Forecasting: principles and practice

Before doing any exercises in R, load the **fpp** package using `library(fpp)`.

1. The function `tsdisplay(data, plot.type="scatter")` is useful for showing a time plot, ACF plot and lagged scatterplot on the same graph. Use it to produce plots of the following time series:

   `bricksq, hsales, ibmclose`

   Can you spot the effects of seasonality, cyclicity and trend?

2. For each of the same series (`bricksq, ibmclose, hsales`):

   (a) Use either the naive or seasonal naive forecasting method and apply it to the full data set.

   (b) Compute the residuals and plot their ACF. Do the residuals appear to be white noise? What did your forecasting method miss?

   (c) Do a Ljung-Box test on the residuals. What do the results mean?

3. For the data set `bricksq`:

   (a) Split the data into two parts using

   ```
   bricks1 <- window(bricksq, end=1987.99)
   bricks2 <- window(bricksq, start=1988)
   ```

   (b) Check that your data have been split appropriately by producing the following plot.

   ```
   plot(bricksq)
   lines(bricks1,col="red")
   lines(bricks2,col="blue")
   ```

   (c) Calculate forecasts using each of the four benchmark methods applied to `bricks1`.

   (d) Compare the accuracy of your forecasts against the actual values stored in `bricks2`. For example:

   ```
   f1 <- meanf(bricks1)
   accuracy(f1,bricks2)
   ```

   (e) Which method does best? Why?

   (f) For the best method, compute the residuals and plot them. For example

   ```
   res <- residuals(f1)
   plot(res)
   hist(res, breaks="FD")
   Acf(res)
   ```

   Do the residuals appear to be uncorrelated and normally distributed?

4. Consider the daily closing IBM stock prices (data set `ibmclose`).

   (a) Produce some plots of the data in order to become familiar with it.
   (b) Split the data into a training set of 300 observations and a test set of 69 observations.
   (c) Try various benchmark methods to forecast the training set and compare the results on the test set. Which method did best?
   (d) For the best method, compute the residuals and plot them. What do the plots tell you?
   (e) Can you invent a better forecasting method than any of the benchmark methods for these data?

5. Consider the sales of new one-family houses in the USA (Jan 1987 – Nov 1995). Data set: `hsales`.

   (a) Produce some plots of the data in order to become familiar with it.
   (b) Split the data into a training set and a test set, where the test set is the last two years of data.
   (c) Try various benchmark methods to forecast the training set and compare the results on the test set. Which method did best?
   (d) For the best method, compute the residuals and plot them. What do the plots tell you?
   (e) Can you invent a better forecasting method than any of the benchmark methods for these data?