Time Series in R: Forecasting and Visualisation

Time series in R

29 May 2017
Outline

1. ts objects
2. Time plots
3. Lab session 1
4. Seasonal plots
5. Seasonal or cyclic?
6. Lag plots and autocorrelation
7. Lab session 2
Time series

Time series consist of sequences of observations collected over time.

We will assume the time periods are equally spaced.

Time series examples

- Daily IBM stock prices
- Monthly rainfall
- Annual Google profits
- Quarterly Australian beer production
A time series is stored in a `ts` object in R:
- a list of numbers
- information about times those numbers were recorded.

### Example

<table>
<thead>
<tr>
<th>Year</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>123</td>
</tr>
<tr>
<td>2013</td>
<td>39</td>
</tr>
<tr>
<td>2014</td>
<td>78</td>
</tr>
<tr>
<td>2015</td>
<td>52</td>
</tr>
<tr>
<td>2016</td>
<td>110</td>
</tr>
</tbody>
</table>

\[ y \leftarrow \text{ts(c(123,39,78,52,110), start=2012)} \]
For observations that are more frequent than once per year, add a `frequency` argument. E.g., monthly data stored as a numerical vector `z`:

```
y <- ts(z, frequency=12, start=c(2003, 1))
```
```markdown
## ts objects and ts function

<table>
<thead>
<tr>
<th>Type of data</th>
<th>frequency</th>
<th>start example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual</td>
<td>1</td>
<td>1995</td>
</tr>
<tr>
<td>Quarterly</td>
<td>4</td>
<td>c(1995,2)</td>
</tr>
<tr>
<td>Monthly</td>
<td>12</td>
<td>c(1995,9)</td>
</tr>
<tr>
<td>Daily</td>
<td>7 or 365.25</td>
<td>1 or c(1995,234)</td>
</tr>
<tr>
<td>Weekly</td>
<td>52.18</td>
<td>c(1995,23)</td>
</tr>
<tr>
<td>Hourly</td>
<td>24 or 168 or 8,766</td>
<td>1</td>
</tr>
<tr>
<td>Half-hourly</td>
<td>48 or 336 or 17,532</td>
<td>1</td>
</tr>
</tbody>
</table>
```
ts objects

- Class: “ts”
- Print and plotting methods available.

<table>
<thead>
<tr>
<th>Year</th>
<th>Qtr1</th>
<th>Qtr2</th>
<th>Qtr3</th>
<th>Qtr4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>4612</td>
<td>4651</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1972</td>
<td>4645</td>
<td>4615</td>
<td>4645</td>
<td>4722</td>
</tr>
<tr>
<td>1973</td>
<td>4780</td>
<td>4830</td>
<td>4887</td>
<td>4933</td>
</tr>
<tr>
<td>1974</td>
<td>4921</td>
<td>4875</td>
<td>4867</td>
<td>4905</td>
</tr>
<tr>
<td>1975</td>
<td>4938</td>
<td>4934</td>
<td>4942</td>
<td>4979</td>
</tr>
<tr>
<td>1976</td>
<td>5028</td>
<td>5079</td>
<td>5112</td>
<td>5127</td>
</tr>
</tbody>
</table>
ts objects

\texttt{start(ausgdp)}

\begin{verbatim}
## [1] 1971 3
\end{verbatim}

\texttt{end(ausgdp)}

\begin{verbatim}
## [1] 1998 1
\end{verbatim}

\texttt{frequency(ausgdp)}

\begin{verbatim}
## [1] 4
\end{verbatim}
ts objects

Residential electricity sales

elecsales

## Time Series:
## Start = 1989
## End = 2008
## Frequency = 1
## [1] 2354 2380 2319 2469 2386 2569 2576 2763 2844
## [10] 3001 3108 3358 3076 3181 3222 3176 3431 3527
## [19] 3638 3655
ts objects

```
start(elecsales)

## [1] 1989  1

end(elecsales)

## [1] 2008  1

frequency(elecsales)

## [1] 1
```
Main package used in this course

> library(fpp2)

This loads:

- some data for use in examples and exercises
- **forecast** package (for forecasting functions)
- **ggplot2** package (for graphics)
- **fma** package (for lots of time series data)
- **expsmooth** package (for more time series data)
Outline

1. ts objects
2. Time plots
3. Lab session 1
4. Seasonal plots
5. Seasonal or cyclic?
6. Lag plots and autocorrelation
7. Lab session 2
ts objects

autoplot(ausgdp)
Time plots

```r
autoplot(a10) + ylab("$ million") + xlab("Year") + ggtitle("Antidiabetic drug sales")
```
Outline

1. ts objects
2. Time plots
3. Lab session 1
4. Seasonal plots
5. Seasonal or cyclic?
6. Lag plots and autocorrelation
7. Lab session 2
Lab Session 1
Outline

1. ts objects
2. Time plots
3. Lab session 1
4. **Seasonal plots**
5. Seasonal or cyclic?
6. Lag plots and autocorrelation
7. Lab session 2
antidotplot(a10) + ylab("$ million") + xlab("Year") + ggttitle("Antidiabetic drug sales")
Seasonal plot

```
ggseasonplot(a10, year.labels=TRUE, year.labels.left=TRUE) +
  ylab("$ million") +
  ggtitle("Seasonal plot: antidiabetic drug sales")
```
Seasonal polar plots

```r
ggseasoneplot(a10, polar=TRUE) + ylab("$ million")
```
Seasonal subseries plots

```r
ggsubseriesplot(a10) + ylab("$ million") + ggttitle("Subseries plot: antidiabetic drug sales")
```
Quarterly Australian Beer Production

```r
beer <- window(ausbeer, start=1992)
autoplot(beer)
```
Quarterly Australian Beer Production

```
ggseasonplot(beer, year.labels=TRUE)
```
Quarterly Australian Beer Production

ggsubseriesplot(beer)
Outline

1. ts objects
2. Time plots
3. Lab session 1
4. Seasonal plots
5. **Seasonal or cyclic?**
6. Lag plots and autocorrelation
7. Lab session 2
Time series patterns

**Trend** pattern exists when there is a long-term increase or decrease in the data.

**Seasonal** pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).

**Cyclic** pattern exists when data exhibit rises and falls that are *not of fixed period* (duration usually of at least 2 years).
Time series patterns

```r
autoplot(window(elec, start=1980)) +
ggtitle("Australian electricity production") +
xlab("Year") + ylab("GWh")
```
Time series patterns

```r
autoplot(bricksq) +
  ggtitle("Australian clay brick production") +
  xlab("Year") + ylab("million units")
```
Time series patterns

```r
autoplot(ustreas) +
ggtitle("US Treasury Bill Contracts") +
xlab("Day") + ylab("price")
```
Time series patterns

```r
autoplot(lynx) +
ggtitle("Annual Canadian Lynx Trappings") +
  xlab("Year") + ylab("Number trapped")
```
Differences between seasonal and cyclic patterns:

- Seasonal pattern constant length; cyclic pattern variable length
- Average length of cycle longer than length of seasonal pattern
- Magnitude of cycle more variable than magnitude of seasonal pattern
Seasonal or cyclic?

Differences between seasonal and cyclic patterns:
- Seasonal pattern constant length; cyclic pattern variable length
- Average length of cycle longer than length of seasonal pattern
- Magnitude of cycle more variable than magnitude of seasonal pattern

The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.
Outline

1. ts objects
2. Time plots
3. Lab session 1
4. Seasonal plots
5. Seasonal or cyclic?
6. Lag plots and autocorrelation
7. Lab session 2
Example: Beer production

```r
beer <- window(ausbeer, start=1992)
gglagplot(beer)
```
Example: Beer production
Lagged scatterplots

- Each graph shows $y_t$ plotted against $y_{t-k}$ for different values of $k$.
- The autocorrelations are the correlations associated with these scatterplots.
Autocorrelation

Results for first 9 lags for beer data:

<table>
<thead>
<tr>
<th>$r_1$</th>
<th>$r_2$</th>
<th>$r_3$</th>
<th>$r_4$</th>
<th>$r_5$</th>
<th>$r_6$</th>
<th>$r_7$</th>
<th>$r_8$</th>
<th>$r_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.102</td>
<td>-0.657</td>
<td>-0.060</td>
<td>0.869</td>
<td>-0.089</td>
<td>-0.635</td>
<td>-0.054</td>
<td>0.832</td>
<td>-0.108</td>
</tr>
</tbody>
</table>

`ggAcf(beer)`
Autocorrelation

- $r_4$ higher than for the other lags. This is due to the **seasonal pattern in the data**: the peaks tend to be 4 quarters apart and the troughs tend to be 2 quarters apart.
- $r_2$ is more negative than for the other lags because troughs tend to be 2 quarters behind peaks.
- Together, the autocorrelations at lags 1, 2, ..., make up the *autocorrelation* or ACF.
- The plot is known as a *correlogram*
Trend and seasonality in ACF plots

- When data have a trend, the autocorrelations for small lags tend to be large and positive.
- When data are seasonal, the autocorrelations will be larger at the seasonal lags (i.e., at multiples of the seasonal frequency)
- When data are trended and seasonal, you see a combination of these effects.
Aus monthly electricity production

elec2 <- window(elec, start=1980)
autoplot(elec2)
Aus monthly electricity production

```r
ggAcf(elec2, lag.max=48)
```
Google stock price

autoplot(goog)
ggAcf(goog, lag.max=100)
Which is which?

1. Daily temperature of cow
2. Monthly accidental deaths
3. Monthly air passengers
4. Annual mink trappings

ACF

A

B

C

D

40
60
80
0 20 40 60
chirps per minute
1974 1976 1978
thousands

thousands

1860 1880 1900
thousands

0.0
0.5
1.0

0.0
0.5
1.0

0.0
0.5
1.0

0.0
0.5
1.0

6 12 18 24

5 10 15

5 10 15

5 10 15
Outline

1. ts objects
2. Time plots
3. Lab session 1
4. Seasonal plots
5. Seasonal or cyclic?
6. Lag plots and autocorrelation
7. Lab session 2
Lab Session 2