Henrik Madsen
Time Series Analysis
Chapman Hall/CRC
ISBN 978-1-4200-5976

Modelling and Predicting the Number of Airline Passengers[1]. Related to Chapter 5 and 7 of the book.

Introduction

An airline company is interested in setting up a model for predicting the number of passengers on a regular basis. The available data for this assignment is the number of monthly airline passengers in the U.S. over a period of approximately seven years.

In this assignment you will use models and methods from Chapter 5 and 6 of the book for building a model and for making the predictions. Furthermore we will a cross validation method which is often recommended. For this method a part of the data is not supposed to be used for model building, but is solely to be used for comparing the predictions with the actual data. By this approach you obtain a measure of the performance of the model on an independent set of data.

Data

The data is found on in the file `assignment3data.txt`[2]. After downloading the file the data can be imported and split in two using the commands in S-PLUS (">" indicates the S-PLUS prompt):

```
> x <- read.table("assignment3data.txt")
> x <- x[,1]
> x <- rts(x, start=1995, freq=12)
> x1 <- x[1:78]
> x2 <- x[-(1:78)]
```

[1]The assignment is found at `www.imm.dtu.dk/~hm/time.series.analysis`
[2]The file is found at `www.imm.dtu.dk/~hm/time.series.analysis`

The third line transforms the numeric data into a regular time series, where the start year and frequency is set. This has only cosmetic effect on the output from S-PLUS and is not required. After import the two sets of data `x1` and `x2` should look like:

```
> x1
          1     2     3     4     5     6     7     8     9    10    11    12
1995: 40878 38746 47103 45282 45961 48561 49883 51443 43480 46651 44712 45068
1996: 41689 43390 51410 48335 50856 51317 52778 54377 45403 49473 44585 48935
1997: 44850 43133 53305 49461 50856 52925 55366 55868 46826 50216 47190 49134
1998: 44705 43742 53050 52255 52692 54702 55841 56546 47356 52024 49461 50483
1999: 45972 45101 55402 53256 53334 56457 59881 58424 49816 54684 52754 50874
2000: 46242 48160 58459 55800 57976 60787 62404 61098 51954 56322 54738 52212
2001: 49390 47951 58824 56357 56677 59515
 start     deltat frequency
  1995 0.08333333        12
> x2
          1     2     3     4     5     6     7     8     9    10    11    12
2001:                                     61969 62654 34365 43895 44442 45316
2002: 42947 42727 53553
  start     deltat frequency
 2001.5 0.08333333        12
```

Tasks

Based on the data in `x1` you must now build an appropriate time series model using the methods described in the book. Then the suggested model is used for prediction corresponding to `x2`, and the predictions are compared with the actual data.

- Plot the data (`x1`)
- Consider stationarity
- Estimate the autocorrelation function and the partial autocorrelation function of `x1` or some series derived from `x1`
- Select a model structure
- Estimate the parameters
- Validate the model

- Consider tests for lower model order
- Use the model you have developed for predicting the number of passengers 9 months ahead and include prediction limits
- Compare with the actual number of airline passengers in the second part of the data (`x2`) – comment on the results

Argue for the choices you make. Remember that the model building process is an iterative process and you should always consider stepping back and consider your choices again. However, it is not intended that you use data from the second part of the period, i.e. `x2`, when developing the model.