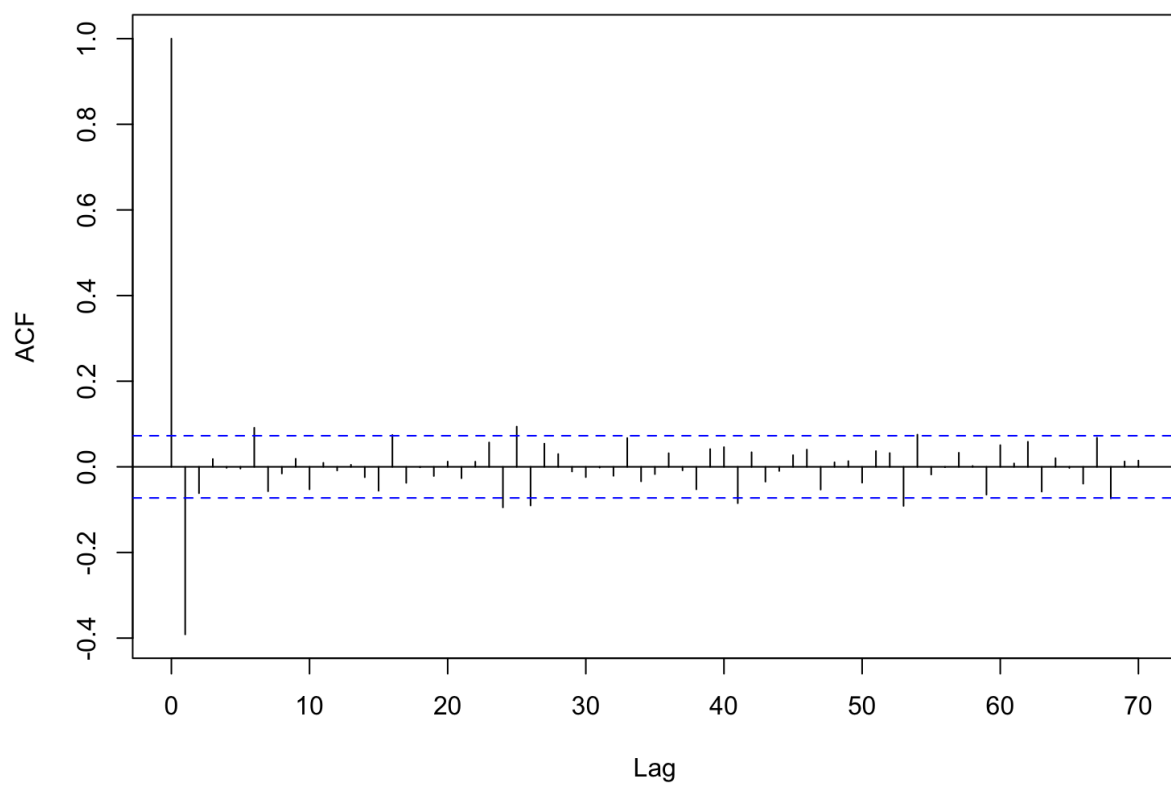
We should consider an integrated (ARIMA) model, so let's calculate and plot the first differences, as well as the associated autocorrelation function:

Hide

```
plot(diff(x$Index),type="l",ylab="Oscillation Index (d=1)")
```
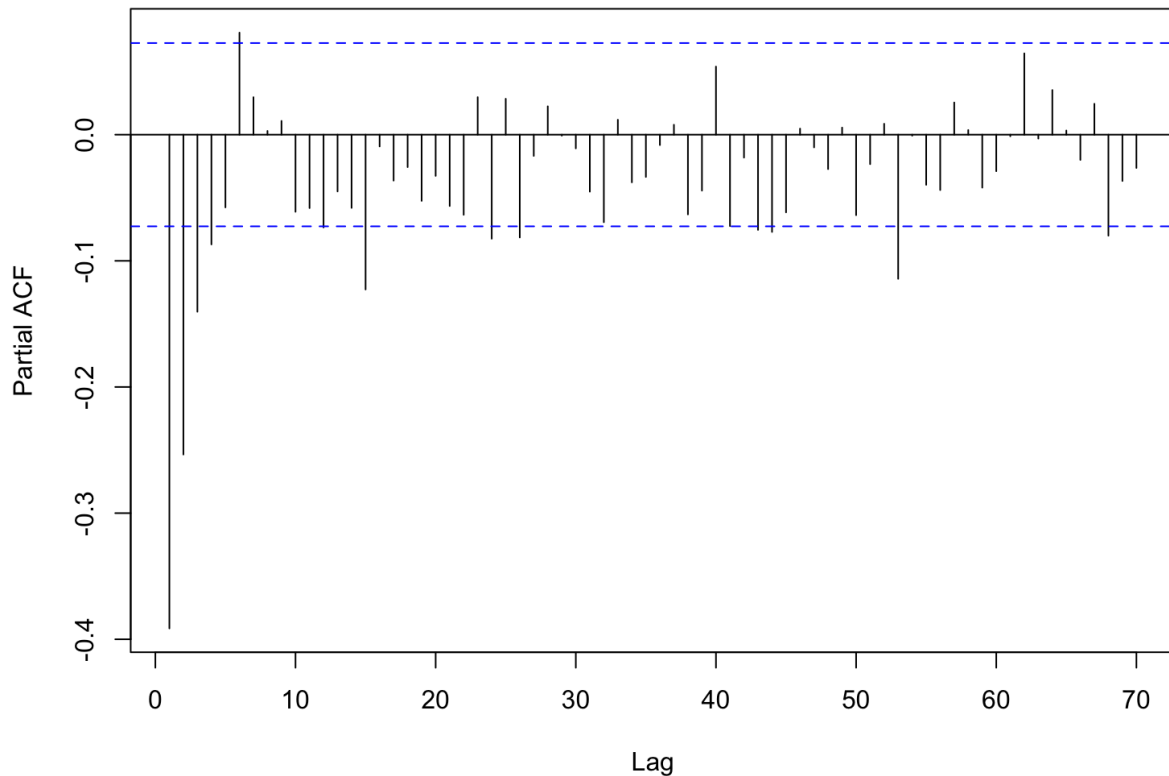
```
acf(diff(x$Index),lag.max=70,main="")
```

That's more promising - there is one large negative peak at lag=1, after which the autocorrelation function decays rapidly and stays small. This indicates that this process is covariance stationary. This also indicates that the Moving Average (MA) part of the model may be of order 1. So an ARIMA(0,1,1) model might be a possibility.

Now let's look at the partial autocorrelation function:

Hide

```
pacf(diff(x$Index),lag.max=70,main="")
```


This also looks promising. There are four negative peaks before the PACF decays below the significance threshold. That indicates that the AutoRegressive (AR) part of the model may have order up to 4.

Now we'll try to create an ARIMA(0,1,1) model:

Hide

```
arima(x$Index,order=c(0,1,1))
```

```
##
## Call:
## arima(x = x$Index, order = c(0, 1, 1))
##
## Coefficients:
##           ma1
##        -0.5579
## s.e.    0.0308
##
## sigma^2 estimated as 52.94:  log likelihood = -2477.98,  aic = 4959.96
```

Note that the standard error of the coefficient indicates significance of the term.

Now try creating other ARIMA models, and compare.

There are a variety of time series datasets in the in-built R "datasets" package. Type `data()` to get a full list. For example, the datasets called `lh`, `ldeaths` and `presidents` are particularly appropriate for this type of analysis. Other datasets also contain time series data, including: `nhtemp`, `lynx`, `Nile`, `co2` and `WWWusage`. Explore such datasets - look at autocorrelation and partial autocorrelation functions, identify whether the datasets are suitable for time series analysis, and try fitting ARIMA models.